Research on container throughput forecast of Shanghai port based on unbiased gray model

Li Chen

Merchant Marine College, Shanghai Maritime University, Shanghai 200120, China

chenlithirteen@gmail.com

Abstract

This paper uses unbiased gray prediction model to predict the container throughput of Shanghai port from 2011 to 2018. After error analysis of the prediction results and original data, it comes to the conclusion that the unbiased gray model has a higher accuracy and better effect on the container throughput of Shanghai port. Finally, this model is used to predict the container throughput of Shanghai port from 2019 to 2023.

Keywords

Gray model,unbiased gray model,Shanghai port,container throughput,forecast.

1. Introduction

As an important index of developing international shipping center, the accurate forecast of container throughput of Shanghai port can provide important reference information for the future development of Shanghai port and the whole Yangtze river delta region.There are many kinds of prediction models that be used widely. Bomin Li et al.[1] used the logistics throughput of Fangchenggang in recent ten years and the linear regression prediction to predict and analyze the cargo throughput of fangchenggang in the next five years. Yiqi Zhao[2] established a product season model by using the throughput of foreign trade goods in China's coastal ports, and predicted the throughput in the next six months after verifying the model. Leiyu Zhang et al.[3] proposed a modeling and prediction method based on the algorithm principle of principal component analysis and support vector machine regression method,and used the statistical data of lianyungang from 1999 to 2016 to predict the container throughput of the port from 2019 to 2020. Junhua Zhang[4] used the grey model to predict the cargo throughput of ningbo port, Shanghai port and other big ports.

Grey model is a kind of method which is suitable for the system with short time series, few statistical data and incomplete information, and uses the establishment of mathematical model and the processing and analysis of a small amount of incomplete information to predict the future development trend of the system, with high modeling accuracy and simple operation. However, since the error exists in the exponential series fitting and parameter selection, the reduction of the error can further improve the accuracy of the prediction results. In order to reduce the error and improve the accuracy, this paper adopts unbiased gray prediction model. This model can effectively avoid the prediction deviation in the gray prediction model, so that the prediction result is more accurate.

2. Establishment of unbiased gray prediction model

2.1 Establishment of grey model

Gray system usually refers to those systems that have the fuzziness of hierarchical and structural relations, the randomness of dynamic changes, and the incompleteness or uncertainty of index data. Grey model is a prediction model for grey system. The process of continuous development and change of things inside the system can be shown through it. The future development of things can be obtained through the establishment of grey model after processing incomplete information.

Steps of grey model modeling[5]:

(1) Establish the original sequence and conduct data processing:

Establish non-negative original sequence:

$$x^{(0)}(i) = [x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)] \tag{1}$$

Carry out first-order accumulation of the data in the original sequence:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1,2,3, \dots, n$$

$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(n)\}$$

Test the sequence smoothness:

$$X^{(k)} = \frac{x_0(k-1)}{x_0(k)}, \mathrm{k} = 2,3, \dots, \mathrm{n} \tag{2}$$

If want to use the data sequence to establish a grey model,we need to guarantee $(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}})$ contains the $X^{(k)}$, if not included,the data need to be processed.

Test the exponential property of the sequence:

$$\pi^{(1)}(k) = \frac{x^{(1)}(k)}{x^{(1)}(k-1)}, k = 3,4, \dots, n \tag{3}$$

If $\pi^{(1)}(k) \in [1,1.5]$,the data sequence has an exponential growth rule.Let the original sequence be a strict exponential sequence,that is:

$$x^{(0)}(k) = Me^{-\eta(k-1)}, k = 1,2,3, \dots, n$$

$$x^{(1)}(k) = M\frac{1-e^{\eta k}}{1-e^{\eta}}, k = 1,2,3, \dots, n \tag{4}$$

establish the data matrix B,Y_n

$$\mathrm{B}=\begin{vmatrix} -\frac{1}{2}[x^{(1)}(1) + x^{(1)}(2)] & 1 \\ -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(3)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(n-1) + x^{(1)}(n)] & 1 \end{vmatrix} \tag{5}$$

$$Y_n = |x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)|^T$$

calculate parameters a and u:

$$\begin{vmatrix} a \\ b \end{vmatrix} = (B^T B)^{-1} B^T Y_n = \begin{bmatrix} \frac{2(1-e^{\eta})}{1+e^{\eta}} \\ \frac{2M}{1+e^{\eta}} \end{bmatrix} \tag{6}$$

final fitting results:

$$\hat{x}^{(0)}(1) = M$$

$$\hat{x}^{(0)}(k) = \frac{-Me^{\eta}(1-e^{a})}{1-e^{a}} e^{-a(k-1)}, k = 2,3, \dots, n \tag{7}$$

2.2 Establishment of unbiased grey model

According to formula (6)：

$$\eta = \ln\frac{2-a}{2+a}, M = ln\frac{2b}{2+a}$$

Among them $\hat{b} = \eta$， $\hat{A} = M$, steps (1) ~ (3) the same as the establishment of gray model steps.

Calculate unbiased grey parameters:

$$\hat{b} = ln\frac{2-a}{2+a},\ \ \hat{A} = \frac{2b}{2+a} \tag{8}$$

Build the original data sequence model:

$$\hat{x}^{(0)}(1) = x^{(0)}(1)$$

$$\hat{x}^{(0)}(k+1) = \hat{A}e^{\hat{b}\hat{k}}, k = 1,2,\dots,n \tag{9}$$

2.3 Model test

In order to ensure that the fitting ability of the established model meets the prediction requirements, the model should be tested.

Relative error test[6]:

The predicted result sequence is as follows:

$$\hat{X}^{(0)} = [\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(n)] \tag{10}$$

First,calculate the residual:

$$\mathrm{E} = [\mathrm{e}(1), \mathrm{e}(2), \dots, \mathrm{e}(\mathrm{n})] = X^{(0)} - \hat{X}^{(0)} \tag{11}$$

Among them, $\mathrm{e(k)} = x^{(0)}(k) - \hat{x}^{(0)}(k), k = 1,2,\dots,n$

Calculate the relative residuals

$$\mathrm{rel(k)} = \frac{e(k)}{x^{(0)}k} \times 100\%, k = 1,2,\dots,n \tag{12}$$

Calculate the average relative error

$$\mathrm{q} = \frac{1}{n}\sum_{k=1}^{n}|rel(k)| \tag{13}$$

Posterior test: Set according to the modeling method for the unbiased grey-forecasting model $\hat{X}^{(0)}$ as shown in formula (10)，the residual is shown in formula (11), s_1^2 and s_2^2 are the variances of original sequence $X^{(0)}$ and residual sequence E, C is the ratio of mean variance of the index, p is the probability of small error.

$$s_1^2 = \frac{1}{n}\sum_{k=1}^{n}[x^{(0)}(k) - x]^2$$

$$s_2^2 = \frac{1}{n}\sum_{k=1}^{n}[e(k) - e]^2 \tag{14}$$

$$\bar{x} = \frac{1}{2}\sum_{k=1}^{n}x^{(0)}(k), \bar{e} = \frac{1}{n}\sum_{k=1}^{n}e(k)$$

$$\mathrm{C} = \frac{s_2}{s_1} \tag{15}$$

$$\mathrm{p} = \mathrm{P}\{|\mathrm{e(k)} - \bar{e}| < 0.6745s_1\} \tag{16}$$

As an important index of the posterior test, in order to make the variance of the original sequence large and the variance of the residual sequence small, the mean variance ratio should be as small as possible, so as to make the original data discrete degree large and the residual discrete degree small. On the contrary, the small error probability, as another important index of the posterior test, should be as large as possible. The larger the small error probability value is, the smaller the difference between the residual error and the average residual error is than the given point $0.6745s_1$.

In order to comprehensively describe the accuracy of the prediction model, relative error, mean variance ratio and small error probability are usually used to analyze and evaluate the prediction model. The accuracy of the model is usually divided into four levels according to these three indicators,details are in the table 1.

Table 1 Accuracy comparison table of unbiased grey prediction model

Model accuracy level	relative error:q	Mean variance ratio:C	Small error probabilit:p
1(Good)	≤0.01	C≤0.35	0.95≤p
2(Qualified)	≤0.05	0.35＜C≤0.5	0.80≤P＜0.95
3(Reluctant)	≤0.10	0.5＜C≤0.65	0.70≤P＜0.8
4(Unqualified)	＞0.20	0.65＜C	P＜0.7

Model accuracy level= max {level of q， level of p， level of C}

3. Container throughput forecast of Shanghai port

Shanghai port is located in the center of China's coastline, the throat of the Yangtze river estuary. Shanghai port area is composed of 66.6 kilometers of the Huangpu river, 106.5 kilometers of the south bank of the Yangtze river estuary south waterway and Yangshan deep-water port. With the completion and operation of Yangshan deep-water port phase iv project, Shanghai international shipping center, the annual container throughput of Shanghai port will be greatly improved. Through the accurate prediction of container throughput of Shanghai port, it will play a reasonable guiding role in the allocation of port resources and the planning of port layout.

Based on the selected data from 2011 to 2018 in Shanghai port container throughput, unbiased grey-forecasting model is set up to fit the historical data,then analyze the error of prediction results , test the applicability of the model,.At last,predict the container throughput data and development trend of Shanghai port in the next five years.

3.1 The selection of data sample

The data samples are composed of the international container throughput data of Shanghai port from 2011 to 2017 contained in the statistical yearbook of Shanghai 2018 issued by Shanghai Bureau of Statistics and the international container throughput data of Shanghai port published by the website of Ministry of Transport of China.AS the table 2 shows:

Table 2 Container throughput of Shanghai port from 2011 to 2018

Year	International standard container throughput(million TEU)
2011	31.739
2012	32.529
2013	33.617
2014	35.285
2015	36.537
2016	37.133
2017	40.233
2018	42.010

3.2 The processing of data

As shown in table 2, the container throughput of Shanghai port has been growing during the eight years from 2011 to 2018. In order to prove that the data can be used to establish grey prediction model, its smoothness and index should be tested. Firstly,the non-negative original data sequence should be established,and than add up the data once.

$$X^{(0)} = \left[X^{(0)}(1), X^{(0)}(2), X^{(0)}(3), \dots, X^{(0)}(8)\right]$$
$$= \{3173.9,3252.9,3361.7,3528.5,3653.7,3713.3,4023.3,4201.0\}$$
$$x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(8)\}$$
$$= \{3173.9,6426.8,9788.5,13317,16970.7,20684,24707.3,28908.3\}$$

Smoothness test and index test were carried out on the data, and the test results were shown in table 3 and table 4:

Table 3 The values of data smoothness test

k	$X^{(k)}$	k	$X^{(k)}$
2	0.9757	6	0.9839
3	0.9676	7	0.9229
4	0.9527	8	0.9577
5	0.9657		

The data level ratio range is (0.7788,1.2840), so it can be seen that this set of data sequence has a good smoothness.

Table 4 The values of index text

k	$\pi(k)$	k	$\pi(k)$
3	1.5231	6	1.2188
4	1.3605	7	1.1945
5	1.2744	8	1.1700

Most of the data obtained from the above table are between [1,1.5], so it can be seen that this set of data sequence has a good index.

3.3 The prediction results of unbiased grey model

First, according to formula (6), the grey differential equation is established and the matrix is obtained as follows:

$$\begin{bmatrix} 31.739 \\ 32.529 \\ \vdots \\ 42.0102 \end{bmatrix} = \begin{bmatrix} -31.739 & 1 \\ -64.268 & 1 \\ \vdots & \vdots \\ -289.083 & 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix}$$

Among them:

$$\begin{vmatrix} a \\ b \end{vmatrix} = (B^T B)^{-1} B^T Y_n = \begin{bmatrix} -0.0426 \\ 30.1932 \end{bmatrix}$$

The development coefficient and coordination coefficient were obtained as follows: a=-0.0426 and b=30.1932. According to formula (8), parameters of unbiased grey model can be obtained as follows:

$$\hat{b} = ln\frac{2-a}{2+a}=0.0426,\ \hat{A} = \frac{2b}{2+a}=30.8504$$

Then through combining computing formula of the unbiased grey-forecasting model and $\hat{b} = 0.0426$, $\hat{A} = 30.8504$ that can forecast results, the results shown in the table below:

Table 5 The values of the prediction

Year	Actual container throughput(million TEU)	Predicted values of unbiased grey mode(million TEU)	Residual	Relative residual (%)
2011	31.739	31.739	0.00	0.00
2012	32.529	32.194	0.3354	1.03
2013	33.617	33.595	0.0217	0.06
2014	35.285	35.058	0.2271	0.64
2015	36.537	36.584	-0.0473	-0.13
2016	37.133	38.177	-1.0441	-2.81
2017	40.233	39.839	0.3937	0.98
2018	42.010	41.574	0.4362	1.04

Testing error: ①The relative error: $q = 0.0084 \leq 0.01$,the result of the prediction is good, ②Mean variance ratio: $C = \frac{s_2}{s_1} = 0.15 \leq 0.35$, the result of the prediction is good, ③ Small error probability: $p = 100\% \geq 0.95$, the result of the prediction is good. Finally, the container throughput of Shanghai port in the next five years is predicted by the unbiased grey model. Table 6 shows the predicted results.

Table 6 Container throughput of Shanghai port under unbiased grey forecast

Year	Actual container throughput(million TEU)	Predicted values of unbiased grey mode(million TEU)
2011	31.739	31.739
2012	32.529	32.194
2013	33.617	33.595
2014	35.285	35.058
2015	36.537	36.584
2016	37.133	38.177
2017	40.233	39.839
2018	42.010	41.574
2019		43.384
2020		45.273
2021		47.244
2022		49.301
2023		51.447

4. Conclusion

In this paper, the unbiased grey model is used to predict the container throughput of Shanghai port from 2019 to the next five years. The sequence obtained by unbiased grey model modeling method was compared with the original sequence, and the posterior test was carried out. After the relative error, mean variance ratio and small error probability were tested respectively, it was found that all the values met the requirements of model accuracy level 1,so,we can come to the conclusion: The unbiased grey model is applicable to the prediction of container throughput at Shanghai port. In addition, according to the prediction results, we can also see that in the next five years from 2019, the container throughput of Shanghai port will keep rising rapidly, and by 2023, the annual container throughput of Shanghai port will reach 51.447 million TEU.

References

[1] B.M.LI,S.S.LI: Forecast and Analysis of Port Logistics Development in Fangchenggang, Foreign Economic and Trade,2019No.3,p.47-49.

[2] Y.Q.Zhao:Port Throughput Prediction based on Time Series Analysis, Market Research, Vol.8(2018)No.13,p.35-37.

[3] L.Y.Zhang,Y.Yang: Research on Container Throughput Prediction Model based on PCA-SVM, Jiangsu Business Theory,2019No.5,p.33-38.

[4] J.H.Zzhang: Forecast and Trend Analysis of Port Throughput in the 13th five-year Plan based on Grey GM(1,1) Model, Logistics Technology,2017No.2,p.65-70.

[5] F.C.Jiang,R.Z.Shi,Q.D.Ma,Q.B.Fan,F.Zhang,H.B.Zou: Application of Unbiased Grey Model in Ship Traffic Flow Prediction of Dongying Port, Journal of Guangzhou Navigation College,Vol.26(2018)No.4,p.45-49.

[6] H.Y.Zheng: Prediction of Cargo Throughput of Ningbo Zhoushan Port based on Grey Model, Market Weekly,2018No,10,p.5-6,13.