ISSN 2354-9114 (online), ISSN 0024-9521 (print)
Indonesian Journal of Geography Vol. 52 , No.1, 020 (36 – 41)
DOI: http://dx.doi.org/10.22146/ijg. 43724 website: https://jurnal.ugm.ac.id/ijg

INDONESIAN JOURNAL OF GEOGRAPHY

RESEARCH ARTICLE

Bandwidth Modelling on Geographically Weighted Regression with Bisquare Adaptive Method using Kriging Interpolation for Land Price Estimation Model

Alfita Puspa Handayani, Albertus Deliar, Irawan Sumarto and Ibnu Syabri
Faculty of Earth Science and Technology, Institut Teknologi Bandung, Indonesia.

Received: 2019-02-21
Accepted: 2019-12-05

Key words:
land price,
kriging,
GWR,
interpolation,
bandwidth

Correspondent email:
alfitapuspa@gmail.com

Abstract Land prices, especially in an urban area, are dynamically changing. To be able to do an evaluation, the right models must have the ability to understand land price characteristics that also dynamically changing. Every land price must attach to a location (spatial based). One of the locations (spatial based) models is Geographically Weighted Regression (GWR). This model can provide a local model based on the concept of attachment between observation and regression points. The main component is the determination of Optimum Bandwidth, which will determine the accuracy of the final GWR model. In the bandwidth process, it is necessary to do trial and error to get the Optimum Bandwidth value. Cross-Validation method commonly used to determine optimum bandwidth on observation point, but this study aims to minimize the process of trial and error in determining optimal bandwidth outside the observation point by using kriging interpolation. The Kriging method can substantially provide better bandwidth usage without having to do a trial process with too many errors.

1. Introduction

Location is acknowledged as an essential attribute of immovable objects such as land and improvements to land (Dziauddin & Idris, 2017). As a highly valuable resource, monitoring of land prices should be considered so land as a commodity and valuable asset would not be used as an object of speculation. The monitoring of land prices is generally conducted with trend series analysis or by using the regression method. Both of these methods resulting in a global model with one equation used in the entire area. As the object that bounded and varied spatially, land prices should be modeled with more than one global model to obtain accurate results.

Geographically Weighted Regression (GWR) is a local weighting regression method which varied spatially that firstly found by Fotheringham (Fotheringham, Brunsdon, & Charlton, 2002). GWR is a regression model that developed for data modeling with continuous response variables and considering the spatial or location aspect (Agnes et al., 2016). GWR model can capture the heterogeneity on a local model (Yu, 2006). The difference between the GWR method with the regression method is on the use of the observation point position (coordinate). Coordinate of observation point on the GWR equation model was used to determine the weighting value given to each parameter. The closer the distance between parameter with the-i observation point, means the higher weight value assigned to that parameter. This statement was appropriate with Tobler's first law of geography, which stated that "Everything is related to everything else, but near thing are more related than the distant thing" (Schabenberger & Gotway, 2004).

This weighting method will produce diverse parameter values for each observation data, depends on proximity distance between the location of each parameter with the location of observation. The maximum weighted value is given to the parameter data located around the location of observation, and this weight value will continuously decrease along with the increase of distance between parameter locations to the location of observation. Giving weight value for each observation point is limited to location restrictions (coverage area) called bandwidth. The boundary of observation coverage area then will determine which parameters that included as an affecting parameter on one of the observation points.

There are spatial function methods that can be used in defining bandwidth (Fotheringham et al., 2002), which are Fixed Spatial Kernel and Adaptive Spatial Kernel. On a fixed spatial kernel model, bandwidth value is defined as a value of distance. While on an adaptive spatial kernel model, bandwidth value is defined as the value of the number of points, which then determine the value of radius distance. Defining bandwidth is the most decisive stage in the GWR method.

Defining bandwidth (b) was more important than choosing the function form of weight (Wand & Jones, 1995).

There was no fixed standard for the best number of bandwidths. Therefore, it needs to use a particular method to compare and set a wide bandwidth that should be used. One method of finding optimum bandwidth is by conducting cross-validation (CV). The formula of CV, according to (Fotheringham et al., 2002) to be implemented in GWR can be seen in equation 1:

$$CV = \sum_{i=1}^{n}[y_i - \hat{y}_{\neq i}(b)]^2 \quad (1)$$

with n is the number of observations, i is regression point of i, Yi is a value of observation in the regression point of i, and $\neq$ ($\neq$i) (b) is prediction value from regression model which in the calculation does not input the observation data in the point of i.

The above formula resulting in the total quadrate value of all residue between the value of observation in i point and value of the predicted result, which the point of data in i was not inputted to the process of regression. Value of CV will show how great the accuracy of the model when data around a point of i has been given weight with desired bandwidth. Then, this validation-cross process is conducted to the entire wide probability of bandwidth that can be implemented in a modeling area. Optimum bandwidth, as shown in Figure 1, is bandwidth with the smallest CV value, and it assumed to produce the best model.In GWR method, it was very likely that each observation points have different optimum bandwidth value. Finding the interpolation technique to estimate optimal bandwidth value in each land area out of a known observation point is one of an important thing to do. According to (Yang, Tong, & Zhu, 2013), bandwidth can be considered as a smoothing parameter with greater bandwidth resulting in higher refinement.

Kriging interpolation is the least-squares interpolation method and estimating the average weight where the weight is a spatial covariance function (Carr, 1994). Useful spatial covariance shows the spatial accuracy of the data. Spatial relationships between data can be shown by a variogram. The variogram is a model that describes the dependence between

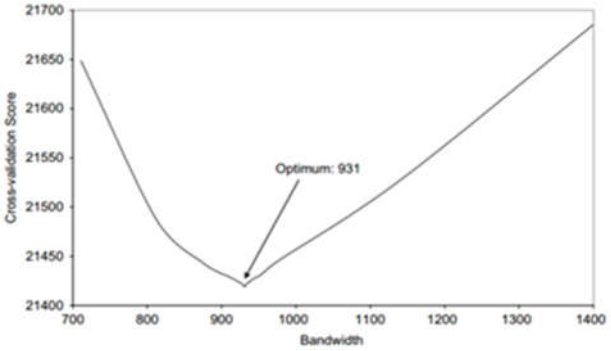

Figure 1. Cross-Validation method to determine optimum bandwidth.

data and distance between data. If the pattern of spatial dependence with that point has been obtained, then the estimated weight for each data can be obtained. In addition to considering the distance between sample points and interpolation points, the weighting of kriging interpolation also takes into account the spatial distribution of the value of the data itself. The basic kriging equation can be seen in equation (2) (Manson, Burrough, & McDonnell, 1999):

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i [z(x_i) - \mu(x_0)] + \mu \quad (2)$$

$Z(x_i)$: *measurement value at i point*

$\hat{Z}(x)$: *estimation value at x point*

λ_i : *weighted of i point*

n : *number of measurement*

Study that using GWR and Kriging as a spatial interpolation method has been done a lot for the last decades. Szymanowski and Kryza (2011) do some study to combine Local regression models for spatial interpolation between GWR and kriging that show a result that GWR is better justified in terms of statistical specification, and the combination between GWR and Kriging interpolation is suggested (Szymanowski & Kryza, 2012). Meng (2014), do some comparison study between GWR and Kriging for spatial interpolation. The study confirms that both methods are powerful in modelling local-spatial prediction. However, regression kriging has more edge on capturing the structure of original data (Meng, 2014). On contrary to this result, Wang et al. (2017) concluded that spatial interpolation and regression analysis models produced by the GWR are more precise compared to the Kriging model when applied to estimate the monthly surface air temperature in China (Wang et al., 2017). Nevertheless, this study is not a comparison between the two methods but instead it will explore the possibility of using kriging interpolation to determine optimum bandwidth for GWR. This study will not be comparing or combine GWR and kriging. This study will explore the possibility of using kriging interpolation to determine optimum bandwidth for GWR.

2. The Methods

Data used in this study are :

1. Administrative Boundary of East Bandung District
2. Polygon of 13.702 land area which divided into some Land Value Zone and Average Indication Value of 2007
3. Parameters with significant influence on the Average Indication Value of East Bandung District, which consists of Toll Gates, Mall, Universities, Hospital, Schools, Stations, Terminal, Public Cemetery, Mosque, Police Office, Market, Road Network, and Digital Elevation Model (DEM).

The distribution point between observation and regres-

sion point can be seen in Figure 2. In this figure, we can also see all the parameter distribution that will be used in this research. The parameters that have a proximity to the observation point will be selected. The closest parameters will then affect the equation formed, while the rest will be ignored.

The number of sample points is determined by the Slovin method with simple random sampling(Sevilla, Ochave, Punsalan, Regala, & Uriarte, 1984). Slovin formula used is in equation 3.

$$n = \frac{N}{1+Ne^2} \quad (3)$$

Description:

n = minimum number of samples

N = number of populations

e = tolerance limit of error

Based on the calculation result added with more substantial size data, then the number of sample data used is 420 land area. Near Analysis method to calculate the nearest distance between the centroid of each sample data on the entire nearest parameters used is in equation 4.

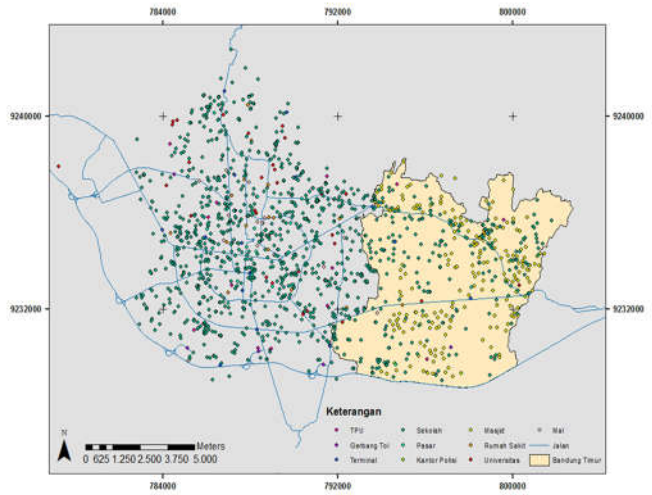

Figure 2. Distribution Point between observation and regression point.

$$d = \sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2} \quad (4)$$

Description:

d = nearest distance

X_2 = abscissa value of 2nd data

X_1 = abscissa value of 1st data

Y_2 = ordinate value of 2nd data

Y_1 = ordinate value of 1st data

The GWR model used in this study is a model with Bi-Square weighting adaptive approach. Every point of land price forms an equation by estimating distance on the nearest surrounding parameter points. This equation was established by giving weight value using the Weighted Least Square equation 5 (Fotheringham et al., 2002).

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (5)$$

Description:

$\hat{\beta}(u_i, v_i)$: matrix of estimation coefficient

i : 1, 2, ... , n (number of observation)

X : matrix of the parameter value

$W(u_i, v_i)$: matrix of weighting diagonal: $0 \leq W(u_i, v_i) \leq 1$

The weighting matrix was formed by using the bi square function. Weight value decreased continuously along with the increase of distance on parameter up to the limit of bandwidth value. The bi square weighting function can be seen in equation 6 (Ward & Gleditsch, 2008).

$$W_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2, & d_{ij} < b \\ 0, & d_{ij} \geq b \end{cases} \quad (6)$$

Description:

d_{ij} : distance of sample points on parameter

b : bandwidth

Results that gained in this stage are optimum bandwidth for the entire sample data. After achieving optimum bandwidth results for the whole of the sample data, then conducting spatial interpolation to estimate the value of bandwidth for the entire land area in the study area. The type of spatial interpolation used is Simple Kriging with this following equation can be seen in equation 7 (Li & Heap, 2008).

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(xi) + [1 - \sum_{i=1}^{n} \lambda_i]\mu \quad (7)$$

Description:

$\hat{z}(x_0)$: fcestimation value on a point of x_0

x_0 : estimated location points

$z(x_i)$: observation value on a point of x_i

x_i : observation location points of-i

λ_i : kriging weighted value of observation of-i

n : 1, 2, ..., n (number of sample data)

μ : constant average value

Weight value was determined by the semivariance value of data using a model of semivariogram spherical function. The equation of the spherical function that been used can be seen in equation 8 (Biswas & Cheng, 2013).

$$y(h) = \begin{cases} c_0 + c\left(\frac{3h}{2\alpha} - \frac{1}{2}\left(\frac{h}{\alpha}\right)^3\right), 0 < h \leq \alpha \\ c_0 + c, \quad h > \alpha \end{cases} \quad y(0) = 0 \quad (8)$$

Description:

$y(h)$: semi variance value

h : the distance between sample data

c_0 : *nugget* (distance between samples is 0)

c : *sill* (maximum semi variance value)

a : maximum limit of the distance value

After all of the process, quality control is conducted by calculating the RMSE value of the bandwidth model in determining land prices with an equation that can be seen in equation 9 (Nanja & Purwanto, 2015).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}} \quad (9)$$

Description:

y_i : real value

$\hat{y}_i$: the value of calculation results

i : 1, 2, ..., n

n : number of data

This study follows processes utilized in GWR 4 Software because it takes into account the GWR model calibration. Nakaya (2007) suggests that the algorithm in GWR 4 can be used to determine geographical relationship between dependent/response variables and independent/explanatory variable (Kemp, 2008; Nakaya, 2007).

3. Result and Discussion

In processing GWR by using an adaptive method, the equation is formed based on the number of points of data samples that will give an effect into models (bandwidth). Data input that required in the process of processing the data using the GWR are:

1. Sample point ID
2. Sample point coordinates (X and Y)
3. NIR as the dependent variable
4. The distance between the point of sample to each parameter as an independent variable
5. Choice of the type of kernel used (adaptive bi-square
6. Size of bandwidth

Results obtained from each model are the calculated land price (y'). The estimated land price is then compared to the actual land price to get the error value. GWR processing is done several times with different bandwidth sizes to obtain the optimum bandwidth. Bandwidth optimum can be determined based on the value of the error that is generated, the smaller the error value, the more precise the bandwidth size. The GWR processing is carried out at each sample point to obtain the optimum bandwidth for all sample data. This process is a different process from the GWR process, which generally uses the cross-validation (CV) or AICC method (Fotheringham et al., 2002). Figure 3 will show optimum bandwidth from every sample point. The distribution of optimum bandwidth is random with a dominant number of 14 all over the area.

Results of optimum bandwidth for several sample data of GWR equation results have a range of values between 14 to 420. The lowest bandwidth value shows the number of 14 because it needs a minimum of 14 observation data to complete the equation, which consists of 11 parameter values, one intercept value, and two other values. Bandwidth value of 420 becomes the upper limit based on the number of used

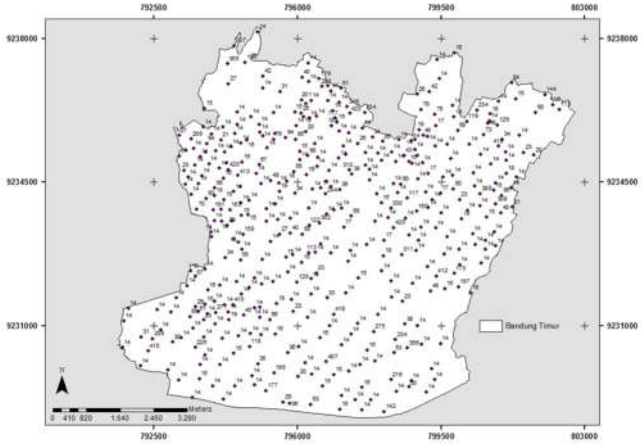

Figure 3. Optimum Bandwidth distribution.

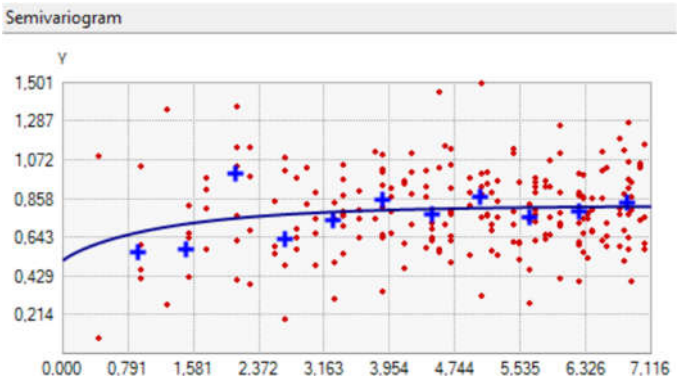

Figure 4. Semivariogram shape of the result sample data.

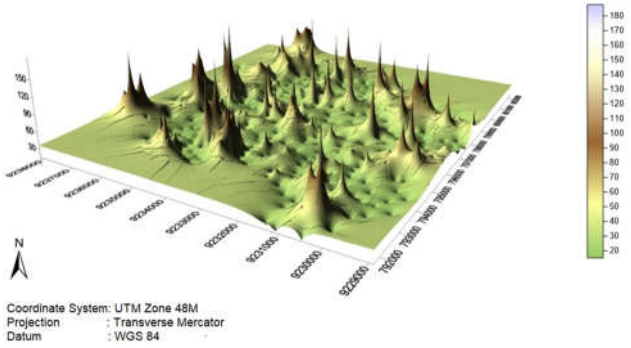

Figure 5. Model of Bandwidth Value on 3D Surface.

observation data, which is 420 sample data. From the result of the best bandwidth value, it can be seen that the best bandwidth value can be obtained when the sample point has the smallest bandwidth size and when the sample point has the most significant bandwidth size (it was affected by the whole of observation data).

It can be concluded that the optimum bandwidth in this study is obtained when the sample was taken is the sample that is closest to the sample point and when the sample used is all population data from the sample. These two conditions can look like different conditions. The first condition qualifies how the nearest point will give more significant weighting (Tobler, 1970) and nearest neighbor, which states that the closest point will have the most significant influence on the point sought. While the second condition meets the statistical requirements, which show that more data is involved, better results will be obtained.

The optimum bandwidth for each sample data is then interpolated to be able to estimate the bandwidth value of all non-sampled points of land (population). The selection of the kriging method is based on the characteristics of the sample data values that are known through the stages of the normal distribution test, statistical stationarity, and data trends (Krivoruchko, Gribov, & Krause, 2011). If the data characteristics do not meet these requirements, it can be overcome using a Normal Score Transformation. This transformation method works by ranking data from the lowest to the highest value, then matching it to the ranking generated by the normal distribution (ESRI, 2018). Normal Score Transformation is part of the simple kriging spatial interpolation method. This interpolation method performs the transformation of the data until it is normally distributed to be subsequently used in estimating values at other points.

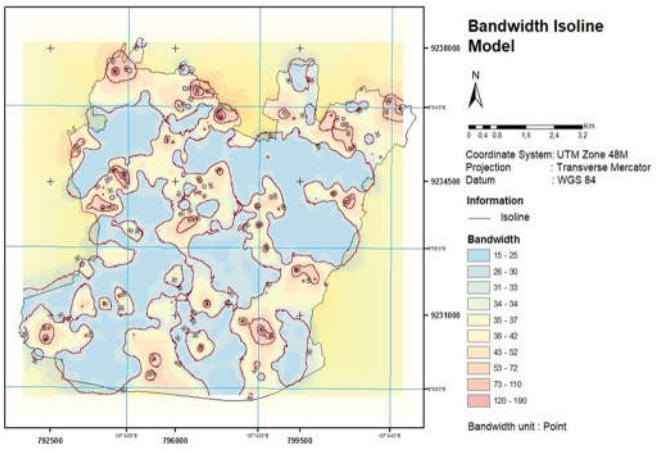

Figure 6. Isoline Model.

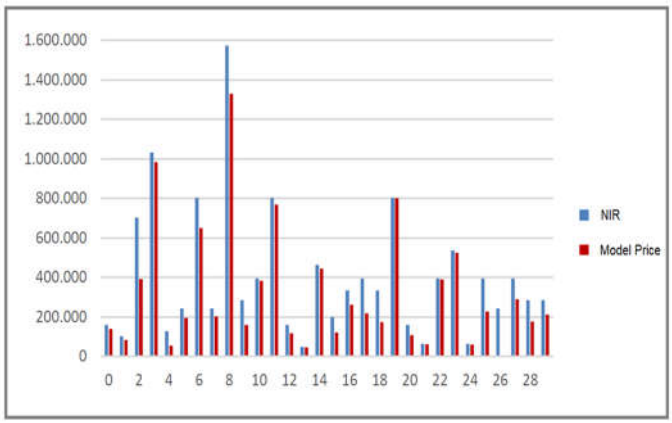

Figure 7. Comparison of data from NIR and model price.

Based on the results of normal data distribution tests that have been carried out, the spatial interpolation method that is suitable for use in the optimum bandwidth value data is simple kriging. After conducting the Normal Score Transformation to make the data normally distributed, this method forms a semivariogram based on the value of semivariance and the distance between the sample data held. Semivariogram models obtained a circular curve with a nugget value of 0.5 and a sill value of 0.819. The amount of the nugget value indicates the magnitude of the error rate in the selection of sample point locations. Semivariance value increases with the increasing value of the distance between sample data to the range of 474.4 m. The semivariogram shape of the result sample data can be seen in Figure 4.

The value change of the sample data interpolation results is caused by the simple kriging method uses the average value in the estimation process. This condition causes the results of interpolation to have a high degree of accuracy in data with values close to average and a low level of accuracy in data with values that move away from the average. Based on this concept, interpolation will obtain good results if the sample data values are normally distributed. The model of the bandwidth value of the entire land area, which resulting from Kriging spatial interpolation, can be seen in Figure 5, while a model of bandwidth value on isoline can be seen in Figure 6.

The result of kriging is each plot of land in the East Bandung area has a bandwidth measure that determines the price of land. A comparison of the magnitude of NIR to land prices calculated at 30 test points can be seen in Figure 7.

Based on the error value obtained at the test point, the reliability level of the land price bandwidth model can be determined using the GWR equation bi square method with kriging interpolation. The RMSE value obtained through the calculation of the error value of 30 test points is 114,600 rupiahs per meter square. RMSE resulting from the interpolation is not much better than the results of RMSE earlier.

What must be considered from this result is that this study interpolates the optimum bandwidth. Simple Kriging is used to interpolate data bandwidth from bandwidth results using GWR to get the optimum bandwidth value of the population so that the trial and error effort to get the optimum bandwidth can be minimized. This study does not compare or combine GWR with Kriging, as conducted by Szymanowski and Kryza (2011) and Meng (2014).

4. Conclusion

Kriging interpolation is built based on the value of the relationship between observational data expressed in a semivariogram form. Semivariogram can describe the value of the relationship between two data to the number of distance values in all observational data and will ultimately reflect the magnitude of errors in the selection of data samples. In this study, the selection of data, the type of interpolation, and error propagation from determining bandwidth interpolation play a very significant role in the resulting model.

From the results of the validation test, it was found that the prediction of land prices from the GWR results on the observation point was better than the predicted land prices from the estimation Kriging interpolation result outside the observation point. RMSE value of GWR results on observation point is 33,011 rupiahs per meter square while RMSE value of overall modeling results using Kriging spatial interpolation is increased significantly to be 114,600 rupiahs per meter square, this shows that kriging interpolation was not good enough to use for modeling optimum bandwidth interpolation for population data.

By looking at these results, further research will still be needed to study other possible spatial interpolation methods that can be tried to better model optimum bandwidth. Simple spatial interpolation methods or other methods such as Triangular Irregular Network (TIN) and Inverse Distance Weighting (IDW) can be tried to be applied in further research.

Acknowledgment

The authors would like to thank the Institute for Research and Community Service (LPPM), Bandung Institute of Technology, which has provided research funding through Research Program, Community Service and Innovation - *Program Penelitian, Pengabdian kepada Masyarakat dan Inovasi* (P3MI) ITB 2018.

References

Agnes, D., Nandatama, A., Isdyantoko, B. A., Nugraha, F. A., Ghivarry, G., Aghni, P. P., … Widayani, P. (2016). Remote sensing and GIS-based site suitability analysis for tourism development in Gili Indah, East Lombok. *IOP Conference Series: Earth and Environmental Science*, *47*(1), 1–4. https://doi.org/10.1088/1755-1315/47/1/012013

Biswas, A., & Cheng, B. (2013). Model Averaging for Semivariogram Model Parameters. *Advances in Agrophysical Research*. https://doi.org/10.5772/52339

Carr, J. R. (1994). *Numerical Analysis for the Geological Sciences* (1st ed.). London,UK: Pearson College Div.

Dziauddin, M. F., & Idris, Z. (2017). Use of geographically weighted regression (gwr) method to estimate the effects of location attributes on the residential property values. *Indonesian Journal of Geography*, *49*(1), 97–110. https://doi.org/10.22146/ijg.27036

ESRI. (2018). Geographically Weighted Regression (GWR). Retrieved January 1, 2019, from http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/geographically-weighted-regression.htm

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (1st ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Kemp, K. K. (2008). *Encyclopedia of Geographic Information Science*. Thousand Oaks, CA: SAGE Publications.

Krivoruchko, K., Gribov, A., & Krause, E. (2011). Multivariate areal interpolation for continuous and count data. *Procedia Environmental Sciences*, *3*, 14–19. https://doi.org/10.1016/j.proenv.2011.02.004

Li, J., & Heap, A. D. (2008). *A Review of Spatial Interpolation Methods for Environmental Scientists* (1st ed., Vol. 16). https://doi.org/10.1097/phh.0b013e3181e31d09

Manson, S. M., Burrough, P. A., & McDonnell, R. A. (1999). Principles of Geographical Information Systems: Spatial Information Systems and Geostatistics. *Economic Geography*, *75* (4), 422. https://doi.org/10.2307/144481

Meng, Q. (2014). Regression Kriging versus Geographically Weighted Regression for Spatial Interpolation. *International Journal of Advanced Remote Sensing and GIS*, *3*(1), 606–615.

Nakaya, T. (2007). Geographically weighted regression. In K. K. Kemp (Ed.), *Encyclopedia of Geographical Information Science* (pp. 179–184). Thousand Oaks, CA: SAGE Publications.

Nanja, M., & Purwanto, P. (2015). Metode K-Nearest Neighbor Berbasis Forward Selection Untuk Prediksi Harga Komoditi Lada. *Jurnal Pseudocode*, *2*(1), 53 – 64.

Schabenberger, O., & Gotway, C. A. (2004). *Statistical Methods for Spatial Data Analysis* (1st ed.). Boca Raton, FL: Chapman & Hall/CRC.

Sevilla, C. G., Ochave, J. T., Punsalan, T. G., Regala, B. P., & Uriarte, G. G. (1984). *An introduction to research methods*. Manila, Phillipines: Rex Book Store.

Szymanowski, M., & Kryza, M. (2012). Local regression models for spatial interpolation of urban heat island-an example from Wrocław, SW Poland. *Theoretical and Applied Climatology*, *108*(1–2), 53–71. https://doi.org/10.1007/s00704-011-0517-6

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46* (June 1970), 234–240. https://doi.org/10.2307/143141

Wand, M. P., & Jones, M. C. (1995). *Kernel smoothing* (1st ed.). London,UK: Chapman & Hall.

Wang, M., He, G., Zhang, Z., Wang, G., Zhang, Z., Cao, X., … Liu, X. (2017). Comparison of spatial interpolation and regression analysis models for an estimation of monthly near surface air temperature in China. *Remote Sensing*, *9*(12), 1–16. https://doi.org/10.3390/rs9121278

Ward, M. D., & Gleditsch, K. S. (2008). *Spatial Regression Models (Quantitative Applications in the Social Sciences)* (1st ed.). Thousand Oaks, CA: SAGE Publications.

Yang, Y., Tong, X., & Zhu, J. (2013). A geographically weighted model of the regression between grain production and typical factors for the Yellow River Delta. *Mathematical and Computer Modelling*, *58*(3–4), 582–587. https://doi.org/10.1016/j.mcm.2011.10.062

Yu, D.-L. (2006). Spatially varying development mechanisms in the Greater Beijing Area: a geographically weighted regression investigation. *The Annals of Regional Science*, *40*(1), 173–190. https://doi.org/10.1007/s00168-005-0038-2