Prediction Of Arrival Volume Based On Improved Markov SCGM(1,1)c Model

Mengying FAN 1  Mingxia ZHENG 2  Hu LIU 3

1 Zhejiang Ocean University Port & Transportation School, Zhoushan Zhejiang 316022
China

Corresponding author’s e-mail: 1119719820@qq.com

Abstract. Aiming at the fact that the arrival gauges of port ships have the characteristics of increasing trend but abnormal fluctuation, and combined with the prediction methods commonly used for random events, this paper proposed a Markov SCGM(1,1)c model for a single-factor system, and the Markov SCGM(1,1)c model is improved by using the method of equal dimension gray number and sliding transfer probability matrix. The results show that the residual mean of the improved Markov SCGM(1,1)c prediction model is reduced to a certain extent, that is, the fitting degree is improved, the prediction accuracy is improved, and the ability to modify the mutation data is enhanced, which meets the actual requirements and provides theoretical support for port anchorage planning.

1. Introduction
Zhoushan port is a deep-water hub port, which is located in the international ocean trunk line. It will become an important port in the Asia-Pacific region and even the world as a result of the "One Belt and One Road" strategy. According to statistics, from January to October of 2017, the cargo handling capacity of Zhoushan port area was 392.95 million tons, increasing by 9.3%, and the container handling capacity was 940,300 TEU, increasing by 25.4%. It can be seen that the throughput of Zhoushan port shows an obvious growth trend, and the amount of ships arriving at the port also increases accordingly, and the contradiction between the large demand of ships arriving at the port and the anchorage facilities with lagging development intensifies. Therefore, the prediction of arrival quantity of port ships based on the prediction theory is of great importance to the rational planning of port anchorage and efficient management of waterway traffic. For the prediction method of gray GM (1,1), the background value of GM (1,1) model is a smooth formula, which is suitable for the prediction of small time intervals and small data. The accuracy of this prediction model can guarantee the random events with an increasing relationship to some extent, while the accuracy will be greatly reduced if there are decreasing outliers. SCGM(1, 1)c model has been successfully applied in many fields as the most powerful development of GM(1, 1) model. SCGM(1, 1)c model can search for useful information from its own time data series to a greater extent and explore its internal laws. Compared with other gray prediction models, it has a more solid theoretical basis and features of less required information, simple calculation and high accuracy. A large number of scholars have combined Markov prediction with gray prediction and applied it in traffic accident prediction, background value optimization [9]-[11] and other aspects. Markov prediction takes time series as a random process, studies the initial probability of different states of things and the probability of the transition between states, and determines the change trend of the future state of things. Markov chain prediction can make up for the defects of gray prediction and improve the accuracy of its combination with gray prediction.
This paper combined with the advantages of single factor system cloud ash model (SCGM(1,1)c) and markov chain, the markov SCGM(1,1)c model improved by the two methods is adopted. Finally, the number of ships arriving at port is forecasted in order to get more realistic forecasting results.

2. Establishment and improvement of markov SCGM(1,1)c model

2.1. Establishment of system cloud gray SCGM(1,1)c model

In view of the two defects of small number of original data samples and incomplete statistical information of ships entering and leaving the port, it is difficult to use general mathematical statistical methods. SCGM(1,1)c model of the system cloud gray does not need to take into account many influencing factors of ship arrival. Based on the background of system cloud, it is constructed according to the grey dynamic modeling principle of integral generation transformation and trend relational analysis. The steps of model establishment are as follows:

2.1.1. **Data handling.** The natural logarithm method is adopted to process the original data, and the processed data is taken as the original sequence for modeling.

- Let the original sequence be \( X(\omega)(k) \), in which
  \[
  X(\omega)(k) = \{ X(\omega)(1), X(\omega)(2), X(\omega)(3), X(\omega)(4), \ldots, X(\omega)(n) \}.
  \]

- Integrate the transformation \( X(\omega)(k) \), from
  \[
  X(\omega)(k) = \sum_{m=1}^{k} \bar{X}(m), \quad k = 2,3,\ldots,n
  \]
  The following is obtained:
  \[
  \bar{X}(1)(k) = \{ X(\omega)(1), X(\omega)(2), X(\omega)(3), X(\omega)(4), \ldots, X(\omega)(n) \}.
  \]
  In which
  \[
  \bar{X}(1)(k) = \frac{1}{\omega} (X(\omega)(k) + X(\omega)(k-1)).
  \]

2.1.2. **Set up a response function.** Set the integral generating sequence of the original time sequence of the number of ships entering and leaving the port \( \{ \tilde{X}^{(1)}(k) \} \) and non homogeneous exponential discrete function \( f_{\omega}(k) = b \cdot e^{\alpha(k-1)} - c \) then system cloud gray SCGM(1, 1)c the mode is:

- \[
  \frac{d\tilde{X}^{(1)}(k+1)}{dk} = a \tilde{X}^{(1)}(k+1) + U
  \]
  The first response function is
  \[
  \tilde{X}^{(1)}(1) = \left( \tilde{X}^{(1)}(1) + \frac{U}{a} \right) e^{\alpha k} - \frac{U}{a}
  \]
  \[
  a = \ln \left( \frac{\sum_{k=1}^{n} \tilde{X}^{(1)}(k-1)X(\omega)(k)}{\sum_{k=3}^{n} (\tilde{X}^{(1)}(k-1))^2} \right)
  \]
  \[
  b = \frac{(n-1) \sum_{k=2}^{n} e^{\alpha(k-1)} X(\omega)(k) - \sum_{k=2}^{n} e^{\alpha(k-1)} \sum_{k=2}^{n} X(\omega)(k)}{(n-1) \sum_{k=3}^{n} e^{\alpha(k-1)} - \sum_{k=3}^{n} e^{\alpha(k-1)} \sum_{k=3}^{n} X(\omega)(k)}
  \]
  \[
  c = \frac{1}{n-1} \left( \sum_{k=1}^{n} e^{\alpha(k-1)} b - \sum_{k=2}^{n} \tilde{X}^{(1)}(k) \right)
  \]
  \[
  \text{Then } \tilde{X}^{(1)}(k) = b - c, \quad U = ac
  \]

2.1.3. **Reduction treatment.** Conduct reduction treatment for \( \tilde{X}^{(1)}(k) \), then get the raw data in the system cloud gray SCGM(1, 1)c. The prediction model is:

- \[
  \hat{X}^{(1)}(k+1) = a^{\alpha(k-1)} \frac{2b(1-e^{-\alpha})}{(1+e^{-\alpha})}
  \]
  \[
  \text{In (6): When } k = 1,2,\ldots,n, \quad \{ \hat{X}^{(1)}(k) \} \text{ is the fitting sequence of original data sequence } \{ X^{(1)}(k) \}; \text{ when } k > n, \quad \{ \hat{X}^{(1)}(k) \} \text{ is the forecast sequence of the raw data sequence } \{ X^{(1)}(k) \}.
  \]

2.1.4. **Model accuracy tes.** Calculate the mean value of the original data sequence \( \bar{X}^{(0)}(k) = \frac{1}{n} (X^{(0)}(1) + X^{(0)}(2) + \cdots + X^{(0)}(k)) \) and variance \( S_0^2 = \frac{1}{n} \left[ (\bar{X}^{(0)}(1) - \bar{X}^{(0)} \right)^2 + (\bar{X}^{(0)}(2) - \bar{X}^{(0)})^2 + \cdots + (\bar{X}^{(0)}(k) - \bar{X}^{(0)})^2 \right] \text{; residual sequence mean } \bar{e}^{(0)} = \bar{X}^{(0)}(k) - \bar{X}^{(0)}(k) \text{ and variance } S_\epsilon^2. \text{ Let the ratio of variance be } C = S_\epsilon^2 S_0^2. \text{ The probability of small error is:}

- \[
  P = \left| \bar{e}^{(0)}(k) - \bar{e}^{(0)} \right| < 0.6745 S_\epsilon
  \]
  Table 1 is the judgment standard of model accuracy level:
Table 1. SCGM(1, 1)c model accuracy level judgment standard

| P value | C value | Model accuracy level |
|---------|---------|----------------------|
| >0.95   | <0.35   | 1 level              |
| >0.80   | <0.50   | 2 level              |
| >0.70   | <0.65   | 3 level              |
| <=0.70  | >=0.65  | 4 level              |

3. Markov chain and its improvement

3.1. Markov chain

The essence of system cloud gray SCGM(1, 1)c prediction is to use exponential curve to fit the original data, the predicted result is a relatively smooth curve. Therefore, the fitting of the data with large fluctuation is poor, and the accuracy of the prediction results using system cloud gray SCGM(1,1)c alone was poor. While for Markov theory: The future state is only related to the present, and the transition probability between the system states is used to predict the future development of the system, which can solve the problem of large random fluctuation of data and improve the prediction accuracy. The main steps of constructing Markov chain are as follows:

3.1.1. State division. Calculate the relative residuals \( \omega(k) \) between the predicted results and the real values, and divide the change interval according to the relative residuals. If a state is set for each interval, the corresponding state of the sequence of relative residuals can be obtained.

3.1.2. Prediction result optimization. The probability of state \( N_i \) transition to state \( N_j \) after \( n \) steps is \( p_{ij}^{(n)} = \frac{d_{ij}^{(n)}}{d_i} \), in which, \( d_{ij}^{(n)} \) is the number of the state \( N_i \) transition to \( N_j \). Then \( n \) step state transition matrix is:

\[
P^{(n)} = \begin{bmatrix}
p_{11}^{(n)} & \cdots & p_{1n}^{(n)} \\
p_{21}^{(n)} & \ddots & \vdots \\
p_{n1}^{(n)} & \cdots & p_{nn}^{(n)}
\end{bmatrix}
\]  

(8)

3.1.3. Prediction result optimization. According to the year to be predicted, the initial state changes to the state of the predicted year after one or more state transitions. Combined with the relative residual formula \( \omega^{(t)}(k) = \frac{x^{(t)}(k) - x^{(0)}(k)}{x^{(0)}(k)} \), the relative residuals for that year can be obtained. The optimized predicted value can be obtained after deformation:

\[
X^{(0)}(k) = \frac{x^{(t)}(k)}{1 - \omega^{(t)}(k)}
\]  

(9)

3.1.4. Markov chain improvement. The transition matrix is constructed with the year \( t \) as the initial state, the state distribution of the year \( t+1 \) is obtained, and the predicted value of the year is calculated. Remove the oldest data in the original sequence, construct the matrix with the state of the year \( t+1 \) as the initial state, and obtain the state probability distribution of the year \( t+2 \).

4. Application examples and comparative analysis

In this paper, the port entry data of Laotangshan port area of Ningbo-Zhoushan port in Zhejiang province from 2008 to 2017 are selected for analysis, because the ship arrival season is different between busy and free. Therefore, the monthly forecast is conducted, and the forecast results for 2018 were inferred based on the monthly forecast results Markov SCGM (1,1)c model was used for the first prediction, and the model improved by the equal-dimension gray filling was used for the second prediction. Take the data from 2008 to 2015 as the original data sequence, and finally predict the number of ships entering and leaving ports in 2018.
4.1. First prediction
Data sequences in May from 2008 to 2017 are selected as samples in this paper. For the original data, See Table 2 for details.

| Year    | 2008 | 2009 | 2010 | 2011 | 2012 |
|---------|------|------|------|------|------|
| Number  | 298  | 325  | 380  | 398  | 432  |
| of vessels entering or leaving the port |       |      |      |      |      |

| Year    | 2013 | 2014 | 2015 | 2016 | 2017 |
|---------|------|------|------|------|------|
| Number  | 454  | 459  | 542  | 548  | 921  |
| of vessels entering or leaving the port |       |      |      |      |      |

4.1.1. Establishment of SCGM(1, 1)c prediction model. For original sequence \(X^{(0)}(k) = \{298,325,380,398,432,454,459,542,548,921\}\) the natural logarithm can be adopted. The data sequence \(\{5.6971, 5.7838, 5.9402, 5.9865, 6.0684, 6.1181, 6.1291, 6.2953, 6.3062, 6.8255\}\) is obtained after processing. For the processed data, the model can be set up according to \("(1) - (6)\", so as to establish SCGM(1, 1)c of ship arrivals. The model as follows:

\[a = 0.01299962 \quad b = 441.6697432 \quad X^{(0)}(k+1) = e^{a(k-1)} \frac{b(1-e^{-a})}{(1+e^{-a})} \quad (k=1,2,3,...,n)\]

4.1.2. Model accuracy test . The mean value of the original data sequence \(\bar{X}^{(0)}(k) = 6.002297616\) is calculated according to formula (7) using the post-difference test method, variance \(S^2 = 0.03308371\). Residual sequence mean \(\bar{e}^{(0)} = -0.00908874\), variance \(S^2 = 0.001446093\). The variance ratio \(C = \frac{s_1}{s_0} = 0.20906964\); Small probability error \(0.6745 \times S_0 = 0.122684319\). While \(\{ |e^{(0)}(k) - \bar{e}^{(0)}| < 0.122684319\} = 1\), Table 1 shows that the model is level 1 for accuracy.

4.1.3. Using Markov to revise the 2016 date. From the above data, the relative residual distribution region of the number of ships arriving at the port from 2008 to May 2017 can be obtained as \((-9\%, 5\%)\), then the fitting sequence is divided into three states according to the relative residuals, \(N_1=(-9\%, -3\%)\), \(N_2=(-3\%, 1\%)\), \(N_3=(1\%, 5\%)\), the status of each year is shown in table 3.

| Year | Relative residual | State | Year | Relative residual | State |
|------|-------------------|-------|------|-------------------|-------|
| 2008 | -0.046979866      | N_1   | 2012 | 0.018518519       | N_3   |
| 2009 | -0.033846154      | N_1   | 2013 | -0.011013216      | N_2   |
| 2010 | 0.044736842       | N_3   | 2014 | -0.082788671      | N_1   |
| 2011 | 0.015075377       | N_3   | 2015 | 0.005535055       | N_2   |

Then the state probability distribution in 2016 is:

\[P(1) = P_0(N_1)P_1^{(N)}(0)P_1^{(N)}(1,0,0) = \left(\begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right) \left(\begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right) = \left(\begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right)\]

In which \((11): P_1^{(N)}(0)\) is the status in 2016, it can be any one of \(N_1\), \(N_2\), \(N_3\), and the probability is equal to \(\frac{1}{3}\), while \(P_1^{(N)}\) is the one-step state transition probability matrix of 2016 data.

Calculate the average relative residual of 2016 data: \(\omega^{(0)}(1) = \frac{x^{(0)}(k)-\bar{e}^{(0)}(k)}{x^{(0)}(k)} = -0.0134\)
According to formula (9), the ship arrivals and departures in 2016 after correction are 578.

4.2. The second forecast

4.2.1. Model refinement. On the basis of the first prediction result for the method of equal dimension gray number, the first data in the original data sequence \( \{X^{(0)}(k)\} \) is removed, add \( \hat{X}^{(0)}(n + 1) \) to \( \{X^{(0)}(k)\} \), and keep the dimension of data sequence unchanged to obtain a new data sequence: \( \{X^{(0)}(k), X^{(0)}(k + 1), \ldots, X^{(0)}(n), \hat{X}^{(0)}(n + 1)\} \). Repeat the initial modeling steps until the final prediction goal is reached.

4.2.2. Data analysis. According to the method of equal dimension gray number complement, the predicted 2016 data "578" is added into the original data sequence as the known data, and the 2008 data "298" is removed, keeping the original data dimension unchanged. The second prediction is made in the same way: Establish model \( (a = 0.012750266, b = 458.6391608) \), the forecast for 2017 is 650, after the Markov correction, it is 695. Similarly, the forecast data for 2018 is 802. The results of the two predictions are shown in table 4.

| Year | Original data | First prediction SCGM(1, 1)c+ Markov chain | Second prediction SCGM(1, 1)c+ Markov chain |
|------|---------------|---------------------------------------------|---------------------------------------------|
| 2008 | 298           | 312                                         | 347                                         |
| 2009 | 325           | 336                                         | 374                                         |
| 2010 | 380           | 363                                         | 392                                         |
| 2011 | 398           | 392                                         | 403                                         |
| 2012 | 432           | 424                                         | 436                                         |
| 2013 | 454           | 459                                         | 471                                         |
| 2014 | 459           | 497                                         | 510                                         |
| 2015 | 542           | 539                                         | 552                                         |
| 2016 | 548           | 578                                         | 599                                         |
| 2017 | 921           |                                             | 695                                         |

4.3. Result analysis

According to the results in table 3 and table 4, the model accuracy and prediction results are mainly analyzed. The prediction accuracy is shown in table 5, and the prediction results of the number of ships entering and leaving the port are shown in figure 1.

| Prediction          | The mean residual | Average error% | C       | P  |
|---------------------|-------------------|----------------|---------|----|
| First prediction    | 0.023425817       | 0.015116589    | 0.20906964 | 1  |
| Second prediction   | 0.010367017       | -0.009823487   | 0.20044315 | 1  |
It can be seen from figure 1 that the predicted value of markov SCGM(1, 1)c model shows a basic linear upward trend, and the trend is basically the same as that of the actual value curve. For the mutation number of "921" in 2017, the predicted value becomes more close to the actual value after the model is modified by the method of equal dimension gray number supplement, which has more reference value. And this paper adopts the improved combination model to predict the number of ships in 2018 as 802.

5. Conclusion
The traditional SCGM(1,1)c and the improved SCGM(1,1)c prediction model are used to predict and compare the accuracy of ship arrivals in Laotangshan port area of Ningbo-Zhoushan port in May from 2008 to 2017. Results show that the double improved SCGM (1, 1)c model predicted value and actual value are basically in accordance with, the residual mean and average relative error is also a certain degree of decline, the fitting degree has a certain degree of improvement, also can largely overcome the influence of abnormal fluctuations in the data, proved that the improved SCGM (1, 1)c the precision of the prediction model is improved obviously.

Project support:
This paper was supported by Zhejiang University Student Science and Technology Innovation Program and New Miao Talent Program, No. xj2018064, No. 2018R411015.

References:
[1] Liu W.F., Wang J.,(2018), The Development Road of International Strong Port of Ningbo Zhoushan Port under the Background of "One Belt and One Road" [J]. Zhejiang Economy, 2018(13):23-24.
[2] Xu , D.Q., (2017), Analysis of Container Transportation in Ningbo Zhoushan port in 2016 and Prospects for 2017 [J].China Ports,2017(03):56-58.
[3] Wang Q., Wang Z.M., (2013), Forecasting of Maritime Traffic Accidents Based on the Improved SCGM (1, 1) c—Markov Model [J]. Navigation of China, 36(04):119-124.
[4] Zhang S.H. , Cao J.Y. , Chen M.Y.(2004), Grey Markov SCGM (1, 1) prediction model [J]. Journal of Foshan University (Natural Science Editi on),2004(01):16-19
[5] Wang X., Liu X.Y., (2017), Research of traffic accident prediction based on Gray Markov model [J]. Technology & Economy in Areas of Communications, 19(04):9-13.