Modified estimation of land values with spatial weight in Bandung city

N T Sugito*, I Soemarto2, S Hendriatiningsih2, B E Leksono2

1Department of Geography Education, FPIPS, Universitas Pendidikan Indonesia
2Institut Teknologi Bandung
*naniatrianawati@upi.edu

Abstract. Land valuation is one aspect of a cadastral system that is integrated with land use and land tenure. Land can be valued based on the benefits that can be given. High and low land values are influenced by many factors, including economic, social, government, and physical factors. Based on factors that can determine the value of land, it is known that land use (zoning) can also affect the formation of land values. Several studies have stated that land use is the most dominant factor affecting land values. This study uses geostatistical analysis in modeling soil values. The purpose of this research is to modify the mathematical model of soil values from geostatistical analysis by including spatial weights. The mathematical model of the modified land value is expected to reflect the actual value, which can then be used in creating a Land Value Zone.

1. Introduction
The phenomenon of land valuation is still found to be a problem, including: (1) the assessment requires hard work, long time, and expensive costs associated with the number of plots of land is very much; and (2) the results of the assessment achieved do not reflect the true value. Statistical analysis of land value models in general still uses simple regression. The analysis does not take into account the influence of location (X, Y). Through this research, a new assessment method can be found that can be a solution to the problems that occur. Land value models are generated through geostatistical analysis. Geostatistics is the development of mathematics and statistics to measure the distribution of an event based on space. In this case, the term spatial includes variables on the surface of the earth such as topographical conditions, vegetation, water, and others. The fields of geology and mining have applied a lot of geostatistical models. Geostatistics was developed to evaluate problems related to geological symptoms to the mine, starting from reserve estimation to grade control problems [1].

Geostatistical Method is an interpolation method that is able to model spatial trends and spatial correlations between several samples being measured. The existence of trends and spatial correlations are in line with Law I of Geography, that is, those who are closer have a closer relationship than those who are farther away. Through geostatistical analysis it can produce estimated parameter values in places where data is unknown. Geostatistical interpolation methods in addition to making the predicted area can also include some measure of the level of uncertainty or error of the predicted results [2]. In geostatistics there is a term known as spatial autocorrelation. Spatial autocorrelation shows the existence of similarities of objects in a space. Spatial autocorrelation is indicated by a systematic pattern in the distribution of a variable. Estimated values in one area will correlate with values in other regions that are in an adjacent position. This study analyzes data using the geostatistical method which will be applied in modeling land values. The mathematical model of soil values from the geostatistical analysis...
will be modified to include spatial weights. The modified land value mathematical model is expected to reflect the actual value and can be used in making the land value zone.

2. Methods
Sample data obtained through a process of measurement or observation that is assumed to represent a population. Sampling is done by purposive sampling technique. The sampling technique is done by the method of purposive sampling which means that the population that is used as the sample of the study is the population that meets certain sample criteria according to the research objectives and what must be represented depends on the judgment or consideration of the researcher [3]. In this study, sampling was carried out in stages, namely: establishing all sub-district areas in each development area to be used as research locations; the number of samples is at least 30% of the total land parcels in each district; and samples pay attention to variations in existing land use. The data used in this study is the average indication value with the number of sample points as much as 30% with the cluster sampling method with the use of purposive sampling in each development area to then be combined into one Bandung City sample that a total of 30% of data is obtained in the areas of development Cibeunying, Gedebage, Ujungberung, Tegalega, Karees, and Bojonegara to then be combined into one sample data Bandung. The sampling scheme is presented in Figure 1.

3. Results and discussion
3.1. Spatial autocorrelation testing
Spatial autocorrelation is indicated by Moran’s scatterplot which places a lot of observations in the HH quadrant and LL quadrant will tend to have positive spatial autocorrelation values while the scatterplot which places many observations in the HL and LH quadrants will tend to have negative spatial autocorrelation values [4]. Figure 2a shows a lot of data plots gathered in quadrant I (HH), which means the area that has a high average indication Value is surrounded by areas that have a high average indication value. This proves that the average indication value in Bandung City tends to have a positive spatial autocorrelation value (Figure 2b).
3.2. Geostatistical Analysis

In geostatistics there is a basic tool for the visualization, modeling and exploitation of spatial autocorrelation of reoriented variables called semivariograms [5]. Semivariogram is a mathematical model of semivariance as a function of lag distance. In the cartesian axis cross, the semivariogram visualized the Y axis as semivariance and the X axis as the lag distance [6,11]. Semivariogram is used to model the correlation of two points in space [7]. The model is usually determined by the least squares regression method. Semivariance (γ) is calculated to describe the expected value of the difference in sample value (z) as a function of the lag distance (h) between pairs of sample points. This semivariogram is used to measure spatial correlation in the form of error variance at a location [8]. In geostatistics there are two types of semivariogram, namely experimental semivariogram and theoretical semivariogram. Semivariograms generated from measurement data are called experimental semivariograms [9]. Theoretical semivariogram has a certain mathematical formula. There are several types of theoretical semivariograms that are often used, namely the Stable, Spherical, Exponential, and Gaussian models. In order to calculate the estimated land value, the experimental semivariogram must be replaced with a theoretical semivariogram [10]. In this study, the position of each plot of land that was originally in the form of polygons must be converted to a point (centroid). Each centroid represents the position (X, Y) of each plot of land. average indication value will be attached to each centroid, namely as a Z value. Furthermore, a geostatistical analysis is performed in which the semivariogram of the Stable, Spherical, Exponential, and Gaussian models is visualized in the following Figure 3.

![Figure 3. Semivariogram models.](image)

### Table 1. Standard Deviations in development areas

| Development area | Standard deviation |
|------------------|--------------------|
|                  | Initial conditions | Conditions after geostatistical analysis |
| Bojonagara       | 739932.993         | 562890.723          |
| Cibeunying       | 888063.730         | 744987.516          |
| Gedeage          | 474332.906         | 446668.038          |
| Karees           | 1034827.257        | 828027.723          |
| Tegalega         | 623251.237         | 501741.570          |
| Ujungberung      | 487273.218         | 464831.108          |
A model can be said to be good if the cross validation model error parameters meet several requirements: (1) Mean Error approaches 0; (2) Root-Mean-Square Standardized approaches 1; (3) Mean Standardized approaches 0; and (4) the amount of Root-Mean-Square is almost the same as the Average Standard Error. Based on these requirements, each development area has an estimated land value model. The Gaussian semivariogram model is the best model for estimating land values in the Bojonagara development area. The Exponential semivariogram model is the best model for estimating land values in the Cibeunying, Gedebage, and Tegalega development areas. The Stable semivariogram model is the best model for estimating land values in the Karees development area. Circular semivariogram model is the best model for estimating land values in the Ujungberung development area (see Figure 4-9). The following is the visualization of estimated land values for each development area. Based on the estimated land values generated from the semivariogram model, there is a decrease in the standard deviation values for each development area as follow Table 1.

![Figure 4. Estimated land value of the Gaussian type in Bojonagara development area.](image)

![Figure 5. Estimated land value of the Exponential type in Cibeunying development area.](image)
Figure 6. Estimated land value of Exponential type in Gedebage development areas.

Figure 7. Estimated land value of the Stable type in Karees development area.

Figure 8. Exponential land value estimated in Tegalega development area.
3.3. Spatial Weight

Furthermore, the estimated land values from the best semivariogram model are processed using multiple linear regression. The dependent variable (y) is in the form of estimated land values obtained from geostatistical analysis. The independent variable (x) consists of the centroid distance to the primary center and secondary center based on the Bandung City Regional Spatial Plan 2004-2013, including X1 as centroid distance to the secondary center of Setrasari, X2 as centroid distance to the primary center of the square, X3 as centroid distance to Kopo Kencana secondary center, X4 as centroid distance to Sadang Serang secondary center, X5 as centroid distance to Turangga secondary center, X6 as centroid distance to the secondary center of Arcamanik, X7 as centroid distance to Margasari secondary center, and X8 as centroid distance to the primary center of the Gedebage.

Table 2. Regression parameters.

| Development area | Multiple R | R2 (R-square) | Adjusted R square | Standard error | p-value |
|------------------|------------|---------------|-------------------|----------------|---------|
| Bojonagara       | 0.863410985 | 0.74547853    | 0.745116901       | 524120.7181    | 2.99912E-10 |
| Cibeunying       | 0.854675343 | 0.730469941   | 0.730284205       | 583270.3902    | 0.106509052 |
| Gedebage         | 0.859582864 | 0.738882701   | 0.738652591       | 386357.8604    | 0.037348279 |
| Karees           | 0.845379092 | 0.714665809   | 0.714362035       | 683235.3901    | 0.008592341 |
| Tegallega        | 0.841634176 | 0.708348086   | 0.707942653       | 464345.9826    | 0.026512717 |
| Ujungberung      | 0.76942962  | 0.592021941   | 0.591770844       | 447769.1071    | 0.092613484 |

Multiple R is a measure of the level (closeness) of the linear relationship between the dependent variable and all independent variables together (see Table 2). In all development areas it can be seen that the level of closeness of the relationship of land values with the centroid distance to the primary center and secondary center is quite high. This shows that the distance of the plot of land to the primary center and secondary center affects the estimation of land value. R2 (coefficient of determination) can be used to measure the goodness (goodness of fit) of the regression equation. The value of R2 lies between 0 - 1, and the compatibility of the model is said to be better if R2 gets closer to 1. The Ujungberung development area has the lowest R2 value (away from 1). The Standard Error will be compared with the standard deviation of the Average Indication Value (conditions after geostatistical...
analysis). The standard error generated by the multiple regression analysis is lower than the standard deviation of the average indication value (conditions after geostatistical analysis). This shows that the regression model is more appropriate in estimating land values. The results of the spatial weights in the development area of Bandung shows in Table 3. The estimated standard deviation of the land value after the spatial weighting has decreased when compared to the standard deviation of the conditions after the geostatistical analysis. Comparison of standard deviations is presented in Table 4.

Table 3. Spatial weights in the development area of Bandung City.

| Independent Variable | Bojonagara | Cibeunying | Gedebage | Karees | Tegalega | Ujungberung |
|----------------------|------------|------------|-----------|--------|----------|------------|
| X1                   | -299.7603162 | -109.2974136 | -131.2078469 | -551.8231376 | -295.8046601 | -889.4181581 |
| X2                   | -1404.315868     | -2622.930626     | -3063.89428     | -1105.936334  | -323.1673385  | -16938.74208 |
| X3                   | 909.2813374      | 2806.116782      | 1867.814404     | 1461.059499   | 534.6947216   | 10595.85082  |
| X4                   | 1515.192494      | 263.2895919      | 768.9073826     | 423.0904148   | -1375.94639   | 4321.516804  |
| X5                   | 5435.526955      | 339.6014725      | 7238.9717387    | -61.4223205   | -1242.248717  | 3584.055667  |
| X6                   | -7277.193912     | -1362.790557     | -175.633576     | 811.8059404   | 4377.69678    | -157.149081  |
| X7                   | -15252.51731     | -1474.855224     | -32.44322124    | 532.1307105   | -367.0544625  | -465.7771427 |
| X8                   | 16010.76364      | -33.50194479     | 186.7117588     | -854.3833221  | -1812.421714  | 9.071094696  |

Table 4. Comparison of standard deviations in development areas.

| Development area | Conditions after geostatistical analysis | Conditions after weighting for spatial planning |
|------------------|------------------------------------------|-----------------------------------------------|
| Bojonagara       | 562890.723                               | 206050.986                                    |
| Cibeunying       | 744987.516                               | 463713.954                                    |
| Gedebage         | 446668.038                               | 223807.453                                    |
| Karees           | 828027.723                               | 468350.081                                    |
| Tegalega         | 501741.570                               | 190226.369                                    |
| Ujungberung      | 464831.108                               | 124716.337                                    |

4. Conclusions
The estimation model of land values based on geostatistical analysis has been produced for all development areas in the city of Bandung and spatial weight is very influential on the estimated land value. Mathematical models of land values can be used in making the creation of land value zones to further be used as a single value for various purposes. After spatial weights are generated for the entire development area, the next step is the validation of the estimated land value model. This is done to determine the level of accuracy of the estimation model of the resulting land value.

References
[1] Chilès J P and Delfiner P 2012 Geostatistics: Modeling Spatial Uncertainty (New York: John Wiley & Sons, Inc.)
[2] Sari D K, Nugroho H and Hendriawaty S 2010 Pemodelan harga tanah perkotaan menggunakan metode geostatistika Jurnal Rekayasa 14 (2): 60-71.
[3] Topp, L., Barker, B., & Degenhardt, L. (2004). The external validity of results derived from ecstasy users recruited using purposive sampling strategies J. Drugalcdep 73: 33-40.
[4] Jacquez J M 2014 SpaceStat 4.0 User Manual: Software for the Space-Time Analysis of Dynamic Complex Systems (New York: BioMedware).
[5] Patil A P, Gething P W, PieI F B and Hay S I 2011 Bayesian geostatistics in health cartography: The perspective of malaria Trends in Parasitology 27 (6): 246-253.
[6] Uyan M and Cay T 2013 Spatial analyses of groundwater level differences using geostatistical modeling Environmental and Ecological Statistics 20: 633-646.
[7] Van Beurden S H and Riezebos H T 1988 The application of geostatistics in erosion hazard mapping Soil Technology 1 (4): 349-364.
[8] Arslan H 2014 Estimation of spatial distribution of groundwater level and risky areas of seawater...
intrusion on the coastal region in Çarşamba Plain, Turkey using different interpolation methods *Environ. Monit. Asses*. 186 (8): 5123-5134.

[9] Martiniano E and Silveira D O 2015 Change Detection in Brazilian savannas using semivariograms derived from NDVI images *Ciênc. Agrotec*. 39 (2): 103-109.

[10] Pham T D and Wagner M 1998 A geostatistical model for linear prediction analysis of speech *Pattern Recognition* 31 (12): 1981-1991.

[11] Widiawaty M A, Dede M and Ismail A 2018 Comparative Study modeling of ground water using geographic information system in Kayuamhon Village, Bandung Barat Regency *Jurnal Geografi Gea* 18 (1): 63-71.