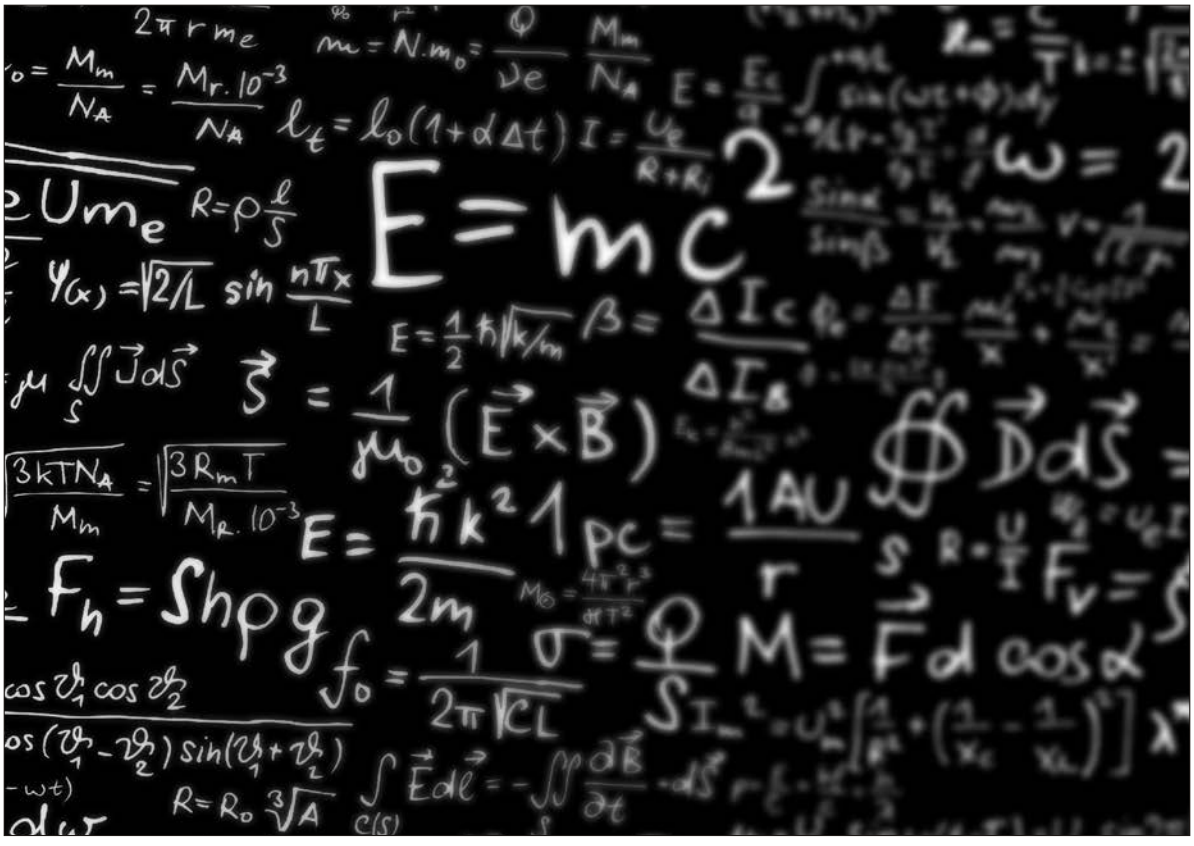

Asian Journal of Mathematics & Statistics

ISSN 1994-5418

science alert

ANSInet
an open access publisher
http://ansinet.com

Asian Journal of Mathematics and Statistics 5 (4): 132-141, 2012
ISSN 1994-5418 / DOI: 10.3923/ajms.2012.132.141

Land Price Model Considering Spatial Factors

[1]Asep Saefuddin, [2]Yekti Widyaningsih, [3]Ardinata Ginting and [4]Mustafa Mamat
[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Indonesia
[2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Indonesia, Indonesia
[3]Department of Geograhpy, Faculty of Mathematics and Natural Sciences, University of Indonesia, Indonesia
[4]Department of Mathematics, Faculty of Science and Technology, University Malaysia Terengganu, Malaysia

Corresponding Author: Asep Saefuddin, Department of Statistics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Malaysia

ABSTRACT

Many studies have highlighted that Ordinary Least Square (OLS) regressions lack the ability to consider spatial dependency including spatial non-stationarity which then lead to bias and inefficient estimations. Land prices are usually depending on locations yielding the prices vary from place to place. Therefore, estimates obtained from the OLS ignoring spatial factors may be inappropriate. Geographically Weighted Regression (GWR) is an alternative model considering spatial non-stationarity. In addition to its appropriateness, GWR produces local specific parameter estimates which then are very useful for the policy makers to avoid a misleading judgment. Some geographic social characteristics and related infrastructures are often used as the determinants or the explanatory variables of applied models for the land prices. In this study, the residuals of both GWR and OLS models were contrasted to obtain the best model. The maps of coefficients of the determinants varied from place to place. Also, we found that the most influential factors were the distance to the high-way gate, the distance to the artery-road and the distance to public facilities.

Key words: Ordinary least square, geographically weighted regression, spatial non-stationarity

INTRODUCTION

Land is a human essential need in the earth. Everyone needs land for living, work and many other activities. The role of land is more important than other economic resources due to its necessary economic condition as well as a place to stay. As an economic resource, appropriate land prices are important for government to provide fair decisions. To analyze the land price data, statistical approaches, such as regression models, are very useful to have right recommendations. Many studies have highlighted that Ordinary Least Square (OLS) regression lack the ability to consider spatial effect which then lead to biased and inefficient estimations (Lochl and Axhausen, 2010). On the other side, Geographically Weighted Regression (GWR) provides good alternative for analyzing spatial data including land prices. The purpose of this research is to compare the OLS and GWR approaches applied on the land price data in South Tangerang, Indonesia.

Land prices, similar to house prices in general are variable depending on the locations. Land price determinant is associated with many factors, not only the environmental conditions but also government policies and the factors of socioeconomic values. Some of the factors are the status

of the land (certified, not certified), the distance from high-way gate, the distance from artery-road, developed/not-developed land, the social characteristic of area (safeness, proper/not-proper, crime rate) residence, the distance from market, the distance from public health facility, the distance from security facility and the distance from education facility. Many studies also indicated that land prices were highly correlated with land use and the growth of regions (Luo and Wei, 2009).

Ignoring factors including areas determining land prices is ridiculous. Usually, a simple approach is to run a linear regression model involving some possible determinants as covariates, such as the distance from high way gate and/or many social characteristics of the area. In the literature, this approach is known as hedonic model of land or house prices (Cheshire and Sheppard, 1995; Lochl and Axhausen, 2010). The technique produces single parameter estimate for each covariate which then called a global model. Although the model provides useful information based on the parameter estimates, it is not precise enough. Luo and Wei (2009) found that the use of GWR improved significantly the global model in terms of model goodness-of-fit and residual autocorrelation. Also Saefuddin *et al.* (2011) found that the GWR performed better than the OLS when applied on poverty data in Indonesia. We, then implemented GWR to analyse land prices in South Tangerang, Province of Banten, one of the urban areas in Indonesia which is very fast moving.

The developments in Jakarta and its surroundings area called Bodetabek (Bogor, Depok, Tangerang and Bekasi) have showed a very rapid improvement since last decade. Physically, as the hinterland of Jakarta, Bodetabek has prime quality for residential area. The rapid development in Jakarta also has accelerated the urbanization process yielding the growth rate of economic investments, trade and industry in Bodetabek. Hence, Jakarta and its hinterland (Bodetabek) region have been continuing to develop into a Mega City Urban. Land prices in these areas are then very crucial to be analyzed.

This study is aimed at modeling the land price in South Tangerang area as a part of BODETABEK. The analysis was based on several explanatory variables including the status of the land (certified/not certified), the distance from high-way gate, the distance from artery-road, developmental status (developed/not-developed land), the characteristic of area (rural, rural-urban, urban), the social characteristics (proper/not-proper residence), the distance from market, the distance from public health facility, the distance from security facility and the distance from education facility. The data set was analyzed by ordinary regression (OLS) and Geographically Weighted Regression (GWR). The results from both approaches were contrasted using some statistical criteria.

MATERIALS AND METHODS

The data: The study area of this research was in South Tangerang, a district in Banten province, Indonesia. The location is in the border of Jakarta and Banten province. Therefore, South Tangerang has been developing rapidly since the last decade and many developers have expanded their properties in this area. Concerned with property and local government policy, the land prices have become an interesting and important topic to be analyzed. The data was taken from 1050 location-points in South Tangerang using variables listed in the Table 1. The data was obtained by a survey to the land owners and the heads of sub-districts in South Tangerang.

Methods

Geographically weighted regression (GWR): Compared to OLS, GWR is considered more appropriate method for data of non-stationary regions. The data is called in stationary condition if the data is homogenous across the study region, otherwise it is called non-stationary.

Table 1: Variables used in the land price model

No.	Name of variable	Type of distribution	Description
1	land price	Continuous	The land prices (IDR) m^{-2}
2	Status	Binary	The Land status (certified: 1, not certified: 0),
3	Highway	Continuous	The distance from high-way gate (m)
4	Artery	Continuous	The distance from artery-road (m)
5	Developed	Binary	The land development (developed: 1, not-developed area: 0)
6	Landchar	Dummy	The characteristic of area (rural: 1, rural-urban: 2, urban: 3)
7	Proper	Binary	The characteristic of area (proper: 1, not-proper: 0 residence)
8	Market	Continuous	The distance from market (m)
9	Health	Continuous	The distance from public health facility (m)
10	Secure	Continuous	The distance from security facility (m)
11	Educ.	Continuous	The distance from education facility (m)

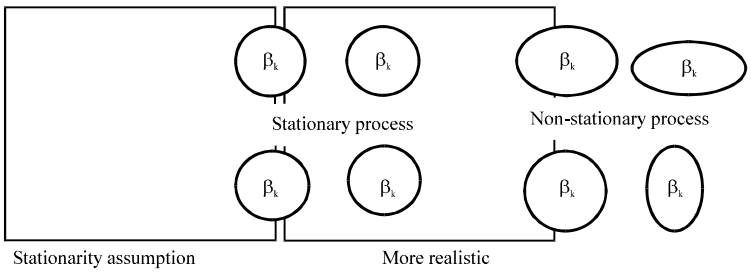

Fig. 1: Stationary versus non-stationary

Fotheringham *et al.* (2002) named this phenomenon is the spatial non-stationarity. Figure 1 indicates parameter of stationary and non-stationary regions.

In the case of spatial non-stationarity, the same stimulus (determinant) provokes different response in different parts of the region which then the approach is termed local specific models as opposed to global models. Fotheringham *et al.* (2002) listed differences between local and global models which are important to consider in analyzing spatial data.

Global model is a linear regression in which the simple one with formula as $y_i = \beta_o + \beta_1 x_1$ is used for all locations assuming regional stationarity. The parameter estimates do not change across areas, have single-valued statistics, cannot be mapped, Geographic Information System (GIS) unfriendly and aspatial (Fotheringham *et al.*, 2002; Yu and Wei, 2004). On the other hand, GWR has the formula as $y_i = \beta_{io} + \beta_{i1} x_1$ in which the betas are not the same for each location and hence, called as local models. Every location has its own parameter estimates. This local statistics are different across space, has multi-valued statistics and can be mapped using GIS.

Let global regression have an equation as follows:

$$y_i = \beta_o + \sum_k \beta_k x_{ik} + \varepsilon_i$$

where, y_i is the value of dependent variable for observation i, β_o is the intercept, β_k is the parameter for variable k, x_{ik} is the value of the k-th variable for i and ε_i is the error term. Instead of calibrating a single regression equation, GWR generates a separate regression equation for each observation. Each equation is calibrated using a different weighting of the observations contained in the data set. Each GWR equation may be expressed as:

$$y_i = \beta_o(u_i, v_i) + \sum_k \beta_k (u_i, v_i)x_{ik} + \varepsilon_i$$

where, (u_i, v_i) captures the coordinate location of i (Fotheringham *et al.*, 1998). Following Tobler (1970) observations nearby one another have a greater influence on one another's parameter estimates than observations farther apart. The weight assigned to each observation is then based on a distance decay function centered on observation i. In the case of areal data, the distance between observations is calculated as the distance between polygon centroids.

Parameter estimation: In linear regression, the parameter estimates obtained in the calibration of a model are constant over space:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

while in GWR which the equation is as follows:

$$y_i = \beta_o(i) + \beta_1(i)x_{1i} + \beta_2(i)x_{2i} + ... + \beta_n(i)x_{ni} + \varepsilon_i$$

with the estimator $\hat{\beta} = (X^T W(i) X)^{-1} X^T W(i) Y$ where, W(i) is a matrix of weights specific to location i such that observations nearer to i are given greater weight than observations further away. An example of W(i) is as the following:

$$W(i) = \begin{pmatrix} W_{i1} & 0 & 0 & ... & 0 \\ 0 & W_{i2} & 0 & ... & 0 \\ 0 & 0 & W_{i2} & ... & 0 \\ ... & ... & ... & \ddots & ... \\ 0 & 0 & 0 & ... & 0 \\ 0 & 0 & 0 & ... & W_{in} \end{pmatrix}$$

where, w_{in} is the weight given to data point n for the estimate of the local parameters at location i. Numerous weighting schemes can be used although they tend to be Gaussian or 'Gaussian-like' reflecting the type of dependency found in most spatial processes.

Weighting schemes can be either fixed or adaptive. A fixed weighting schemes has w_{ij} as:

$$W_{ij} = \exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right]$$

where, d_{ij} is the distance between locations i and j and h is the bandwidth -as h increases, the gradient of the kernel becomes less steep and more data points are included in the local calibration. For fixed weighting schemes, h is fixed for all location i. We need to find the optimal value of h in the GWR routine.

A spatially adaptive weighting function has w_{ij} as:

$$W_{ij} = \left\{\left[1-\frac{d_{ij}^2}{h^2}\right]^2\right\}; \quad \text{if j is one of the Nth nearest neighbours of i otherwise}$$

The optimal value of h has to be found by minimizing a cross-validation score (CV) or the Akaike Information Criterion (AIC).

The distance decay function which may take a variety of forms is modified by a bandwidth setting in which distance the weight rapidly approaches zero. Bandwidth is a distance between center of a polygon or grid to a particular border. There are two types bandwidth, i.e., fixed or adaptive. The bandwidth may be manually chosen by the analyst or using an algorithm that seeks to minimize a cross-validation score:

$$CV = \sum_{i=1}^{n} (y_i - y_{i\neq i})^2$$

where, n is the number of observations and observation i is omitted from the calculation so that in areas of sparse observations the model is not calibrated solely on i. Alternatively, as mentioned before, the bandwidth may be chosen by minimizing the Akaike Information Criteria (AIC) score, given as:

$$AIC_c = 2n\log_e(\tilde{\sigma}) + n\log_e(2\pi) + n\left\{\frac{n + 2\,tr(S)}{n - 2\,tr(S)}\right\}$$

where, tr(S) is the trace of the hat matrix. The AIC method has the advantage of taking into account the fact that the degrees of freedom may vary among models centred on different observations. In addition, the user may choose a fixed bandwidth for every observation or a variable bandwidth that expands in areas of sparse observations and shrinks in areas of dense observations. Optimal bandwidth selection is a trade-off between bias and variance. Too small a bandwidth leads to large variance in the local estimates and too large a bandwidth leads to large bias in the local estimates.

Because the regression equation is calibrated independently for each observation, a separate parameter estimate, t-value and goodness-of-fit is calculated for each observation. These values can thus be mapped, allowing the analyst to visually interpret the spatial distribution of the nature and strength of the relationships among explanatory and dependent variables (Fotheringham *et al.*, 2002). The map of parameter estimates was facilitated in R-software developed by Bivand and Yu (2011).

Result of ordinary regression: This section gives the results of OLS and GWR computation. The outcome (land price) was transformed using natural logarithmic to be normally distributed.

Figure 2 is the normal Q-Q Plot of the dependent variable (land price). The result of natural logarithmic transformation (Fig. 2b) indicated its normality. Furthermore, modeling analysis used natural logarithmic (ln) of land prices as the dependent variable. The analysis was implemented based on the logarithmic scale.

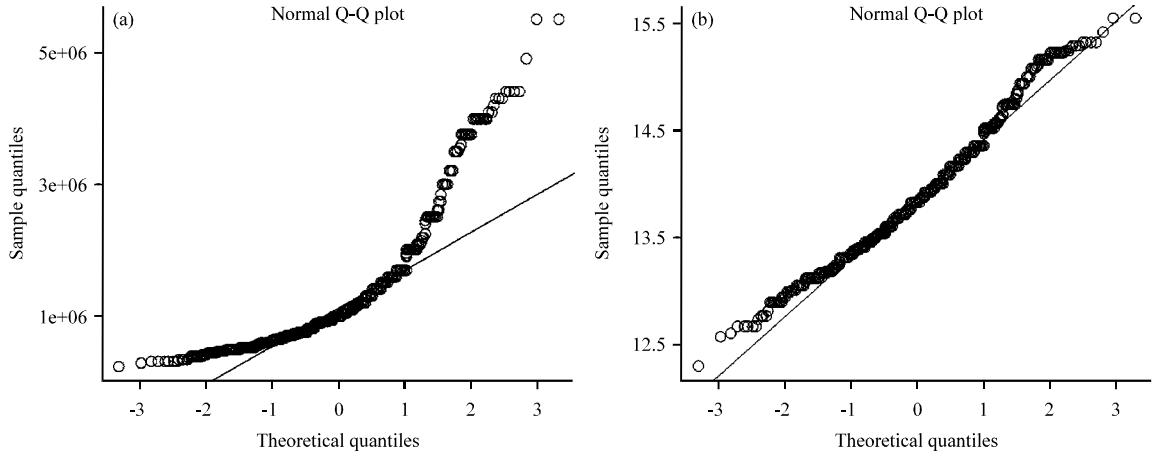

Fig. 2(a-b): The normal Q-Q plot of the dependent variable (land price and land for natural logarithmic (land price)

Table 2: OLS parameter estimation for all explanatory variables

Variable	Estimate	Std. error	t-value	Pr(>|t|)	Significance
Intercept	1.42E+01	1.28E-01	110.872	< 2E-16	***
Status	-2.63E-02	4.24E-02	-0.62	0.5352	#
Highway	-4.39E-05	9.61E-06	-4.574	5.37E-06	***
Artery	-1.29E-04	3.62E-05	-3.559	0.000389	***
Developed	-1.12E-01	4.62E-02	-2.429	0.015315	*
Landchar 2	1.47E-01	1.05E-01	1.399	0.162185	#
Landchar 3	1.68E-01	1.00E-01	1.68	0.093354	-
Proper	3.99E-01	9.65E-02	4.133	3.87E-05	***
Market	-2.77E-05	3.65E-05	-0.758	0.448474	#
Health	-8.65E-05	1.91E-05	-4.532	6.51E-06	***
Secure	3.08E-05	1.98E-05	1.561	0.118854	#
Educ.	-8.41E-05	5.66E-05	-1.485	0.137903	#

Significant at #1, -: 0.1, *0.5, **0.01, ***0.001

Table 2 was the result of OLS showing that not all explanatory variables were statistically significant. Furthermore, to simplify the analysis, regression modeling was continued based on the significant variables, i.e., the distance from high-way gate, the distance from artery-road, developmental characteristic (developed/not-developed area), social characteristic (proper/not-proper residence) and the distance from public health facility. The result was presented in Table 3. This model was used to predict the land price in the study area shown by Fig. 4a. However, since it is a global model, the parameter estimate for each explanatory variable were the same over the region. Hence, map was only available for land price prediction, but not for the parameter estimates. On the hand, GWR approach may produce map for parameter estimates as shown the section of GWR result.

The result of GWR: Similar to the previous model, GWR of land prices used only determinant (explanatory variables) having statistically significant effect obtained by the OLS method. The

Table 3: OLS parameter estimation for significant explanatory variables

Variable	Estimate	Std. error	t-value	Pr(>|t|)	Significance
(Intercept)	1.44E+01	6.01E-02	238.703	< 2e-16	***
Highway	-4.90E-05	8.59E-06	-5.702	1.54E-08	***
Artery	-1.41E-04	3.45E-05	-4.097	4.52E-05	***
Developed	-1.11E-01	4.06E-02	-2.725	0.00655	**
Proper	2.79E-01	5.25E-02	5.317	1.29E-07	***
Health	-7.99E-05	1.73E-05	-4.613	4.47E-06	***

Valuer are significant at **0.01 and ***0.001

Table 4: The Summary of GWR coefficient estimates

Parameter	Min.	1st Qu.	Median	3rd Qu.	Max.	Global
X. Intercept	1.30E+01	1.39E+01	1.41E+01	1.43E+01	1.61E+01	14.3555
highway	-3.27E-04	-5.56E-05	-1.15E-05	5.03E-05	4.20E-04	0
artery	-1.03E-03	-3.80E-04	-2.58E-04	-1.05E-04	2.56E-04	-0.0001
developed	-8.42E-01	-1.44E-01	-6.74E-02	-5.95E-02	-2.91E-01	-0.1107
proper	-4.81E-01	1.66E-01	2.82E-01	3.97E-01	9.96E-01	0.2789
Health	-2.98E-04	-6.87E-05	4.95E-06	8.28E-05	3.98E-04	-0.0001

explanatory variables in the model were the distance from high-way gate (meter), the distance from artery-road (meter), developmental status (developed/not-developed area), social characteristic (proper/not-proper residence) and the distance from public health facility (meter).

As an advantage of analysis using GWR is mappable output. The map provides specific information for each area including parameter estimates which impossible in the case of global models. In this study, we created maps for the parameter estimates and land price predictions. The usefulness of GWR in terms of expressing parameter estimates using GIS-map was discussed by Mennis (2006). Similar to this study, ignoring spatial non-stationarity may blur the differences of parameter estimates. Hence, we realized the effect of spatial non-stationarity was not trivial The summary of the parameter estimates obtained by GWR was listed in Table 4 with the Bandwidth is 1159.857 of the CV score equals 240.9669. AIC: 1307.714, Residual sum of squares: 195.3455 and Quasi-global R-square: 0.3809504. The GWR procedure followed Charlton *et al*. (2003) and the software obtained from Bivand and Yu (2011).

Figure 3 is the maps of local regression parameters of GWR for (a) highway, (b) artery, (c) developmental status, (d) social properness and (e) health, while the land prices may be seen in Fig. 3f. The figures indicated that coefficient parameters of highway, artery, properness and health are not the same for all locations. The coefficient parameters of highway were high in the north-west area, but pretty low in the west area. The coefficient parameters of artery were high in the center and west of South Tangerang, but low in the north-west area. The coefficient parameters of properness were high in the center and north-east area and they were low in the south-east area. The coefficient parameters of health were high in the east and south-east area and lower in other areas. Meanwhile, the coefficient parameters of developmental land status were almost uniform high in all locations.

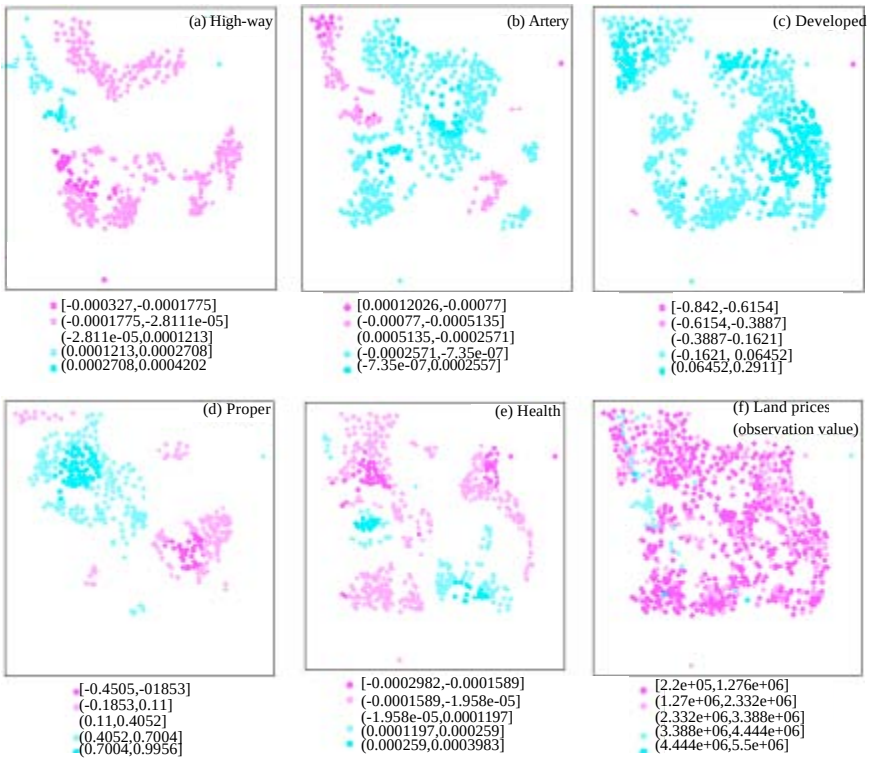

Fig. 3(a-f): Maps of (a, b, c, d, e) local regression parameters (GWR) and (f) values of land prices

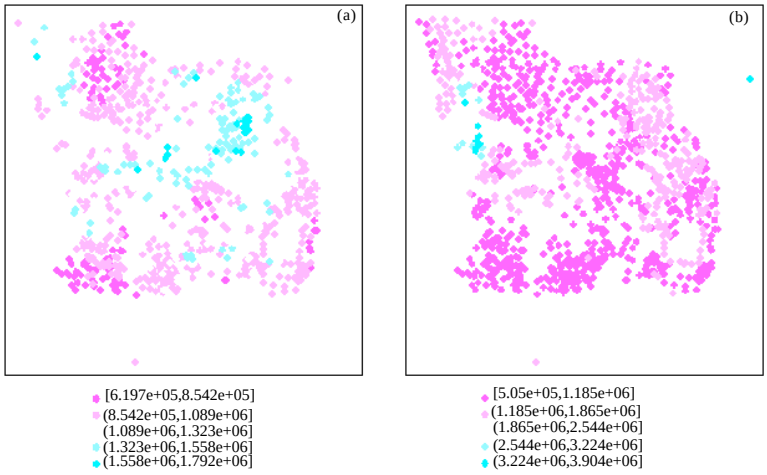

Fig. 4(a-b): (a) Map OLS and (b) GWR land price estimations

Figure 4a was the predicted values based on OLS, while Fig. 4b was based on GWR. These Fig. 4a and b showed that the GWR based land prices were better than OLS based, in which the former had small differences with the real prices. The predicted land prices based on OLS overestimated in all areas. The result was also supported by the residual analysis of OLS and GWR

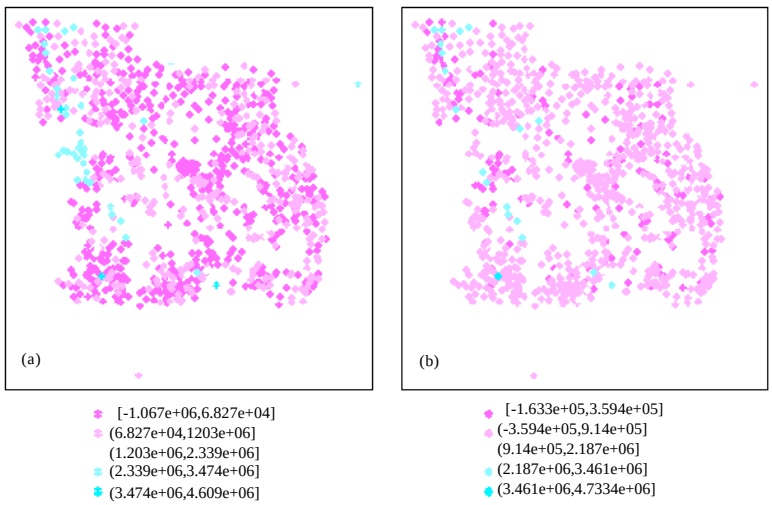

Fig. 5(a-b): (a) Map of OLS and (b) GWR residuals

showed by Fig. 5. The colors of locations in the Fig. 5b are lighter than the color in the Fig. 5a indicating that residuals in Fig. 5b are nearer to zero than residuals in Fig. 5a.

CONCLUSION

The map of parameter estimates provided clear explanation about spatial non-stationarity expressed well by GWR. In addition, the model using GWR provided smaller error than that from the OLS. In this study, the OLS based land price prediction were overestimated. GWR approach of land price models provided better performance statistically compared to the OLS one. Therefore, in the analysis hedonic models of land price, GWR was recommended.

The land price in South Tangerang was influenced by the distance from high-way gate, the distance from artery-road, developmental characteristic (developed/not-developed area), social characteristic (proper/not-proper residence) and the distance from public health facility. The distance from the market facility did not have a significant effect to the land price. Following to a geographic research, the district has many mini markets spread in almost every 2 kilometers and hence the market centers were equivocal. The negative sign of land development indicated that buyers preferred an empty land compared to land with private houses. The local government may use the information in order to have a realistic land price regulation.

ACKNOWLEDGMENT

Dr. Bivand and Yu are acknowledged for allowing us to use the GWR software.

REFERENCES

Bivand, R. and D. Yu, 2011. Package spgwr. R Software Package. http://cran.r-project.org/web/packages/spgwr/spgwr.pdf

Charlton, M., S. Fotheringham and C. Brunsdon, 2003. GWR 3: Software for geographically weighted regression. Spatial Analysis Research Group, Department of Geography, University of Newcastle, UK.

Cheshire, P. and S. Sheppard, 1995. On the price of land and the value of amenities. Economica, 62: 247-267.

Fotheringham, A.S., C. Brunsdon and M. Charlton, 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester, UK.

Fotheringham, A.S., M.E. Charlton and C.F. Brunsdon, 1998. Geographically weighted regression: A natural evolution of the expansion method for spatial data analysis. Environ. Plann. A, 30: 1905-1927.

Lochl, M. and K.W. Axhausen, 2010. Modeling hedonic residential rents for land use and transport simulation while considering spatial effects. J. Transport Land Use, 3: 39-63.

Luo, J. and Y.H.D. Wei, 2009. Modeling spatial variations of urban growth patterns in Chinese cities: The case of Nanjing. Landscape Urban Plann., 91: 51-64.

Mennis, J., 2006. Mapping the results of geographically weighted regression. Cartographic J., 43: 171-179.

Saefuddin, A., N.A. Setiabudi and N.A. Achsani, 2011. On comparison between ordinary linear regression and geographically weighted regression: With application to Indonesian poverty data. Eur. J. Sci. Res., 57: 275-285.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economics Geography, 46: 234-240.

Yu, D. and Y.D. Wei, 2004. Geographically weighted regression. CSISS, Department of Geography, UWM. http://pages.csam.montclair.edu/~yu/GISDay_GWR.ppt