Short-term Demand Forecasting of Shared Bicycles Based on Seasonal Grey Markov Model

Hengzi Liu1, Yulong He1*, Tailong Song1 and Peng Xu1
1Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing, 100124, China
*Corresponding author’s e-mail: ylhe@bjut.edu.cn

Abstract. According to the characteristics of periodicity, nonlinearity and randomness of shared bicycle riding data, a seasonal grey Markov model was proposed. In the new prediction model, the seasonal GM (1,1) model was used to get the prediction results firstly. Then, Markov model was used to modify the prediction residual. In the process of residual correction, the residual sequence was selected according to the new information priority principle, and the residual was modified by the expectation of the median value in the state interval, which improves the prediction accuracy of the model. Finally, the model was applied to the demand forecast of the Citi Bike shared bicycles in New York, the first three weeks of June 2019 during the Saturday peak hours. The numerical results show that the mean absolute percentage error (MAPE) of predicted value of seasonal grey Markov model is 2.72%, which is superior to the traditional GM (1,1) model, seasonal GM (1,1) model and grey Markov model.

1. Introduction
Under the development of the sharing economy, shared bicycles have appeared in the streets and alleys of cities. Green, low-carbon and healthy bicycle travel methods are being generally accepted by residents. The shared bicycle system is a slow traffic system that solves the short-distance travel problem of users, and is an important part of the urban traffic system. Precise demand forecasting for shared bicycles can provide a reference for shared bicycle scheduling, thereby improving the operating efficiency of the shared bicycle system.

At present, many experts and scholars have carried out research on the demand forecast of shared bicycles. For example, Cao D.D. et al. [1] used the LSTM neural network model to predict the hourly demand for shared bicycles in New York City; Gao W. et al. [2] based on the LSTM neural network model, considering geographic and meteorological characteristics to optimize the prediction results, and improve the prediction performance of the model; Westland J.C. et al. [3] used a deep learning model to predict the demand for shared bicycles and achieved good prediction results; Fournier H. et al. [4] established a sine model for estimating seasonal bicycle demand, which can also estimate the monthly average daily demand and the annual average daily demand of bicycles; Gao X.H. et al. [5] proposed a model based on moments and a new hybrid method, with a time interval of 2 hours to predict the demand for shared bicycle rental. The demand for shared bicycles is affected by factors such as time and space conditions, cyclists’ psychological and physiological conditions, climate and environmental conditions, which is random and uncertain. There is uncertainty in the shared bicycle system, and the information obtained is usually incomplete, so it can be modeled as a grey system, and then the grey prediction theory is used to study the demand for shared bicycles.
The main model in grey forecasting theory is the GM(1,1) model. The GM(1,1) model has the advantages of simple modeling process and low requirements for data sequence distribution patterns, and has been widely used in the prediction of complex systems [6]. The traditional GM(1,1) model is directly used for prediction, and the accuracy is not high. Many experts and scholars have proposed a series of improved models through initial value optimization [7], background value optimization [8], residual improvement [9] and combination optimization [10,11]. Xiao X.P. et al. [12] used a cyclic truncated accumulation operator to weaken the volatility of the data, and then proposed a seasonal GM(1,1) (SGM(1,1)) model based on the idea of the original GM(1,1) model.

From the perspective of prediction results, the residual sequence of the traditional GM(1,1) model has strong volatility, while the Markov chain is suitable for data sequences with random volatility [13]. The Markov model is a time series model with no aftereffect. It can improve the random flexibility of the residual sequence after correcting the residual error, and the prediction result is more scientific and reasonable [14].

For the forecast of shared bicycle demand with cyclical fluctuations, the research results show that the SGM(1,1) model has better adaptability than the GM(1,1) model. In order to further improve the prediction accuracy of SGM(1,1) model, this paper uses Markov model to correct the residual error of SGM(1,1) model prediction. By calculating the shared bicycle riding data provided by the Citi Bike shared bicycle website in New York, the first three weeks of June 2019 on Saturday peak hours, the prediction results show that the seasonal grey Markov model proposed in this paper has high accuracy and can be used to forecast short-term demand for shared bicycles.

2. The traditional GM (1,1) model

In this part, the traditional grey prediction model is introduced and discussed. The grey prediction, originally proposed by Professor Deng Julong, is characterized by the prediction of the first order accumulating data sequence rather than the original data. The most widely used is the GM (1,1) model, which is introduced as follows [15]:

Assume an original, non-negative sequence is

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

its first-order accumulated generating operation(1-AGO) sequence is

\[ x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)) \],

where \( x^{(1)}(k) = \sum_{j=1}^{k} x^{(0)}(i), k = 1,2,\ldots,n \).

The basic form of GM(1,1) is

\[ x^{(0)}(k) + az^{(1)}(k) = b \],

where \( z^{(1)}(k) \) is the background value and defined as:

\[ z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k - 1) \],

the parameter \( a, b \) can be solved by the least square method: \( p = (a,b)^T = (B^TB)^{-1}B^TY \), where

\[ Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \]

The whitening differential equation corresponding to Eq.(3) is:

\[ \frac{dx^{(1)}(k)}{dt} + ax^{(1)}(k) = b \].

Substitute the parameters \( a, b \) into Eq.(4), the time response sequence of GM (1,1) model can be solved by taking \( x^{(0)}(1) \) as the initial condition.
\[
\hat{x}(k+1) = (\hat{x}(0)(1) - b/a)e^{-ak} + b/a, k = 1,2,\cdots, n. \tag{5}
\]

The prediction value of the original sequence is obtained by reduction:
\[
\hat{x}^{(0)}(k+1) = \hat{x}(k+1) - \hat{x}(k), k = 1,2,\cdots, n. \tag{6}
\]

3. The seasonal grey Markov model

3.1. SGM (1,1) model

Hourly ride data of shared bicycles usually have complex nonlinearity, randomness and volatility, and present seasonal characteristics. Therefore, the seasonal GM (1,1) (SGM (1,1)) model that accumulates the original sequence seasonally, which weakens the volatility of data, is proposed by Xiao X.P. et al. [12] as follow:

Assume that \( x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)) \) is an original, non-negative sequence. The cycle \( q \) is taken for rolling accumulation to get:
\[
y^{(0)}(k) = CTAGO(x^{(0)}(k)) = \sum_{j=1}^{q} x^{(0)}(k+j-1), \forall k = 1,2,\cdots, n-q+1. \tag{7}
\]

The 1-AGO sequence of \( y^{(0)} \) is \( y^{(1)} = (y^{(1)}(1), y^{(1)}(2), \cdots, y^{(1)}(r)) \), where
\[
y^{(1)}(k) = \sum_{i=1}^{k} y^{(0)}(i) = \sum_{j=1}^{q} x^{(0)}(k+j-1). \tag{8}
\]

Substitute the \( y^{(0)} \) sequence into the traditional GM (1,1) model, then the corresponding grey differential equation is:
\[
y^{(0)}(k) + az^{(1)}(k) = b. \tag{9}
\]

Set
\[
p = \begin{bmatrix} a \\ b \end{bmatrix}, \quad Y = \begin{bmatrix} y^{(0)}(2) \\ y^{(0)}(3) \\ \vdots \\ y^{(0)}(r) \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix}_{(r-1)\times n}, \quad C = \begin{bmatrix} x^{(0)}(1) \\ x^{(0)}(2) \end{bmatrix}.
\]

Then the differential equation (9) can be expressed as \( CX = BP \). The least square method is used to obtain the parameters: \( p = (a, b)^T = (B^TY)^{-1}B^TX \). Where
\[
B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(r) & 1 \end{bmatrix} = \begin{bmatrix} -0.5 & -0.5 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 \\ 0 & -0.5 & \cdots & 0 & 0 & 1 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -0.5 & -0.5 & 1 & 1 & \cdots & 1 \end{bmatrix}_{(r-1)\times (r+2)} \quad \Delta = B_1 \cdot A \cdot G \cdot M
\]

The definition of the whitening differential equation corresponding to the SGM (1,1) model is:
\[
\frac{dy^{(1)}(k)}{dt} + ay^{(1)}(k) = b. \tag{10}
\]

Similarly, we get the time response sequence of SGM (1,1) model by solving the whitening differential equation (10):
\[
y^{(1)}(k+1) = (y^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}, k = 1,2,\cdots, n-q+1. \tag{11}
\]

The regressive sequence is:
\[
\hat{y}^{(0)}(k+1) = \hat{y}^{(1)}(k+1) - \hat{y}^{(1)}(k), k = 1,2,\cdots, n-q+1. \tag{12}
\]

The prediction value is reduced to:
\[
\hat{x}^{(0)}(k+1) = \hat{x}^{(0)}(k-q+2) + \hat{y}^{(0)}(k) + x^{(0)}(k-q+1)
= \hat{y}^{(0)}(k-q+2) - \hat{y}^{(0)}(k-q+1) + x^{(0)}(k-q+1), k = q, q+1, \cdots, n-1. \tag{13}
\]
3.2. Markov model correction prediction residual

The Markov chain is a kind of time series model with no aftereffect. The residual sequence predicted by SGM (1,1) model is positive and negative with strong randomness. The Markov chain is used to modify the residual sequence. It can improve the randomness of the residual sequence, and the prediction result is more scientific and reasonable. The specific steps are as follows:

1) Residual sequence is obtained from preliminary prediction results

\[ \varepsilon^{(0)}(n) = (\varepsilon^{(0)}(1), \cdots, \varepsilon^{(0)}(n)) \]  

where \( \varepsilon^{(0)}(k) = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1, 2, \cdots, n \). The residual sequence is divided into \( s \) states, which \( E = \{E_1, E_2, \cdots, E_s\} \) determines the state of each element in the sequence.

2) State transition probability matrix is calculated

\[ p^{(m)} = \begin{bmatrix} p_{11}^{(m)} & p_{12}^{(m)} & \cdots & p_{1s}^{(m)} \\ p_{21}^{(m)} & p_{22}^{(m)} & \cdots & p_{2s}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{s1}^{(m)} & p_{s2}^{(m)} & \cdots & p_{ss}^{(m)} \end{bmatrix} \]  

where \( p_{ij}^{(m)} = \frac{M_{ij}^{(m)}}{M_i^{(m)}}, M_i^{(m)} = \sum_{j=1}^{s} M_{ij}^{(m)}, i, j = 1, \cdots, s \). \( p_{ij}^{(m)}, M_{ij}^{(m)} \) respectively represents the probability and times of state \( E_i \) transferring to \( E_j \) state through \( m \) steps; \( M_i \) is the times of state \( E_i \) appearing.

3) Correct the residual sequence. According to the new information priority principle, the state of the first \( s \) elements of the residual sequence to be modified is taken as the original state, and the 1, 2, \cdots, \( s \) steps are transferred respectively according to their distance from the elements to be modified. In the transfer matrix corresponding to the number of transfer steps, the row vector corresponding to the original state is taken to form a new probability matrix. Sum the column vectors of the new probability matrix, get the probability matrix of the elements to be modified in each state interval \( [1, \cdots, s] \), and get the modified residual value by weighted average

\[ \varepsilon''^{(0)}(k) = \sum_{i=1}^{s} \lambda_i v_i, \]  

where \( \lambda_i = \frac{p_i}{\sum_{i=1}^{s} p_i} \), \( v_i \) is the center of the \( E_i \) state interval, which is the average of the two end points.

4) Optimize the prediction results to get the final prediction value

\[ \begin{align*} \hat{x}''^{(0)}(k) &= \hat{x}^{(0)}(k), k = 1, 2, \cdots, s \\ \hat{x}''^{(0)}(k) &= \hat{x}^{(0)}(k) + \varepsilon''^{(0)}(k), k = s + 1, \cdots, n. \end{align*} \]

3.3. Model steps

Step1: Set the original sequence \( x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)) \). Take \( q \) as a cycle, the 1-CTAGO sequence generated by seasonal accumulation of the original sequence is \( y^{(0)} \).

Step2: The first-order accumulation of sequence \( y^{(1)} \) is carried out, and the parameters \( a \) and \( b \) are solved.

Step3: Substitute parameters into the SGM (1,1) model to obtain the time response series, and calculate the preliminary predicted values [16].

Step4: According to the preliminary prediction results, the residual sequence \( \varepsilon^{(0)} \) is obtained, which is modified by Markov model, and the optimized prediction value \( \hat{x}''^{(0)}(k) \) is obtained.
4. Case study

4.1. Data sources
The shared bicycle riding data used in this study comes from the website of Citi Bike in New York, USA. The number of shared bicycle rides per hour is taken as an example, take the data from 8:00-18:00 on Saturdays of the first three weeks of June 2019 for fitting and forecasting research. After sorting the data, the original data is shown in Figure 1.

![Figure 1. Original cycling data of shared bicycles](image)

4.2. Error index
The mean absolute percentage error (MAPE) can directly see the number of relative errors of the model for the overall data sample, which is an estimate of the overall data error. The absolute error, the relative error and the maximum absolute relative error can only evaluate a single data point, and MAPE is also commonly used to compare the prediction results of different models horizontally. MAPE calculation is as equation (18):

$$MAPE = \frac{1}{M} \sum_{i=1}^{M} \left| \frac{X_i - \hat{X}_i}{X_i} \right| \times 100\%$$  \hspace{1cm} (18)

The smaller the value of MAPE is, the higher the precision is. That is to say, if the value of MAPE is lower than 10\%, it is considered that the precision of the model is high [17]. If the range of MAPE is 11-20\%, the precision is relatively high. If it is 21-50\%, the precision is general. If it is greater than or equal to 51\%, the precision is poor.

4.3. Forecast result
The study selects the hourly cycling data of Citi Bike shared bikes in New York, the first three weeks of June 2019, from 8:00 to 18:00 on Saturdays as the original data, and the original data is substituted into the traditional grey GM(1,1) model, seasons GM(1,1) model (SGM(1,1) model), grey Markov model and seasonal grey Markov model, Matlab R2020b is used to calculate the predicted values of the four models to verify the accuracy of the model. The predicted values of the short-term demand for shared bicycles for the four models on the third week of 2019, Saturday, June 22, 8:00-18:00, are shown in Table 1.

| Time     | Actual number of rides/vehicles | GM(1,1) model Predictive value | MAPE  | Grey Markov model Predictive value | MAPE  | SGM(1,1) model Predictive value | MAPE  | Seasonal grey Markov model Predictive value | MAPE  |
|----------|---------------------------------|---------------------------------|-------|---------------------------------|-------|---------------------------------|-------|---------------------------------|-------|
| 8:00-9:00| 2072                            | 5830.5                          | 1.8139| 5830.5                          | 1.8139| 1928.3                          | 0.0694| 1928.3                          | 0.0694|
| 9:00-10:00| 3190                            | 5883.5                          | 0.8444| 3899.3                          | 0.2224| 2738.1                          | 0.1417| 2958.7                          | 0.0725|
| 10:00-11:00| 4408                            | 5937                            | 0.3469| 4467.5                          | 0.0135| 3936.3                          | 0.1070| 4289.1                          | 0.0270|
| 11:00-12:00| 5261                            | 5991                            | 0.1388| 5364.7                          | 0.0197| 4894.5                          | 0.0697| 5217.3                          | 0.0083|
The results in Table 1 show that the average MAPE of the seasonal grey Markov model is 2.72%, which is better than 33.15% of the traditional GM(1,1) model, 21.98% of the grey Markov model and 6.10% of the SGM(1,1) model.

In the forecast of the seasonal grey Markov model, the new information priority principle is used to select the residual sequence. Since the residual sequence is distributed, the sequence is divided into 5 states based on the sample mean and mean square deviation, which are respectively $[-40.3, \bar{x} - 0.5\bar{x})$, $[\bar{x} - 0.5\bar{x}, \bar{x} + 0.5\bar{x})$, $[\bar{x} + 0.5\bar{x}, \bar{x} + \bar{x})$, $[\bar{x} + \bar{x}, 671.3]$, the corresponding state transition probability matrices are as follows:

\[
P^{(1)} = \begin{bmatrix}
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 2/3 & 0 & 0 \\
0 & 0 & 1/2 & 1/2 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 
\end{bmatrix} \quad P^{(2)} = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1/2 & 1/2 & 0 \\
0 & 0 & 1/2 & 0 & 1/2 \\
0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 
\end{bmatrix} \quad P^{(3)} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 
\end{bmatrix}
\]

The comparison of the short-term demand forecasting results of shared bicycles of the four models are shown in Figure 2 and Figure 3.

![Figure 2. Comparison of the original data and the prediction results of the four models](image1)

![Figure 3. Error comparison of prediction results of four models](image2)

It can be seen from Figure 2 and Figure 3 that the forecast results of the seasonal grey Markov model are volatile and closer to the original data, and its forecast errors fluctuate within a lower level range, and the overall forecasting effect is better than the other three models.
5. Conclusions
In view of the periodicity, nonlinearity and randomness of shared bicycle riding data, the SGM(1,1) model is selected as the research object, and at the same time, considering the volatility and randomness of the residual sequence, seasonal grey Markov model is proposed. Research shows that for the short-term demand forecast of shared bicycles with cyclical fluctuations, the SGM(1,1) model has better adaptability than the GM(1,1) model. The accuracy of the GM(1,1) model and SGM(1,1) model modified by the Markov model has been improved to a certain extent, and the forecasting effect of the seasonal grey Markov model is significantly better than that of the traditional grey Markov model.

Acknowledgments
We acknowledge that the public data is provided by the Citi Bike website. This work is partially supported by the National Key Research and Development Program (2017YFC0803903). Technical support from Beijing Key Laboratory of Traffic Engineering is acknowledged.

References
[1] Cao D.D., Fan S.R., Zhang Y., Xia K.W. (2020) Short-term demand forecasting of shared bicycles based on long short-term memory neural network model. Science Technology and Engineering, 20: 8344-8349.
[2] Gao W., Meng Z.H., Li D.Z., Chen Z.Y. (2019) Short-term rent forecast method in bicycle sharing system. Computer Engineering and Design, 40: 1796-1801.
[3] Westland J.C., Mou J., Yin D. (2019) Demand cycles and market segmentation in bicycle sharing. Information Processing & Management, 56: 1592-1604.
[4] Fournier N., Christofa E., Knodler Jr M.A. (2017) A sinusoidal model for seasonal bicycle demand estimation. Transportation Research Part D: Transport and Environment, 50: 154–169.
[5] Gao X.H., Lee G.M. (2019) Moment-based rental prediction for bicycle-sharing transportation systems using a hybrid genetic algorithm and machine learning. Computers & Industrial Engineering, 128: 60-69.
[6] Mao S.H., Gao M.Y., Xiao X.P., Zhu M. (2016) A novel fractional grey system model and its application. Applied Mathematical Modelling, 40: 5063-5076.
[7] Dang Y.G., Liu S.F., Liu B. (2005) The GM models that $x^{(1)}(n)$ be taken as initial value. Chinese Journal of Management Science, 13: 133-136.
[8] Gao N., Gao C.Y. (2017) A new method of constructing background value of GM(1,1) model by using nonhomogeneous exponential function. Statistics and Decision Making, 24-27.
[9] Yang X.G., Li D.S., Ren G.Y. (2018) Prediction and analysis of road traffic accidents in Beijing based on residual correction grey forecasting mode. Mathematics in Practice and Theory, 2018, 48: 159-166.
[10] Cao J., Shi S.L., Cao H.J., Li Y., Wang Y., Chen X.Y. (2019) Study on prediction and relationship of road transportation accidents of dangerous chemicals and traffic accidents based on GM(1,1)-Markov model. Journal of Safety Science and Technology, 15: 26-31.
[11] Qian W., Yang K.L., Yang H.H., Xu Q.Z. (2015) Short-term traffic flow prediction based on combined models. Computer Simulation, 32: 175-178+193.
[12] Xiao X P, Yang J W, Mao S H, Wen J.H. (2017) An improved seasonal rolling grey forecasting model using a cycle truncation accumulated generating operation for traffic flow. Applied Mathematical Modelling, 51: 386-404.
[13] Tang Y.Y., Li L., Qin Y. (2018) Highway freight price prediction based on improved grey Markov model. Journal of Transportation Engineering and Information, 16: 38-43+48.
[14] Wang X., Liu X.Y. (2017) Research on traffic accident prediction based on grey Markov model. Transportation Technology and Economy, 19: 9-13.
[15] Si S.K., Sun X.J. (2011) Mathematical modeling algorithm and application. National Defense Industry Press, Beijing.
[16] Shen Q.Q., Liu H.Z., Wang Y., Huang Y. (2018) Seasonal grey Fourier model with applications to short-term traffic flow forecasting. Journal of Nantong University (Natural Science Edition), 17: 30-35.

[17] Shen Q.Q., Wang Y., Huang Y., Liu H.Z. (2020) Grey Verhulst-Markov model with improved initial value and its application. Statistics and Decision Making, 36: 30-33.