# NONPARAMETRIC CORRELOGRAM TO IDENTIFY THE GEOGRAPHIC DISTANCE OF SPATIAL DEPENDENCE ON LAND PRICES

*Winsy* **WEKU**[1,2], *Henny* **PRAMOEDYO**[1], *Agus* **WIDODO**[1], *Rahma* **FITRIANI**[1]
[1]Brawijaya University, Malang, Indonesia; [2]Sam Ratulangi University, Manado, Indonesia

**Abstract:** The spatial autocorrelation measurement of land prices uses a covariance function to describe the spatial dependence and it can be identified as a geographic distance on the correlogram. The geographic distance of spatial dependence can state that land prices are interdependent to each other and scattered in the research area. Therefore, the purpose of this research is to define the geographic distance of spatial dependence on land prices using a nonparametric correlogram. A nonparametric approach to covariance functions using the composition of Bessel and Gaussian-type functions are adopted because they correspond to the positive definite characteristics. The cubic spline interpolation is used to refine the curve fitting, while the intersection between the nonparametric correlogram value C(h) against the horizontal axis is determined using the Jenkins Traub algorithm. The results showed that the nonparametric correlogram identified a geographic distance of land prices smaller than the correlogram used so far. A small distance means that the land price in a location is greatly affected by the neighbors compared to a larger distance.

**Key Words:** *geographical distance, land prices, nonparametric correlogram, spatial dependence.*

## Introduction

A collection of empirical and theoretical studies have been conducted to model the effect of environmental impacts on land prices using the hedonic method. Some of these studies show that land prices are influenced by the infrastructure and socio-economic features (Barreca et al. 2018), by the geographic and geometric accessibility (Morales et al. 2019), and by structure and accessibility variables (Chica-Olmo et al. 2019). These studies have examined the spatial effect on urban land prices from the influence of externalities. However, this research has not shown how important the geographical distance is as a spatial influence in modeling urban land prices.

In the analysis of urban land prices, they can be considered as variables that have a spatial dependence. Spatial dependence is a form of assessing the correlation of a variable connected with the spatial location. This means that the ability of spatial dependence is a measure that states the relationship between the variations in land price properties and the spatial proximity or the geographic distance can be expressed by the continuous function of numerical differences in property compared to the distance. Thus, one can see that the closer the two locations are, more differences will be in the weak property (or the greater the similarity). This function also allows dependence on land prices where distance plays an important role in the concept of geographic distance and distance decay functions. This concept can be identified using a variogram or correlogram in calculating the spatial dependence. It should be noted the use of the words spatial dependence and spatial autocorrelation – both are closely related, as a spatial autocorrelation is a common form of spatial dependence. In its development, spatial dependence has its definition of distance (Legendre and Legendre 2012).

Some research on land prices with regards to geographic distance has been observed by Chica-Olmo et al. (2019) using the variogram, Morales et al. (2019) with the multivariate regression, Barreca et al. (2018) using the spatial autocorrelation, while Crosby et al. (2018) performs the variogram fitting on land prices, and Shaker (2018) shows that Conditional Autoregressive (CAR) residuals were assessed by Moran's I spatial correlogram, etc. The basic problem of this research of the spatial dependence is how to clearly and analytically determine the spatial dependence at a certain distance. Usually, the concept of this problem can be described through a correlogram that uses the covariance function. Correlogram as a plot of covariance function that changes with the distance against locations is very important to describe the spatial dependency.

Before a statistician performs spatial the predictions, traditionally under stationary principles, two main things must be considered to estimate the correlation (Gorsich and Genton 2000). First, it is the estimation of correlation values at a certain lag or distance based on the stochastic process. If the process is isotropic, the correlogram is only a function of distance, so it can be estimated using several estimators (Cressie 2015). Second, the correlogram values are estimated and fitted using parametric models. This can be carried out because the estimated point does not guarantee the definite positive nature of the correlogram or the continuous nature of the correlogram is unknown. The disadvantage of this method is that researchers must choose their preferred model and determine their respective parameters (Genton and Gorsich 2002).

For fundamental reasons like this, a nonparametric estimator for a correlogram is very important so the selection of parametric models is no longer needed. In development, several studies have been carried out that attempt to eliminate the selection of models and parameters on the correlogram. Such as the nonparametric curve fitting by Shapiro and Botha (1991), based on the spectral representation of isotropic or definite positive functions by Yaglom (1957). However, Hall et al. (1994) criticize the approach taken by Spahiro and Botha because it turns out that the definite positive estimator only applies discretely and not continuously, so they propose their estimators through the kernel method approach. The realization of the kernel method approach is made clear by Bjørnstad and Falck (2001) by proposing the use of a nonparametric kernel estimator where the asymptotic kernel function is the B-Spline cubic, which is called a correlogram spline. One important thing to remember is that not all functions can be used as candidate covariance functions because they must fulfill the positive properties of semidefinite. The nonparametric method here is based on the isotropic spectral representation of the positive definite function properties derived by Yaglom (1957), based on Bochner's theorem.

The use of the correlogram itself has been applied to various fields, such as the detection of: TB case notification rates (CNR) in Bangladesh (Rood et al. 2019); neighborhood and spillover effects on rice farmers (Villanueva et al. 2017); landscape in Moldova (Shaker 2018); copper mining area (Nguyen et al. 2016); land prices (Jiao and Liu 2012); genetics (Diniz-Filho et al. 2009); semiconductors (Jeong et al. 2008) and ecology (Bjørnstad and Falck 2001). In the research on land prices, Jiao and Liu (2012) used an index on the spatial correlogram (Moran Index and Geary's Index) derived from monoton variance plots. However, the use of Moran Index as a correlogram function has the disadvantage of not having definite positive characteristics. Though not all data has a variance plot that is monotonous but non-monotonous, where the plot of variance decreases to a minimum then rises, or vice versa. But unfortunately, for non-monotonous cases, the availability of literature and research on the spatial dependence analysis, more specifically to the value of using a correlogram, is still difficult to find. The covariance function to describe the case of the non-monotone hole effect is found in several previous studies, for example Ye et al. (2015) and Weku et al. (2019) which use the Bessel and Gaussian-type functions, and Yang and Shao (2018), using Bessel and Gaussian functions. Although the correlogram is fundamentally not the main key in the concept

of spatial statistics, it is also useful in conducting exploration and as a descriptive tool. For this reason, the correlogram can help the variogram to provide richer information.

Therefore, in this study, in order to improve the ability of previous correlograms, it is proposed to use the non-parametric covariance function with the general class. The nonparametric covariance function also applies to variogram classes for all dimensions and it is a definite positive. Since the form used is nonparametric, we refer to this modification as a correlogram nonparametric. To test the effectiveness of the nonparametric correlogram, it is applied to the land price data in the city of Manado. This is intended to determine the geographical distance from the spatial dependence of land prices from the locations that are mutually influential and have similarities.

*Literature review*

The spatial correlogram is very good for checking patterns of spatial autocorrelation in data or residual models. It also shows how the correlation of spatial observations when the lag increases. Based on the direction, the spatial correlogram is divided in two. First, non directional spatial autocorrelation, such as the traditional spatial correlates using the autocorrelation index (Moran I or Geary c) will be plotted against the distance or the Spline Correlogram with non-parametric covariance functions with distances plotted kernels have also been introduced by Bjørnstad and Falck in 2001. Second, the directional spatial autocorrelation such as windrose correlogram. This procedure calculates directional correlation using a method introduced by Oden and Sokal (1986). In the traditional spatial correlogram, the pairs of points are inserted into separate classes/bin based on the distance between the points, while the correlational windrose, class/bin is based on distance and direction. For example, a point pair that is 50 km away with an angle of 60° between them will be placed in a separate bin from a point pair that is 50 km away at an angle of 20°. However, windrose correlograms require large samples rather than nondirectional correlograms. To handle small-sized samples, a bearing correlogram is introduced using a bearing procedure that is combined with a windrose correlogram (Rosenberg 2000).

*Correlogram Estimation*

Spatial data is data obtained from the measurement results of a location. Spatial data comes from different spatial locations that indicate dependencies between the measurement values and the location of lands. Expressed that {Z(s): s∈D} is a spatial process for *D* of a particular nature and $D \subset R^2$, Euclidean space of the two dimensions and *s* is the position of the location (Cressie 2015).

The classic estimator for variogram proposed by Matheron in 1962 used the moment method as follows (Weller and Hoeting 2015, Luo et al. 2018):

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} \left( Z(s_i) - Z(s_j) \right)^2$$

(1)

where $N(h) \equiv \left\{ (i,j) : s_i - s_j = h \right\}$ and $|N(h)|$ limited number of elements from $N(h)$.

Like the variogram, the covariogram can also be estimated using (Bjornstad and Falck 2001):

$$\hat{C}(h) = \frac{1}{n} \sum_{N(h)} (Z(s_i) - \bar{Z}) \left( Z(s_j) - \bar{Z} \right)$$

(2)

where $\bar{Z} = \frac{1}{n} \sum_{i=1} Z(s_i)$ is the average sample of land prices from observation. Equation (2) .

is more often used because it guarantees positive definite form estimates on $\mathbb{R}^1$

## *Nonparametric Covariance Function*

The property of the definite positive function is (Zastavnyi and Porcu 2017):

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\lambda_j C(h) \geq 0$$

(3)

for all $\lambda_i$ , h and n, then *C(h)* is a covariance function.

One of the main keys in nonparametric modeling for a correlogram is the theorem given by Bochner. He provides an approach through spectral representation for any positive definite function. Bochner's theorem states that a function can be used as a covariance function if it conforms with the positive definite properties and can be solved in a fourier transformation from *F(w)* as a positive measure unbounded to finite, written as follows (Yao 1999, Weller and Hoeting 2015):

$$C(h) = \int_{\mathbb{R}^d} e^{2\pi i h \cdot w} dF(w)$$

(4)

where *h* represents the distance.

It is assumed that the correlogram is isotropic, meaning that it depends only on the distance and not on the direction of the vector lag *h*. Nonparametric estimations of isotropic correlograms use the series given by Yaglom (1957) as a representation of Bochner's theorem, as follows:

$$C(h) = \sum_{j=1}^{\infty} p_j \Omega_d(b_j h)$$

(5)

where $p_j$ is a positive coefficient, scalar $b_j$ reppresents the nodes and $\Omega_d$ is the basis for the function on $\mathbb{R}^d$ expressed as (Golinskii et al. 2018):

$$\Omega_d(h) = \left(\frac{2}{h}\right)^{\frac{d-2}{2}} \Gamma\left(\frac{d}{2}\right) J_{\frac{d-2}{2}}(h)$$

(6)

$\Gamma\left(\frac{d}{2}\right)$ is gamma function dan $J_v$ is first kind of Bessel function with *v* orde.

If $C(h)$ isotropic, then Bochner's theorem can be written as follows (Gorsich and Genton 2000):

$$C(h) = \int_0^\infty \Omega_d(bh) F(dt)$$

(7)

Here $F(t)$ is a limited-not increase function and $\Omega_d(h) = \left(\frac{2}{h}\right)^{\frac{d-2}{2}} \Gamma\left(\frac{d}{2}\right) J_{\frac{d-2}{2}}(h)$. When *d = 1,* then $C(h) = \cos(bh)$, when *d = 2,* then $C(h) = J_0(bh)$; when d = 3 then $C(h) = \sin(bh)/bh$, when *d → ∞,* then $C(h) = exp\{-(bh)^2\}$ (Gorsich and Genton 2000, Ploner and Dutter 2000).

The periodicity function has a weak hole effect structure when *d* increases and becomes a Gaussian function when *d → ∞.*

For two dimensional random field *(d = 2),* when $J_v(\cdot)$ which is the first form Bessel function with order *u* and $v = \left(\frac{d}{2}\right) - 1$, then the idea becomes $v = \left(\frac{d}{2}\right) - 1 = 0$. Therefore $C(h) = J_0(bh)$ which can be expressed as:

$$C(h) = J_0(bh) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!^2} \left(\frac{bh}{2}\right)^{2k}$$

(8)

where *b* is number of sign.

## Methodology

### *Proposed Model: Nonparametric Correlogram*

To identify the geographical distance of urban land prices, we develop a nonparametric model of covariance functions in place as a correlogram. This Correlogram works based on equation (8) which uses the Bessel function to fit the covariogram. Because of the properties of the linear combination $C(h) = C_1(h)\, C_2(h)$, which $C_1(h) = J_0(bh) = \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{kbh}{2}\right)^{2k}}{(k!)^2}$ and $C_2(h) = exp\left(-\left(\frac{h}{r}\right)^m\right)$.

Our multiplication correlogram model can be written as:

$$C(h) = C_1(h)C_2(h) = \sum_{k=1}^{p} J_0(kbh)\, exp\left(-\left(\frac{h}{r}\right)^m\right)$$

(9)

Note, that the nonparametric correlogram depends on the bin size. The more bin is chosen, the more smooth the curves will be, but this will cause the pair to decrease in some places (keeping in mind that there is a limit of at least 30 sample pairs for each bin). It is important to choose the bin carefully.

### *Research Area*

This research was carried out in the city of Manado which is the capital of the province of North Sulawesi (Fig. 1). Geographically, Manado is located on the Manado Bay and it is surrounded by a mountainous terrain. The population in 2017, based on BPS data, is of 430 133

inhabitants. The large population in the city of Manado caused a high population density. With an area of 157.26 km$^2$, the population density reaches 2736 people/km$^2$. The city of Manado is located at the edge of the northern peninsula of the island of Sulawesi, at a geographical position of 124°40′ - 124°50′ East and 1°30′ - 1°40′ North.



**Fig. 1** – **Localization of Manado City, North Sulawesi Province, Indonesia**

*Data preparation*

In this study, we use processed data obtained from the BP2RB (Regional Tax and Retribution Management Agency) of Manado which is listed in the Source of Taxpayers' Association according to the book category of Manado municipality for 2018. In this, processed data with attributes consisted of: taxpayers interests, the object tax number, the taxpayer's address, the object of land and building tax, each of which is divided into the object area, object class and NJOP (Nilai Jual Objek Pajak – Tax Object Selling Value), as well as the property tax provisions.

There are 150 spatial locations that were selected to be used in this research (Fig. 2). The locations of land prices to be observed are considered as dependent variables. The legend on the map illustrates the land price in rupiahs, where the red color indicates the high land price (30000-3500000), while the bright color states that the land price is low (0-500000). High land prices are near shopping and service areas, while low land prices are on the outskirts of Manado.
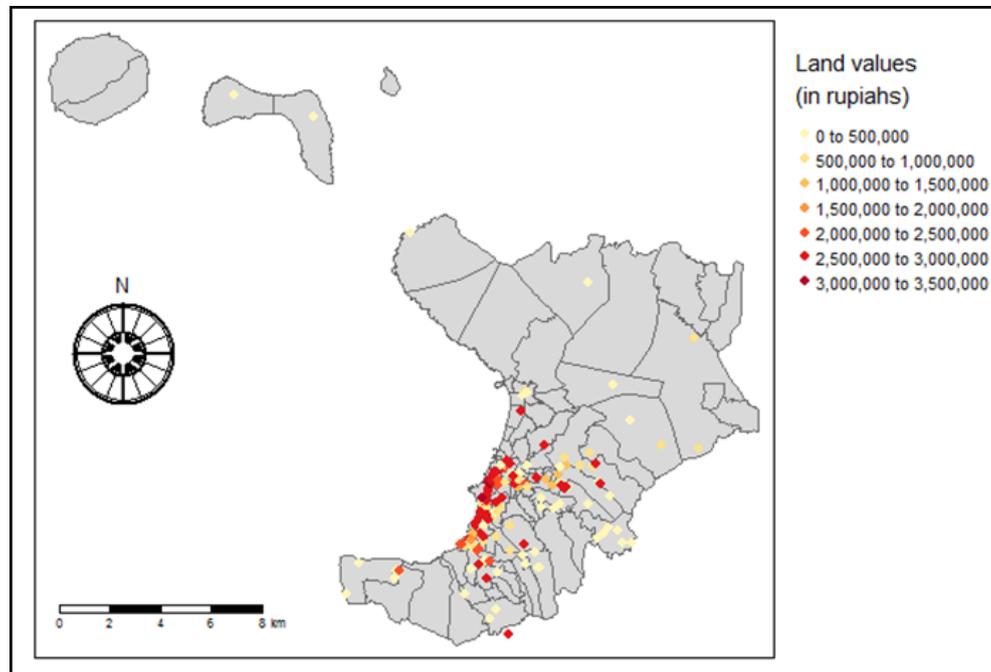
*Fig. 2 –* **Administrative map of Manado city and the 150 observation points**

## Results

The first step is to estimate the covariogram. Based on equation (2), a calculation is made to estimate the covariogram, where the results are as shown in Table 1 and Fig. (3a). It appears that the estimation curve oscillates around C(h) = 0, it decreases to local minima at C(h) = -0.1043 and it rises again pass through C(h) = 0 while it reaches the local maxima at C(h) = 0.1146. In Table 1, it is stated that the first distances located at the interval [1577.7304, 2022.0978] have a strong correlation. This value of C(h) was carried out through the principle of curve fitting using the nonparametric covariance functions.

The second step is fitting the covariance value C(h) with the correlogram. There are two correlograms that are often used so far – the spatial correlogram and the correlogram spline (Fig 3b, 3c). The spline correlogram with non-parametric covariance functions using kernel was plotted against the distance, while the traditional spatial correlogram using autocorrelation indices (Moran I or Geary c) has also been plotted against the distance. Both of them shows that the land price correlation decreases when the lag increases and it has intercepts at a certain distance when the correlation is 0. Note that both the monoton spline and the spatial correlogram cannot fit the estimation curve given by equation (2) to the maximum distance of (9174.1743 m) at class 21. These correlograms may identify only one intercept of C(h) = 0, even though after a distance of 4000 meters, the correlogram should fluctuate over C(h) = 0.

**Table 1**
**Classes, data pairs, distances and covariance values based on equation (2)**

| Bin Classes | np | dist | C(h) | | Class | np | Dist | C(h) |
|---|---|---|---|---|---|---|---|---|
| 1 | 383 | 274.9404 | 0.0877 | | 12 | 466 | 5168.0557 | -0.1043 |
| 2 | 676 | 678.4604 | 0.1455 | | 13 | 442 | 5621.5478 | -0.1559 |
| 3 | 768 | 1119.8843 | 0.0459 | | 14 | 352 | 6061.8550 | -0.0129 |
| 4 | 799 | 1577.7304 | 0.0576 | | 15 | 198 | 6505.1697 | -0.0681 |
| 5 | 822 | 2022.0978 | -0.0001 | | 16 | 179 | 6977.5815 | 0.1146 |
| 6 | 817 | 2477.9712 | 0.0676 | | 17 | 139 | 7391.2285 | 0.0089 |
| 7 | 951 | 2919.6496 | -0.0741 | | 18 | 137 | 7891.1404 | 0.0408 |
| 8 | 915 | 3363.5502 | -0.1420 | | 19 | 141 | 8326.1946 | 0.0170 |
| 9 | 702 | 3821.0236 | -0.0367 | | 20 | 135 | 8772.5338 | 0.2610 |
| 10 | 637 | 4263.9529 | 0.0358 | | 21 | 66 | 9174.1743 | -0.0599 |
| 11 | 566 | 4729.4027 | -0.2497 | | | | | |
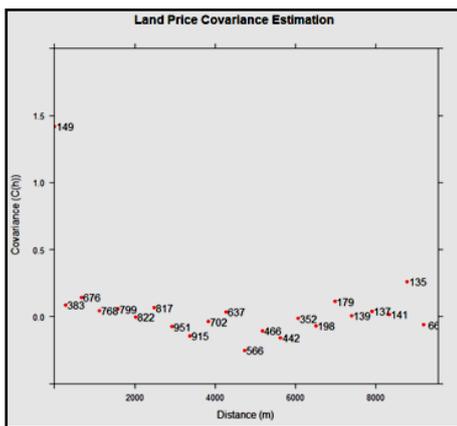


a
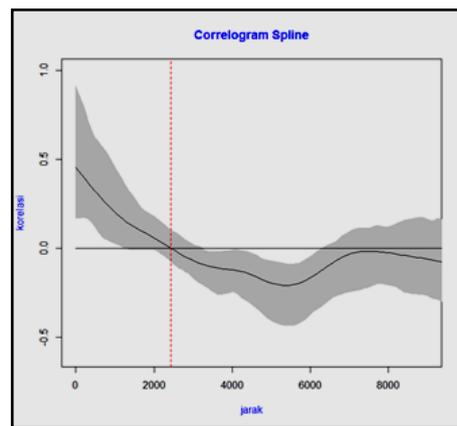


b
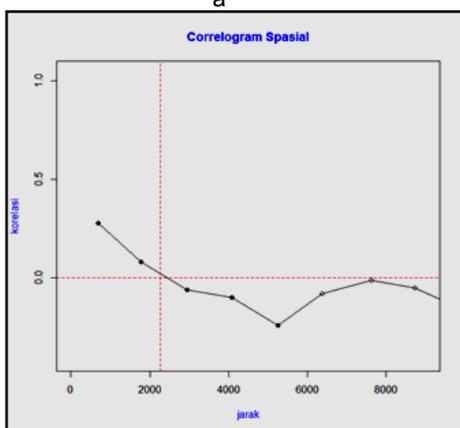


c

*Fig. 3* **– (a) The covariogram estimation for land prices using equation (2) in** $\mathbb{R}^1$ **; plot on the relationship between the spatial correlation to distance for land price data**
(using the R software-gstat package) with (b) Spline Correlogram; and (c) Spatial Correlogram

Then we applied the nonparametric correlogram model to overcome the periodicity problem on the covariogram. This model is a combination of multiplication of Bessel and Gaussian-type functions as stated in equation (8). The use of spline cubic interpolation to approach the results of the combination approach looks much better than the use of smoothing curves using LOWESS on spatial correlograms and spline correlograms with cubic B-spline as asymptotic kernel functions. By taking p=1, 2, ... , 5, in equation (2), the results of estimation and curve fitting is shown as in Table 2 and Fig. 4 with a periodicity which passes C(h) = 0.

According to the calculations made with the R program, the intersection between the correlogram curve with the horizontal axis occurs at h = 2433.812 (correlogram spline) and h = 2265.832 (spatial correlogram). This means that the correlation at the distances will be very strong, while when the distance is increasing, then the correlation of the location of land prices will be weaken. Where the locations are close to each other in the range of the distance, this will give a strong influence on the land prices.



a                          b                          c



d                                        e
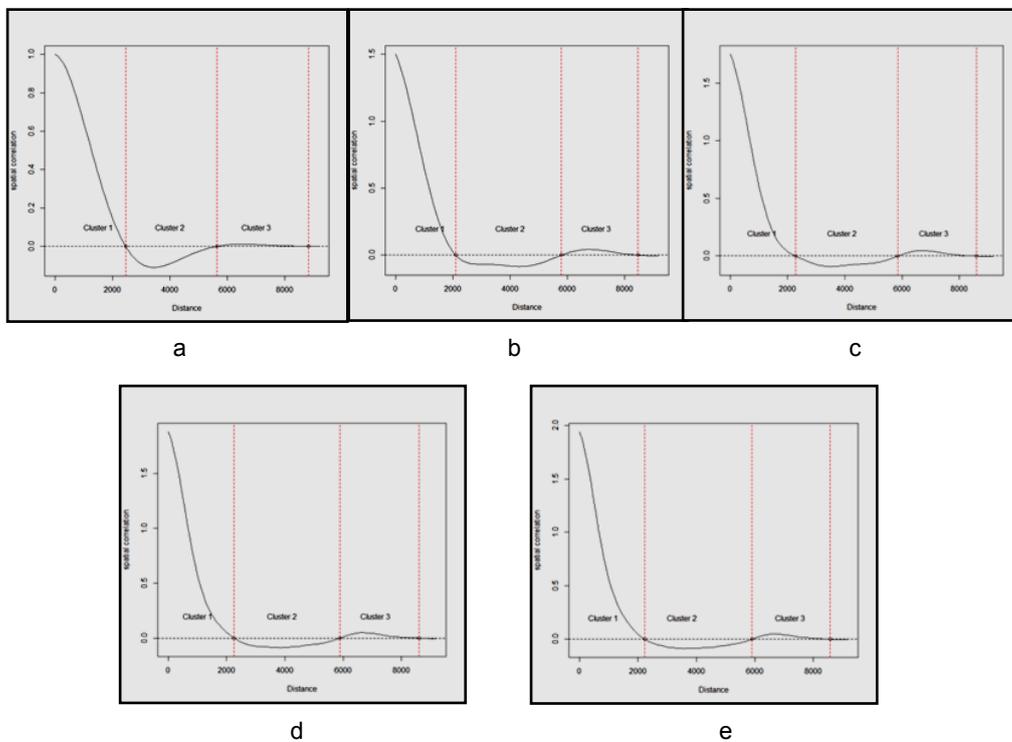
**Fig. 4 – Nonparametric Correlogram with different bases to plot the correlation of land prices against the geographic distance**

The availability of function *C(h)* and data pairs (distance, correlation) make it easy to determine the intersection. There are 3 points intercept at *h* when *C(h) = 0* which divides the distance into 3 parts. Intersections can occur together with the changes in the value of spatial autocorrelation from positive to negative and from negative to positive. The nonparametric Correlogram shows the existence of three distance clusters that affect the high and low prices of land. The first cluster shows a strong distance correlation, while the second cluster and the third cluster show a less strong correlation because the lag distance increases.

**Correlogram values for each model of 21 bin classes**

| Bin Classes | Lag distance (m) | p=1 | p=2 | p=3 | p=4 | p=5 |
|---|---|---|---|---|---|---|
| 1 | 274.9404 | 0.8975 | 1.3218 | 1.5144 | 1.5978 | 1.6317 |
| 2 | 678.4604 | 0.7147 | 0.9565 | 1.0017 | 0.9892 | 0.9716 |
| 3 | 1119.8843 | 0.4994 | 0.5382 | 0.4787 | 0.4489 | 0.4485 |
| 4 | 1577.7304 | 0.2895 | 0.2028 | 0.1603 | 0.1752 | 0.1836 |
| 5 | 2022.0978 | 0.1204 | 0.0174 | 0.0351 | 0.0471 | 0.0393 |
| 6 | 2477.9712 | -0.0060 | -0.0553 | -0.0232 | -0.0356 | -0.0334 |
| 7 | 2919.6496 | -0.0811 | -0.0678 | -0.0662 | -0.0699 | -0.0670 |
| 8 | 3363.5502 | -0.1146 | -0.0690 | -0.0890 | -0.0800 | -0.0841 |
| 9 | 3821.0236 | -0.1150 | -0.0768 | -0.0857 | -0.0861 | -0.0841 |
| 10 | 4263.9529 | -0.0940 | -0.0842 | -0.0746 | -0.0805 | -0.0797 |
| 11 | 4729.4027 | -0.0609 | -0.0772 | -0.0676 | -0.0652 | -0.0673 |
| 12 | 5168.0557 | -0.0285 | -0.0515 | -0.0541 | -0.0508 | -0.0493 |
| 13 | 5621.5478 | -0.0003 | -0.0135 | -0.0213 | -0.0238 | -0.0239 |
| 14 | 6061.8550 | 0.0187 | 0.0203 | 0.0188 | 0.0173 | 0.0163 |
| 15 | 6505.1697 | 0.0286 | 0.0396 | 0.0444 | 0.0465 | 0.0475 |
| 16 | 6977.5815 | 0.0302 | 0.0406 | 0.0434 | 0.0436 | 0.0433 |
| 17 | 7391.2285 | 0.0260 | 0.0299 | 0.0280 | 0.0263 | 0.0259 |
| 18 | 7891.1404 | 0.0172 | 0.0132 | 0.0103 | 0.0108 | 0.0114 |
| 19 | 8326.1946 | 0.0086 | 0.0022 | 0.0026 | 0.0036 | 0.0032 |
| 20 | 8772.5338 | 0.0007 | -0.0034 | -0.0012 | -0.0018 | -0.0018 |
| 21 | 9174.1743 | -0.0045 | -0.0048 | -0.0039 | -0.0046 | -0.0042 |

The geographical distance can be determined by observing the fluctuations that occur around the zero horizontal axis. For p=1, the geographic distance is between class 5 and class 6 at interval [2022.0978, 2477.9712], and after that the autocorrelation is negative until the classes 13 and 14 at interval [5621.5478, 6061.8550], then there is a rise with positive autocorrelation until classes 20 and 21 at interval [8772.5338, 9174.1743]. For p = 2,3,4,5, it gives slightly different results. In particular, the last fluctuation occurred between classes 19 and 20 at interval [8326.1946, 8772.5338].

### *Optimal Geographic Distance*

The measurement of geographical distance is considered important because it always appears in the urban analysis and population geography. Our next concern is how to determine the optimal geographical distance which is located at each interval previously mentioned. The geographical distance can be calculated using the nonparametric correlogram as the smallest correlation value for *h*, such that *C(h) = 0*. In general, there are two patterns of autocorrelation that are formed, such as the positive and negative autocorrelation. If the land price has similarities with its neighbors, then there will be a positive autocorrelation, and vice-versa for a negative autocorrelation.

The optimal geographic distance measurement is calculated from the correlogram *C(h)* as the smallest value for *h* such that *C(h) = 0*. The determination of this distance can be established when the correlogram curve passes through the abscissa for ordinate correlation at *0*. For example, when p = 1, the geographical distance is between 2000 and 3000, and after that distance of the autocorrelation is negative and then it fluctuates around the zero line. The

fluctuation of the curve around the zero axis *(C(h) = 0)* can be considered as a nonlinear equation, so that we can determine the h from the equation. The Jenkins-Traub algorithm can be used to solve nonlinear equations in order to obtain the h or the optimal geographical distance.

The geographic distance plot shows that the low-pass component of the land price is the most important thing to analyze. The geographic distance in the first cluster shows that there is a positive autocorrelation until it reaches the upper bound and it forms the periodicity (Table 3). For example, for base *p=1*, it has 3 clusters of geographic distance, where the first cluster (0<h<2451.376) and the third (5626.905<h<8820.882) are positive autocorrelations, while the second cluster (2451.376<h<5626.905) is a negative autocorrelation. The first geographical distance cluster shows that there is a strong correlation between the locations at a distance of 2451.376 to land prices, then the correlation of locations to land prices will decrease slowly after that geographical distance. An interesting finding is that the geographical distance will approach the correlogram spline for p = 1, while p = 3 causes the geographical distance to approach the spatial correlogram.

*Table 3*

**Radius estimation (meters) with nonparametric, spline and spatial correlogram**

| Cluster | Nonparametric Correlogram | | | | | Spline | Spatial |
|---------|------|------|------|------|------|--------|---------|
|         | base1 | base2 | base3 | base4 | base5 | | |
| 1 | 2451.376 | 2091.351 | 2269.686 | 2251.173 | 2221.306 | 2433.812 | 2265.832 |
| 2 | 5626.905 | 5782.948 | 5850.338 | 5880.357 | 5892.046 | | |
| 3 | 8820.882 | 8455.558 | 8601.896 | 8608.763 | 8578.312 | | |

The important thing about the nonparametric correlogram is the determination of the autocorrelation value as a geographical distance. When the geographical distance decreases, this means that there is a close autocorrelation between the locations that have similarities. Another consequence is that there will be a certain distance that does not have a significant correlation between the locations of observations separated by that distance. A further application to analyze land prices, this autocorrelation value is useful for determining the geographic radius of land prices that are close together or, in other words, the maximum distance where the spatial interaction of land prices is still significantly influential.

*Spatial Dependency with Geographic Distance Clustering*

The next step is to visualize the geographic distance cluster based on *p*. The geographic distance is important in looking at the effect of distance in the spatial dependency so that land prices are interrelated. The distance with a positive spatial autocorrelation is a different form of geographic distance with a negative spatial autocorrelation and it will form individual clusters (Fig. 5). The first distance cluster (colored in red) is represented by a geographic distance and a positive spatial autocorrelation; it has several points of location for the value of isolated land that does not have a spatial influence. This first distance cluster is the optimal cluster because it is close to the origin. This states that the location of land prices at a distance of 0 meters to 2451,376 meters has a high land price that is affected by the adjacent locations with a high land price. The second distance cluster (blue) is represented by a distance and negative spatial autocorrelation, located at a distance of 2451,479 meters to 5627,115 meters. This cluster indicates that high land price are influenced by the low land price, and vice-versa. And the third distance cluster (green), with radius and positive spatial autocorrelation, is located at a distance of 5626,905 meters up to 8820,882 meters. This cluster indicates that high land prices are influenced by the high land price, and vice-versa. This third distance cluster is not optimal because it is far from the origin point compared to the first distance cluster.
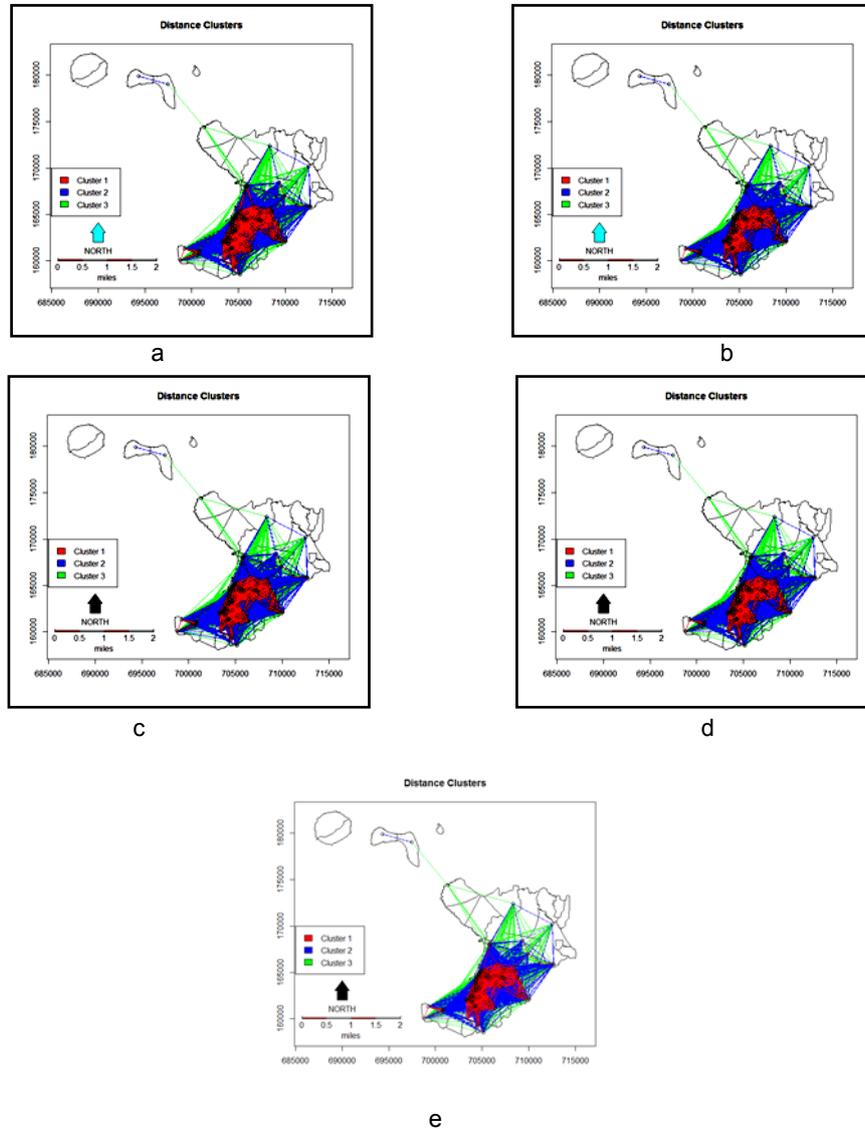
a



b



c



d



e

*Fig. 5* **– Distance clusters of mutually influential observation locations**

The last stage is how to determine the optimal geographical distance. In this section we use the Moran Index which can determine the level of strength for the spatial autocorrelation. The size of the distance will be calculated by the Moran Index in exploring the size of the correlation object observation. In other words, this index will measure how the land price object in one location is similar to the other around it (Table 4).

*Table 4*
**Moran I and linear gradient forming the Moran scatter plot**

| Basis | p=1 | p=2 | p=3 | p=4 | p=5 |
|---|---|---|---|---|---|
| **Distance** | 2451.3760 | 2091.3510 | 2269.6860 | 2251.1730 | 2221.3060 |
| **Moran I** | 0.1853 | 0.2267 | 0.1939 | 0.1952 | 0.2075 |
| **Gradien** | 0.0721 | 0.0524 | 0.0656 | 0.0650 | 0.0566 |

For a larger base ($p \longrightarrow 5$), it is known that the distance is getting smaller but not so with Moran's I. From all bases, we can see that the largest distance is 2451.3760 meter at base p = 1, while the smallest distance is 2091.3510 at base p = 2. Furthermore, by considering the largest Moran Index value (0.2267), the distance = 2091.3510 meters is chosen with the base model p = 2 to illustrate the correlation between the pairs of spatial observations when the lag increases. This means that the distance = 2091.3510 meters states that there is a strong correlation between the location of the land price of the adjacent land compared to the other geographic distances.

## Discussion

As we already know in the analysis of spatial dependence, the price of the land of a location is influenced by the price of the surrounding neighbors. The core of this research is to find out how far the geographical distance in Manado, so that all areas within the geographical distance, will have a strong influence on each other in determining the land prices.

In the results that have been done using a nonparametric correlogram, the optimal geographical distance is 2091.3510 meters. This means that each location of land within this geographical distance has a strong influence or correlation to one another. On the other hand, each location outside the geographical distance has a less powerful influence in determining land prices.

In the discussion of this study, the relationship between the geographical distance and land prices is analyzed. The main findings of this relationship can be explained as follows:

1. Accessibility (distance from the economic center), meaning that the easier access to the economic center to be reached, the more valuable the land will be. For example, the city of Manado has the highest land price in the city center or what is known as the city's business district. This is due to the closeness between the community and economic activities and the workplace. The factors that influence land price movements are the most important in making various models of land prices.

2. Infrastructure/facilities, here means that the price of land is influenced by the availability and closest distance to the facility, for example offices, education, health facilities, traditional or modern markets. Proximity to this facility will save time in accessing the location of the facility, resulting in an increase in the price of land in locations close to the infrastructure.

3. A property that is in a geographical area with high infrastructure facilities or high economic activity will increase land prices. For example, the distance to transportation facilities and the existence of a highway will increase the efficiency of community work, and it will reduce the level of congestion.

These findings have many implications for the urban society. In the Central Business District, it

is seen that high land prices (red colors) dominate in the region because the centers of economic activity, services, shopping, health, etc. are indeed centralized in the area. While low land prices (bright colors) are on the edge of the city of Manado. Therefore, for people who have high income levels, it is not a problem to choose to live in that geographical distance. Conversely, people who have a low income will choose to stay on the edge of the city of Manado that exceeds the geographical distances due to low land prices. For people who have activities close to the city center, for example the government employees, the service providers, or the infrastructure (offices, schools, modern and traditional markets, health centers), they will choose to stay within the geographical distance. So it is not surprising, if the land price in the downtown of Manado is so high and mutually influential among its neighbors.

People who have high income levels or have jobs that are close to the Central Business District can buy land to work or live in the areas marked in red (within geographical distances that have strong correlations). Meanwhile, those with a low economic capacity can live in areas far from the CBD or on the edge of the city of Manado, which are marked by green areas.

Because the outermost area of the city is a location with low land prices, so the region is still less populated, this could be due to the lack of infrastructure. Therefore, it can be a concern for the city government to equalize the infrastructure development so that land prices are evenly distributed throughout the city of Manado. For the development of the city in the future, it is strongly recommended to carry out development in the suburbs. Because land is still widely available at relatively lower prices compared to land prices close to the city center.

## Conclusions

In this study, Bessel and exponential covariance functions have been used which play a very important role in estimating the correlogram in a nonparametric manner. The form of multiplication and additive composition applied has met the requirements of positive definite functions in the spectral representation. The form of periodicity (oscillation) can be well fulfilled by the composition of covarince functions. With this nonparametric correlogram model, a researcher has the advantage of not having to choose the model first. The nonparametric correlogram curve fitting using the spline cubic interpolation produces a smooth curve so it can fit the results of the covariogram method of the moment, things that cannot be done by spatial correlograms or correlogram splines. The larger the bin size, the more curved the resultant as the smaller bin size. The geographical distance occurs when C(h) = 0 is determined by using the Jenkins Traub algorithm.

In applying for a land prices dataset, the nonparametric correlogram provides the optimal geographical distance for land prices given when the covariance value is zero. The oscillating covariance with a distance interval can be considered as a cluster form. The formed cluster illustrates the important relationship between the geographical distances with a positive spatial autocorrelation and a negative spatial autocorrelation. The Moran index can be used to determine the strength of the land price autocorrelation that occurs at a distance = 2091.3510 meters with a composition model of Bessel base p = 2 and a Gaussian-type function. At this optimal distance, it can be said that two locations in Manado with identical attributes will have similar prices if they are adjacent to each other rather than if they are far apart (larger than 2091.3510 m). The location of land which is within the geographical distance has a dependency or correlation to each other in order to determine and change land prices. The use of spatial dependence with geographical networks can represent a basis for determining land prices in the future.

# References

BARRECA A., CURTO R., ROLANDO D. (2018), *Housing Vulnerability and Property Prices: Spatial Analyses in the Turin Real Estate Market*, Sustainability 10, 3068.

BJØRNSTAD O. N., FALCK W. (2001), *Nonparametric spatial covariance functions: Estimation and testing*, Environmental and Ecological Statistics 8, 53-70.

CHICA-OLMO J., CANO-GUERVOS R., CHICA-RIVAS M. (2019), *Estimation of Housing Price Variations Using Spatio-Temporal Data*, Sustainability 11, 1551.

CRESSIE N. (2015), *Statistics for Spatial Data*, Wiley, New York.

CROSBY H., DAMOULAS T., CATON A., DAVIS P., DE ALBUQUERQUE J. P., JARVIS S. A. (2018), *Road distance and travel time for an improved house price Kriging predictor*, Geo-Spatial Information Science 21 (3), 185-194.

DINIZ-FILHO J. A. F., NABOUT J. C., TELLES M. P. D. C., SOARES T. N., RANGEL T. F. L. V. B. (2009), *A Review of Techniques for Spatial Modeling in Geographical, Conservation and Landscape Genetics*, Genetics and Molecular Biology 32 (2), 203-211.

GENTON M. G., GORSICH D. J. (2002), *Nonparametric Variogram and Covariogram estimation with Fourier-Bessel Matrices*, Computational Statistics & Data Analysis 41 (1), 47-57.

GOLINSKII L., MALAMUD M., ORIDOROGA L. (2018), *Radial Positive Definite Functions and Schoenberg Matrices with Negative Eigenvalues*, Transactions of the American Mathematical Society 370 (1), 1-25.

GORSICH D. J., GENTON M. G. (2000), *Variogram Model Selection via Nonparametric Derivative Estimation*, Mathematical Geology 32 (3), 249-270.

HALL P., FISHER N. I., HOFFMAN B. (1994), *On the Nonparametric Estimation of Covariance Functions*, The Annals of Statistics 22 (4), 2115-2134.

JEONG Y.-S., KIM S.-J., JEONG M. K. (2008), *Automatic Identification of Defect Patterns in Semiconductor Wafer Maps Using Spatial Correlogram and Dynamic Time Warping*, IEEE Transactions on Semiconductor Manufacturing 21 (4), 625-637.

JIAO L., LIU Y. (2012), *Analyzing the Spatial Autocorrelation of Regional Urban Datum Land Price*, Geo-Spatial Information Science 15 (4), 263-269.

LEGENDRE P., LEGENDRE L. (2012), *Numerical* Ecology, Elsevier, Amsterdam.

LUO J., FRISKEN S., MACHADO I., ZHANG M., PIEPER S., GOLLAND P., TOEWS M., UNADKAT P., SEDGHI A., ZHOU H., MEHRTASH A., PREISWERK F., CHENG C. C., GOLBY A., SUGIYAMA M., WELLS W. M. (2018), *Using the variogram for vector outlier screening: application to feature-based image registration*, International Journal of Computer Assisted Radiology and Surgery 13 (12), 1871-1880.

MORALES J., FLACKE J., ZEVENBERGEN J. (2019), *Modelling residential land values using geographic and geometric accessibility in Guatemala City*, Environment and Planning B: Urban Analytics and City Science 46 (4), 751-776.

NGUYEN T. T., VUA D. T., TRINHB L. B., NGUYEN T. L. H. (2016), *Spatial Cluster and Outlier Identification of Geochemical Association of Elements: A Case Study in Juirui Copper Mining Area*, Bulletin of the Mineral Research and Exploration 153, 159-167.

ODEN N. L., SOKAL R. R. (1986), *Directional Autocorrelation: An Extension of Spatial Correlograms to Two Dimensions*, Systematic Zoology 35 (4), 608-617.

PLONER A., DUTTER R. (2000), *New directions in geostatistics*, Journal of Statistical Planning and Inference 91 (2), 499-509.

ROOD E., KHAN A. H., MODAK P. K., MERGENTHALER C., VAN GURP M., BLOK L., BAKKER M. (2019), *A Spatial Analysis Framework to Monitor and Accelerate Progress towards SDG 3 to End TB in Bangladesh*, ISPRS International Journal of Geo-Information 8 (1), 14.

ROSENBERG M. S. (2000), *The Bearing Correlogram: A New Method of Analyzing Directional Spatial Autocorrelation*, Geographical Analysis 32 (3), 267-278.

SHAPIRO A., BOTHA J. D. (1991), *Variogram fitting with a general class of conditionally nonnegative definite functions*, Computational Statistics & Data Analysis 11 (1), 87-96.

SHAKER R. R. (2018), *Examining sustainable landscape function across the Republic*

*of Moldova*, Habitat International 72, 77-91.

VILLANUEVA D. B., PEDE V. O., RODRIGUEZ U. E., SUMALDE Z. M., GARCIA Y. T. (2017), *Assessment of Neighborhood and Spillover Effects on Technical Efficiency of Irrigated Rice Farmers*, Asian Journal of Agriculture and Development 14 (2), 103-125.

WEKU W., PRAMOEDYO H., WIDODO A., FITRIANI R. (2019), *Non-Monoton Nonparametric Variogram to Model the Land Price of Manado City with Hole Effect Periodicity Structure*, IOP Conf. Ser.: Materials Science and Engineering 546, 052083.

WELLER Z. D., HOETING J. A. (2015), *A Review of Nonparametric Hypothesis Tests of Isotropy Properties in Spatial Data*, Statistical Sciences 31 (3), 305-324.

YAGLOM A. M. (1957), *Some classes of random fields in n-dimensional space, related to stationary random procesess*, Theory of Probability & Its Applications 2 (3), 273-320.

YANG Y., SHAO C. (2018), *Spatial Interpolation for Periodic surfaces in manufacturing using a Bessel additive variogram model*, Journal of Manufacturing Science and Engineering 140 (6), 061001.

YAO T. (1999), *Nonparametric cross-covariance modeling as exemplified by soil heavy metal concentrations from the Swiss Jura*, Geoderma **88 (1-2)**, 13-38.

YE J., LAZAR N. A., LI Y. (2015), *Nonparametric variogram modeling with hole effect structure in analyzing the spatial characteristics of fMRI data*, Journal of Neuroscience Methods 240, 101-115.

ZASTAVNYI V. P., PORCU E. (2017), *On Positive Definiteness of Some Radial Functions*, Lobachevskii Journal of Mathematics 38, 386-394.

Correspondence: Brawijaya University, Veteran Street, 65145, Malang, Indonesia.

Email: winsyweku@gmail.com