Research on Logistics Demand Forecast of Port Based on Combined Model

Peng Yan¹, Lin Zhang¹*,², Zhiyun Feng¹,², Jing Zhang¹,²

¹College of Civil and Architectural Engineering, North China University of Science and Technology, Tangshan, Hebei 063210, China
²Hebei Province Earthquake Engineering Research Center, Tangshan, Hebei 063000, China

*Corresponding author’s e-mail: 3259415398@qq.com

Abstract. In order to improve the accuracy, practicability and prediction accuracy of the single prediction model, a logistics demand combination forecasting model based on cargo throughput is established. Based on the original statistical data of throughput, the two models of gray and exponential smoothing are established respectively. On this basis, the weighted assignment method of variance reciprocal is used to construct the combined forecasting model. According to the prediction results of different prediction models analyzed by the three evaluation indexes of average relative error, maximum fitting error and minimum fitting error, the three evaluation indexes of the combined prediction model are respectively 6.344%, 16.345% and 0.343%, which are smaller than the single item model. It indicates that the established combined forecasting model can effectively improve the accuracy of the throughput prediction model based on overcoming the shortcomings of the single-term throughput prediction model.

1. Introduction

In recent years, in the process of China's economy showing medium-high speed development, the modern logistics industry is also in the period of rapid development of gold. The research on logistics demand is more and more valued by most scholars, how to reasonably predict the logistics demand becomes one of the core issues for the research of logistics demand. Many domestic experts and scholars have applied different research methods from different angles to explore and study this issue. Shi H Y, Lu X[1] and Wang H F[2] analyzed the intrinsic relationship between the logistics demand of Yingkou City and Hangzhou City and the different factors such as economy and industrial structure, and constructed a prediction model of logistics demand in the two cities based on neural network. According to the fitting results of the prediction model, the practical significance of the prediction method was expounded. Based on GDP and the correlation between different indicators such as industrial added value and the turnover of goods, Liang Y M[3] established a forecast model of logistics demand regression analysis using cargo turnover as a measure. On the basis of fully analyzing the advantages and disadvantages of different forecasting methods, Yang Z, Cao Z Q, Li J H [4] established a forecasting model for agricultural products cold chain logistics demand based on gray GM(1,1) model, in order to provide a certain theoretical support and guidance for the good development of the cold chain logistics of agricultural products. Different from the above research methods, Cheng Z X[5] proposed a SPA clustering logistics demand forecasting method. The method...
used the industrial added value speed of China's first, second and third industries as the basic indicators to predict the growth rate of China's logistics demand, and then validated the model according to the fitting effect.

The logistics demand forecasting method proposed in this paper is different from the research methods adopted by other relevant research scholars. It takes port cargo throughput as an important indicator of logistics demand forecasting, and uses the combined model established by gray model and exponential smoothing model to make reasonable scientific prediction. In addition, the average relative error, the maximum fitting error and the minimum fitting error are used to measure the fitting of different prediction models to analyze and verify whether the established prediction model’s practicability and prediction accuracy meet the requirements of throughput prediction, then select the optimal throughput prediction model.

2. Principle of single prediction model

2.1 Gray model

Set $X^{(0)}_1 = (x^{(0)}_1, x^{(0)}_2, \ldots, x^{(0)}_n)$ to the original data sequence, $X^{(1)}_1$ is a first order accumulation sequence of $X^{(0)}_1$, $Z^{(1)}_1$ is $X^{(1)}_1$’s immediate sequence of mean generation, then

$$X^{(0)}_1 (k) + a x^{(0)}_k (k) = b$$

The above formula is GM(1,1) gray differential equation, $z^{(1)}_1 (k) = 0.5 \cdot x^{(1)}_1 (k) + x^{(1)}_1 (k-1)$, $k = 1, 2, 3, \ldots, n$.

$-a$ is the system development coefficient and $b$ is the drive coefficient.

Hypothesis $\xi = [a, b]^T$ is a parameter column, and $Y = \left[ x^{(0)}_1 (2), x^{(0)}_1 (3), x^{(0)}_1 (4), \ldots, x^{(0)}_1 (n) \right]^T$.

$$B = \begin{bmatrix} -z^{(1)}_1 (2) & 1 \\ -z^{(1)}_1 (3) & 1 \\ \vdots & \vdots \\ -z^{(1)}_1 (n) & 1 \end{bmatrix}$$

The parameter column $\xi = [a, b]^T$ calculates $\hat{\xi} = B^T B^{-1} Y$ with a least squares estimate. Then the approximate time response of the GM(1,1) model is

$$x^{(1)} (k+1) = x^{(0)} (1) \cdot e^{-ak} + \frac{b}{a}, \quad k = 1, 2, 3, \ldots, n$$

2.2 Exponential smoothing model

The quadratic exponential smoothing method not only has the advantages of simple calculation operation, easy understanding, and less original data required, but also eliminates other contingency factors while improving the accuracy of model prediction.

Assuming the original sequence is $Y_1, Y_2, Y_3, \ldots, Y_n$, then the quadratic exponential smoothing model is

$$Y_{t+1} = a Y_t + b_t * T$$

$$S^{(2)}_t = a S^{(1)}_t + (1-a) S^{(2)}_{t-1}$$

$$a_t = 2 S^{(1)}_t - S^{(2)}_t$$

$$b_t = \frac{a}{1-a} (S^{(1)}_t - S^{(2)}_t)$$

In the formula: $S^{(2)}_t$ — Quadratic exponential smoothing value of the t-th cycle;
The initial value of the quadratic smoothing index prediction model is determined. If the number of data items in the actual prediction sequence is more than 15 items, the observation value of the first stage of the prediction data or the previous stage is usually used as the initial value. If the number of actual predicted sequence items is less than 15, the average of the first three data is generally used as the initial value of the predicted data. The formula is

\[ S^{(1)}_t = S^{(2)}_{t-1} = \frac{Y_1 + Y_2 + Y_3}{3} \]  

(5)

The determination of the smoothing coefficient \( \alpha \) should be based on the stationarity of the original data sequence. If the sequence changes relatively stable, the smoothing coefficient \( \alpha \) is chosen to be 0.1~0.3; if the sequence changes greatly, the smoothing coefficient \( \alpha \) should be selected as 0.7~0.9. Later, according to the principle of minimum RMS (mean square error) index and high prediction accuracy to select the optimal smoothing coefficient \( \alpha \).

3. Case Analysis

The cargo throughput of Port from 2008 to 2016 is the original data (the data in Tab. 1 comes from the Statistical Yearbook), and the cargo throughput of the port is predicted by the gray GM (1,1) model and the exponential smoothing model. In addition, the weights of the individual models are solved by the variance reciprocal method and a combined prediction model is established. The validity and practicability of the combined model are verified by three evaluation indexes: average relative error, maximum fitting error and minimum fitting error.

| Table 1. The original data of cargo throughput |
|------------------------------------------------|
| Time  | 2008  | 2009  | 2010  | 2011  | 2012  | 2013  | 2014  | 2015  | 2016  |
| Throughput/Ten thousand tons | 10854 | 17559 | 24608 | 31263 | 36504 | 44620 | 50064 | 49285 | 52051 |

3.1 Gray model prediction

According to Tab. 1, the original data column is

\[ X^{(0)}_1 = (10854, 17559, 24608, 31263, 36504, 44620, 50064, 49285, 52051) \]  

(6)

Based on 2008 throughput data, according to the model solving step, you can get

\[ a = -0.1241, \ b = 21196 \]

Get the gray model of Tangshan Port cargo throughput as

\[ x^1(k+1) = 181651.74 e^{0.1241k} - 170797.74 \]  

(7)

By entering different values, you can solve the throughput prediction values for different years.

3.2 Exponential smoothing model prediction

According to the statistical software, we can get the curve estimation of Port throughput original data. It can be seen that the correlation coefficient \( R^2 \) of the original data change of the throughput which in line with the linear variation trend is 0.96 and the fitting degree of the exponential change trend is 0.88. Therefore, we should establish the quadratic exponential smoothing model to predict throughput.
According to the data in Fig. 1, the change of the throughput data is more obvious. First, the smoothing coefficient should be selected between the intervals 0.7 and 0.9. Secondly, the optimal smoothing coefficient is selected according to the magnitude of the mean square error for throughput prediction. The actual value of the throughput and the fitted value of the different smoothing coefficients are shown in Tab. 3. According to the principle of minimum mean square error, according to Table 3, the optimal smoothing coefficient is $\alpha = 0.9$. Therefore, the throughput quadratic exponential smoothing prediction model is established as

$$Y_{t,T} = 52028.69 + 2275.33 \times T$$  \hspace{1cm} (8)

Table 2. Comparison of actual and predicted throughput values of Tangshan Port with different smoothing coefficients

| Year | Actual throughput value (Ten thousand tons) | $\alpha = 0.7$ | $\alpha = 0.8$ | $\alpha = 0.9$ |
|------|---------------------------------|----------------|----------------|----------------|
| 2008 | 10854                           | 17674          | 17647          | 17647          |
| 2009 | 17559                           | 17513          | 17490          | 17467          |
| 2010 | 24608                           | 27390          | 28805          | 30228          |
| 2011 | 31263                           | 36233          | 37220          | 37782          |
| 2012 | 36504                           | 41931          | 42130          | 42011          |
| 2013 | 44620                           | 51147          | 51711          | 52201          |
| 2014 | 50064                           | 56400          | 56266          | 55962          |
| 2015 | 49285                           | 52677          | 51233          | 49820          |
| 2016 | 52051                           |                |                |                |

Mean square error

$$1396.7 \quad 1343.7 \quad 1338.4$$

3.3 Combined model prediction

The key to combining different models into one model is how to determine the weight coefficients of each model. There are many methods to determine the weight coefficients. The optimal weights of each model are solved from the perspectives of linear programming, nonlinear programming and positive definite matrix. Different from other methods, in order to improve the prediction accuracy of the combined model and reflect the good stability of the combined model, this paper starts from two different angles of variance and sum of squared errors, and determines the weight coefficient of each model according to the variance reciprocal method.

If the actual value sequence of an indicator is $X_1, X_2, X_3, \ldots, X_{n-1}, X_n$, the optimal weight coefficient matrix corresponding to the K prediction methods of this index is
The K prediction methods have a fitted value of \( y_{i1}, y_{i2}, y_{i3}, \ldots, y_{in} \), \( i = 1,2,3,\ldots,n \), and the weight of each model is \( \omega_i = M_i / A \) (\( A = \sum_{i=1}^{k} M_i, M_i = \sum_{j=1}^{n} (x_j - y_{n})^2 \)). Then the predicted value of the indicator according to the combined model is

\[
y_t = \sum_{i=1}^{k} \omega_i y_{ti}
\]

According to the above steps, by analyzing the throughput data of Port can calculate that the weights of the gray model and the quadratic exponential smoothing model are \( \omega_1 = 0.32 \) and \( \omega_2 = 0.68 \). By comparing the actual and fitting values of throughput of the gray model, the exponential smoothing model, and the combined model, we can get Tab. 3.

### Table 3. Comparison of different model throughput prediction results

| Time  | Actual throughput value (Ten thousand tons) | combined model fitting value | Relative error | Gray model fitting value | Relative error | Exponential smoothing model fitting value | Relative error |
|-------|---------------------------------------------|------------------------------|----------------|-------------------------|----------------|-------------------------------------------|---------------|
| 2009  | 17559                                      | 19691                        | 12.14%         | 24034                   | 36.876%        | 17647                                     | 0.501%        |
| 2010  | 24608                                      | 20586                        | 16.35%         | 27213                   | 10.586%        | 17467                                     | 29.019%       |
| 2011  | 31263                                      | 30415                        | 2.712%         | 30813                   | 1.439%         | 30228                                     | 3.311%        |
| 2012  | 36504                                      | 36857                        | 0.966%         | 34890                   | 4.421%         | 37782                                     | 3.501%        |
| 2013  | 44620                                      | 41209                        | 7.644%         | 39505                   | 11.463%        | 42011                                     | 5.847%        |
| 2014  | 50064                                      | 49811                        | 0.506%         | 44732                   | 10.650%        | 52201                                     | 4.268%        |
| 2015  | 49285                                      | 54262                        | 10.10%         | 50650                   | 2.769%         | 55962                                     | 13.547%       |
| 2016  | 52051                                      | 52229                        | 0.343%         | 57350                   | 10.181%        | 49820                                     | 4.268%        |

### 3.4 Evaluation of different model prediction results

In order to reasonably evaluate the fitting effect of the prediction model, this paper adopts three indicators: average relative error, maximum fitting error and minimum fitting error to evaluate the model prediction results.

Suppose \( e_i = |y_i - y^*| \) is the i-th group absolute error and \( y^* \) is the i-th group prediction value of each model. Therefore, the evaluation index expressions are:

- **average relative error**:
  \[
  \bar{\xi} = \frac{1}{m} \sum_{i=1}^{m} \frac{e_i}{y_i} = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{y_i - y^*}{y_i} \right|
  \]

- **maximum fitting error**:
  \[
  \xi_{\text{min}} = \min \left\{ \frac{1}{m} \sum_{i=1}^{m} \frac{y_i - y^*}{y_i} \right\}
  \]

- **minimum fitting error**:
  \[
  \xi_{\text{max}} = \max \left\{ \frac{1}{m} \sum_{i=1}^{m} \frac{y_i - y^*}{y_i} \right\}
  \]

Comparing the average relative error, the maximum fitting error and the minimum fitting error of different prediction models, the errors of the gray model and the exponential smoothing model are higher than that of the combined model, so they are not suitable for predicting the future throughput. It can be seen that the optimal prediction model is a combined model method.

### Table 4. Comparison of different model evaluation indicators

| Model                          | Combined model | Gray model | Exponential smoothing model |
|--------------------------------|----------------|------------|-----------------------------|

5


|                          | Average relative error | Maximum fitting error | Minimum fitting error |
|--------------------------|------------------------|-----------------------|----------------------|
|                          | 6.344%                 | 11.048%               | 8.035%               |
|                          | 16.345%                | 36.876%               | 29.019%              |
|                          | 0.343%                 | 1.439%                | 0.501%               |

4. Conclusion

(1) The combined prediction model to a certain extent not only reduces the randomness of the original data, but also effectively improves the accuracy, prediction accuracy and practicability of the prediction model.

(2) According to the three evaluation indexes of average relative error, maximum fitting error and minimum fitting error, the prediction results of different model prediction models are analyzed. The three evaluation indicators of the model are 6.344%, 16.345% and 0.343%, respectively. This shows that compared with the single prediction model, the predicted value of the combined prediction model is the closest to the actual value. On the basis of overcoming the shortage of the single throughput prediction model, the accuracy and prediction accuracy of the prediction model are further improved.

Acknowledgments

This work was supported by National Natural Science Foundation of China(51378171), and also supported by Foundation Sciences North China University of Science and Technology(Z201423).

References

[1] Shi H Y, Lu X.(2017)Forecast of logistics demand in yingkou city based on neural network. J.Journal of Chifeng University(Natural Science Edition,33:171-173.

[2] Wang H F.(2017)Forecast of logistics demand in hangzhou city circle based on BP neural network.J.Modern Business Trade Industry,29:46-47.

[3] Liang Y M.(2015) Logistics demand forecast based on regression analysis method-taking anhui province as an example.J.Journal of Guangxi Normal University for Nationalities,32:65-69.

[4] Yang Z, Cao Z Q, Li J H.(2017) The demand forecast of agricultural products cold chain logistics in guangxi based on grey forecasting method.J.Logistics Engineering and Management,39:86.

[5] Cheng Z X.(2014)Forecast and analysis of logistics demand scale in China based on SPA clustering algorithm.J.Market Modernization,27:50-52.

[6] Zhang J G, Ye X Y, Gong Y.(2017) BForecast research on trend of rural e-commerce industry clusterbased on grey neural network combination model.J.Journal of Commercial Economics,1:192-194.

[7] Wang N.(2017)OWHA-based logistics demand combination improvement and forecasting model construction.J.Journal of Commercial Economics,1:192-194.

[8] Guan Z P, Zou W Y, Wu Y N.(2010) Forecast analysis of tax revenue in China based on combination model.J.Journal of Commercial Economics,5:82-83.

[9] Jin X, Qian Z.(2009)WConstruction and empirical analysis of sales portfolio prediction model.J.Journal of Commercial Economics,25:29-31.