Inverse Distance Weighting interpolation on the optimum distribution of kernel - Geographically Weighted Regression for land price

A P Handayani¹, A Deliar¹, I Sumarto¹ and I Syabri²

¹Geodetic and Geomatics Engineering, Institut Teknologi Bandung, Ganesha 10 Bandung
²Regional and City Planning, Institut Teknologi Bandung, Ganesha 10 Bandung

Abstract. Land, as a commodity, has high economic value that was rapidly changing. It requires a model of land price estimation that can counterbalance the demand. The model of land prices can be formed by calculating the proximity distance between land parcels to parameters that gave an effect on surrounding price changes, such as infrastructure and other public facilities. GWR (Geographically Weighted Regression) is a spatial analysis technique based on a regression method that used in this study. Using GWR, the boundaries of the surrounded land area will determine in the form of bandwidth values. The bandwidth is analogous to the radius of a circle, with the point of the land that will be estimated as the center. This bandwidth plays an important role to determine how far the distance or how many points will influence the price of a land area. The estimation of land price is conducted by assign a weight to the price of other parcels which are within the bandwidth range following the Kernel weighting function, Bi square, and Gauss. The land price bandwidth obtained is then interpolated using the IDW (Inverse Distance Weighting) method to form the bandwidth model of the entire study area. Weighting with Adaptive kernel Bi-Square gives better results compared to Adaptive Gauss kernel.

1. Introduction
All fulfillment of human needs occurs on land. The limited amount of land is accompanied by an increase in various factors such as an increase in population and economic growth, making the land a valuable resource and will continue to have a high value. An assessment of land prices also needs to be carried out for various purposes, including investment purposes, or collateral valuation by banks, and to set a fair land price for an area commonly called the Tax Object Selling Value (NJOP-Nilai Jual Objek Pajak). The assessment process will continue to be carried out because land prices will continue to change. The valuation of land prices can be done in various ways, one of which is a regression method that models the relationship between land prices and multiple variables that are thought to contribute to forming the land price. Variables forming the price of land can come from internal characteristics such as slope and fertility, or external characteristics related to how closely the features of the environment such as the location of the city center, or other major facilities. Thus, land that has the same characteristics intuitively will have relatively the same price. This regression method is commonly used in mass land valuation to produce accurate and efficient assessments.
In modeling land prices, in addition to paying attention to the price-forming variables, the land also has a spatial identity, and each has a unique tendency based on its location. However, the linear regression method does not take into account the spatial variation of the modeled land price. To overcome these shortcomings, one method that can be applied is the regression method using geographical weighting or GWR (Geographically Weighted Regression).

In this research, bandwidth price modeling in the East Bandung region was carried out using the GWR method with Adaptive Gaussian and Bi Square kernel type, then IDW (Inverse Distance Weighting), which is interpolation that takes distance as weight, will be used so that bandwidth prices can be modeled throughout the East Bandung region. Modeling the bandwidth of land prices can help to obtain information on land to be controlled for housing, economy, and others.

The purpose of this study is to determine the reliability level of bandwidth price modeling in the land using GWR with Gauss and Bi Square method based on IDW (Inverse Distance Weighting) interpolation.

2. Data and Methods
The input data used in this study was obtained from Bandung Regional Revenue Service (DISPENDA-Dinas Pendapatan Daerah) in the form of data on the Average Indication Value (NIR-Nilai Indikasi Rata-rata) and other data derived from aerial photographs of Bandung City at a scale of 1: 1000 from Bandung City Spatial Directorate (DISTARU-Dinas Tata Ruang). The data include:

1. Data on the administrative area of East Bandung
2. East Bandung road data
3. Data on determining land prices
4. 2007 NIR Data in East Bandung

2.1. Samples
Samples are parts of a population that are used to infer or describe a population. The selection of samples with the right method can describe the actual condition of the population and can save energy, time, and costs. A frequently asked question in the sampling method is how many samples are needed to represent a phenomenon. One method that can be used to determine the number of samples is using the Slovin formula [13], as follows. Sample distribution can be seen in figure 1.

\[ n = \frac{N}{1+Ne^2} \]  

\( n \): Number of samples  
\( N \): Total population  
\( e \): Error tolerance limit
2.2. Geographically Weighted Regression
GWR is a regression method that uses weighting based on location. The difference between the GWR method and the usual linear regression method is the mathematical model. In ordinary linear regression, all samples or data to be searched for parameter values can use one model. In case of GWR, it utilizes the weight that represents the distance between the estimated point location and the observed location. Therefore, each point has a different equation, or in other words, each point has its mathematical model. The GWR equation is stated as follows [7].

\[ y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \epsilon_i \]  

\((u_i, v_i)\) : shows the coordinates of the i-point.
\(\beta_0(u_i, v_i)\) : shows the realization of a continuous function at the point i.
\(\epsilon_i\) : indicates an error value.
\(x_{ik}\) : shows the independence variable.

2.3. Adaptive Spatial Kernel
The adaptive spatial kernel uses the nearest neighborhood as bandwidth to determine the number of points that have the same characteristics as the field to be searched for [7]. The adaptive spatial kernel can be seen in Figure 2.

[Diagram of GWR with Adaptive Spatial Kernels]

Spatial weighting functions that can be used on Adaptive spatial kernels are:

a) Adaptive Gaussian


\[ w_{ij} = \exp \left( -d_{ij}^2 / \theta_{i(k)}^2 \right) \]  

(3)

\( w_{ij} \): Weight Value  
\( d_{ij} \): Euclidean distance between i and j  
\( \theta_{i(k)} \): Adaptive bandwidth in the distance to the closest neighbor to k with metric units

b) Adaptive Bi-Square

\[
 w_{ij} = \begin{cases} 
 (1 - d_{ij}^2 / \theta_{i(k)}^2)^2 & d_{ij} < \theta_{i(k)} \\
 0 & d_{ij} > \theta_{i(k)} 
\end{cases} 
\]  

(4)

\( w_{ij} \): Weight Value  
\( d_{ij} \): Euclidean distance between i and j  
\( \theta_{i(k)} \): Adaptive bandwidth in the distance to the closest neighbor to k with metric units

2.4. Inverse Distance Weighting Interpolation

IDW (Inverse Distance Weighting) interpolation is deterministic interpolation that assumes that the interpolation value will be more similar to sample data that is nearby than remote sample data. IDW interpolation uses distance as a weight, so the sample points that are close by will have greater weight, and the amount of weight will decrease as the sample point distance increases. The mathematical equations of Inverse Distance Weighting interpolation are as follows [4].

\[
 Z_j = K_j \sum_{i=1}^{n} \frac{1}{d_{ij}} Z_i 
\]  

(5)

\[
 K_j = \sum_{i=1}^{n} \frac{1}{d_{ij}} 
\]  

(6)

\( Z_j \): Estimated value  
\( d_{ij} \): Distance to point I to j  
\( Z_i \): Sample point value  
\( n \): Number of data  
\( d_{ij}^p \): Distance from the point I to j with power  
\( K_j \): Distance strength value

The data validation utilized 30 test points that can be seen in figure 3.
2.5. Indicated Parameters

The indicated parameters that influence the land price of East Bandung District consist of Toll Gates, Malls, Universities, Hospitals, Schools, Stations, Terminals, Public Cemeteries, Mosques, Police Offices, Markets, Road Networks, and DEM (Digital Elevation Model). The distribution of all parameters can be seen in Figure 4.

3. Results and Discussion

3.1. Results of Geographically Weighted Regression Modeling

The result of modeling that is important in this study is the difference between the original NIR value and the NIR value generated by the model. The difference becomes the error value needed to get the optimum bandwidth value. The bandwidth value that produces the model with the smallest error becomes the optimum bandwidth. The error value for each sample data is obtained by calculating the difference between the NIR data and the NIR as a result of data processing in GWR4 with Adaptive Gaussian kernel types and Adaptive Bi Square. The minimum error value is an error value that is closest to zero.
From the sample data amounting to 420 points, the processing of using Adaptive Gaussian obtained the minimum error average value of 1161.51 rupiah/m², the lowest minimum error was -2.87 rupiahs/m² and the highest minimum error value was 1,248,599 rupiah/m². The calculation is done using Adaptive Bi Square; the average value of the error value of the sample data is Rp. -100. Whereas the standard deviation value is 33,000 Rupiah. There is an error value because of the different methods of determining the NIR data and the NIR calculated. NIR calculated results obtained from the Geographically Weighted Regression equation. While NIR data is determined based on Circular of the Ministry of Finance of the Republic of Indonesia, Director General of Tax Number KEP 55 / PJ.6 / 1999 concerning Technical Guidelines for Determining NIR Value, determination of NIR value must go through the preparation stage, selling price data collection, data compilation, data recapitulation and plot transaction data on ZNT work maps, data analysis, ZNT map making, also ZNT and NIR analysis books. So that with the different methods in determining the value of land, the resulting value is different. Besides, there may be other factors that determine the value of the land at a certain point so that it has a relatively large error value.

The modeling results using sample points on Adaptive Gauss produce a Root Mean Square Error (RMSE) value of 223,469 rupiah/m². This RMSE value shows the quality of the model generated from the GWR process with Adaptive Gaussian kernel types. While the results of modeling using sample points on Adaptive Bi Square produce a value of Root Mean Square Error (RMSE) of Rp. 33,011 rupiah/m². This RMSE value shows the quality of the model generated from the GWR process with the Adaptive Bi Square kernel type.

### 3.2. Inverse Distance Weighting (IDW) Interpolation Results

The interpolation result shows continuous value represented by raster data. In this study, the results of IDW interpolation with Adaptive Gaussian were generated, which can be seen in Figure 5.

![Adaptive Gaussian IDW interpolation results](image)

**Figure 5.** Adaptive Gaussian IDW interpolation results.

The results of interpolation based on the IDW method will be in the range of values used for interpolation. The value of the interpolation results is not smaller than the minimum value and is not higher than the maximum value of the input data used in interpolation. Therefore, there is no hilltop or deepest valley that can be described.

The model could be converted into an isoline map that describes the bandwidth. The continuous isoline shows the optimum bandwidth in the entire East Bandung region with an isoline hose of thirty so that the results can be seen in Figure 6.
The result of interpolation is a continuous value represented by raster data. The other results of IDW interpolation were produced with Adaptive Bi Square can be seen in Figure 7. From the raster, the model can also be made into an isoline map that describes the bandwidth using the isoline. The continuous isoline shows the optimum bandwidth in the entire East Bandung region with an isoline hose of thirty that can be seen in Figure 8.
3.3. Results of Model Validation with TEST Point
The validation needs to be done to determine the quality of the model. The validation is conducted by using 30 test points. The quality of the model can be obtained by calculating the Root Mean Square Error (RMSE). RMSE is the average difference between the value of the quadratic value derived from the NIR value of the test point with NIR test points based on Geographically Weighted Regression modeling based on Adaptive Gaussian and Adaptive Bi Square with a bandwidth that refers to the results of IDW interpolation. RMSE results using thirty test points produced RMSE of 306,611.5 rupiah/m² using Adaptive Gaussian and 84,733.0 rupiah/m² with Adaptive Bi Square. The interpolation used one-meter pixel size. Comparison of RMSE results at the sample point with RMSE at the test point can be seen in the following Table 1.

![Figure 8. Map of isoline bandwidth for land prices in East Bandung.](image)

| No. | Type, Function, Interpolation | RMSE (sample point) | RMSE (Test Point) |
|-----|-------------------------------|---------------------|-------------------|
| 1   | Adaptive, Gauss, IDW          | 223.469             | 306.622           |
| 2   | Adaptive, BiSquare, IDW       | 33.011              | 84.733            |

4. Conclusions and Suggestions
4.1. Conclusions
The weighting type used for modeling land prices using the Geographically Weighted Regression (GWR) method has a substantial effect on results. Weighting with Adaptive kernel Bi-Square gives better results compared to Adaptive Gauss kernel. One of the characteristics of Bi-Square Weighting is that this weighting can eliminate the influence of data that is outside of bandwidth. The value of the y-axis or weight value will be close to zero if the value of the x-axis or distance approaches the bandwidth value. The graph is different from the figure of gauss weighting, where the value of the y-axis still has not reached zero or is always high when the value of the x-axis approaches the bandwidth value. The nature of Bi-Square Weighting can be used to find out which sample data is included in the calculation.
in modeling one sample data with a particular bandwidth value. These characteristics have made weighting using adaptive kernel Bi square will give better results.

4.2. Suggestions
Further research is needed to calculate the level of reliability and increase the level of reliability of the Inverse Distance Weighting (IDW) Interpolation process in applying the Average Indication Value (NIR) model to the test data so that the RMSE value from the test data is free from errors caused by the process interpolation. Further research is needed regarding the causes of changes in the model and error value of modeling one sample data when changing the adaptive type bandwidth value used.

5. References
[1] Anderson S, 2001 An evaluation of spatial interpolation methods on air temperature in Phoenix, AZ [paper] Department of Geography, Arizona State University Tempe.
[2] Blair, Edward & Johnny Blair. 2015 Applied Survey Sampling United States of America: SAGE Plications, Inc
[3] Dale P F & McLaughlin J. 1999 Land Administration New York: Oxford University Press
[4] De Smith M J, Goodchild M F and Longley P A 2015 Geospatial Analysis: A Comprehensive Guide to Principles, Techniques, and Software Tools. 5th Edition. http://www.spatialanalysisonline.com/HTML/
[5] Ernawari R 2006 Studi Pemodelan Nilai Tanah di Kota Tulungagung Kabupaten Tulungagung. Malang: Universitas Brawijaya
[6] Fink Arlene 2006 How to Conduct Survey: A Step-by-Step Guide United State of America: Sage Publications, Inc
[7] Fotheringham A S, Brunsdon C & Charlton M 2002 Geographically Weighted Regression, the Analysis of Spatially Varying Relationships England: John Wiley & Sons, Ltd
[8] Gruver J A 2017 Penn State College of Earth and Mineral Science. The Pennsylvania Stae University
[9] Harjanto W H 2003 Konsep Dasar Penilaian Property Yogyakarta: BPPE-Yogyakarta
[10] Margono 2004 Metodologi Penelitian Pendidikan Jakarta: PT Rineka Cipta.
[11] Nursalam. 2008 Konsep dan Penerapan Metodologi Penelitian Ilmu Keperawatan Jakarta: Salemba Medika.
[12] Pangestuti D R 2012 Aplikasi Teknik Principal Component Analysis Untuk Pemodelan Nilai Tanah Bandung: Teknik Geodesi dan Geomatika, ITB
[13] Sevilla, Consuelo G. et al. 2007 Research Methods. Rex Printing Company. Quezon City
[14] Watson D F & Philip G M 1985 A Refinement of Inverse Distance Weighted Interpolation Geo-Processing 2: 315-327