Habitat International 67 (2017) 69–78

Contents lists available at ScienceDirect

Habitat International

journal homepage: www.elsevier.com/locate/habitatint

Spatial and hedonic analysis of housing prices in Shanghai

Zezhou Huang [a], Ruishan Chen [a, b, *], Di Xu [c, **], Wei Zhou [d]

[a] Key Laboratory of Geographic Information Science, Ministry of Education, and School of Geographic Sciences, East China Normal University, Shanghai 200241, China
[b] Key Laboratory of the Coastal Zone Exploitation and Protection, The Ministry of Land and Resource of China, Nanjing 210023, China
[c] Urban Development Research Institution, Shanghai Normal University, Shanghai 200234, China
[d] School of Public Administration, Hohai University, Nanjing 210098, China

A R T I C L E I N F O

Article history:
Received 1 March 2017
Received in revised form
2 July 2017
Accepted 5 July 2017

1. Introduction

Housing prices in Chinese cities, particularly in Shanghai, Beijing, Guangzhou and Shenzhen, are among the highest in the world. The average price for an apartment in Shanghai was around $10,000 per square metre in 2014 (Shanghai Bureau of Statistics, 2015), which is clearly beyond affordability for most families and strongly affects the quality of life in urban populations. According to Tobler's first law of geography, the closer the distance, the greater the influence between two geographical entities (Tobler, 1970). Accordingly, the price for one house is likely to be most strongly affected by the price of neighbouring houses, although the degree of influence remains to be explored. Hedonic models may be the most appropriate tools to understand this (Rosen, 1974). Hedonic models deconstruct housing prices into their component attributes such as their structural, neighborhood and accessibility attributes, acquire the estimated values of these attributes, and evaluate housing prices according to their attributes (Hui, Zhong, & Yu, 2012; Liao & Wang, 2012; Wen, Xiao, & Zhang, 2017).

Shanghai, one of the leading industrial centers in China, suffers greatly from the problem of high housing prices. In 2003, the average housing price in Shanghai reached 5118 Chinese Yuan (618.34 US dollar at that time) per square meter which even surpassed the average housing price in Beijing. Since then, Shanghai had become one of the cities with the highest average housing price in China (National Bureau of Statistics of China, 2003). While United Nations Center for Human Settlements estimated that a rational ratio of housing price and revenue is about 3:1, that in Shanghai in 2004 was amazingly 14.8:1. Even if the residents' average income in Shanghai was fewer than one-tenth of that in the US, the average housing price in Shanghai was higher than that in New York and Chicago (Gu, 2005). In 2010, with the ratio of housing price and revenue at about 17:1, middle-income residents in Shanghai absolutely lacked the housing affordability through bank loans (Luan, Zhou, & Yi, 2012). Under such circumstances, it is meaningful for us to further explore the spatial pattern and determining factors of housing prices in Shanghai.

A number of different approaches have been adopted to improve the accuracy of housing price predictions. Among these, Ordinary Least Squares (OLS) regression is perhaps the commonly employed hedonic pricing approach (Craven & Islam, 2011). OLS has several distinctive features, such as simplicity of calculation, fast computation, and easy interpretation. However, since many hedonic problems have been generated by the variables, researchers have challenged the validity of the (OLS) regression (Gao & Li, 2011).

One problem of OLS is that the distribution of its residuals over space is characterized by regularity and significant spatial correlation (Des Rosiers, Thériault, & Villeneuve, 2000). Spatial autocorrelation, the correlation between features based upon replicated realizations of the geographic distribution of some attributes, persists in the result of OLS regression (Dunse, Jones, Orr, & Tarbet, 1998). The main shortcoming of spatial auto-correlation is that it can make the *t*-test values of OLS regression unreliable (Bowen, Mikelbank, & Prestegaard, 2001). The resultant t-values make it impossible to decide whether some explanatory variables are significant in modeling the housing prices (Miron, 1984). Small sampling variability is also a potential weakness in OLS regression (Basu & Thibodeau, 1998). An additional problem of the method is spatial

* Corresponding author. Present/permanent address: Key Laboratory of Geographic Information Science, Ministry of Education, and School of Geographic Sciences, East China Normal University, Shanghai 200241, China.
** Corresponding author. Present/permanent address: Urban Development Research Institution, Shanghai Normal University, Shanghai 200234, China.
E-mail addresses: yellow32@126.com (Z. Huang), chenrsh04@gmail.com (R. Chen), xudicugb@gmail.com (D. Xu), 20060064@hhu.edu.cn (W. Zhou).

http://dx.doi.org/10.1016/j.habitatint.2017.07.002

heterogeneity, whereby there is continuous change and uneven distribution of spatial data in a study area. The relevance of these independent and dependent variables is likely to vary within the study area (Gillen, Thibodeau, & Wachter, 2001).

In order to model spatially heterogeneous processes, the Geographically Weighted Regression (GWR) model is an appropriate alternative (Cleveland & Devlin, 1988; Wang et al., 2016). GWR provides a local model of variables to calculate the regression equations. It incorporates both independent and dependent variables within the study area to build local equations (Fotheringham, Brunsdon, & Charlton, 2003). However, GWR also has its limitations and its credibility has been challenged (Wheeler, 2014, pp. 1435–1459). Problems such as multicollinearity (Wheeler & Tiefelsdorf, 2005), extreme coefficients (Cho, Lambert, Kim, & Jung, 2009) and dependencies between spatial errors (Fotheringham et al., 2003) have been uncovered in GWR. However, in general, GWR is an appropriate model to analyze the spatial nonstationarity of housing price distribution (Páez, Long, & Farber, 2008).

There have been many studies on the housing prices in China in the last ten years. For example, Zheng, Kahn, and Liu (2010) used hedonic models to examine the relationship between housing prices, investment, wages, and pollution and found that Chinese cities are undergoing a transition from "producer cities" to "consumer cities". Hou (2010) analyzed housing market prices in Beijing and Shanghai and concluded that the housing price bubble seems to have appeared in Beijing from 2005 to 2008 and in Shanghai from 2003 to 2004. Li and Ge (2008) used their models to evaluate inflation capabilities of housing assets in Shanghai and found that, even if hedging against anticipated and unanticipated inflation was not a feature of the Shanghai real estate market, the actual yield ratio was still positive. However, while many studies have analyzed the real estate market on the macro level, studies of housing prices on the micro level in Shanghai are still limited.

Although the theory of hedonic price models has been widely applied to analyze housing prices in countries like the US (Sander & Polasky, 2009), France (Gouriéroux & Laferrère, 2009), Norway (Osland, 2010), Japan (Shimizu, Takatsuji, & Nishimura, 2010), Austria (Helbich, Brunauer, Vaz, & Nijkamp, 2013), and Netherlands (Özyurt, 2014), similar studies in China are still scarce. Questions such as how housing prices vary over space in Shanghai, or what are the most important factors determining the market value of a single property are still unanswered. Therefore, the main purpose of this article is to estimate housing prices in light of their relations with a series of independent variables including the size, quality, neighborhood, and accessibility of the property.

We applied big data techniques, more specifically, a large set of data on house location, characteristics and price extracted from real estate website, to examine the spatial distribution of housing prices in Shanghai and its determining factors.

The methodology and analysis of results may help government administration to better understand the spatial pattern and determining factors of housing prices in Shanghai so that they can, accordingly, formulate policies on land use and urban planning. The performances of both OLS and GWR models are examined in our analysis which contribute to theoretical development. Although the analysis of housing prices is limited to Shanghai in this article, the basic method can be applied to further analyze the housing prices in other major cities around the world.

This article consists of 4 sections. In section 2, the data set and the hedonic pricing theory are described. In section 3, the results of the spatial auto-correlation analysis, hot spot analysis, and the hedonic models estimated by OLS and GWR are presented. In section 4, discussion and conclusion are provided.

2. Data and method

2.1. Housing price and attributes data

This article uses exactly 12,732 individual pieces of data from a Shanghai real estate website (http://sh.centanet.com/ershoufang) which owned by Centaline Property Agency to analyze the spatial relation of housing prices in Shanghai in early May 2016. Centaline Property Agency, founded in 1978, is the largest real estate intermediary business agency in Shanghai. Its data have high credibility. Because thousands of records are uploaded one day on the website and they are always presented in a mixed and disorderly way, it is difficult to gather these information in a good order manually. However, web crawler, an Internet bot which systematically browses the Internet and gathers the useful information, can collect, clean, and filter data automatically, which greatly saves manpower. In this study, we used the web crawler to acquire all the data. The distribution of the housing data collected is shown in Fig. 1.

Data collected by the web crawler include the price, the area, the address, the orientation, the floor, the decoration condition, and the age of houses that were sold in early May of 2016 in Shanghai. Each address of the houses is assigned a latitude and a longitude using a software program called Xgeocoding. It is evident that the features are concentrated in the center of Shanghai and are not uniformly distributed. The socioeconomic data of different districts including the GDP per capita, male-female ratio, population density and average wage in 2010 are also collected from 2010 Shanghai population census. Shanghai has been readjusting and optimizing its own industrial structure since the development and opening of the Pudong District in 1992. In recent years, the spatial distribution of industry in Shanghai has been relatively stable, which results in a relatively stable spatial distribution of the population. Therefore, data such as the male-female ratio and population density in Shanghai in 2010 can still be used to explain the housing prices in 2016. However, the division of districts in Shanghai has changed from 2010 to 2016: the old Huangpu District and Luwan District merged into new Huangpu District in 2011, and the old Jing'an District and Zhabei District merged into new Jing'an District in 2015. In this article, we use the division of districts in 2010.

The housing prices in Shanghai is analyzed with Geographical Information System (GIS). GIS provides a powerful tool of hedonic price model under Spatial Statistics Tools in ArcToolbox, It has been applied by many scholars in recent years. GIS can effectively present all the geographic entities vividly by positioning them on the map according to their coordinates which is conducive to further analysis (Din, Hoesli, & Bender, 2001). Besides, GIS also has advantages of integrating data and analyzing spatial entities efficiently.

2.2. Hedonic price theory

In the past 3 decades, the hedonic price model was widely used to evaluate the value of houses worldwide. Hedonic price model decomposes its object of study into component attributes and acquires the estimated values of these attributes (Lancaster, 1966; Rosen, 1974). In the field of real estate, hedonic price model commonly uses regression analysis to estimate the effects of various housing attributes including structural attributes, accessibility attributes and neighborhood attributes on housing prices (Wilhelmsson, 2002). Among them, the accessibility attributes and neighborhood attributes contain the main location factors. However, traditional hedonic price model is difficult to capture the location factors. In order to accurately reflect the impact of location factors on the housing prices, the spatial effects should be

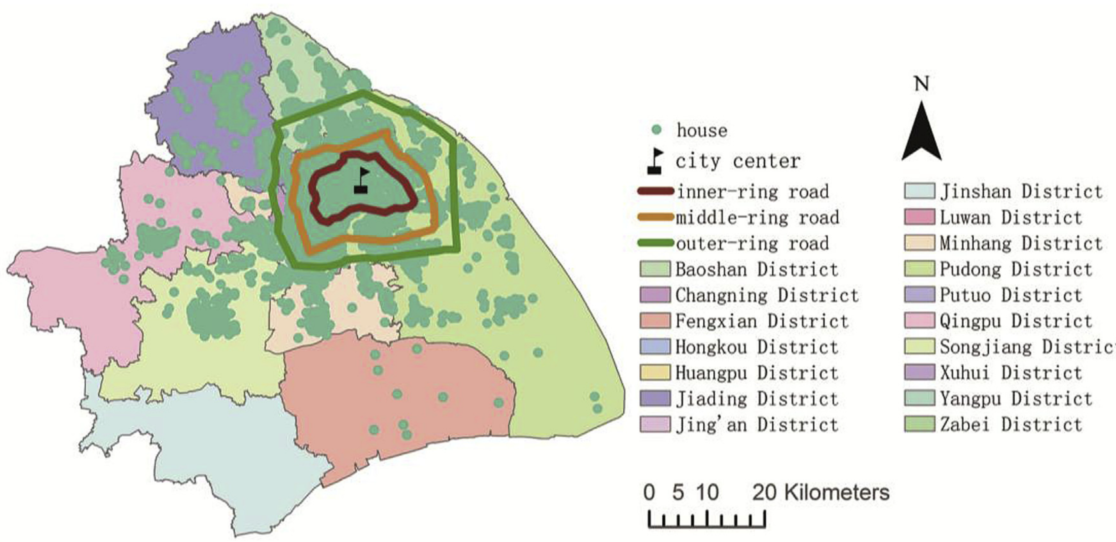

Fig. 1. The distribution of extracted housing data in Shanghai (Chongmin and other islands are not included).

considered.

To resolve this problem, many scholars began to further apply the spatial hedonic price model. One important issue of spatial hedonic price model construction is the integration of spatial effects into the model. If the spatial effects are not properly identified, series of problems like multicollinearity, spatial correlation, and spatial heterogeneity will exist (Des Rosiers, Thériault & Villeneuve, 2000). In the early literature of spatial hedonic price model, Dubin (1998) uses the geostatistical method- Kriging-to estimate the covariance structure of the model. Can (1990) uses a special lag model with variable coefficients to capture the neighborhood effects.

In recent years, scholars begin to integrate hedonic price model with Geographical Information System (GIS). GIS makes it possible to calculate various neighborhood and location attributes in hedonic price model. These real state studies which have considered the spatial effect prove a great improvement in terms of price forecasting and statistical inference (Bowen et al., 2001; Dubin, 1998). Since houses are fixed on the geographical space, and almost all the housing information is essentially spatial information, it is suitable to apply GIS to explore the real estate (Belsky, Can, & Megbolugbe, 1998). Besides, GIS can also synthesize all kinds of spatial data and their analysis comprehensively. GIS has already been widely applied to analyze real estate in different countries and regions (Thrall, 1998).

2.3. Hedonic price models

Hedonic housing prices model is a method to analyze the relation between housing prices and their attributes. Traditional hedonic price models set every attribute of housing as an explanatory variable and set housing prices as the dependent variable (Chau & Chin, 2002). Generally, the housing price can be classified as:

$$P = f(S, A, N) \tag{1}$$

Where S is the structural attributes, A is the accessibility attributes, and N is the neighborhood attributes. Ordinary Least Squares (OLS) is usually used to estimate housing prices in traditional hedonic price models. OLS assumes that error terms are independent and identically distributed, samples are independent of each other, and explanatory variables are independent and exogenous. OLS model can be defined as:

$$Y = \alpha + X\beta + \varepsilon \tag{2}$$

Y is a n vector of the housing prices, α is a n vector of parameters, β is a k vector of the parameters, X is the $n \times k$ matrix of the combination of attributes and ε is an $n \times 1$ error term. By examining their parameters, the effect of each attribute can be analyzed.

The main spatial effects in hedonic models are spatial autocorrelation and spatial heterogeneity (Des Rosiers, Thériault & Villeneuve, 2000). If spatial effects exist in the data of nearby houses, the hedonic housing price models will not consistent with normal distribution which brings forth challenges to spatial statistics. If OLS is still used to estimate housing prices regardless of the spatial effects, the precision of OLS prediction and the significance of the model estimate will be impaired.

The extent of spatial dependency among geographic entities can be measured and analyzed in the Spatial auto-correlation analysis (Anselin & Bera, 1998). There are many ways to examine spatial auto-correlations, such as Geary's C (Geary, 1954), Moran's I (Moran, 1950) and Getis-Ord Gi* (Ord & Getis, 1995). Among them, Moran's I is the most common method which is defined as:

$$I = \frac{\mathrm{n}}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \overline{X}\right)\left(x_j - \overline{X}\right)}{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2} \tag{3}$$

And Getis-Ord Gi* is defined as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{ij}x_j - \overline{X}\sum_{j=1}^{n} w_{ij}}{\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{ij}^2 - \left(\sum_{j=1}^{n} w_{ij}\right)^2\right]}{n-1} \times \left[\frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\overline{X}\right)^2\right]}} \tag{4}$$

n is the sample size, x_i is the housing price in i location, x_j is the housing price in a specified area, $\overline{X}$ is the mean of the attribute and w_{ij} is the spatial weight matrix.

If both spatial auto-correlation and spatial heterogeneity exist simultaneously in the same data, the problem will be far more difficult. De Graaff, Florax, Nijkamp, and Reggiani (2001) listed three obstacles in handling these spatial effects: the difficulty in distinguishing the spatial auto-correlation and spatial heterogeneity through observation, the special form of spatial heterogeneity caused by spatial auto-correlation, and the difficulty in distinguishing the spatial auto-correlation and the spatial heterogeneity through an empirical method.

Different from OLS, Geographically Weighted Regression (GWR) is a local regression method whose parameters varying spatially (Cleveland & Devlin, 1988). It uses a series of weight matrices to integrate spatial structure and geographic information into the linear regression model. GWR then combine all the dependent and independent attributes in the bandwidths of target characteristics and build different equations. GWR can effectively solve the spatial non-stationarity of housing price distribution (Páez et al., 2008), although issues like multicollinearity (Wheeler & Tiefelsdorf, 2005), extreme coefficients (Cho et al., 2009) and the dependence between spatial errors (Fotheringham et al., 2003) are still unavoidable.

2.4. Explanatory variables and their expected effects

Explanatory variables that influence the housing prices have been widely discussed when applied hedonic pricing models. Structural attribute, which refers to the internal attributes of the house structure, is a key factor that decides the housing price (Tu, Yu, & Sun, 2004). The living area (Baumont, Ertur, & Gallo, 2004; Helbich et al., 2013; Opoku & Abdul-Muhmin, 2010; Wilhelmsson, 2002) and room or toilet quantity (Baumont et al., 2004; Helbich et al., 2013; Opoku & Abdul-Muhmin, 2010) are widely recognized as the most important structural attributes. The types of houses like whether it has a terrace, garage and basement (Baumont et al., 2004; Helbich et al., 2013) are often considered by the buyer. The age of the house (Baumont et al., 2004; Helbich et al., 2013) and its decoration conditions (Opoku & Abdul-Muhmin, 2010) are also emphasized by the residents.

Neighborhood attribute, which is the attribute of house environment and its surrounding area, is another kind of explanatory variables that worth noting. The unemployment rate and average wage are both key neighborhood attributes which indicate the development degree of the area (Baumont et al., 2004; Helbich et al., 2013). The education level and social class of the population who lived in the neighborhood is also an appropriate reflection of the surrounding environment (Baumont et al., 2004).

Accessibility attribute is also considered as a key factor by prospective residents. Residents always prefer to live in an area that is close to various amenities such as parks (Montgomery & Curtis, 2006). Distance to the nearest public facilities such as school, commercial center, hospital, workplace (Baumont et al., 2004) and nearby transportation (Wang & Jiang, 2016) will also be the major consideration from the house purchasers.

2.4.1. Structural attributes

The housing data on Centaline Property Agency's website included the attributes of the price, area, address, floor, orientation, decoration and age of houses. The housing price is recorded in units of ten thousand Chinese Yuan (min = 32, max = 8000, mean = 476 and std = 448). The housing prices have a positive correlation with its area as under similar conditions, larger houses are more expensive than smaller ones. The area is recorded in units of square meter (min = 16, max = 916, Mean = 121 and std 84). Besides, their market values also depend greatly on the orientation, the floor, the decoration, and the age of houses.

Houses facing to the west or east will be exposed to the sun too much especially in summer. However, the environment in houses facing to the north will be cold and damp because of the lack of sunshine. Houses facing to the south are the best because they will receive enough sunshine without direct solar radiation. Therefore, houses facing to the south will have higher housing prices than others.

Houses in the ground floor of a building may be cold and damp, and their visions are always restricted. If the building is surrounded by woods, the harassment of mosquitoes and other insects are also unbearable in the lower floor. However, houses in the higher floor often suffer from the bad weather such as the extreme heat and air pollution. In a building without an escalator, it is also weary to climb stairs. Therefore, houses in the intermediate floors of a building are always the most expensive.

A fully furnished house is usually more expensive than a rough house, and a newly built house is less expensive than an old one. Both the extent of decoration and the age of the houses should be considered as the structural attributes of housing prices.

Each house is given variables of 'price' and 'area', and these variables are respectively assigned with the log of housing price in ten thousand Chinese Yuan and the log of total floor area in square meter. Dummy variables like 'decoration', 'orientation', 'floor' and 'age' are also applied in each house, and those of houses that are well furnished, facing the south, in the intermediate floors, or was built before 1990 will be assigned the value 1 while others are assigned the value 0. The distribution of housing price and total floor area in Shanghai are shown in Fig. 2.

2.4.2. Neighborhood attributes

The first neighborhood attribute used in the hedonic analysis is GDP per capita (min = 7387 CNY, max = 41,726 CNY, mean = 13,948 CNY and std = 6328) in different districts which reflect the living standards of people. The second neighborhood attribute is the average wage in different communities which evaluates the people's economic levels. Generally, the higher people's living standards and economic levels are, the more they are willing to pay for a comfortable accommodation. The average wage is recorded in units of Chinese Yuan per year (min = 36,628, max = 97,125, mean = 61,836 and std = 7603).

The population density in every community is the third neighborhood attribute. Neighborhoods that are densely populated are expected to be more competitive. Housing price spontaneously rises as the demand for the houses rises. The population density is recorded in units of number of people per square meter (min = 0.000, max = 1.057, mean = 0.107 and std = 0.096).

Male-female ratio (min = 0.015, max = 14.062, mean = 0.844 and std = 0.398) is also considered a neighborhood attribute in our study. Maybe male-female ratio doesn't affect housing prices as directly as previous attributes do. However, in China, it is always men need to buy houses. If there are more males than females in some places, males tend to be more eager to buy their houses due to the pressure from the society and housing prices will be higher there.

Each house is given variables 'gdppc', 'avewage', 'density' and 'gender', and these variables are respectively assigned with the GDP per capita, the average wage in Yuan per year, the population density in people per square meter and the male female ratio. The distribution of GDP per capita, average wage, population density and male-female ratio in Shanghai are shown in Fig. 3.

2.4.3. Accessibility attributes

The distance from houses to the downtown area is one of the most important accessibility attributes. Houses that near downtown areas provide more convenience to people for shopping, dining, and entertainment. In Shanghai, there are nine downtown areas that worth mentioning as shown in the Fig. 4: Wujiaochang in Yangpu district, Quyang in Hongkou district, Daning in Zhabei district, Ever-bright Center in Zhabei district, Dapuqiao Station in

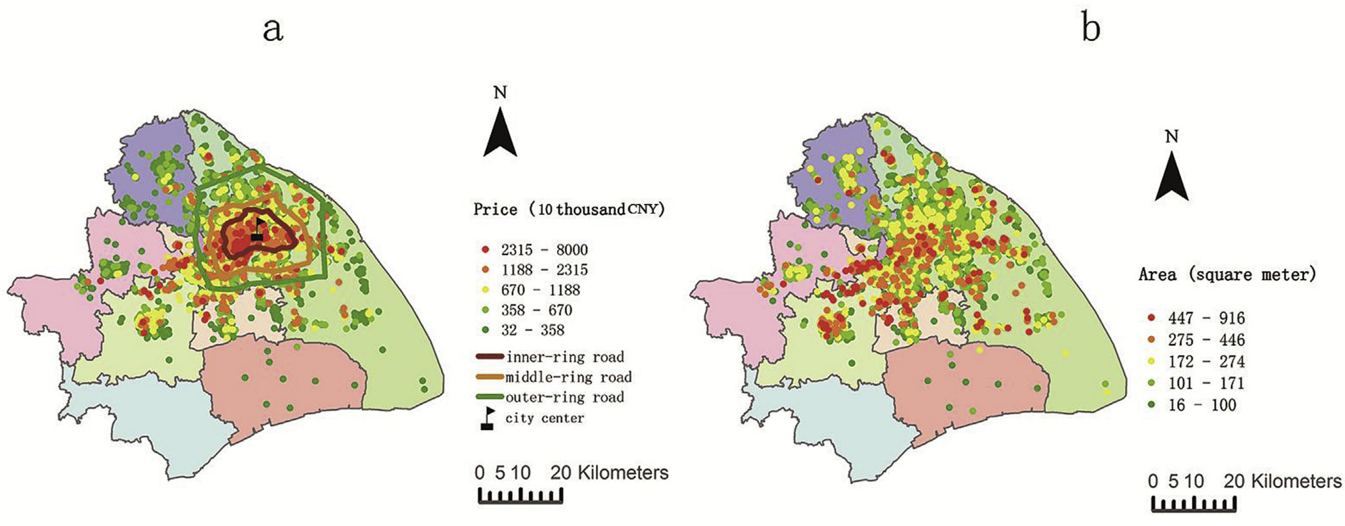

Fig. 2. The distribution of (a) housing price and (b) total floor area.

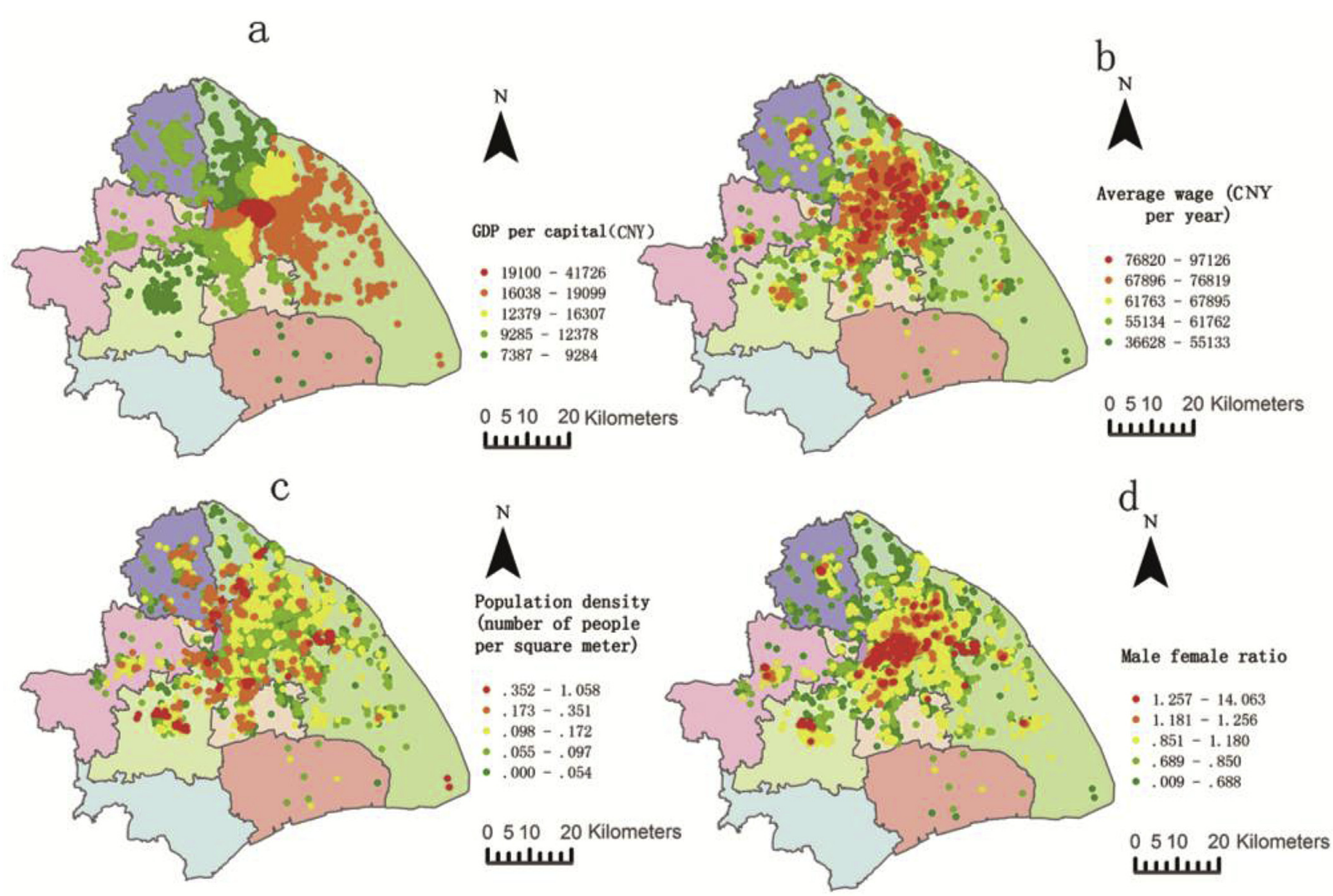

Fig. 3. The distribution of (a) GDP per capita, (b) average wage, (c) population density and male female ratio in Shanghai.

Huangpu district, Jing'an Temple in Jing'an district, Xujiahui in Xuhui district, Changfeng ecological business district in Putuo district and Central Business District in Putuo district. Each house is given a variable 'distance' and the variable is assigned with the euclidean distance from houses to the downtown area in meters.

Also, houses near schools are favored by families with children. We marked the top 36 key primary schools in Shanghai with GIS as shown in Fig. 4, and houses that are within buffer zone with a radius of 1 km are assumed as houses near the primary schools in this article. Each house is given a dummy variable 'school', and variables of houses that are near primary schools are assigned the value 1, while others are assigned the value 0.

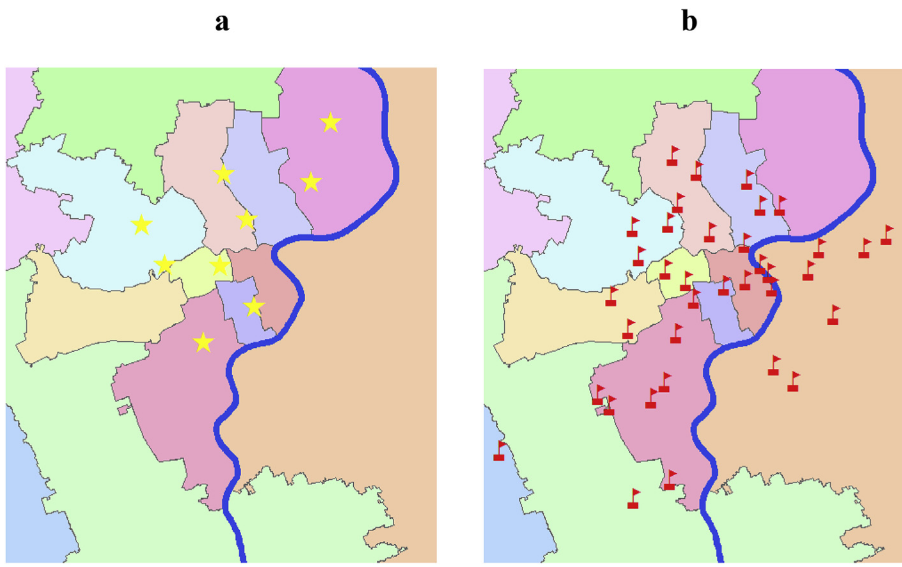

Fig. 4. The distribution of (a) downtown areas and (b) key primary schools.

The descriptions of all the explanatory variables which will be used in the hedonic models are shown in Table 1.

3. Results

3.1. Spatial auto-correlation

In GIS, Spatial auto-correlation is mainly evaluated by calculating Moran's I. After normalized by variance, values of Moran's I are in the range of −1 to 1. The positive values of Moran's I mean that a high value tends to attract other high value but repel low values and the negative ones are just the opposite. The index will approximate zero if the features' values are the results of the random processes. The absolute value of Moran's I indicates the degree to which values are attracted or repelled. The p-value is the possibility whether the features' values are randomized and it is always associated with z-score. These features' high and low values can be deemed as spatially clustered rather than randomly distributed if their p-value is low and its z-score is positive.

Calculating Moran's I of the housing prices in Shanghai leads to the following results (Table 2). The p-value, which is almost zero, shows that it is almost impossible for these data to be a result of a random spatial process. The Moran's I, which is 0.916, means that houses which have high prices are likely to cluster together, so do the cheap ones. The spatial auto-correlation in Shanghai housing prices may affect the subsequent analysis.

Table 1
Descriptions of the explanatory variables.

variable	description
price	Log of housing price (ten thousand Chinese Yuan)
area	Log of total floor area (square meter)
decoration	Decoration of the house (1 = furnished; 0 = not furnished)
orientation	Orientation of the house (1 = south; 0 = others)
floor	Floor of the house (1 = intermediate; 0 = others)
old	Whether the house was built before 1990 (1 = yes; 0 = no)
gdppc	GDP per capita
school	Whether the house is near the primary schools (1 = yes; 0 = no)
gender	Male female ratio
avewage	Average wage (Chinese Yuan per year)
density	Population density (number of people per square meter)
distance	Distance from the house to the downtown area (meter)

Table 2
The result of auto-correlation analysis.

Moran's Index	0.916
Expected Index	−0.001
Variance	0.002
z-score	22.941
p-value	0.000

3.2. Hot spots analysis

Hot Spot Analysis tool in GIS can identify whether houses with high or low prices cluster together. The Getis-Ord Gi* statistic is used in the Hot Spot Analysis to recognize the statistically significant spatial clusters of values. z-scores and p-values are also included in the output table which are used to ensure that these results are not results of a random process. False Discovery Rate (FDR) is then used to avoid multiple testing and spatial correlation. Having a high value is not enough for one point to be deemed as statistically significant. Other points with high values are also required to be found near a statistically significant hot spot. We use fixed distance band to conceptualize the spatial relationships. The threshold distance will guarantee these points have at least one neighbor. The output features are shown in Fig. 5.

3.3. Ordinary Least Squares

A summary report of OLS was obtained by inputting the attributes into Ordinary Least Squares tool in GIS (Table 3). The adjusted R-square in OLS model is 0.70 which indicates that almost 70% of the variation in logged prices can be explained. Attributes with an asterisk in probability are unlikely to be zero in essential. The results are based on the incorrect assumption of spatially uncorrelated model residuals which may lead to inefficient parameter estimations.

There are some variables worthy of attention in Table 3. Firstly, the coefficients of these independent variables are most important because they show the relation between dependent variables and independent variables. Besides, the standard errors also need to be

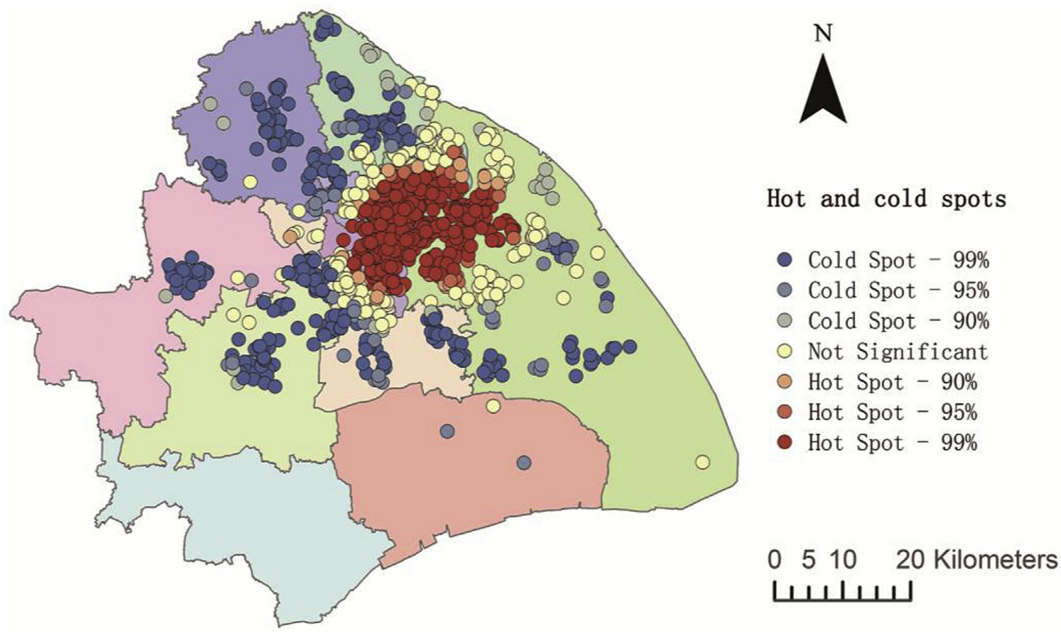

Fig. 5. The distribution of hot spots of housing prices in Shanghai.

Table 3
The result of Ordinary Least Squares.

variable	coefficient	std error	t-statistic	probability
intercept	1.156	0.023	56.893	0.000*
area	1.078	0.003	373.630	0.000*
decoration	0.026	0.003	8.585	0.000*
orientation	0.015	0.000	4.764	0.000*
floor	0.003	0.376	0.885	0.000*
old	−0.117	0.004	−22.134	0.000*
gdppc	0.001	0.000	34.045	0.000*
school	0.193	0.007	28.823	0.000*
gender	0.026	0.008	3.121	0.002*
avewage	0.001	0.000	9.603	0.000*
density	0.789	0.095	7.839	0.000*
distance	−0.001	0.000	−173.108	0.000*
Multiple R-Squared	0.700			
Adjusted R-Squared	0.700			

noticed because it can tell you the possibility of getting same values after resampling and recalculating, and decide the significance of the results. Thirdly, t-statistics can evaluate the statistical significance of these independent variables. Finally, a low p-value can demonstrate that these coefficients are not likely to be zero in essential.

All the structural attributes like decoration, orientation, and floor show a positive correlation with housing prices as expected. The age of the houses indeed has a great effect on the housing prices. Most of the old houses have short property right period and are also poor in appearance which keeps the young customers away, and a number of problems often emerge in their infrastructures. The neighborhood attributes like population density, male-female ratio, average wage and school proximity all promote the housing prices to some extent.

The distance to the downtown area is also an essential attribute of housing prices as expected. Houses near downtown areas are favored by families with children and office workers because of their great convenience and accessibility.

3.4. Geographically weighted regression

To improve the estimation of housing prices in Shanghai, we use Geographically Weighted Regression in GIS. In GWR, the Gaussian kernel is used to distinguish between global and local effects by assigning a high value to densely-distributed feature and vice versa. Due to the high demand of variables in GWR, we choose total floor area, GDP per capita, distance to downtown areas, and male-female ratio to be the explanatory variables from the structural, neighborhood and accessibility attributes of the housing prices. A fixed distance is used to calculate the bandwidths of the Gaussian kernel and Akaike Information Criterion. The adjusted R-Squared in this GWR analysis is 0.826. However, the result is still restricted to multicollinearity (Wheeler & Tiefelsdorf, 2005), extreme coefficients (Cho et al., 2009) and the dependence between spatial errors (Fotheringham et al., 2003).

The local adjusted R-Squared or predicted value is shown in Fig. 6 a. The local adjusted R-Squared ranges from 0.633 to 0.884 throughout Shanghai. Housing prices in downtown areas such as Jing'an District and Huangpu District are best explained. Some houses in the suburban areas such as Songjiang District and Jiading District are also well explained. Fig. 6 b shows the different effects of explanatory variables on housing prices in Shanghai.

3.4.1. The effect of the total floor area

A positive effect of the total floor area on housing prices has been shown in Fig. 7a. The strongest effect is shown in the downtown area while the weakest one is shown in suburban areas. Because of the convenience of the downtown area and the rarity of its land, citizens are more willing to pay more for houses with a larger area there. However, land in suburban areas is adequate so they are not in high demand.

3.4.2. The effect of population density

The effect of population density on housing prices is relatively complicated in Shanghai (Fig. 7b). In traditional downtown areas like the old Jing'an District and Huangpu District, available lands for construction become difficult to find which results in the scarcity of

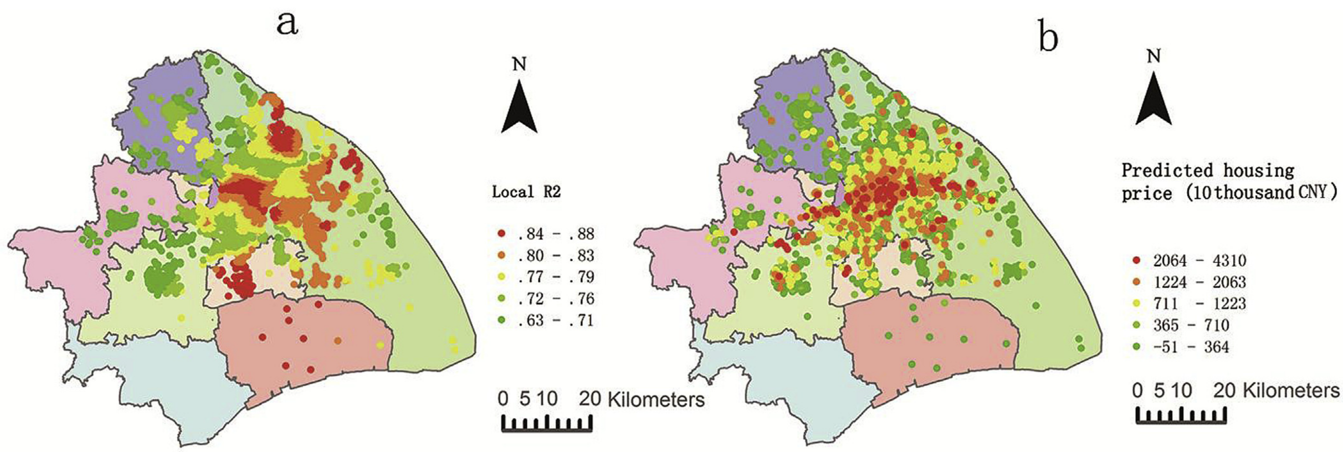

Fig. 6. The distribution of (a) GWR performance and (b) predicted housing prices.

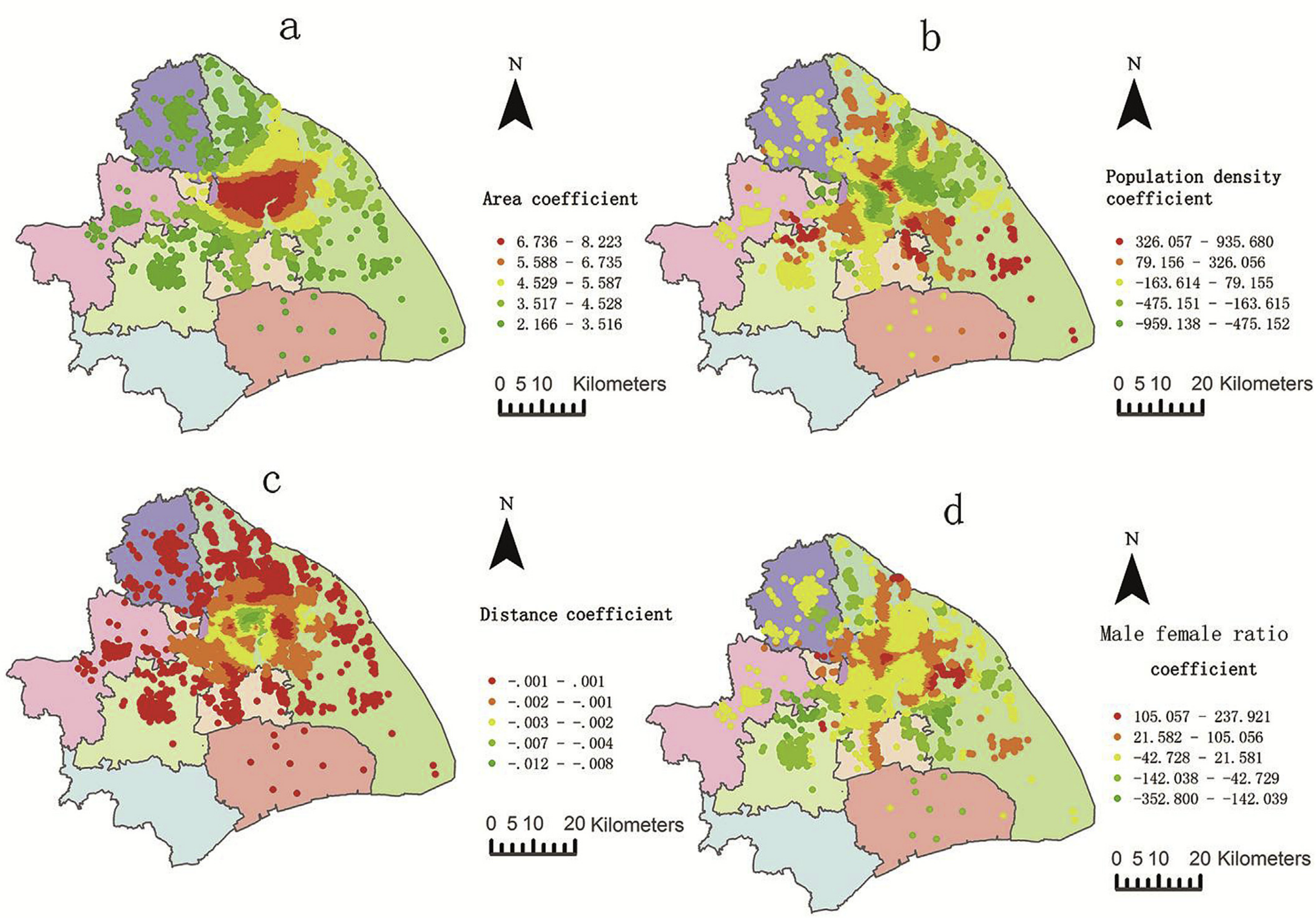

Fig. 7. The effects of (a) area, (b) population density, (c) distance and (d) male female ratio.

commercial houses. Besides, many high-rise office buildings cluster in these areas which appeal to a huge number of high-income groups to live and work there. Finally, some historical and cultural factors also attract lots of investments and make these areas over populated. Therefore, under this type of supply-demand relation, a positive effect of population density on housing prices is shown in these areas.

Although lands in Xuhui District and Changning District are relatively ample, and population density there is not so high, housing prices there are still not lower than those in areas like old Jing'an District, and Huangpu District, which results in a negative effect of population density on housing prices. One of the main reasons for the high housing prices in areas like Xuhui District and Changning District is the high educational and cultural level there.

The housing demands of high-income groups also contribute to the high housing prices to some extent.

A negative effect of population density is shown in Yangpu District and Hongkou District because of the high proportion of the old public houses and traditional Chinese dwellings-Shikumen there. While lots of workers in traditional industries cluster in these areas, their demands of high-price houses are relatively low. Hence, housing prices are not so high in these areas.

When it comes to areas like Minhang District and Baoshan District, a positive effect of population density is shown because of both the low population density and the low housing prices. While plenty of lands can be used for construction, housing prices there is relatively low. Even if a lot of population come to areas like Minhang District and Baoshan District every year, the large size of these areas keeps the population density at a low level.

3.4.3. The effect of the distance to downtown areas

The distance to downtown areas has less effect on housing prices as the distance increases, and almost drops to zero in suburban areas. It is because in downtown areas, the cultural and athletics facilities are well established, the medical treatments and health organizations are clustered, and amusement functions are well found. As the life there is very convenient and the demands for housing are large, housing prices are relatively high. However, the suburban areas in Shanghai are relatively independent. Residents in these areas can meet their needs by the suburban area's own facilities which have no direct relation to the distance to downtown areas.

3.4.4. The male-female ratio

While male-female ratio has a positive effect on housing prices in many areas in Shanghai, it has a negative effect in areas like old Jing'an District. It is because a large number of high-income white-collar women tend to work in the high-rise office buildings in these areas and the female population even surpasses the male population there. At the same time, their demands of houses also promote the housing prices which result in a negative effect of male-female ratio on housing prices.

4. Discussion and conclusion

This article examines the dynamics of housing prices in Shanghai based on the 12,732 pieces of housing data in Shanghai in early May 2016 with the help of GIS. The data has house attributes including location, age, area, floor, orientation, the level of decoration and price, and covers all the districts of Shanghai expect Jinshan and Chongming district. Our study shows that the spatial distribution of high or low housing prices is so spatially clustered that it can't be deemed as the results of random processes based on the value of Moran's I at 0.916. Getis-Ord Gi* statistic also shows that the hot spots are spatially concentrated. In the global specification, OLS analysis with the adjusted R-squared of 0.7 has well detected the relation between the attributes of the houses and the housing prices as expected. GWR model with the adjusted R-squared of 0.826 has greatly improved the estimation of housing prices, however, with the local adjusted R-squared ranges from 0.633 to 0.884 shows the heterogeneity of the relationship. The housing prices in downtown areas can be explained better than suburban areas. The result indicates that the effects of the attributes are different over space. The coefficient of total floor area and distance to the downtown area are distributed more regular than of population density and male-female ratio.

Since studies on the micro structure of housing prices in Shanghai are relatively limited, our spatial and hedonic analysis of housing prices in Shanghai is significant to understand its heterogeneity and complexity. We agree that both the physical attributes and the geographical attributes of these houses contribute to the final housing prices (Widlak, Waszczuk, & Olszewski, 2015). Among all the structural attributes, the area and the age of the house affect the housing prices most, especially in the downtown areas. Meanwhile, the effect of education quality is also worthy of attention. More high-quality schools should be established, and educational resources can be more equalized to ease the tension. Furthermore, it is also necessary to control the population density to meet the different needs of people from different classes.

In face of spatial auto-correlation in hedonic models, different models are used previously such as estimated generalized least squares (Basu & Thibodeau, 1998) and spatial lag model (Cohen & Coughlin, 2008). In this article, geographically weighted regression is used to deal with the spatial auto-correlation. By calculating geographical entities using regression equations and constructing local model equations, GWR can prevent the observed weakness of fixed effects in OLS. The result of adjusted R-Squared in GWR model is 0.826 which demonstrates that GWR is effective in exploring the spatial non-stationarity of housing price distribution.

Since the development and openness of Pudong District in 1992, Shanghai has become one of the global cities in the world. Spatial and hedonic analysis of housing prices in Shanghai has important implications for policies on land use and urban planning in Shanghai. The downtown areas have the highest housing prices because they share the best public services there. The analysis helps government administration better understand the spatial pattern and determining factors of housing prices in Shanghai so that relative policies on land use, industry planning, public services configuration can be set to promote the healthy, harmonious, and sustainable development of economy and society.

However, GWR model also has its limitation and is not appropriate for every kind of data. Some variables are held constant across the study area and others vary spatially. In that case, mixed or semi-parametric models can be applied to analyze these variables. Some variables may not be quite Poisson distributed and a Negative Binomial model will be more useful (Charlton, Fotheringham, & Brunsdon, 2009). Considering the limitations and advantages of the methods can improve the estimation of housing price in terms of its relations to more variables. Currently, there are many property agencies which control different housing information in Shanghai, most of them have websites to provide housing information for the advertisement. It is better to integrate all these data to take advantage of the big data and identify more meaningful results on the spatial distribution and determining factors of housing prices in the near future.

Acknowledgement

We sincerely thank editor and anonymous reviewers for constructive comments on the manuscript. We gratefully acknowledge the financial support from the National Natural Science Foundation of China (41401094) and program of the key laboratory of the coastal zone exploitation and protection, the Ministry of land and resource of China(2015CZEPK05).

References

Anselin, L., & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. *Statistics Textbooks and Monographs, 155*, 237–290.
Basu, S., & Thibodeau, T. G. (1998). Analysis of spatial autocorrelation in house prices. *The Journal of Real Estate Finance and Economics, 17*(1), 61–85.
Baumont, C., Ertur, C., & Gallo, J. (2004). Spatial analysis of employment and population density: The case of the agglomeration of Dijon 1999. *Geographical*

Analysis, 36(2), 146–176.
Belsky, E., Can, A., & Megbolugbe, I. (1998). A primer on geographic information systems in mortgage finance. *Journal of Housing Research, 9*(1), 5–31.
Bowen, W. M., Mikelbank, B. A., & Prestegaard, D. M. (2001). Theoretical and empirical considerations regarding space in hedonic housing price model applications. *Growth and Change, 32*(4), 466–490.
Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. *Economic Geography, 66*(3), 254–272.
Charlton, M., Fotheringham, S., & Brunsdon, C. (2009). *Geographically weighted regression*. White paper. National Centre for Geocomputation. National University of Ireland Maynooth.
Chau, K. W., & Chin, T. L. (2002). A critical review of literature on the hedonic price model. *International Journal for Housing Science and Its Applications, 27*(2), 145–165.
Cho, S., Lambert, D. M., Kim, S. G., & Jung, S. (2009). Extreme coefficients in geographically weighted regression and their effects on mapping. *GIScience & Remote Sensing, 46*(3), 273–288.
Cleveland, W. S., & Devlin, S. J. (1988). Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596–610.
Cohen, J. P., & Coughlin, C. C. (2008). Spatial hedonic models of airport noise, proximity, and housing prices. *Journal of Regional Science, 48*(5), 859–878.
Craven, B. D., & Islam, S. M. (2011). Ordinary least-squares regression. In L. Moutinho, & G. D. Hutcheson (Eds.), *The SAGE dictionary of quantitative management research* (pp. 224–228). Bangalore, India: SAGE Publications.
De Graaff, T., Florax, R. J., Nijkamp, P., & Reggiani, A. (2001). A general misspecification test for spatial regression models: Dependence, heterogeneity, and nonlinearity. *Journal of Regional Science, 41*(2), 255–276.
Des Rosiers, F., Thériault, M., & Villeneuve, P. Y. (2000). Sorting out access and neighbourhood factors in hedonic price modelling. *Journal of Property Investment & Finance, 18*(3), 291–315.
Din, A., Hoesli, M., & Bender, A. (2001). Environmental variables and real estate prices. *Urban Studies, 38*(11), 1989–2000.
Dubin, R. A. (1998). Spatial autocorrelation: A primer. *Journal of Housing Economics, 7*(4), 304–327.
Dunse, N., Jones, C., Orr, A., & Tarbet, H. (1998). The extent and limitations of local commercial property market data. *Journal of Property Valuation and Investment, 16*(5), 455–473.
Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically weighted regression: The analysis of spatially varying relationships*. Sussex, England: John Wiley & Sons Ltd.
Gao, J. B., & Li, S. C. (2011). Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using geographically weighted regression. *Applied Geography, 31*, 292–302.
Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician, 5*(3), 115–146.
Gillen, K., Thibodeau, T., & Wachter, S. (2001). Anisotropic autocorrelation in house prices. *The Journal of Real Estate Finance and Economics, 23*(1), 5–30.
Gouriéroux, C., & Laferrère, A. (2009). Managing hedonic housing price indexes: The French experience. *Journal of Housing Economics, 18*(3), 206–213.
Gu, W. (2005). The comparison of housing prices in Shanghai and America. *Price: theory & Practice (01)*, 50–51 (in Chinese).
Helbich, M., Brunauer, W., Vaz, E., & Nijkamp, P. (2013). Spatial heterogeneity in hedonic house price models: The case of Austria. *Urban Studies, 51*(2), 390–411.
Hou, Y. (2010). Housing price bubbles in Beijing and Shanghai? A multi-indicator analysis. *International Journal of Housing Markets and Analysis, 3*(1), 17–37.
Hui, E. C., Zhong, J. W., & Yu, K. H. (2012). The impact of landscape views and storey levels on property prices. *Landscape and Urban Planning, 105*(1), 86–93.
Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 132–157.
Liao, W. C., & Wang, X. (2012). Hedonic house prices and spatial quantile regression. *Journal of Housing Economics, 21*(1), 16–27.
Li, H., & Ge, C. (2008). Inflation and housing market in Shanghai. *Property Management, 26*(4), 273–288.
Luan, G., Zhou, W., & Yi, W. (2012). Resaerch of residents' housing affordability based on the housing price to income ratio-take the middle-income residents in Shanghai as an example. *China Opening Journal, 02*, 39–43 (in Chinese).
Miron, J. (1984). Spatial autocorrelation in regression analysis: A beginner's guide. In G. L. Gaile, & C. J. Willmott (Eds.), *Spatial statistics and models* (pp. 201–222). Boston, USA: D. Reidel.
Montgomery, M., & Curtis, C. (2006). *Housing mobility and location choice: A review of the literature. Impacts of transit led development in a new rail Corridor(2006).* working paper #2 http://urbanet.curtin.edu.au/local/pdf/ARC_TOD_Working_Paper_2.pdf.
Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika, 37*(1/2), 17–23. http://www.stats.gov.cn/ztjc/ztfx/fxbg/200403/t20040312_14529.html (in Chinese).
National Bureau of Statistics of China. (2003). *The national real estate development market prosperity status reports in 2003*.
Opoku, R. A., & Abdul-Muhmin, A. G. (2010). Housing preferences and attribute importance among low-income consumers in Saudi Arabia. *Habitat International, 34*(2), 219–227.
Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis, 27*(4), 286–306.
Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research, 32*(3), 289–320.
Özyurt, S. (2014). *Spatial dependence in commercial property prices: Micro evidence from The Netherlands*. Working Paper Series #1627. European Central Bank.
Páez, A., Long, F., & Farber, S. (2008). Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. *Urban Studies, 45*(8), 1565–1581.
Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55.
Sander, H. A., & Polasky, S. (2009). The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA. *Land Use Policy, 26*(3), 837–845.
Shanghai Bureau of Statistics. (2015). *Shanghai statistics yearbook 2015*. Beijing: China Statistics Press.
Shimizu, C., Takatsuji, H., Ono, H., & Nishimura, K. G. (2010). Structural and temporal changes in the housing market and hedonic housing price indices: A case of the previously owned condominium market in the Tokyo metropolitan area. *International Journal of Housing Markets and Analysis, 3*(4), 351–368.
Thrall, G. I. (1998). GIS applications in real estate and related industries. *Journal of Housing Research, 9*(1), 33–59.
Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(sup1), 234–240.
Tu, Y., Yu, S. M., & Sun, H. (2004). Transaction-based office price indexes: A spatiotemporal modeling approach. *Real Estate Economics, 32*(2), 297–328.
Wang, C., Du, S., Wen, J., Zhang, M., Gu, H., Shi, Y., et al. (2016). Analyzing explanatory factors of urban pluvial floods in Shanghai using geographically weighted regression. *Stochastic Environmental Research and Risk Assessment*, 1–14. http://dx.doi.org/10.1007/s00477-016-1242-6.
Wang, Y., & Jiang, Y. (2016). An empirical analysis of factors affecting the housing price in Shanghai. *Asian Journal of Economic Modelling, 4*(2), 104–111.
Wen, H., Xiao, Y., & Zhang, L. (2017). Spatial effect of river landscape on housing price: An empirical study on the Grand Canal in Hangzhou, China. *Habitat International, 63*, 34–44.
Wheeler, D. C. (2014). *Geographically weighted regression. In handbook of regional science*. Berlin Heidelberg, German: Springer.
Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems, 7*(2), 161–187.
Widlak, M., Waszczuk, J., & Olszewski, K. (2015). *Spatial and hedonic analysis of house price dynamics in Warsaw*. NBP Working Paper #197 http://econpapers.repec.org/paper/nbpnbpmis/197.htm.
Wilhelmsson, M. (2002). Spatial models in real estate economics. *Housing, Theory and Society, 19*(2), 92–101.
Zheng, S., Kahn, M. E., & Liu, H. (2010). Towards a system of open cities in China: Home prices, FDI flows and air quality in 35 major cities. *Regional Science and Urban Economics, 40*(1), 1–10.