Geographically Weighted Regression Model (GWR) Based Spatial Analysis of House Price in Shenzhen

Jijin Geng[1]*, Kai Cao[2], Le Yu[3], Yong Tang[1]
[1]Development Center of Land and Real Estate Valuation in Shenzhen, Shenzhen, China;
[2]Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong;
[3]Center for Earth System Science, Tsinghua University, Beijing, 100084, China)
*Corresponding author, e-mail: gengjijin@gmail.com

***Abstract*—Through applying spatial statistical analysis, Geographical Weighted Regression (GWR) model and GIS technology, this study aims at finding the relationship between the effects of various factors and spatial distribution of residential house price. The traditional regression models are reviewed firstly, the model without the consideration of spatial characteristics cannot reach very nice precision to simulate the spatial distribution of the house price. In this study, the spatial statistical model, coupled with GIS as well as GWR model, is developed. The proposed model is validated using the house price data in Shenzhen, China, when considering these factors such as the land price, transportation, the distance to the commercial center, the distance to hospital, school, the house type, the brand of the house etc. It is demonstrated that our approach provides an effective model to present the distribution of the residential house price and serve as a tool for house price appraisal during the property tax levy process.**

Keywords- GWR; House Price; Spatial Analysis; Shenzhen

I. INTRODUCTION

The trend of analyzing and predicting the house price precisely is very meaningful and remains a challenge to the publics, investors as well as the government [1]. Recently, the property tax changes to be the hot topic in China, which will contribute a lot to the local fiscal revenue and adjust the overheated real estate market in China. Of which, one of the essential issues is to valuate the house price accurately, so as to ensure the levy process as well as the social equity.

Generally, the most popular applied modeling method to valuate the real estate is based on the general linear regression methods, under the consideration of these potential influencing factors such as the land price, floor area, the direction, plot ration etc. Sometimes, the location factors have also been considered for the valuation of the house price, such as the distance to the commercial centers, distance to the hospitals, schools, roads network etc. To some extent, it could serve as the valuation model that could bring acceptable result. However, from the perspective of the spatial characteristics of the distribution of the house price, the spatial heterogeneity, the traditional least square (OLS) estimation method could not reflect the entire price surface precisely. In order to improve the valuation precision, non-linear method has also been used for calculate the complicated house prices surface, however, the complexity of computation along with the increasing of exponent also brings another problem for the valuation process.

Geographical Weighted Regression (GWR) model was firstly built by Fotheringham [5] when he was trying to do the study with respect to the spatial heterogeneity. After that, Brunsdon and Fotheringham [2] have studied the relationships between the house price and the area, and mentioned several questions GWR faced: the selection of the variables, bandwidth, and the spatial autocorrelation of error. Zhang [12] utilized GWR model to study the height of crown and the result showed that GWR could bring better significance and residual error than OLS. In China, there are also some successful applications of GWR in different fields, Xuan [11] analyzed the spatial characteristics of precipitation that changes along with the elevation. Lv [9] also made use of GWR to study the spatial factors that influence the land price etc. In this study, due to its advantage on spatial statistical field, the GWR model would be utilized to analyze the distribution of house price in Shenzhen and discuss its potential as the tool of house price appraisal for the property tax levy.

II. METHODOLOGY

A. Principles of GWR

GWR model could be represented as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i)x_{ik} + \varepsilon_i \tag{1}$$

Where: (u_i, v_i) is the spatial coordinates of sample i, $\beta_0(u_i, v_i)$ is the value of i on the continuous function $\beta_k(u_i, v_i)$. If $\beta_k(u_i, v_i)$ for all these locations are the same, the model would be a global regression model.

In the model mentioned above, the weight of an observation is decided by the proximity to i. Hence, the weight of an observation will change along with change of i.

978-1-61284-848-8/11/$26.00 ©2011 IEEE

The formula is as follows:

$$\beta_0(u_i,v_i) = (X^T W(u_i,v_i)X)^{-1} X^T W(u_i,v_i)Y \qquad (2)$$

Where:

$$X = \begin{bmatrix} 1 & x_{11} & \ldots & x_{11} \\ 1 & x_{21} & \ldots & x_{21} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & \ldots & x_{n1} \end{bmatrix}, \; Y = \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{bmatrix}$$

$$W(u_i,v_i) = W(i) = \begin{bmatrix} w_{i1} & 0 & \ldots & 0 \\ 0 & w_{i2} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & w_{in} \end{bmatrix} W = ,$$

$$\beta = \begin{bmatrix} \beta_0(u_1,v_1) & \beta_1(u_1,v_1) & \ldots & \beta_k(u_1,v_1) \\ \beta_0(u_2,v_2) & \beta_1(u_2,v_2) & \ldots & \beta_k(u_2,v_2) \\ \ldots & \ldots & \ldots & \ldots \\ \beta_0(u_n,v_n) & \beta_1(u_n,v_n) & \ldots & \beta_k(u_n,v_n) \end{bmatrix},$$

β_0 is the estimation value of β, n is the number of samples, k is the number of the variables, w_{in} is the weight of n according to the location i.

B. Weight function and Bandwidth

The weight function is generally presented by Gauss function, the formula is as follows:

$$W_{ij} = e^{-\frac{1}{2}(\frac{d_{ij}}{b})^2} \qquad (3)$$

Where:

b is the bandwith, if the data of location i is observed, then the weights of other location will decrease along with the increase of distance according the to Gauss function. The more the b and the distance are, the less the weight of this location will be. From another perspective, given b, the location where is far away will be close to 0.

The bandwith could be decided by the method as follows:

$$cv(b) = \sum_{i=1}^{n} \left[y_i - y(b) \right]^2 \qquad (4)$$

Where:

y_i is the observation value of y at (u_i,v_i). (b) is the match value of y at (u_i,v_i) by the subtraction of b by the observation values based on the estimated method of OLS.

Choose b and make:

$$cv(b_0) = \min_{b>0} cv(b) \qquad (5)$$

The b is the bandwidth value.

In this study, the weights indicate how much the factors inside the bandwidth influence the house price, the less the distance is, the more the weight will be. We could know that the weight varies inversely to distance.

C. Reasons for the application of GWR on the house price study

The spatial dependence of house prices is the core reason we apply GWR in this study. Firstly, it is clearly that the house price is instable related to the factors. From another perspective, the principle of Market Sales Comparison as traditional valuation method, is close to the principle of GWR, which could reflect spatial dependence. Besides, the representation of GWR by mapping is another advantage that could present the house price distribution and help valuate the unknown houses' price intuitively.

III. CASE STUDY

A. Introduction to Research Area and Data

Shenzhen is located at the southern tip of the Chinese ainland on the eastern bank of the mouth of the Pearl River and neighbors Hong Kong. Occupying an area of 1,953 square kilometers, the city has a subtropical marine climate with plenty of rain and sunshine and is rich in tropical fruit. By the end of 2009, there were 8.9123 million permanent residents, of whom 2.4145 million had hukou (household registration).

In less than 30 years, Shenzhen, a tiny border town of 30,000 people in 1979, has grown into a modern metropolis. It established many firsts in the history of world industrialization, urbanization and modernization.

A gateway to the world for China, Shenzhen is one of the most developed cities in China and a city that has grown the fastest. Its per capita GDP ranks first among China's major cities. At the meantime, the house price also increases very fast in the past decade.

Figure 1. Research area- Shenzhen

The main data considered inside the modeling process include the Map of Shenzhen, total 653 average market prices of housing estates during the first quarter of 2010 and their XY coordinates, attributes of plot ratio, green ratio, brand etc; the distribution of schools in Shenzhen in 2010, distribution of hospitals, parks, gas stations etc. in Shenzhen; transportation networks of roads, subways etc. Here is the housing estate points map as follows:

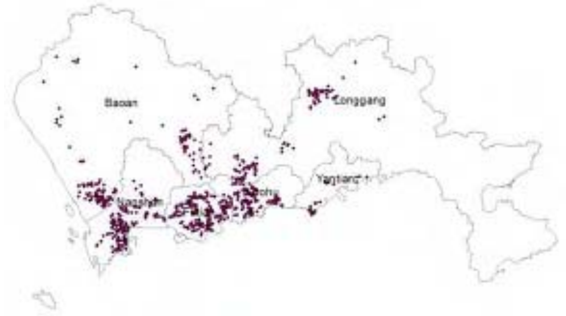

Figure 2. Housing estate sampling points

The other proximity factors could be seen as follows:

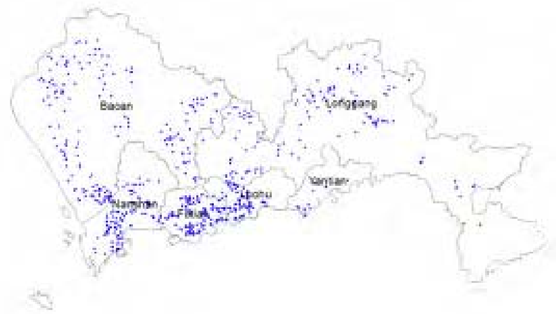

Figure 3. Distribution of schools

Figure 4. Distribution of hospitals

Figure 5. Distribution of parks

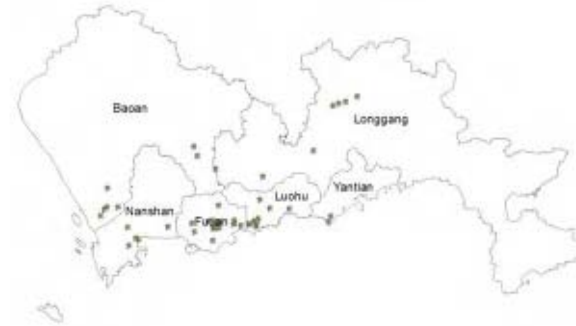

Figure 6. Distribution of commercial centers

Figure 7. Distribution of gas stations

Figure 8. Distribution of roads and subway stations

B. *Construction of GWR Model*

1) *Factors considered*

With respect to the traditional valuation methods of house price, as well as the data we could use, we divided the factors into two aspects, one is the inherent factors, mainly including the brand of the housing estate, the plot ratio, green ratio, area, building structure, the other one is the proximity factors, including the distance to the nearest school, distance to the nearest hospital, distance to the nearest park, distance to the nearest commercial centers, distance to the nearest gas stations, distance to the nearest roads, and distance to the nearest subway station. Euclidean distance function is utilized to compute the distance to each proximity factor. The list of these factors is in the following table.

TABLE I. FACTORS AND REPRESENTATION

Factors	Representation
Brand	BRA
Plot Ratio	PR
Green Ratio	GR
Area	AR
Building Structure	BS
Distance to the Nearest School	DS
Distance to the Nearest Hospital	DH
Distance to the Nearest Park	DP
Distance to the Nearest Commercial Center	DC
Distance to the Nearest Gas Station	DG
Distance to the Nearest Road	DR
Distance to the Nearest Subway Station	DSS

2) *Model construction*

Based on these factors mentioned above, the GWR model could be constructed below:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik}(BRA, PR, GR, AR, BS, \\ DS, DH, DP, DC, DG, DR, DSS) + \varepsilon_i \tag{6}$$

Where y_i is the house price in location (u_i, v_i).

C. *Validation of the Model*

Through the regression computation, the GWR based regression model for analyzing the distribution of house price and valuation of house price did decrease the sum of the residual error, the R square has also been improved from 0.56 to 0.79, which could prove the advantage of GWR as the regression model in house price valuation problem and the improvement than the general regression model.

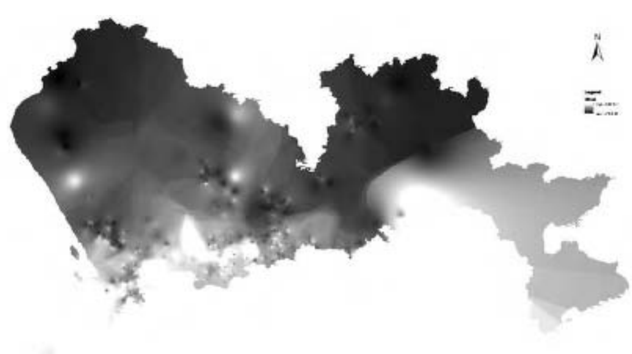

Figure 9. Simulation surface of house price

With respect to the simulation surface of house price in 2010 in Shenzhen, it could bring better precision than general regression model, however, as can be seen from the Fig. 9, there are also some places under the mis-estimation situation due to unbalance of data. Thus, it is also necessary to combine the appraiser's judgment before the final house price valuated get public.

IV. CONCLUSION

In this study, GWR based regression model is constructed to simulate the distribution of the house price and valuate the house price quickly for property tax levy. The case study in Shenzhen is performed under the consideration of inherent factors and proximity factors including: brand of the housing estate, the plot ratio, green ratio, area, building structure, the distance to the nearest school, distance to the nearest hospital, distance to the nearest park, distance to the nearest commercial centers, distance to the nearest gas stations, distance to the nearest roads, and distance to the nearest subway station. Totally 653 samples, which are collected from the commercial transaction prices in 2010, are utilized to estimate the parameters with regard to the model we constructed. And the regression result did bring better precision than general regression model with better residual error and R square. It really shows the potential to be the house price appraisal tool with its better precision and quick computation process for the understanding of house market and property tax levy.

ACKNOWLEDGMENT

The study attributes to the excellent work of the team in the Development Center of Land and Real Estate Valuation in Shenzhen, Shenzhen, China and The Chinese University of Hong Kong. Especially thanks to Miss He her contribution in this study.

REFERENCES

[1] A. S. Adair, J. N. Berry, and W. S. McGreal, "Hedonic modeling, housing submarkets and residential valuation," Journal of Property Research, 13(17), 1996, pp. 67-83.

[2] C. Brunsdon and A. S. Fotheringham, "Some notes on Parametric significance tests for geographically weighted regression," Journal of Regional Science, 39(3), 1999, pp. 497-524.

[3] A. Can and I. Megbolube, "Spatial dependance and house price index construction," Journal of Real Estate Finance and Economics, 1997(14), pp. 203-222

[4] A. S. Fotheringham, C. Brunsdon, and M. Charlton, Geographically Weighted Regression. Chichester, UK: John Wiley and Sons, 2002.

[5] A. S. Fotheringham, M. E. Charlton, and C. Brunsdon, "The geography of parameter space: an investigation of spatial non-stationarity," International Journal of Geographical Information Science, 10(5), 1996, pp. 605-627.

[6] Y. Huang and Y. Leung, "Analyzing regional industrialization in Jiangsu proveince using geographically weighted regression," Journal of Geographical Systems, 2002(4), pp. 233-249.

[7] A. E. Gelfand, H. J. Kim, C. F. Sirmans, and S. Banerjee, "Spatial modeling with spatially varying coefficients processes," Journal of the American Statistical Association, 98(462), 2003, pp. 387-396.

[8] Y. Leung, C. Mei, and W. Zhang, "Statistical tests for spatial nonstationarity based on the geographyicaly weighted regression model," Environment and Planning A, 32(1), 2000, pp. 9-32.

[9] P. Lv and H. Zhen, “Affecting factors research of Beijing residential land price based on GWR model,” Economic Geography, 30, 2010, pp. 472-478.

[10] R. K. Pace, R. Barry, and C. F. Sirmans, “Spatial statistics and real estate,” Journal of Real Estate Finance and Economics, 17(1), 1998, pp. 5-13.

[11] H. Xuan and S. Luo, “The relationship between elevatlion and pericipitation in China,” Chinese Statistical Workshop, 2007, pp. 110-113.

[12] L. J. Zhang, “Modeling spatial variation in tree diameter-height relationships,” Forest Ecology and Management, 189(1), 2004, pp. 317-329.