Research on Forecast of Beijing Housing Price based on Spatial Gray Markov Model

Xingxing Chen1**, Juliao Qiu2** and Yanke Wu1*

1Faculty of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang, Guangdong, 524088, China
2College of Ocean and Meteorology, Guangdong Ocean University, Zhanjiang, Guangdong, 524088, China
*Corresponding author’s e-mail: yanke.wu@163.com
**These authors contributed to the work equally and should be regarded as co-first authors.

Abstract. In recent years, with the continuous advancement of the economy, the real estate industry has developed rapidly, which has led to the persistent rise of housing prices. How to understand the changing trend of housing prices reasonably and formulate corresponding policies to effectively control them has become a social problem for the government departments. However, to maintain the housing development, it is critical to monitor and predict housing prices accurately so as to support the decision-making process in the housing field. Based on the actual data of housing prices from January 2018 to March 2019 in Beijing, this research has built a model named Spatial Gray Markov Model, through combing the gray GM (1,1) prediction, and other models, to predict the housing price in Beijing. And the result shows that the Spatial Gray Markov Model is much better than the other models in forecasting, and we get a better outcome that the error and accuracy of the predicted has improved and its accuracy is increasing of 0.31% compared to that of the Grey GM (1, 1) prediction value.

1. Introduction
Since the Reform and Opening, China's estate industry has developed rapidly. In recent years, the price of China's estate industry has the trend to rise, which has exceeded the economic range that ordinary people can afford, and the excessive and continuous rise in housing prices has brought a series of social problems to China, also making a great impact on China's economy. What's more, for the first-tier cities such as Beijing, Shanghai, Guangzhou, and Shenzhen, due to the lowest price having been reached 40 to 50 thousand per square meter recently, most working-class people can hardly afford such a high housing price [1].

As for Beijing, the capital of China, after entering the 21st century, the growth rate of residential housing prices has been increasing rapidly. With the improvement of transportation modes such as the subway, the immigration of foreign populations, the rapid development of the economy, and its unique political and economic center in China, Beijing's housing prices far exceed those of Shanghai, Guangzhou and Shenzhen. In August 2019, according to the statistics of 58 Anjuke Inc. Housing Research Institute, as the economic center area, Chaoyang District’s average listing price has reached 73,678 yuan/m²; Haidian District has reached to 90,640 yuan/m²; as the most centrals district, Xicheng District is 121,352 yuan/m², and Dongcheng District is 104,709 yuan/m² which are for the four central
areas of Beijing economy. In response to such high housing prices, the Beijing Government has successively introduced some relevant housing security policies to help migrant workers coming to Beijing to settle down. For example, implementing some measures such as strengthening policy innovation and optimizing the safeguard system [2].

However, owing to the current policy measures are not enough to meet the housing problems of most residents as well as the volatility of the housing market and population growth, the relevant departments have so many difficulties to the formulation of housing policies. Consequently, based on the historical data of Beijing house prices, this paper attempts to establish a housing price forecasting model, Spatial Gray Markov Model, to predict the housing prices in Beijing and through the models to improve the accuracy of the forecast so as to provide some substantial references to people wanting to buy houses in Beijing, as well as provide the theoretical basis for the regulation of housing prices for the Beijing Municipal Commission of Housing and Urban-rural Development and other relevant departments.

2. Literature Review

Now that housing prices are affected by multiple factors such as political factors, economic factors, administrative factors and natural factors, they tend to grow with time, thus increasing the difficulty and reducing the accuracy of housing price forecasts. Domestic and foreign scholars have used prediction methods such as Gray System [3], Time Series Prediction [4], Markov Chain [5], and Neural Network [6], but they are unable to predict the trend and true value of the housing price accurately and effectively. In reality, each forecasting method has its advantages and disadvantages. Therefore, we cannot simply predict the trend and true value of housing prices through a single method or model.

Grey prediction is a kind of short-term prediction, with the benefit of less sample data points, time series data, but it has a poor prediction effect on volatility data; Markov model is an accurate and effective prediction to analyze and predict the volatility data. In this regard, this paper will adopt the connection between gray GM (1,1) prediction model with the Markov model, to form a Spatial Gray Markov Model, so it can fully highlight their advantages and weaken the respective shortcomings of the individual models and improve the accuracy and effectiveness of housing price forecasts.

3. Research Subject and Data Collection

3.1. Research Subject

As the Chinese central city, Beijing, after entering the 21st century, has a rapid growth rate of residential housing prices. With the high-speed development of transportation, economy and other aspects, as well as the immigration of foreign populations, it can be seen from historical data and information that in order to control the increasing growth of housing prices, the government has introduced a series of related housing policies to alleviate Beijing's growing housing pressure and control the housing price growth more effectively. As researched, the study of Beijing housing prices mainly includes macroscopicity and inaccuracy [7].

3.1.1. Macroscopicity.

Beijing's housing price forecast should be studied from the development of various industries, economic stability and the government's welfare policy. If only unilateral research is carried out, its main influencing factors will be ignored [8].

3.1.2. Inaccuracy.

Seeing that there are many factors affecting Beijing's housing prices, such as economic factors, policy factors, social factors and human factors, we can't predict housing prices more accurately.
3.2. Data Acquisition
This paper obtained the data of the housing prices from January 2018 to March 2019 in Beijing through the Anjuke Inc. website, https://www.anjuke.com/fangjia/beijing2019/. And then draw a figure as shown in Figure 1.

![Figure 1. The Actual Housing Price data from Jan. 2018 to Aug. 2019, in Beijing.](image)

It can be seen from the trend of data in Figure 1 that housing prices are not simply rising or falling, but rising in fluctuation in view of various factors. The following is to make an accurate prediction of housing prices by establishing the Spatial Gray Markov Model [9].

4. Research Method

4.1. Gray forecasting model
Taking Beijing’s housing price sequence as an original data sequence \( x^{(0)} = (x^{(0)}(0))_{k=0}^{n} \) of Gray forecasting model, after verifying it is suitable for gray prediction by the class-compare verification, our establishing gray differential equation is shown in follow

\[
x^{(0)}(k) + ax^{(0)}(k) = b, k = 2,3,\ldots,n
\]

After solving the above equation and getting the predicted value, we will get

\[
\hat{x}^{(0)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, k = 0,1,\ldots,n-1,\ldots
\]

And then find out the predicted value of \( \hat{x}^{(0)} \), as shown in follow

\[
\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k = 1,2,\ldots,n-1,\ldots
\]

Among them, parameter \( a \) reflects the trend of \( \hat{x}^{(1)} \) and \( \hat{x}^{(0)} \), called the development parameter, and parameter \( b \) reflects the alternative relation of data, called gray action, containing gray information.

4.2. Modify the predicted value with Markov model
For the Gray GM (1,1) prediction model, it can make a more accurate prediction of the object by using less data, because the gray forecasting model is a first-order differential equation model. It has a good effect on the exponential and linear growth data series, but it has a poor fitting and prediction effect on the data series with randomness, uncertainty and volatility [10]. Compared to it, the Markov model is suitable for randomly fluctuating data sequences, and has better fitting and prediction effects. So the
Markov model is introduced to correct and compensate for the shortcomings caused by the Gray forecasting model, thus improving the accuracy of the prediction values. This paper combines two single models into the Spatial Gray Markov Model, which combines the advantages of both and makes up for the shortcomings of both [10].

The following calculation steps are based on the Gray forecasting model. The Markov model is used to correct the prediction results in order to make it more accurate.

4.2.1. Calculate relative residuals
Definition: The relative residual is the ratio of the prediction error to the actual value, as shown in below [11]

\[ \epsilon(t) = \frac{x^{(0)}(t) - \hat{x}^{(0)}(t)}{x^{(0)}(t)}, t = 1, 2, ..., n \]  

among them, \( x^{(0)}(t) \) is the actual value, \( \hat{x}^{(0)}(t) \) is the predicted value.

4.2.2. Dividing the state space
In order to reduce the interference of random errors, a reasonable criterion is selected to divide the relative residuals into states and divide them into \( n \) states. Any state interval can be expressed as [11]

\[ E_i = [e_{i1}, e_{i2}], i = 1, 2, ..., n \]  

among them, \( e_{i1}, e_{i2} \) is the upper and lower limits of the state \( E_i \), the set of relative values is

\[ E = (E_1, E_2, ..., E_n) \]  

The number of state divisions is very important. The state division should be based on the number of samples and the error range of prediction. The number of divisions should be reasonable. When there is sampler data, the number of state divisions should be more. Otherwise, the state difference is not obvious, losing the meaning of the adjustment of the fluctuations, but if the state is too much, it will be obviously disorganized, so as to set the number of division states to be generally 3 to 5[1].

4.2.3. Calculate the state transition probability
The state transition probability refers to the probability that something moves from one state to another [9], and the transition probability that the state is transferred to by one step is [11]

\[ P_{ij} = \frac{M_{ij}}{M_i} \]  

among them, \( M_{ij} \) indicates the number of states in the sample that are transferred from state \( E_i \) to state \( E_j \), \( M_i \) indicating the number of times the state \( E_i \) appears in the sample. If it appears at the end, it is not counted in at all [1].

4.2.4. Calculate the state transition probability matrix
The State Transition Probability Matrix refers to the conditional probability matrix when the state in which the system is in time \( t \) is changed to the state of time \( t+1 \), and the 1-step State Matrix of the random event is expressed as [11]

\[ P_{ij} = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix} \]
It is easy to know that the sum of each row element of the state transition matrix is equal to 1, and
the direction of the next state transition of the object is generally judged by the 1-step transition
probability matrix [11].

4.2.5. Determine the object’s transfer state
In the Markov model, the next step is independent of the past state and only related to the current state.
If the predicted object is in the state $E_k$, only does the transition probability of the state transition
probability matrix’s row $k$ need to be considered. If the column $j$ is the maximum number in the row,
the next time the predicted object is most likely to be turned from the state $E_k$ to the state $E_j$. If the
maximum numbers of row $k$ are more than one, then the second step, the third step, or even the n step
need to be calculated until the maximum element of the row has only one. The column in which it is
located is the state of the next transfer of the predicted object [11].

The step $k$ of state transition probability matrix is expressed as:

$$
P^{(k)} = P \cdot P \cdots P = 
\begin{bmatrix}
P_{11}^{(k)} & P_{12}^{(k)} & \cdots & P_{1n}^{(k)} \\
P_{21}^{(k)} & P_{22}^{(k)} & \cdots & P_{2n}^{(k)} \\
\vdots & \vdots & \ddots & \vdots \\
P_{n1}^{(k)} & P_{n2}^{(k)} & \cdots & P_{nn}^{(k)}
\end{bmatrix}
$$

(9)

4.2.6. Determine the corrected predicted value
The modification of the predicted value is related to the next transfer state. If the predicted object is
transferred to the state [12], the modified formula of the Gray predicted value is as follow

$$
\hat{y}(k) = \frac{\hat{x}^{(0)}(k)}{1 \pm 0.5 |e_{i_j} + e_{j_j}|}
$$

(10)

among them, $e_{i_j}$, $e_{j_j}$ are the upper and lower limits, when the predicted value is larger than the actual
value, take "+"; on the contrary, the predicted value is "-" when the actual value is smaller than the
actual value.

5. Result Analysis

5.1. Gray Prediction Model
After substituting the model, we can get the Gray GM (1,1) forecast value for Beijing's housing prices
from January 2018 to August 2019, as shown in Table 1.

And then, we draw a figure about the actual value of housing price and Gray GM (1,1) predicted
value in Beijing, as shown in Figure 2.
Figure 2. The Actual Value of Housing Price and Gray GM (1,1) Predicted Value in Beijing.

Through the residual test, the correlation test and the posterior test, we can know that the error of the model is small, and the accuracy of the prediction is high. However, it is known from Figure 2 that the predicted value of the Gray GM (1,1) model is a straight line or the exponential growth line. Seeing the error is small, it can not describe the change of the actual value fluctuations well. Therefore, we add the Markov model to correct the predicted value of the Gray GM (1,1) model as below.

5.2. Spatial Gray Markov Model

Through the obtained gray GM (1,1) prediction value, the state division of each relative error and the equation (3), the Spatial Gray Markov Model value of the housing price in each month in Beijing is calculated, as shown in Table 1.

Table 1. The forecast value and error of Gray GM (1,1) and Spatial Gray Markov Model for housing prices from Jan. 2018 to Aug. 2019 in Beijing.

| Time    | The real value | Gray GM (1,1) | Spatial Gray Markov Model |
|---------|----------------|---------------|--------------------------|
|         | Pred. value    | Error/%       | Pred. value              | Error/%       |
| 2018.01 | 51732          | 51732         | 51732                    | 0             |
| 2018.02 | 52077          | 52642         | 52240                    | -0.31         |
| 2018.03 | 52240          | 52757         | 52354                    | -0.22         |
| 2018.04 | 52685          | 52871         | 52467                    | 0.41          |
| 2018.05 | 52850          | 52986         | 52581                    | 0.51          |
| 2018.06 | 53350          | 53101         | 53806                    | -0.85         |
| 2018.07 | 53850          | 53216         | 53922                    | -0.13         |
| 2018.08 | 54233          | 53322         | 54030                    | 0.37          |
| 2018.09 | 54199          | 53447         | 54156                    | 0.08          |
| 2018.10 | 53768          | 53563         | 54274                    | -0.94         |
| 2018.11 | 53716          | 53680         | 54094                    | -0.70         |
| 2018.12 | 53205          | 53796         | 53388                    | -0.34         |
| 2019.01 | 53678          | 53913         | 53504                    | 0.32          |
| 2019.02 | 53763          | 54030         | 53620                    | 0.27          |
| 2019.03 | 53866          | 54147         | 53736                    | 0.24          |
The average error of the predicted value of Gray GM (1,1) calculated by Table 1 is 0.69%, the maximum error is 1.66%. However, the average error of the predicted value of Spatial Gray Markov Model is 0.38%, and the maximum error is 0.94%. The error of the Spatial Gray Markov Model is 0.31% lower than that of the Gray GM (1,1) predictor, and the variance of the error of the gray Markov predictor is also smaller [13], which fully reflects the excellence and applicability of the Spatial Gray Markov model. And the comparison of the prediction results of the two methods [13], as shown in Figure 3.

It can be seen from Figure 3 that the actual value of Beijing housing prices is volatility, while the Gray GM (1,1) forecast (the Red Line) is almost linear or exponential growth, which cannot reflect and predict the fluctuation of house prices. On account of dividing the state and establishing the Markov model to correct the Gray GM (1,1) prediction value [14], the final predicted value can better reflect the fluctuation of the real housing price in Beijing and also do a good job in prediction.

Figure 3. The Actual Value of Housing Price, Gray GM (1,1) Predicted Value and Spatial Gray Markov Model Predicted Value in Beijing.

6. Conclusion
This paper establishes the Spatial Gray Markov Model to predict the housing price data from January 2018 to August 2019 in Beijing. First, through the histogram, it is found that the housing price is fluctuating, and the average error of the predicted value obtained by establishing the Gray GM (1,1) prediction model is 0.13% and the maximum is 1.53%, which is a bit large. Then, after drawing a scatter plot of the predicted data comparison to the real value, we know that the model can not predict accurately. In order to improve the prediction accuracy, the Markov model is built to correct the obtained prediction value [19]. Therefore, the new model, the Spatial Gray Markov Model, is obtained which average error is 0.38% and the maximum error is only 1.04%. In comparison with the error of Gray GM (1,1), which witnesses a high rate and high variance, the rate in Spatial Gray Markov Model has a decrease of 0.49% and much less variance. Besides, it can be seen that the Spatial Gray Markov Model can well predict the trend of housing prices. In general, this model can help the government and relevant departments to better understand the trend of housing prices in order to formulate some policies to control the housing prices in a reasonable way.

7. Lack of research
In this paper, when the Markov model is used to correct the GM (1,1) prediction value, the state is divided into four parts, using the interval aliquot method based on relative residuals. Therefore, it is affected by some subjective factors, thus don't knowing if this division method is the best choice to be
used. Consequently, we can make a further study in this field.

Acknowledgments
This research was financially supported by Faculty of Mathematics and Computer Science, Guangdong Ocean University. We acknowledge the helpful suggestions from the anonymous reviewers. We also acknowledge the data provider. The housing prices data used in this paper is from Anjuke Inc., which can be downloaded from the website, https://www.anjuke.com/fangjia/beijing2019/.

References
[1] Chen Ming. Application of Grey Markov Model in Forecasting the Number of Tourists in Guangdong[J]. Corporate report. 2012(5): 143-144.
[2] Beijing Housing and Urban and Rural Construction Development White Book in 2019[Z]. 2019—08.
[3] Hu Liuxing, Wu Jiefei. Comparison of Two Price Forecasting Methods Based on Grey System Theory[J]. Statistics and Decision. 2010(07): 22-24.
[4] Ding Yuezhi. House Forecasting and Fluctuation Analysis Based on Time Series Model[D]. Shandong University, 2018.
[5] CanceGu Xiujuan, Li Chao, Research on House Price Forecast Based on Markov Chain[J]. Consumer Economics. 2012(5): 40-42.
[6] Gao Yuming, Zhang Renjing. Analysis of House Price Prediction Based on Genetic Algorithm and BP Neural Network[J]. Computer Engineering. 2014.40(04): 187-191.
[7] Xiao Xinping. The parameters in gray models[J]. Journal of Grey System, 2000,13(4): 73-78
[8] Xie Naiming, Liu Sifeng. Research on Grey System Modeling Technology[D]. Nanjing Aerospace University. 2008
[9] Shan Chuanpeng. Population Prediction in jiangxi Province Based on Grey Markov Model[J]. Jiangxi University of Finance and Economics, 2014, (4): 95-100
[10] Shu Fuhua, Zhang Shiqi. Prediction of China’s Paper Consumption Based on Gray Markov Model[J]. Journal of Xuzhou Institute of Technology, 2017, (32): 28-33.
[11] Li Jindan, Xu Mei. Forecasting Financial Fluctuations Based on Grey Markov Model[J]. Journal of Wuhan University of Technology. 2014(36):130-131.
[12] Liu Miao Grey Forecasting Model and Its Application in Electric Power Demands[D]. Xian University of Architecture and Technology. 2012
[13] Zhang Yufen. The Study of Markov Property and Its Test[J]. Journal of Heibei University, 2007, (2): 312-313
[14] He Chenggang. Research on Markov Model Prediction Method with Its Application[D]. Anhui University, 2011
[15] Cheng Zhou, Xie Shiqian, Pan Chengyi. Probability Theory and Mathematical Statistics[M]. Zhejiang: Higher education press. 2010.
[16] Jing Xin. Mathematical modeling[M]. Shanghai: Tongji University Press. 2017
[17] China's housing price trends. Baidu. https://www.docin.com/p-2113503541.html. 2019.5.19
[18] Liu Sifeng, Lin Y. Grey information theory and practical [M]. London: Springer-Verlag, 2006:2-4
[19] Deng Julong. The Problem of gray systems[J]. Systems & Control Letter, 1982, 35(12): 288-294