

Port Cargo Throughput Forecasting Based On Combination Model

Zhizhen Chen ,Yan Chen, Taoying Li

College of Transportation Management, Dalian Maritime University, China

*dlmu_brenda@163.com

Keywords: GM(1,1); Pearl Curve; exponential smoothing; combination forecasting; cargo throughput

Abstract. Port cargo throughput forecasting is an essential issue in port planning and management. Owing to that cargo throughput is affected by many factors; a single model is often difficult to get an accurate prediction. On the basis of discussing single modeling principle, the combination forecasting model based on Minimum-variance is therefore proposed for forecasting cargo throughput. By comparing the performance evaluation indexes, the results of examples imply that the combination model has higher precision than a single model.

1 Introduction

Port cargo throughput refers to the total amount of cargo via waterways into and out of the port area [1]. Port cargo throughput is an important indicator to measure port management level and operation scale. Forecasting cargo throughput accurately plays an essential role in adjusting the direction of port development, preparing a reasonable port operation plan, planning the construction of port infrastructure and enhancing the throughput capacity of the port. It provides a quantitative basis for the scientific management and decision-making.

Based upon the summary of a great amount of literature data, the quantitative forecasting method of port throughput mainly includes time series method [2], causal analysis method [3], combination forecasting method [4], etc. Due to the diversity of the factors affecting the throughput, it is difficult to meet the needs to use a single linear or curve model when the data to be predicted is fluctuant. In the past few years, researchers and scholars have put up with various methods to go up the forecasting performance. Through the observation of container throughput of Ningbo port in recent years, Shi et al. [5] used the combined model of GM (1,1) and cubic exponential smoothing method to optimize the container throughput forecast of Ningbo port in the next few years. Considering Wenzhou port in the past 20 years of historical data throughput, Sun et al. [6] employed the time series analysis and the Markov chain correction model to update the weights of the combination model. The successful application of combination forecasting in many fields proves that combination model is an effective method for information synthesis.

In this paper, we put forward a combination model on the principle of Minimum-variance to optimize the forecasting method, combined with the Pearl curve model, GM(1,1) and quadratic exponential smoothing model.

2 Brief Introductions of Models

2.1 Single Forecasting Models

Pearl Curve Model. A frequently used growth curve is the Pearl curve, named after Raymond Pearl, an American biologist, regarded as one of the founders of bio gerontology [7]. The Pearl curve model finds applications in a range of fields, including biology, economics, sociology, political science, and statistics. The equation for the Pearl curve is

$$y(t) = \frac{L}{1 + ae^{-bt}} \quad (1)$$

Where y is a non-linear function of t , ranges from zero to the upper limit L ; $a > 0$, $b > 0$. The solving methods of the parameters are described as follows: firstly, construct equal-time settlement time series data by Cubic Spline interpolation, divide the time series sample with a same interval into three sections, if the remainder is not zero then the sample formerly removes remainder terms: $1, 2, 3, \dots, r; r+1, r+2, r+3, \dots, 2r; 2r+1, 2r+2, \dots, 3r$. Define S_1 , S_2 and S_3 as the sum of reciprocals of data:

$$\begin{cases} S_1 = \sum_{t=1}^r \frac{1}{y(t)} \\ S_2 = \sum_{t=r+1}^{2r} \frac{1}{y(t)} \\ S_3 = \sum_{t=2r+1}^{3r} \frac{1}{y(t)} \end{cases} \quad (2)$$

Reformulate Pearl curve model as a reciprocal form with equation:

$$\frac{1}{y(t)} = \frac{1}{L} + \frac{ae^{-bt}}{L} \quad (3)$$

Then, we have the following expressions:

$$\begin{cases} S_1 = \sum_{t=1}^r \frac{1}{y(t)} = \frac{r}{L} + \frac{a}{L} \sum_{t=1}^r e^{-bt} = \frac{r}{L} + \frac{ae^{-b}(1-e^{-rb})}{L(1-e^{-b})} \\ S_2 = \sum_{t=r+1}^{2r} \frac{1}{y(t)} = \frac{r}{L} + \frac{ae^{-(r+1)b}(1-e^{-rb})}{L(1-e^{-b})} \\ S_3 = \sum_{t=2r+1}^{3r} \frac{1}{y(t)} = \frac{r}{L} + \frac{ae^{-(2r+1)b}(1-e^{-rb})}{L(1-e^{-b})} \end{cases} \quad (4)$$

So the calculation formula for each parameter is:

$$b = \frac{\ln \frac{(S_1 - S_2)}{(S_2 - S_3)}}{r} \quad (5)$$

$$L = \frac{r}{S_1 - \frac{(S_1 - S_2)^2}{(S_1 - S_2) - (S_2 - S_3)}} \quad (6)$$

$$a = \frac{(S_1 - S_2)^2 (1 - e^{-b}) L}{[(S_1 - S_2) - (S_2 - S_3)] e^{-b} (1 - e^{-rb})} \quad (7)$$

GM(1,1) Model. Grey system theory is a kind of analysis method proposed by the Chinese scholar Professor Deng Julong in 1982 [8], which is a forecasting method for the long term description of the development of things by using a small amount of incomplete information and the

establishment of the grey differential prediction model. The grey forecasting method has been widely used in many fields such as transportation, logistics, agriculture, commerce and so on. The analysis approaches of grey system model GM (1, 1) is used in this paper. Set the original nonnegative time series as $X^{(0)} = (X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n),)$, get accumulated generating operation sequence

$$X^{(1)}(i) = \sum_{k=1}^i X^{(0)}(k), k = 1, 2, \dots, n \quad (8)$$

The differential equation-formed mathematical model of for GM(1,1) is:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = b \quad (9)$$

Use the least square method to solve the parameters of the above whitening equation. There exists the following equation:

$$(a, b)^T = (B^T B)^{-1} B^T Y \quad (10)$$

where B is a matrix and Y is a vector:

$$B = \begin{bmatrix} -0.5(X^{(1)}(1) + X^{(1)}(2)) & 1 \\ -0.5(X^{(1)}(2) + X^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -0.5(X^{(1)}(n-1) + X^{(1)}(n)) & 1 \end{bmatrix} \quad Y = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ \vdots \\ X^{(0)}(n) \end{bmatrix}$$

Take a, b into Eq. 9, establish time response equation:

$$\hat{X}^{(1)}(k+1) = (X^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a} \quad (11)$$

Then take formula $\hat{X}^{(0)}(i) = \hat{X}^{(1)}(i) - \hat{X}^{(1)}(i-1)$ to calculate the direct prediction value.

Double Exponential Smoothing Model. Exponential smoothing method is a special case of the weighted moving average method, which predicts values with a weighted average of past time series data. When the changes of time series present a linear trend, we need to do double exponential smoothing on the basis of the first one so as to correct the lag error. The simplest form of exponential smoothing is given by the Eq. 12:

$$S_t^{(1)} = \alpha y_t + (1 - \alpha)S_{t-1}^{(1)} \quad (12)$$

Where α is the smoothing factor, and $0 < \alpha < 1$. In other words, the smoothed statistic $S_t^{(1)}$ is a simple weighted average of the current observation y_t and the previous smoothed statistic $S_{t-1}^{(1)}$. And the double exponential smoothing equations are:

$$S_t^{(2)} = \alpha S_t^{(1)} + (1 - \alpha)S_{t-1}^{(2)} \quad (13)$$

$$\hat{y}_{t+T} = a_t + b_t T \quad (14)$$

In Eq. 14, set t as current period; T as the interval period between current and forecasting period; \hat{y}_{t+T} as the fitted value at period $t+T$. We calculate the intercept a_t and the slope b_t as follows:

$$a_t = 2S_t^{(1)} - S_t^{(2)}, b_t = \frac{\alpha}{1-\alpha} (S_t^{(1)} - S_t^{(2)}) \quad (15)$$

2.2 Combination Forecasting Model.

The core idea of combination forecasting is according to the results of various forecasting methods to determine the weight for each method and thus to forecast the future data. The least square sum of squared error is defined as the objective function of the combined model to minimize the square sum of the prediction error. The optimal set of weighting coefficients is the solution of the following programming problem. Suppose k is the count of single forecasting models; n is the number of samples. Set f_{it} as the fitted value of method i at cycle t , w_i as the weight of method i , \hat{y}_t as the combination forecasting value. Obviously, we hope the difference between the predicted value and the actual value the smaller the better. The regression analysis method of weight coefficient is to use Q to represent the error sum of squares between the actual value and combination forecasting value and to minimize it.

$$\begin{cases} \min Q = \min \sum_{t=1}^n [y_t - \hat{y}_t]^2 = \min \sum_{t=1}^n [y_t - \sum_{i=1}^k w_i f_{it}]^2 \\ \text{s.t. } \sum_{i=1}^k w_i = 1, w_i \geq 0 \end{cases} \quad (16)$$

3 Experiment of Port Cargo Throughput

Result of Single Forecasting Model. In order to test and evaluate the effectiveness of the model, we selected 13 sets data from 2000 to 2012 which came from the website of the Dalian municipal bureau of statistic. According to the data, we made a case study of cargo port throughput with the models as shown in Table 1.

Table 1. Single Forecasting Model Formula

Pearl Curve	$y(t) = \frac{70306}{1 + 7.3630e^{-0.1748t}}$
GM(1,1)	$\hat{X}^{(1)}(k+1) = (X^{(0)}(1) + 80525.45)e^{0.1138k} - 80525.45$
Exponential Smoothing	$\hat{Y}_{2012+T} = 36957.97 + 2855.97T$

Result of Combination Forecasting Model. According to the calculation step of combining the model, set Pearl model, GM (1,1) model, double exponential smoothing model respectively y_1, y_2, y_3 get the weight through Excel then construct model as $y = 0.5767y_1 + 0.3956y_2 + 0.0277y_3$.

With throughput of 2013 as the validation data, the relative error statistics between the three models and the combination forecasting data and the actual data are obtained, as shown in Table 2.

Table 2. data and forecast results of various models

Year	Actual Value	Pearl Curve		GM(1,1)		Exponential Smoothing		Combination Model	
		Fitted Value	Relative Error	Fitted Value	Relative Error	Fitted Value	Relative Error	Fitted Value	Relative Error
2000	9699.0	8406.8	13.32%	9904.0	2.11%	2686.3	72.30%	8840.4	8.85%
2001	10519.0	9789.0	6.94%	10897.4	3.60%	5542.3	47.31%	10109.7	3.89%
2002	11188.0	11356.8	1.51%	12210.6	9.14%	8398.2	24.94%	11612.5	3.79%
2003	12601.7	13121.3	4.12%	13682.1	8.57%	11254.2	10.69%	13291.3	5.47%
2004	14516.2	15089.7	3.95%	15330.9	5.61%	14110.2	2.80%	15158.0	4.42%
2005	17085.2	17264.3	1.05%	17178.4	0.55%	16966.2	0.70%	17222.1	0.80%
2006	20046.0	19640.8	2.02%	19248.5	3.98%	19822.1	1.12%	19490.7	2.77%
2007	22286.0	22207.5	0.35%	21568.1	3.22%	22678.1	1.76%	21967.6	1.43%
2008	24588.4	24944.4	1.45%	24167.2	1.71%	25534.1	3.85%	24653.3	0.26%
2009	27202.8	27823.6	2.28%	27079.6	0.45%	28390.1	4.36%	27545.0	1.26%
2010	31399.3	30809.3	1.88%	30342.8	3.36%	31246.0	0.49%	30636.9	2.43%
2011	33690.8	33860.1	0.50%	33999.4	0.92%	34102.0	1.22%	33921.9	0.69%
2012	37426.2	36930.5	1.32%	38096.6	1.79%	36958.0	1.25%	37392.5	0.09%
2013	40746.2	39973.91	1.90%	42687.47	4.76%	39813.94	2.29%	41042.87	0.73%

Result Analysis. In order to make a comparative analysis of the prediction accuracy and the validity of the model, this paper uses the following evaluation indicators to evaluate the effectiveness of the prediction.

(1) SSE(Sum Square Error)

$$SSE = \sum_{t=1}^n (y_t - \hat{y}_t)^2 \tag{17}$$

(2) MAE(Mean Absolute Error)

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \tag{18}$$

(3) MAPE(Mean Absolute Percentage Error)

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{19}$$

Among them, y_t is the actual value, \hat{y}_t is the forecasting value, and n is the time length of the performance test of the prediction model. Based on the three prediction performance indexes mentioned above, the prediction accuracy of the three single prediction models and the combination model is shown in Table 3.

Depending on the Table 3, the smaller the model prediction error is, the better the performance is. From Table 3, we can note that the three indicators of the combination model keep in a lower and smoother level than single forecasting models. The average relative error is 2.78%. It shows that the combination model which integrates multiple approaches has higher accuracy and this model can be applied to simulate the actual port cargo throughput.

Table 3. Comparison of prediction accuracy of the model

models	SSE	MAE	MAPE
Pearl Curve	4167212.92	475.29	2.94%
GM(1,1)	6075355.72	15174.10	3.46%
Exponential Smoothing	86642905.87	1571.79	13.29%
Combination Model	271841.77	420.67	2.78%

4 Conclusion

In this paper, we put forward a combination forecasting method for port cargo throughput based on Pearl Curve Model, GM(1,1) and Exponential Smoothing, and illustrate that the combination forecast model performs better than single forecasting model by examples. It is verified that that the given combination model has better fitting and forecast precision. Therefore it has a widely applicable prospect in port management in the time-series data modeling.

At the same time, the shipping industry has been undergoing continuous changes which bring a lot of uncertain factors. The model proposed in this article may not be the most suitable because the result is based on the historical data, and the future trend of development is not always according to the historical data. Additionally, the selection of individual models in the combination has a certain influence on the modeling of the combined model. As a consequence, in order to improve the accuracy of prediction, the selection of a single model is needed to be optimized for different case applications.

Acknowledgment

This work is partially supported by National Key Technology Research and Development Program of the Ministry of Science and Technology of China (No. 2014BAH24F04), National Natural Science Foundation of China (No.71271034), the National Social Science Foundation of China (Grant No.15CGL031).

References

- [1] Kuang Haibo, "Research on the prediction model of Chinese coastal port throughput," *Science Research Management*, vol.30 NO.3, pp. 187-192, May 2009.
- [2] BROYLES J R, COCHRAN J K, DOUGLAS C M, "A statistical Markov chain approximation of transient hospital inpatient inventory," *European Journal of Operational Research*, 2010.
- [3] Xu Changxin, Yan Yixin, Zhang Ping, "Prediction Model of Port Throughput Based on System Dynamics," *Port & Waterway Engineering*, NO.5, pp.27-40, May 2006.
- [4] Liu Zhijie, Ji lin, Ye Yuling, et al. "Combined Forecast Method of Port Container Throughput Based on RBF Neural Network," *Journal of Tongji University (Natural Science)*, vol.35 NO.5, pp.739-744, May.2007.
- [5] Shi Zejun, Li Kai, "Container Throughput Forecasting Based on Gray Method and Exponential Smoothing Method," *Journal of Chongqing Jiaotong University (Natural Science)*, vol.27 NO.2, pp.302-304, Apr.2008.
- [6] Sun Zhilin, Lu Yaqian, Huang Saihua, "Prediction of port throughput based on Markov chain time series analysis," *Journal of Zhejiang University(Engineering Science)*, vol.46 NO.7, pp.1289-1294, Jul. 2012.

[7] Internet users, “Raymond Pearl, ” Wikipedia, the free encyclopedia,

[8] http://en.wikipedia.org/wiki/Raymond_Pearl.

[9] Ma Kai, “Grey Model in Power Load Forecasting with the improved method,”
TRANSPORT(second half),2008.05.