OPEN ACCESS 22
ISSN 2460-9056
http://socj.telkomuniversity.ac.id/ijoict/
Intl. Journal on ICT
Vol. 8, No. 2, Dec 2022. pp. 22-39
doi: 10.21108/ijoict.v8i2.678

Estimation Rainfall Data in Malang Raya using Ordinary Kriging Method with Jackknife Technique

Novia Nur Rohma [1]*

[1]*STAI Al-Yasini Pasuruan*

[*] novia.nnr45@gmail.com

Abstract

Malang Raya is located between several mountains so that it has very varied geographical conditions, so it is possible that there are non-sampled locations for rainfall data. In general, rainfall in the area around the rain posts cannot be known with certainty because measurements are not made at every location. Ordinary kriging utilizes spatial values at sampled locations and variograms that show correlations between spatial points to predict values at other unsampled locations, where the predicted value depends on their proximity to the sampled locations. The jackknife method can solve the parameter estimation problem with a good degree of accuracy without regard to distribution assumptions. Kusumawardani conducted measurement simulations on the ordinary kriging method with the jackknife technique based on spherical and exponential variograms. Hardiansyah performed accuracy calculations on the ordinary kriging method with the jackknife technique based on exponential variograms. So this research was conducted on data with an unknown distribution assumption based on the ordinary kriging method approach with the jackknife technique based on spherical, exponential and Gaussian variograms. The result, the best exponential semivariogram in April, May, June, July, August, September, November and December. Best spherical semivariogram in February, March, May, June and October. Best gaussian semivariogram in January, May and June. Based on the RMSE value, jackknife kriging is good for interpolating normally and abnormally distributed data.

Keywords: kriging; Ordinary kriging ; Jackknife Kriging; semivariogram; rainfall

I. INTRODUCTION

Data presented in the geographical position of an object, related to its location, shape and relationship in the earth's space is called spatial data. Geostatistics is a science that focuses on spatial data. The purpose of geostatistical analysis is to predict the spatially dispersed subsets of the measurement results so that interpolation can be performed on the data. In geostatistics there is an estimation method for handling variables that have varying values with changing locations or places which are called regionalized variables. The estimation method used to handle regionalized variables is called the kriging method. The kriging method was first introduced by Daniel Krige, a South African mining engineer around 1950 [3].

Ordinary kriging (classical kriging) is a method of calculating sample point estimation values and is the simplest kriging method. Ordinary kriging utilizes spatial values at sampled locations and variograms that show correlations between spatial points to predict values at other unsampled locations, where the predicted value depends on their proximity to the sampled locations. In Ordinary Kriging, estimating a variable value at a certain point is done by observing similar data in other areas. The weight in the ordinary kriging method is influenced by the variogram model, so the accuracy in selecting the

Received on 20 Nov 2022. Revised on 28 Nov 2022. Accepted and Published on 28 Dec 2022.

variogram model will provide a good estimate of the kriging method [4]. The variogram is a graph of variance versus distance (lag), and half of the variogram quantity is called a semivariogram. The semivariogram can be used to measure spatial correlation in the form of the variance of different observations at a location. There are several semivariogram models, the semivariograms that are often used are the spherical, exponential, and Gaussian semivariogram models [5].

Jackknife is a nonparametric and resampling technique that aims to determine the bias, standard error and confidence interval estimates of population parameters, such as: mean, median, proportion, correlation coefficient and regression, without always paying attention to distributional assumptions. In 1949, Quenouille had introduced the jackknife method for estimating the bias of an estimator by removing an observation from the original sample. The sample obtained is used to calculate the value of the estimator. The jackknife method can also solve the problem of parameter estimation with a good degree of accuracy without regard to distribution assumptions [6].

The Greater Malang area is located between several mountains so that it has very varied geographical conditions, so it is possible that there are locations that are not sampled for rainfall data. In general, rainfall in the area around the rain posts cannot be known with certainty because measurements are not made at every location.

II. LITERATURE REVIEW

A. *Geostatistics*

Geostatistics is a statistical method used to see the relationship between variables measured at a certain point with the same variable measured at a point with a certain distance from the first point (spatial data) and used to estimate parameters in places where the data is unknown [7].

B. *Spatial Data*

Spatial data is data that has georeferenced data with various attribute data in various spatial units [8]. There is a dependency between one location and another, so that the spatial data is dependent data because the data is collected from different locations and indicates the dependence between data measurements and location. Dependence will decrease if the locations between observations are more spread out [3]. The geostatistical method is optimal or has a high accuracy value when applied to normally distributed and stationary data [5].

1) *Kolmogorov Smirnov Normality Test*

The Kolmogorov Smirnov test is a test used to decide if there are samples from a population that are spread over a certain distribution. This test will compare the cumulative distribution of the sample with a certain cumulative distribution, in this case the normal cumulative distribution

2) *Stationarity*

According to Suprajitno [9], a spatial data is said to be stationary if the data does not contain a trend. Meanwhile, a spatial data is said to be non-stationary if the spatial data contains a trend, namely where the variables in the spatial data form a curve. Spatial data stationarity can be determined using plots as shown in Fig. 1.

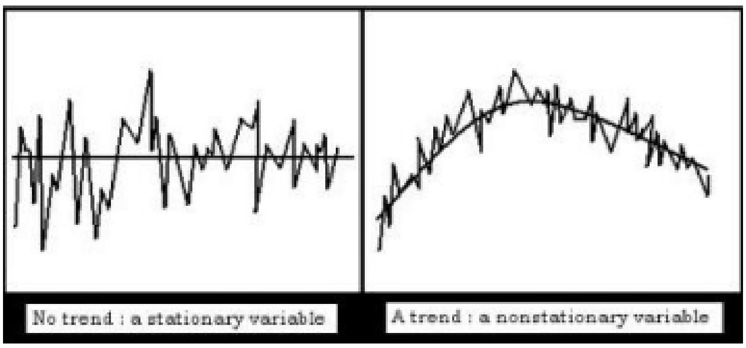


Fig. 1. An example of a Spatial Data Stationarity Plot [9]

C. Kriging Method

Kriging is a geostatistical method that uses known values and semivariograms as weights to predict values at other locations that have not been measured. The predicted value of the kriging method is not the same as the original data, but varies and depends on the proximity of the location of the original data [10].

Kriging provides an estimate of the unknown value at each unsampled point and the error variance. Several steps in the kriging method are (1) analyzing data samples statistically, (2) creating variogram models, (3) creating autocorrelation models and covariance functions to estimate spatial autocorrelation values, (4) making interpolation results by predicting values at location whose value is unknown, and (5) analyze the variance value [11].

Kriging estimator $\hat{Z}$(s) from $Z(s)$ can be written as [5]:

$$\hat{Z}(s) - m(s) = \sum_{i=1}^{n} \lambda_i \left[Z(si) - m(si)\right] \tag{1}$$

where:
s, s_i : location for the estimate and one of the locations of the adjacent data
$m(s)$: expected value of $Z(s)$
$m(s_i)$: expected value of $Z(s_i)$
$Z(si)$: observed value at the point s_i
$\hat{Z}(s)$: kriging estimator of $Z(si)$
λ_i : weight $Z(s_i)$ for location estimation s
n : the number of sample data used for estimation.

The application of the kriging method is carried out with the assumption of stationarity in the average (μ) and variance (σ^2), so that if the stationarity assumption is violated, kriging will produce a less precise predictive value. The purpose of kriging is to determine the value of the weighting coefficient λ_i which minimizes the variance in the stated estimator:

$$\sigma^2(s) = var\left\{\hat{Z}(s) - Z(s)\right\} \tag{2}$$

where:

$\sigma^2(s)$: estimator variance $\hat{Z}(s)$
$\hat{Z}(s)$: kriging estimator of $Z(si)$
$Z(s)$: observed value at the point s_i

with the estimate at each location being the difference in the true value of the estimator value $\hat{Z}$(s) with value $Z(s)$ defined:

$$\sigma^2(s) = Var\left[\sum_{i=1}^{n} \lambda_i Z(s_i) - Z(s)\right] \tag{3}$$

To make predictions with the kriging interpolation method, the two main things to do are to know the pattern of dependency relationships and to make predictions. To do this, kriging goes through two stages. The first stage is to create a variogram and covariance function to estimate the value of the relationship statistically (called spatial autocorrelation) which depends on the autocorrelation model formed. The second stage is predicting the value at a location whose value is not yet known.

D. Variogram

According to Wackernagel [12], a variogram is a method of analyzing spatial data diversity based on distance measurements. Variogram analysis performs calculations at a number of locations and looks at the relationships between observations at various locations. The variogram is used to determine the distance at which the observed data values become independent or uncorrelated. The hypothesis used to determine the variogram is based on intrinsic stationarity. According to Fischer and Getis [13], the average value will change with the difference in the area, then the variance value will also change with the wider area of observation. Thus, the covariance value cannot be obtained because the average value of μ is not known. Even though the average value is not constant, there will still be differences at each change in the lag distance so that the expected value will be 0.

The equation of the variogram can be explained in the following equation:

$$\begin{aligned} 2\gamma(h) &= var[Z(s+h) - Z(s)] \\ 2\gamma(h) &= 2\sigma^2 - 2C(h) \\ \gamma(h) &= C(0) - C(h) \end{aligned} \tag{4}$$

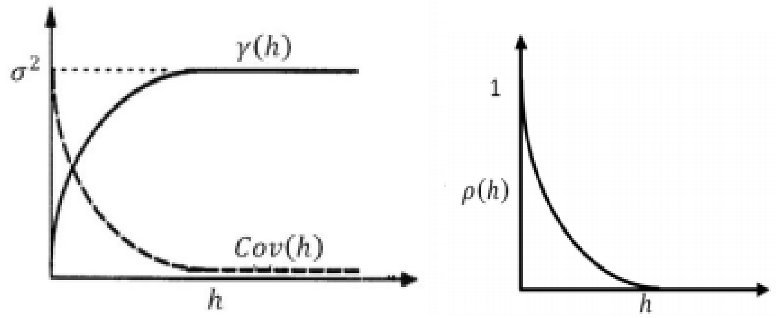
Where the covariance value meets the assumption of second-order stationarity. Then $C(0)$ is the covariance value at or the covariance value at a distance of 0 which is commonly called the variance. The equation of the variance is explained as follows:

$$\begin{aligned} C(0) &= E[\{Z(s) - \mu\}\{Z(s+0) - \mu\}] \\ C(0) &= \sigma^2 \end{aligned} \tag{5}$$

From the information in equations (4) and (5), the autocovariance value can be calculated. The value of autocovariance depends on where Z it is measured. So to calculate the attachment between locations, the correlation value or correlation function can be calculated using the covariogram which is defined as:

$$Corr\big(Z(s), Z(s+h)\big) = \frac{C(s)}{C(0)} = \rho(s) \tag{6}$$

Based on the equation (10), $\gamma(h)$ is the semivariogram value at the h-th lag. The relationship of the correlogram, covariance and semivariogram is that the form of the correlogram will be similar to the shape of the curve of the covariance function and the shape of the semivariogram curve will look like the shape of the inverted covariant curve [5] as shown in Fig. 2 as follows:



(a) (b)

Fig. 2. (a) Relationship between Covariance and Semivariogram, (b) Correlogram

Covariance and semivariogram have an inverse relationship. If the covariance starts from the maximum σ^2 on $h = 0$ descend towards 0, while the semivariogram starts from 0 and reaches a maximum σ^2. In the chorelogram, it can be seen that if the distance increases, the correlation becomes weaker.

E. Semivariogram Modeling

One tool that indicates the existence of spatial autocorrelation between data from a variable and functions as a measure of variance is called a semivariogram. Semivariograms are usually described in graphical form based on mathematical calculations [15]. Semivariogram is divided into 2, is: experimental Semivariogram and theoretical Semivariograms

I. Experimental Semivariogram

Experimental semivariogram is also called cloud semivariogram. The semivariogram is calculated from the measured data and then plotted as a function of distance [16]. For example $Z(s_i)$ is the measured value at the location i, whereas $s_i = (x_i, y_i)$ is a vector containing spatial coordinates x, y dan $h = s_1 - s_2$ is the distance vector between the points s_1 to s_2. So that the experimental semivariogram is formulated as:

$$\gamma_{ij} = 0.5 \ \mathrm{E}[Z(s_i) - Z(s_i)]^2 \quad (7)$$

For all possible distance pairs $\{(s_i, s_j)\}$ for $i, j = 1, 2, 3, .., n$, then plotted as a function of distance $h = s_i - s_j$ formulated:

$$|h| = |s_i - s_j| = [(x_i - x_j)^2 + (y_i - y_j)^2]^{\frac{1}{2}} \quad (8)$$

where:
s_i : location to -i
s_j : location to -j
x_i : longitude location to -i
x_j : longitude location to -j
y_i : latitude location to -i
y_j : latitude location to -j

The difficulty in the experimental semivariogram is to see patterns when the calculation involves up to thousands of points therefore, to overcome the difficulties a grouping stage is carried out on y_{ij} based on the similarity of distances is called the binning process. The grouping process aims to facilitate interpretation [17].

The variogram model is defined as the variance of the difference between two observations at two different locations, while the semivariogram is half of the variogram value [18]. The estimated experimental variogram at a distance h can be written as follows:

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(s_i + h) - Z(s_i)]^2 \quad (9)$$

The experimental semivariogram is written as follows:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i + h) - Z(s_i)]^2 \quad (10)$$

where:
$2\hat{\gamma}$: variogram value with a distance of h
$\hat{\gamma}(h)$: semivariogram value with distance h
$Z(s_i)$: observed value at the point s_i
$Z(s_i + h)$: observed value at the point $s_i + h$
$N(h)$: the number of pairs of points that have a distance h.

II. Theoretical Semivariograms

The experimental semivariogram has an irregular shape, making it difficult to interpret and does not directly use in interpretation. Furthermore, the experimental semivariogram value will be matched with the theoretical semivariogram model for use in estimation in order to obtain a smooth and continuous covariance pattern so that it can be used to derive the covariance matrix in kriging calculations [14].

In calculating the theoretical semivariogram required parameters to be used. The parameters used to find the theoretical semivariogram value are as follows:

a. *Nugget Effect* (C)

The discontinuity at the center of the variogram with respect to the vertical line that jumps from 0 at the center to the variogram value at the smallest distance separation is called the nugget effect [4]. According to Cressie [3], the Nugget effect is the estimation of a semivariogram at a distance of about zero where the phenomenon is discontinuous around the starting point of the semivariogram. Reflects sampling error and analytical error.

b. *Sill* $(C_0 + C)$

Sill is the stable period of a variogram that reaches the range [7]. Sill describes where the variogram becomes a flat area, that is, the variance does not experience an increase. The sill consists of two parts, namely the nugget effect (the intersection of the graph with the y-axis) and the partial sill (the sill that has been reduced by the nugget effect) [3].

c. *Range* (a)

Sill is the stable period of a variogram that reaches the range [7]. Sill describes where the variogram becomes a flat area, that is, the variance does not experience an increase. The sill consists of two parts, namely the nugget effect (the intersection of the graph with the y-axis) and the partial sill (the sill that has been reduced by the nugget effect) $a\sqrt{3}$. The distance at which the variogram reaches the sill value [3]. The theoretical semivariogram image can be seen in Fig 3.

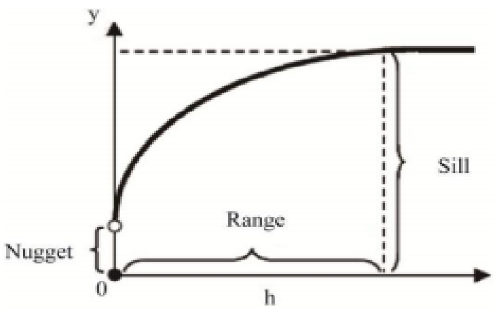


Fig. 3. Characteristics of the Semivariogram

Theoretical semivariogram models that are often used [5]:

- Spherical Model

$$\gamma(h) = \begin{cases} C_0 + C\left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right) \text{for h } \leq \text{ a} \\ C_0 + C \qquad\qquad others \end{cases} \quad (11)$$

where:

h	: distance between samples
$C_0 + C$	: *sill*
a	: *range*

- Exponential Model

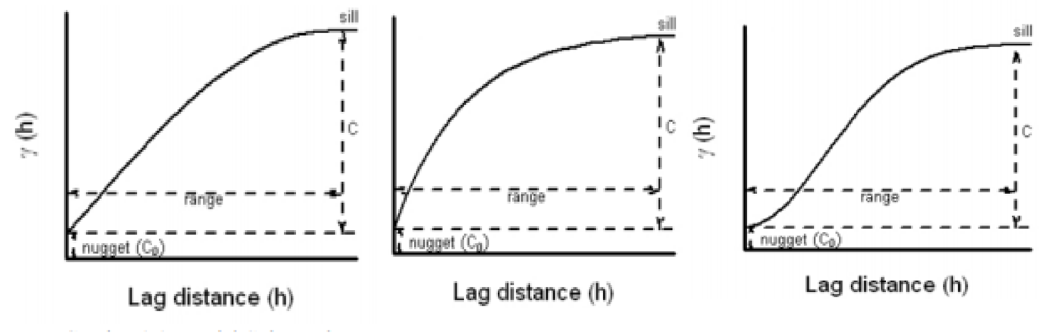
The exponential model semivariogram has a very steep increase. The semivariogram form of the exponential model is formulated as follows:

$$\gamma(h) = \begin{cases} c_0 + c\left(1 - \exp\left(\frac{-3h}{a}\right)\right) & h \neq 0 \\ 0 & h = 0 \end{cases} \quad (12)$$

- Gaussian Model

The Gaussian Model Semivariogram is the quadratic form of the exponential. The semivariogram form of the Gaussian model is formulated as follows:

$$\gamma(h) = \begin{cases} c_0 + c\left(1 - \exp\left(\frac{-3h^2}{a^2}\right)\right) & h \neq 0 \\ 0 & h = 0 \end{cases} \quad (13)$$



(a) (b) (c)

Fig 4. (a) Spherical Model (b) Exponensial Model (c) Gausian Model

The WLS method is the most widely used method because it pays more attention to distances that have a larger number of measurement point pairs, and gives greater weighting to smaller theoretical semivariogram values. Estimation of theoretical semivariogram parameters using the WLS method is carried out by determining the parameters $\boldsymbol{a}$ and $\boldsymbol{\sigma^2}$ so as to minimize the equation [3].

$$\sum_{j=1}^{K}\left|N\big(h(j)\big)\right|\left\{\frac{\hat{\gamma}\big(h(j)\big)}{\gamma\big(h(j);\theta\big)}-1\right\}^2 \quad (14)$$

where:

$\left|\boldsymbol{N}\big(\boldsymbol{h}(\boldsymbol{j})\big)\right|$: the number of pairs of measurement points in lag -$\boldsymbol{j}$, j=1,2,3,..., $\boldsymbol{K}$ where $\boldsymbol{K}$ is the amount of lag

$\hat{\boldsymbol{\gamma}}\big(\boldsymbol{h}(\boldsymbol{j})\big)$: experimental semivariogram values at lag -$\boldsymbol{j}$

$\gamma(h(j):\theta)$: standard semivariogram value at lag -j lag with parameters $\theta = \{a, \sigma^2\}$

F. Ordinary Kriging Method

Ordinary kriging is the simplest kriging method found in geostatistics. Ordinary kriging assumes the population mean is unknown and constant. In ordinary kriging $Z(s)$ is the average of $Z(s) = E\big(Z(s)\big)$, where $E\big(Z(s)\big) = \mu$.

Cressie [3] explains that ordinary kriging relates to spatial predictions with two assumptions:

Model Assumptions:

$$Z(s) = \mu + e(s),\ \ e(s) \in D \subset R^2,\ \mu \text{ unknown} \quad (15)$$

Predictive assumptions

$$\hat{Z}(s) = \sum_{i=1}^{n} \lambda_i\, Z(s_i) \text{ with } \sum_{i=1}^{n} \lambda_i = 1 \quad (16)$$

where:

$\hat{Z}(s_0)$: variable predictive value s

$Z(s)$: actual value of the variable s at location -i

μ : expectations from $Z(s)$

$e(s)$: error value of $Z(s)$

λ_i : weights that determine the size of distances and points

D : random set in R^2

R^2 : real number

n : the number of sample data used for estimation

As for the properties of ordinary kriging, one of the goals of kriging is to produce the best linear unbiased estimator (BLUE). The following is an explanation of the ordinary nature of kriging*:*

1) Linier

Based on the equation (22) $\hat{Z}(s)$ has a linear estimator because the linear function of $Z(s)$. If there are n measurements at a location $1, 2, \ldots, n$ then it can be stated $Z(s_1), Z(s_2), \ldots, Z(s_n)$. Then make an estimate $Z(s)$at the non−sampled locations stated in $Z(s_0)$ based on the sampled data. So that the following equation will be produced :

$$\hat{Z}(s_0) = \sum_{i=1}^{n} \lambda_i Z(s_i) \tag{17}$$

Based on the equation (17) maka $\hat{Z}(s_0)$ can be said to be a linear estimator because the linear function of $Z(s_i)$.

2) Unbiased

Unfamiliarity with kriging by carrying out the expected value will get an error at a certain location which has an expected value equal to zero, the results obtained are as follows:

$$E\left(\hat{e}(s)\right) = E\left(Z(s) - \hat{Z}(s)\right)$$

$$= E(Z(s) - \sum_{i=1}^{n} \lambda_i Z(s_i))$$

$$= E\left(Z(s)\right) - \sum_{i=1}^{n} \lambda_i \left(Z(s_i)\right) \tag{18}$$

Because of value $E\left(\hat{e}(s)\right) = 0$ then you will get it:

$$E\left(\hat{e}(s)\right) = E\left(Z(s)\right) - \sum_{i=1}^{n} \lambda_i E\left(Z(s_i)\right) = 0$$

$$E\left(Z(s)\right) = \sum_{i=1}^{n} \lambda_i E\left(Z(s_i)\right)$$

$$\mu = \sum_{i=1}^{n} \lambda_i \mu$$

$$\sum_{i=1}^{n} \lambda_i = 1 \tag{19}$$

So that the ordinary kriging method produces an estimator that is not biased $\sum_{i=1}^{n} \lambda_i = 1$.

3) Efficient (Minimizing the error range)

The Ordinary kriging method has unbiased properties by minimizing the error range. Assuming that

$$var\left(Z(s_0)\right) = \sigma^2$$

G. The Jackknife method

The Jackknife method was first introduced by Quenouille in 1949 which is a nonparametric and resampling method that aims to estimate standard errors and bias values. Jackknife is a method used to estimate an unknown population distribution with the empirical distribution obtained from the re-sampling process. The principle of the Jackknife method is to remove one piece of data and repeat it as many times as there are samples. From the repetition process, bias and variance can be calculated [6].

For example $X = (x_1, x_2, \ldots, x_n)$ dan $\theta_i = t(x_i)$ is an estimator built from data X to estimate the parameter θ, then from the set $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$ is the jackknife standard error estimator [6]:

$$S_{jack} = \left[\frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}_{(i)} - \hat{\theta}_{(.)}\right)^2\right]^{\frac{1}{2}} \tag{20}$$

where:

$\hat{\theta}_{(.)} \quad = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}$

$\hat{\theta}_{(i)}$ = test *jackknife* ke-i dari $\hat{\theta}$

n = sample size .

H. Ordinary Kriging Algorithm with Jackknife

The ordinary kriging estimator when all sample data is used is

$$\hat{Z}(s) = \sum_{i=1}^{n} \lambda_i Z(s_i)$$

where:

$\sum_{i=1}^{n} \lambda_i = 1$ (unbiased condition for ordinary kriging, on equation 25)

λ_i = Ordinary kriging weights with samples i

Ordinary kriging estimator when there is a sample j unused, with $j = 1,2,\dots,n$ is

$$\hat{Z}(s) - j = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij} Z(s_i) \quad (21)$$

where:

$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij} = 1$ (unbiased conditions for jackknife kriging)

λ_{ij} = ordinary kriging weight i when sample j is not used

λ_{ij} value 0 when $i = j$

The unbiased condition of jackknife kriging can be explained as follows:

$$\begin{aligned} E\left(\hat{e}(s)\right) &= E\left(Z(s) - \hat{Z}(s) - j\right) \\ &= E\left(Z(s) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij} Z(s_i)\right) \\ &= E\left(Z(s)\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij}\, E\left(Z(s_i)\right) \end{aligned} \quad (22)$$

Because of value $E\left(\hat{e}(s)\right) = 0$ then you will get it:

$$\begin{aligned} E\left(\hat{e}(s)\right) &= E\left(Z(s)\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij}\, E\left(Z(s_i)\right) = 0 \\ E\left(Z(s)\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij}\, E\left(Z(s_i)\right) \\ \mu &= \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij}\, \mu \\ \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij} &= 1 \end{aligned} \quad (23)$$

So that the ordinary kriging jackknife method produces an estimator that is not biased $\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij} = 1$

The difference between ordinary kriging and ordinary kriging with a jackknife estimator lies in the weights used. In ordinary kriging the weights obtained come from all data, while the weights in ordinary kriging with jackknife come from data issued one by one as much as n times (the amount of data) so that more than one weight is obtained. For more details, see Table I and Table II.

TABLE I
ORDINARY KRIGING WEIGHTS

i	**1**	**2**	**3**	$\cdots$	**n**
λ_i	λ_1	λ_2	λ_3	$\cdots$	λ_n

From Table II it can be seen that the ordinary kriging weights are in the form of scalar data where λ_i show weight i obtained from all data.

TABLE II
ORDINARY WEIGHT OF KRIGING WITH THE JACKKNIFE TECHNIQUE

λ_{ij}		j					
		1	2	3	4	...	n
	1	0	λ_{12}	λ_{13}	λ_{14}		λ_{1n}
	2	λ_{21}	0				
i	3	λ_{31}		0			
	4	λ_{41}			0		
	$\vdots$	$\vdots$				$\ddots$	
	n	λ_{n1}					0

From Table II it can be seen that the ordinary weights of kriging with the jackknife technique are in the form of vector data where λ_{ij} show weight i obtained when j not used. For example λ_{12} shows weight to 1 when data 2 is not used. Using the jackknife method will eliminate the sample z_j, then the jackknife estimator[19] is

$$Z(s)_j = n\hat{Z}(s) - (n-1)\hat{Z}(s)_{(.)} \quad (24)$$

where:
$\hat{Z}(s)_{(.)} = n^{-1}\sum \hat{Z}(s)_{-j}$

Tukey [20] states that to get a measure of the accuracy of the Jackknife estimator, the following pseudo-values are used

$$Zp_{-j} = n\hat{Z}(s) - (n-1)\hat{Z}(s)_{-j} \quad (25)$$

The jackknife estimator is a pseudo-value average, i.e.:

$$Z(s)_j = n^{-1}\sum_j Zp_{-j} \quad (26)$$

From equations (26) and (25), the jackknife kriging estimator is obtained as follows [21]:

$$\begin{aligned} Z(s)_j &= n^{-1}\sum_j Zp_{-j} \\ &= n^{-1}\sum_j \left\{ n\hat{Z}(s) - (n-1)\hat{Z}(s)_{-j} \right\} \\ &= \frac{n^2}{n}\hat{Z}(s) - \frac{(n-1)}{n}\sum_j \hat{Z}(s)_{-j} \\ &= n\sum_{i=1}^{n} \lambda_i\, z(s_i) - \frac{(n-1)}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_{ij}\, z(s_i) \end{aligned}$$

$$= \sum_{i=1}^{n} \left\{ n\lambda_i - \frac{(n-1)}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij} \right\} z(s_i)$$
$$= \sum_{i=1}^{n} \alpha_i \, z(s_i) \tag{27}$$

where:

$\alpha_i = n\lambda_i - \frac{(n-1)}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij}$

α_i = jackknife kriging weights with samples i

The total weight of the jackknife kriging is

$$\sum_i \alpha_i = n \sum_{i=1}^{n} \lambda_i - \frac{(n-1)}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_{ij}$$
$$= n - \left(\frac{(n-1)}{n} \right) n$$
$$= n - (n-1)$$
$$= n - n + 1 = 1$$

I. Kriging Model Goodness Criteria

The purpose of the Root Mean Square Error (RMSE) is to compare the accuracy between two or more models. The smaller the RMSE value produced by a model, the more accurate the model will be. RMSE is formulated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum\nolimits_{j=1}^{n} \left[Z(s_i) - \hat{Z}(s_i) \right]^2} \tag{28}$$

where:

$Z(s_i)$ and $\hat{Z}$ (s_i) denote the actual and the predicted value of the variable s at location -i, respectively
n is number of observations. The kriging method is said to have good estimation accuracy if it has a small RMSE value [7].

III. RESEARCH METHOD

This study uses secondary data on monthly rainfall in Malang Raya in 2016 in units of mm/month. Secondary data was obtained from UPT Water Resources Management Malang Raya, based on data from UPT Water Resources Management Malang Raya it is known that there are 44 rainfall posts which are spread out, namely 3 rainfall posts in Malang city, 7 rainfall posts in Batu city and 34 rainfall posts in Malang Regency

A. Data Analysis

The analysis carried out in this study is as follows:
1. Prepare digitized maps in SHP format.
2. Explain the stationarity of the data by looking at the descriptive results through plotting the data according to location coordinates (latitude and longitude). If the data does not have an up or down trend of spatial data with latitude or longitude, in other words random, then the ordinary kriging method can be used. However, if there is a pattern trend, the ordinary kriging method cannot be used. If there is a trend, it is necessary to carry out a transformation.
3. Detect the normality of the data with the Kolmogorov Smirnov test.
4. Analyzing data with the ordinary kriging spatial interpolation method
   a. Forming an experimental semivariogram/ semivariogram cloud

b. Form a theoretical semivariogram using three models, namely spherical, exponential, and gaussian
c. Analyzing the accuracy of spherical, exponential, and gaussian variogam models using RMSE
5. Make a map from the best interpolation results

IV. RESULTS AND DISCUSSION

The presentation of results should be simple and straightforward. This section reports the most important findings, including results of statistical analysis as appropriate and comparisons to other research results. Results given in figures should not be repeated in tables. This is where Authors should explain in words what he/she/they discovered in the research. It should be clearly laid out and in a logical sequence. This section should be supported suitable references [6].

A. Stationarity

The ordinary kriging method requires observational data to be stationary, that is, observational data must have a constant average and variance at each location or observational data does not contain a trend. According to [9], a spatial data is said to be stationary if the data does not contain a trend. If there is a trend, the interpolation results of the ordinary kriging method will be inaccurate. Trend checks are carried out by forming plots of annual rainfall data in the Greater Malang Region in the direction of latitude or easting and longitude or northing. To form a plot, coordinates are needed for each rain post point in the Greater Malang Region. Plots for latitude (easting) coordinates are shown in Fig. 5, while plots for longitude (northing) coordinates are shown in Fig. 6.

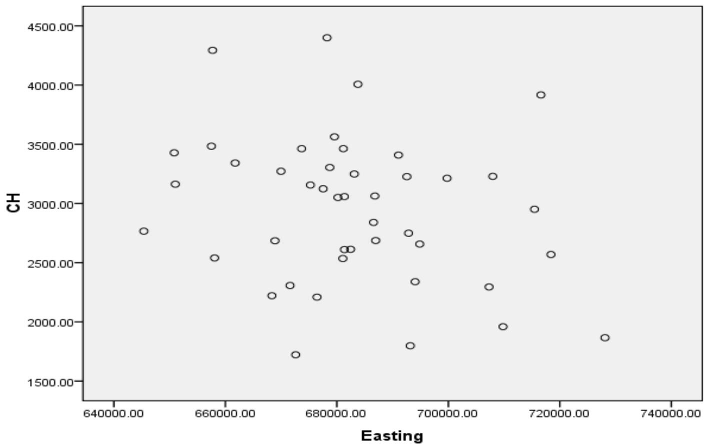


Fig. 5. Plot of Rainfall Value in Malang Raya Region Against Easting

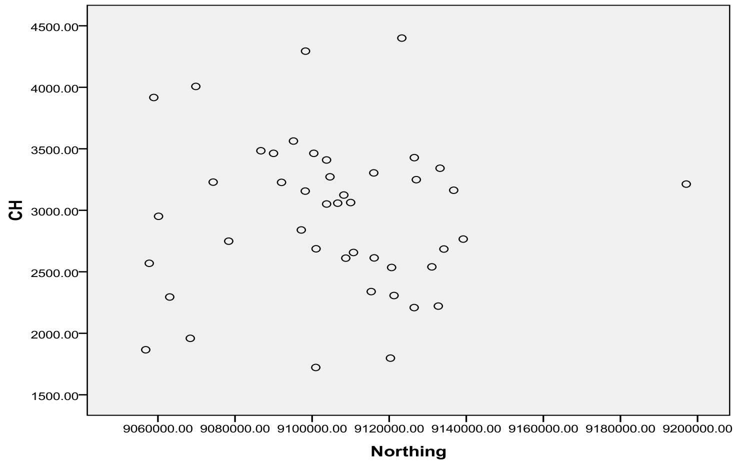


Fig. 6. Plot of Rainfall Value in Greater Malang Region Against Northing

Judging from Fig. 5 and Fig. 6 it is known that there is no trend of either increasing or decreasing rainfall values in the Greater Malang Region for easting and northing, so it can be concluded that there is no trend in the rainfall data in the Malang Raya Region in 2016.

B. Normality Test

Deviations from normality and stationarity can cause problems with poor accuracy, so the best first step is to look at histograms or data plots to check the normality and distribution of data to check trends [5]. The normality test was carried out by looking at the histogram and the Kolmogorov Smirnov test. Kolmogorov Smirnov test results from 44 rainfall posts from January to December 2016 in Malang Raya are shown in Table III.

TABLE III
KOLMOGOROV-SMIRNOV TEST

Month	Significance Value	Information
January	0.308	Normal
February	0.952	Normal
March	0.414	Normal
April	0.634	Normal
May	0.276	Normal
June	0.611	Normal
July	**0.031**	Abnormal
August	0.726	Normal
September	0.576	Normal
October	0.388	Normal
November	0.848	Normal
December	0.439	Normal

In the rainfall data from January to December 2016, it is known that all months are normally distributed except July. According to Bohling [5], the geostatistical method will obtain optimal values or have high accuracy values when applied to normally distributed and stationary data. Then it will be interpolated using ordinary kriging.

C. Ordinary Kriging Interpolation

Spatial interpolation is a method used to predict unknown values based on values obtained from an observation [18]. The first step in carrying out ordinary kriging interpolation is to construct an experimental semivariogram. The experimental semivariogram was calculated from the measured data and then plotted as a function of distance [16]. The results of empirical semivariogram calculations are displayed in graphical form which is often called a semivariogram cloud. The form of the semivariogram cloud of rainfall data for July 2016 is shown in Fig. 7.

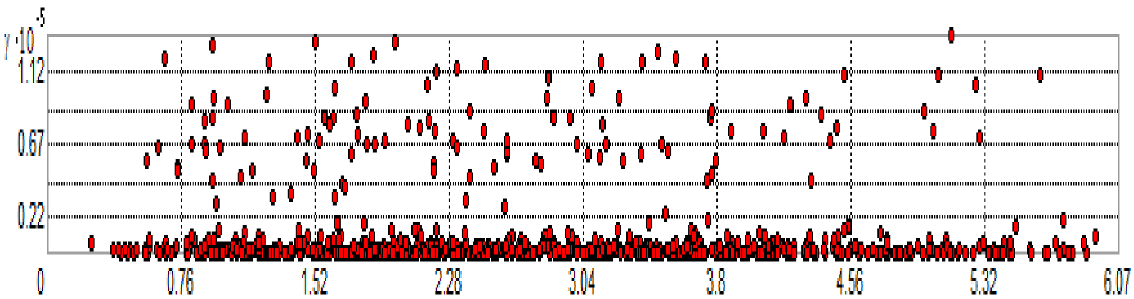


Fig. 7. Plot Result of the Semivariogram Cloud

From the results of the semivariogram cloud in Fig. 7, it is difficult to see patterns in determining the appropriate semivariogram model because it has many observation points, so it is necessary to group it based on distance (binning) to make semivariogram modeling easier by entering data into a range of intervals. The grouping process aims to facilitate interpretation [17]. Semivariogram binning results on rainfall data in Malang Malang Raya in July 2016 is shown in Fig. 8.

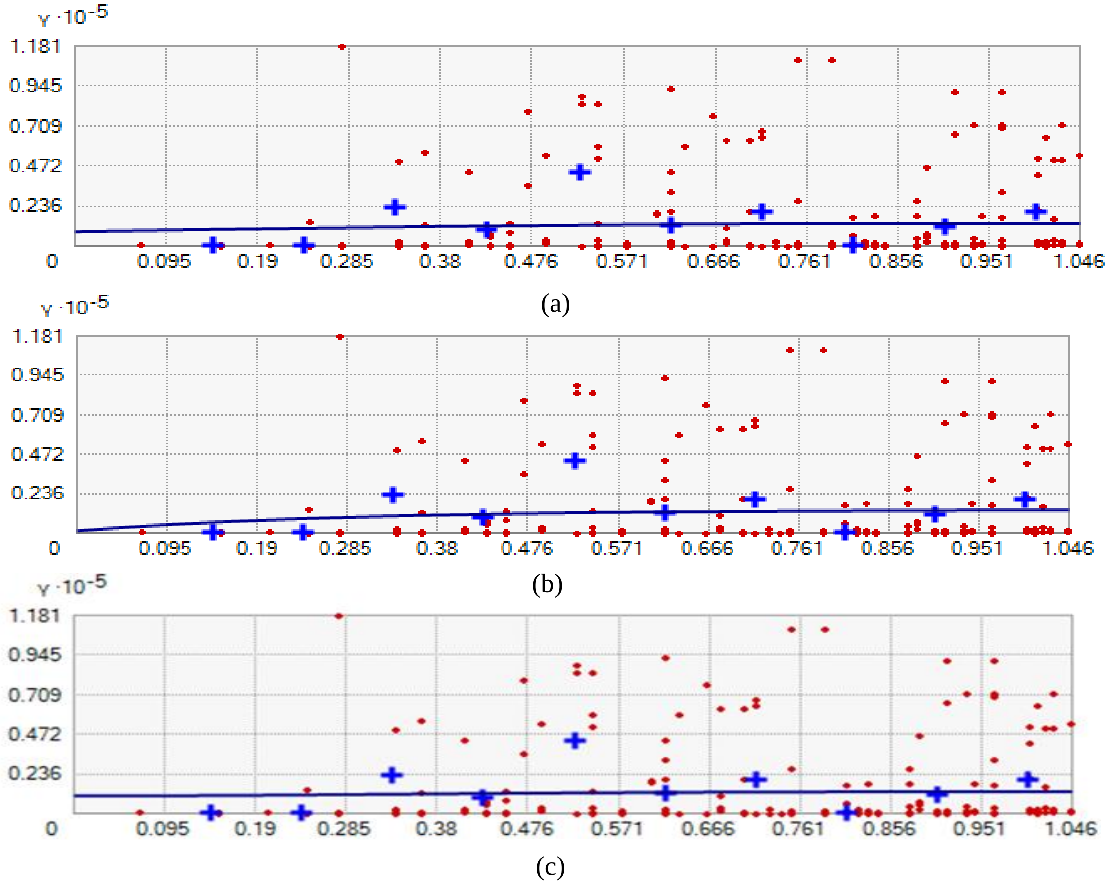


Fig. 8. Rainfall Data Binning Semivariogram in Malang Raya in July 2016 (a) Spherical Semivariogram (b) Exponential Semivariogram (c) Gaussian Semivariogram

Fig. 8. is a semivariogram that has been binned. The red dot is the binned point for each lag, while the blue dot shows the average binned for each lag. After binning, semivariogram modeling is carried out by means of structural analysis, namely by matching experimental semivariograms with theoretical semivariograms for use in estimation in order to obtain smooth and continuous covariance patterns so that they can be used to derive covariance matrices in kriging calculations [15]. The theoretical semivariogram model used is spherical, exponential and gaussian [5]. The following is a semivariogram model in July 2016.

1) *Spherical Model*

$$\gamma(h) = \begin{cases} 13172.8\left(1.5\left(\dfrac{h}{8373.171}\right) - 0.5\left(\dfrac{h}{8373.171}\right)^3\right), for\ \text{h} \leq 8373.171 \\ 13172.8 \qquad Otherwise \end{cases}$$

2) *Exponential Model*

$$\gamma(h) = \begin{cases} 129506.1\left(1 - \exp\left(\dfrac{-3h}{8373.171}\right)\right) & h \neq 0 \\ 0 & h = 0 \end{cases}$$

3) *Gaussian Model*

$$\gamma(h) = \begin{cases} 15014.95\left(1 - \exp\left(\dfrac{-3h^2}{8373.171\ ^2}\right)\right) & h \neq 0 \\ 0 & h = 0 \end{cases}$$

The semivariogram spherical model has a sill estimate of 13172.8 m which indicates that the semivariogram value has reached a stable value, namely the maximum height where the semivariogram value has no correlation between the data. The distance value when the semivarogram reaches the sill is 8373.171 m. After obtaining the semivariogram model, then the process of selecting the best model for the three semivariogram models is carried out based on the best model selection criteria.

D. Interpolation of Ordinary Kriging with the Jackknife Technique

The first step in carrying out spatial interpolation with the Ordinary Kriging method with the Jackknife technique is to delete the 1st data with the jackknife technique and then interpolate with ordinary kriging to obtain jackknife kriging 1 (JK 1) in each semivariogram model. The average is calculated so that an interpolated value will be obtained for each rain post. The overall results of JK interpolation using the exponential model for each rain post in July 2016 are presented in Table IV.

TABLE IV
JACKKNIFE KRIGING EACH RAIN POST

Rain Post	Predictions	Rain Post	Predictions
Bantur	105.574	Pujon	256.051
Blambangan	79.143	Sekar	79.785
Bululawang	76.069	Singosari	78.892
Clumprit	122.246	Sitiarjo	381.950
Dampit	72.977	Sumberpucung	105.631
Dau	65.345	Tajinan	71.378
Donomulyo	138.964	Tangkilsari	65.327
Gondanglegi	106.656	Tumpang	58.908
Jabung	35.188	Tumpukrenteng	98.062
Jombok	135.419	Turen	104.326
Kalipare	133.364	Wagir	63.876
Kantor_Cd	48.265	Wajak	129.148
KarangPloso	104.050	Ngaglik	64.497
KarangSuko	104.175	Ngujung	68.337
Kasembon	100.891	Pendem	72.466
Kdgrejo	115.477	Sbr_Gondo	80.255
Ngajum	47.000	Temas	68.270
Ngantang	206.156	TinjuMoyo	78.093
Pagak	140.165	Tlekung	51.577
Penarukan	59.838	Blimbing	102.487
Pohgajih	118.000	KedungKandang	66.325
Poncokusumo	59.460	Sukun	75.939

To determine the best semivariogram model, cross validation was carried out using the smallest RMSE value among the three models. The RMSE value is obtained from the difference between the observed rainfall data and the prediction results of each model.

Based on Table V, the smallest RMSE values of the exponential semivariogram are found in April, May, June, July, August, September, November and December 2016. The smallest RMSE values of the Spherical semivariogram are in the rainfall data for February, March, May, June and October 2016. The smallest RMSE value is the Gaussian semivariogram in the rainfall data for January, May and June 2016. To find out the variances of the three semivariogram models are the same or different, a statistical test is carried out. The test used is the Bartlet test. Bartlet test results from January to December 2016 are shown in Table VI.

TABLE V
RMSE VALUE OF ORDINARY KRIGING WITH THE JACKKNIFE TECHNIQUE (JK)

Month	Semivariogram Model	RMSE
January	Spherical	59.634
	Exponential	2.972
	Gaussian	**2.188**
February	Spherical	**101.094**
	Exponential	101.166
	Gaussian	116.344
March	Spherical	**83.631**
	Exponential	83.881
	Gaussian	98.231
April	Spherical	46.496
	Exponential	**2.463**
	Gaussian	78.411
May	Spherical	**1.718**
	Exponential	**1.718**
	Gaussian	**1.718**
June	Spherical	**9.235**
	Exponential	**9.235**
	Gaussian	**9.235**
July	Spherical	70.578
	Exponential	**56.825**
	Gaussian	89.311
August	Spherical	20.733
	Exponential	**4.186**
	Gaussian	30.756
September	Spherical	18.256
	Exponential	**1.052**
	Gaussian	59.862
October	Spherical	**73.657**
	Exponential	73.689
	Gaussian	84.708
November	Spherical	117.033
	Exponential	**115.922**
	Gaussian	135.186
December	Spherical	68.948
	Exponential	**62.856**
	Gaussian	80.358

Note: Numbers in bold print indicate the lowest RMSE

Based on Table VI. with a significance level of 5% it is known that March, May, June, September, October and December have values *p-value* > α so H_0 received, so that the variances of the three semivariogram models are the same, which means that the spherichal, exponential and gaussian semivariograms can be used in March, May, June, September, October and December. However, the researchers chose the semivariogram with the smallest RMSE, namely the exponential semivariogram in March, May, June, September, December and the spherical semivariogram in October. In January, February, April, July, August and November *p-value* < α so H_0 received, so that the spherichal, exponential and gaussian semivariogram variances are different. Based on these results, the RMSE can determine the best semivariogram based on the smallest value.

TABLE VI
P-VALUE RESULTS OF THE JACKKNIFE KRIGING BARTLET TEST MODEL SEMIVARIOGRAM

Month	P-Value
January	1 x 10^{-11}
February	1 x 10^{-13}
March	0.918
April	0.010
May	0.997
June	0.998
July	0.020
August	0.007
September	0.290
October	0.290
November	0.012
December	0.534

Fig. 9. show the interpolation results are grouped into 10 classes and each class has a different color. The higher the rainfall interpolation results, the map will be orange. Vice versa, the lower the rainfall interpolation results, the map will be blue. According to BMKG (2013) low rainfall ranges from 0-100 mm, medium rainfall ranges from 101-300 mm, high rainfall ranges from 301-500 mm and very high rainfall is more than 501 mm. Based on the results of ordinary kriging interpolation using an exponential semivariogram, the area around the south of Malang Raya has moderate rainfall.

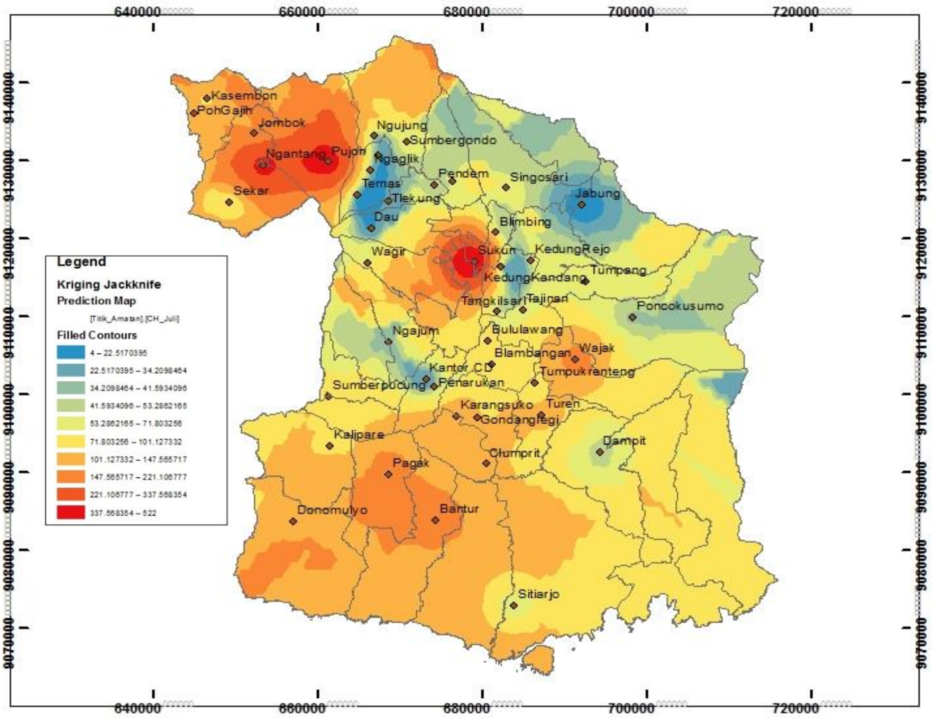


Fig. 9. Interpolation Map of Ordinary Kriging Using Jackknife Technique on Rainfall Data in Malang Raya in July 2016

V. CONCLUSION

Based on the results of the study, it was found that the classification of the 2016 Malang Raya rainfall data was divided into 2 data, namely normally distributed data and abnormally distributed data. Normal distribution data is in January, February, March, April, May, June, August, September, October, November, December. While the data is not normally distributed in July. jackknife kriging has the best exponential semivariogram in April, May, June, July, August, September, November, and December. Best spherical semivariogram in February, March, May, June, and October. Best gaussian semivariogram in January, May, and June. For non-normally distributed data, a relatively small RMSE value is obtained, this indicates that jackknife kriging is good for interpolating non-normally distributed data. For further studies, researchers may be able to use more types of interpolation methods so they can compare the best methods for interpolating normally distributed and abnormally distributed data.

REFERENCES

[1] Kusumawardani, D. S. Simulasi Pengukuran Ketepatan Model Variogram Pada Metode Ordinary Kriging Dengan Teknik Jackknife. Jurnal Gaussian, 3(3), 333 – 342. 2014.

[2] Hardiansyah, J. Strategi Perhitungan Akurasi Pada Metode Ordinary Kriging Dengan Menggunakan Teknik Jackknife. Skripsi, ITB. Bogor. 2001.

[3] Cressie, N.A.C. Statistics For Spatial Data. John Wiley and Sons, Inc. New York . 1993

[4] Isaaks, E.H dan Srivastava, R.M.,. *Applied Geostatistics*. New York: Oxford University Press. 1989.

[5] Bohling, G. (2005). *Kriging*. http://people.ku.edu/~gbohling/cpe940. Tanggal akses 20 April 2016.

[6] Efron, B. dan Tibshirani, R.J. *An Introduction to The Bootstrap.* Chapman and Hall. New York. 1993.

[7] Oliver, M.A., dan Webster, R. *Basic Steps in Geostatistics: The Variogram and Kriging.* London: Springer. 2015.

[8] Nuarsa, I. W. *Menganalisa Data Spasial dengan Arcview GIS 3.3 untuk Pemula.* PT Elex Media Komputindo. Jakarta. 2004.

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

[10] Tatalovich, Z. 2005. *A Comparison of Thiessen Polygon, Kriging, and Spline Models of UV Exposure.* www.ucgis.org/summer2005/studentpapers/tatalovich.pdf. Tanggal Akses 5 Oktober 2016.

[11] Tarboton, K.C. *Kriging of Regional Properties in The Western San Joaquin Valley.* California. 1995.

[12] Wackernagel, H. 2003. *Multivariate Geostatistics : An Introduction with Applications*. Springer-Verlag Berlin Heidelberg. New York

[13] Fischer, M.M., dan Getis, A. *Handbook of Applied Spatial Analysis*. Springer. Jerman. 2010.

[14] Bailey, T. dan Gatrel, A. *Interactive Spatial Data Analysis*. Essex: Pearson Education Limited. 1995.

[15]Gandhi, V. 2010. *Semivariogram Modeling*. http://www-users.cs.umn.edu/~gandhi/course/CS870/g4_semivariogram_second_draft.pdf. Tanggal akses 25 April 2016.

[16] Gaetan, C. dan Guyon, X. *Spatial Statistics and Modeling*. New York: Springer Science+Business Media., LLC. 2010.

[17] Gribov, A., Krivoruchko, K., dan Ver Hoef, J. M. (2001). *Modelling the Semivariogram: New Approach, Methods Comparison and Case Study*. www.esri.com/sofware/arcgis/extensions/geostatistical/about/literature.html-19k-. Tanggal akses 2 September 2016.

[18] Pawitan, G. *Analysis of Agregated Spatial Social data*. Disertasi. Australia: University of Wollongong. 2001.

[19] Efron, B. 1982. *The Jackknife, the bootstrap and Other Resampling Plans*. SIAM. Philadelphian.

[20] Tukey, J. W., 1958. Bias and Confidence in Not-quite Large Samples. *Abstract, Ann. Mat. Statist*., Vol. 29, p.614.

[21] Adisoma, G. S. 1993. *The Aplication of The Jackknife in Geostatistical Resource Estimation: Robust Estimator and Its Measure of Uncertainty*. http://hdl.handle.net/10150/186547. Tanggal Akses 4 Agustus 2017.