

Logistic Forecast Analysis of Sichuan Province on the Basis of Multi-model Combination

Jiaying Zhao*

Department of Pittsburgh, Sichuan University, Chengdu, China

*Corresponding author: 2019141520154@stu.scu.edu.cn

Abstract. Sichuan province serving as the hub of comprehensive transportation in the southwest of our country, has the longest transportation lines in China and its freight transportation maintains a healthy development. In order to give a quantitative reference for the government to formulate logistics improvement and development policies, determining the scale of logistics infrastructure construction and analyzing the situation of the logistics market, this paper establishes several forecasting models to reveal the mechanism and forecast the logistics demand. Freight turnover is used as an indicator to reveal logistics demand and seven variables are summarized as influential factors. Then, multiple regression model, BP neural network model and grey forecasting model are available ways to do the forecast and their forecasting error is measured by several criteria. In order to avoid the technical shortage of a single model, this paper give each of the model a weight ratio and use a combination model to analyze the logistics demand of Sichuan Province. Since the multi-model has low overall error rate, using this combination model to predict the logistics needs of Sichuan in the coming years is pretty feasible.

Keywords: Logistics forecast; freight turnover; Sichuan province; combined prediction model.

1. Introduction

1.1 Background

The logistics industry undertaking the important task of communicating the upstream and downstream of the industry, can be called as the "blood pipeline" of national social and economic development, and it plays an indispensable part in optimizing the industrial structure. To some extent, the potential logistics demand determines how much room is left for the blossom of the logistics industry. Therefore, it is necessary to establish a scientific and reasonable forecasting model to reveal the relevant mechanism and forecast the logistics demand. This will provide a quantitative reference for formulating logistics development policies, determining the scale of logistics infrastructure construction and analyzing the situation of the logistics market. Besides, it is also very essential for the sustainable and continuable development of the logistics industry. Freight volume and freight turnover are important indicators of logistics demand. Thus, this paper uses these two indicators to give a quantitative forecast of the future logistics market demand.

1.2 Literature Review

When considering the logistics forecast of Sichuan Province, Xiaoping Lin and Jie Yuan established a prediction model of cargo and postal throughput in Chengdu Shuang Liu Airport based on the grey system theory [1]. Suqin Dong and Xiaoyan Li speculated on the development of fresh produce logistics under the national live-carry model [2]. Since Xioaping Lin and Jie Yuan did research of Chengdu shuangliu airport, Suqin Dong and Xiaoyan Li focus on fresh produce logistics only, there are no overall logistic forecast of Sichuan Province. Besides, Yang Wang used GM (1,1), cubic exponential smoothing method and multiple linear regression to forecast the highway passenger volume in Sichuan Province [3]. Fangming Tian and Qingfu Liu forecast the railway freight volume and its development trend in Sichuan province for the next 6 years [4]. There are already forecasts of highway passenger volume and railway freight volume of Sichuan province, but there is no research about the overall freight volume, including railway, highway, airway, waterway, of Sichuan province.

Nowadays, there are more and more research on logistics forecast. The most common methods are multiple linear regression, BP neural network and grey prediction method. Jiong Liu did the forecast of AnHui Province by multiple linear regression [5]. Wang Yingchun and Liu Wenjuan used grey forecasting method to make a reasonable inference of the development trend of logistics in Fujian Province according to social and economic statistical data of Fujian Yearbook and the development trend of freight volume. [6]. Under the background of rural e-commerce development, Zeng Minling, Liu Runmin, Gao Ming and Jiang Yizhang forecast the rural logistics quantity required of Guangdong Province using the GM (1,1) model. [7]. In terms of neural network, Yuxin Cui did a study of Logistics quantity required prediction of Hebei Province based on Lasso-BP Neural Network [8]. Xuexue Gao established a BP neural network model and selected the corresponding economic indicators to do the logistics forecast of Hainan Province in the next few years [9]. There has been plenty logistic forecast based on a single prediction method.

1.3 Objection and Motivation

Sichuan province serves as the hub of comprehensive transportation in the southwest of our country, and the freight transportation has maintained a healthy development. The total length of transportation lines of Sichuan has reached 410564 km, ranking the first in China. As important indicators of the transportation industry, freight volume and freight turnover play an important part in the economic development. There has been plenty of research about freight volume and freight turnover of AnHui, FuJian, HeBei etc, but there is few research related with that of Sichuan. The only research related with Sichuan are about fresh food logistics, highway passenger volume and railway freight volume. So, it is necessary to supplement the forecast of freight volume and freight turnover in Sichuan Province.

Besides, most logistic forecasts are based on single prediction method. Since every individual model has its inevitable shortcomings, it's better to use a combination model to improve the overall prediction accuracy. Therefore, this paper first uses grey prediction model, BP neural network model as well as multiple regression model to do the forecast individually. Then determines the corresponding weights of these three models on the basis of their forecasting accuracy and combines them into a new model to make assumption of the logistics trend of Sichuan in the coming years.

2. Data and variables

2.1 Data sources

In order to ensure that all indicators are available and authoritative, this research collects all data from Sichuan Statistical Yearbook [10] published by the Sichuan Provincial Bureau of Statistics from 2011-2021.

2.2 Data description and collection

Industrial structure, population and environment, consumption level and transportation development have essential impact on the logistics demand of SiChuan, so this paper collect some corresponding factors as independent variables, which can be found in Table 1.

Table 1. Independent variables

Industrial structure category	Added value of the primary, secondary and tertiary industries, GDP
Population and environment	Residents' population at year end, The total retail sales of consumer goods
Consumption level	Average disposable income, Urban residents average annual consumption expenditure
Transportation development	Railway mileage, Highway mileage

Volume of freight traffic (the total number of goods transported by transport enterprises within a certain period, the unit is ton) and rotation volume of freight transport (the actual total tonnage of goods transported by the transport enterprise multiply the product of the average transport distance within a certain period, the unit is tonnage kilometers) are chosen as dependent variables to represent the logistics demand of Sichuan.

2.3 Forecasting Methods

2.3.1. Multiple Linear Regression

Multiple linear regression is one of the easiest and most frequently used statistical ways to do the forecast. In multiple linear regression (MLR), a few explanatory variables also called as independent variables are used to forecast the result of a response variable also called as dependent variable. In other terms, MLR examines the linear relationship between multiple independent variables and one dependent variable [11]. The formula of MLR is (1):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \epsilon \dots\dots\dots(1)$$

Where, i is the number of observations, y_i stand for dependent variable, x_i represent explanatory variables, β_p is the slope coefficients for each independent variable, β_0 is y-intercept (constant term) and ϵ is model's error term or residuals.

F test, testing for overall significance, can be used to examine if there exist a huge relationship between the response variable and all explanatory variables. t test, however, is used to examine the significance of every single variable. R-square is a goodness-of-fit test, which measures to what percentage of the variance in the dependent variable can independent variables explain collectively. R-squared use a 0 – 100% scale to evaluate the strongness of relationship between the MLR model and the response variable [12].

2.3.2. BP neural network

Artificial Neural network (ANN) is a new type of information processing and computer system which abstracts, simplifies, and simulates biological structure on the basis of modern neuroscience research. The main purpose of artificial neural network is to simulate and realize the autonomous learning and thinking ability of people, and to dig out the internal connections using limited sample. [13].

BP neural network is currently the most widely used neural network model in practice. Back propagation is the abbreviation of "error back propagation". In this method, the weight of neural network can be fine-tuned according to the error rate of the last period. By improving the generalization ability of the model, properly adjusting the weight can help reduce error rate therefore make the model more convincible. [14].

BP neural network has many layers including input layer, hidden layer and output layer. According to Kolmogorov theorem, there exists a relationship among the number of neurons in these three layers [15]:

$l = \sqrt{n + m} + \alpha$. where l represents the number of neurons in the hidden layer, n and m represent the number of input and output neurons, and α is a constant whose value is from 1 to 10.

2.3.3. Grey prediction models

GM (1,1) model, a special case of grey prediction model, is suitable for the prediction of small sample data. It can take full advantage of the insufficient data and give high-precision prediction results.

2.4 Evaluation metrics

The difference between the actual value and the model's estimate result is called as prediction error or residual. MSE, RMSE, MAE, MAPE are four ways to measure the residual, as shown in Table 2. For each measurement, the magnitude of the metric can present whether these models are performing well. Small residuals point to good predictive ability, while large ones suggest otherwise.

Table 2. Four evaluation metrics

Abbreviation	Name	Formula
MSE	Mean Square Error	$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$
RMSE	Root Mean Square Error	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$
MAE	Mean Absolute Error	$MAE = \frac{1}{n} \sum_{i=1}^n \hat{y}_i - y_i $
MAPE	Mean Absolute Percentage Error	$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left \frac{\hat{y}_i - y_i}{y_i} \right $

3. Results

3.1 Deletion and selection of dependent variable and independent variables

3.1.1. Initial Variables

Initially, there are two dependent variables and ten independent variables.

Dependent variables: Volume of freight traffic(ten thousand tons), Rotation volume of freight transport(Million tons*kilometer).

Independent Variables: Value-added of primary industry(billion yuan), Value added of the secondary industry(billion yuan), Value added of the tertiary industry (billion yuan), GDP(billion yuan), Total retail sales of social consumer goods (\$ 100 million), Resident population at year end(ten thousand people), Average disposable income(yuan per person), Urban residents average annual consumption expenditure (yuan per person), Railway mileage (ten thousand kilometer), Highway mileage(ten thousand kilometer).

3.1.2. Correlation Analysis and variables selection

Correlation analysis is done between variables (Fig.1-3). In the actual modeling process, whether it is multiple regression model, neural network, or grey prediction model, only when high correlation occurs between dependent variables and independent variables, using model analysis to find the relevant specific form is meaningful. Therefore, this paper selects freight turnover as the measurement index of logistics demand scale in Sichuan Province.

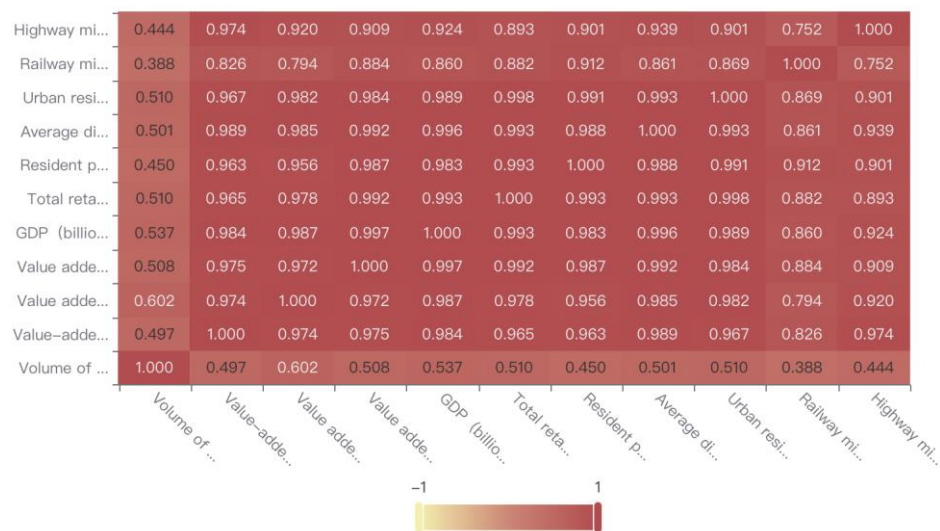


Fig.1 Correlation between volume of freight traffic and all independent variables
(Photo credit: Original)

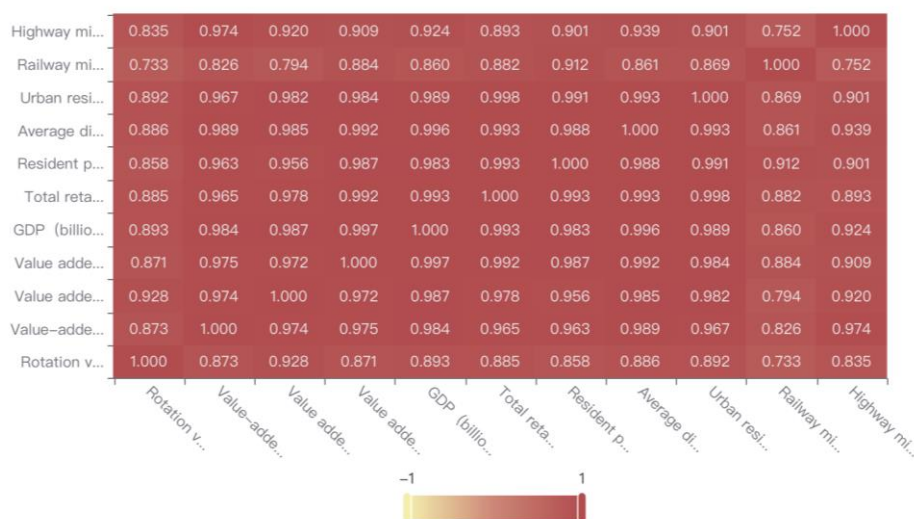


Fig.2 Correlation between Rotation volume of freight transport and all independent variables
(Photo credit: Original)

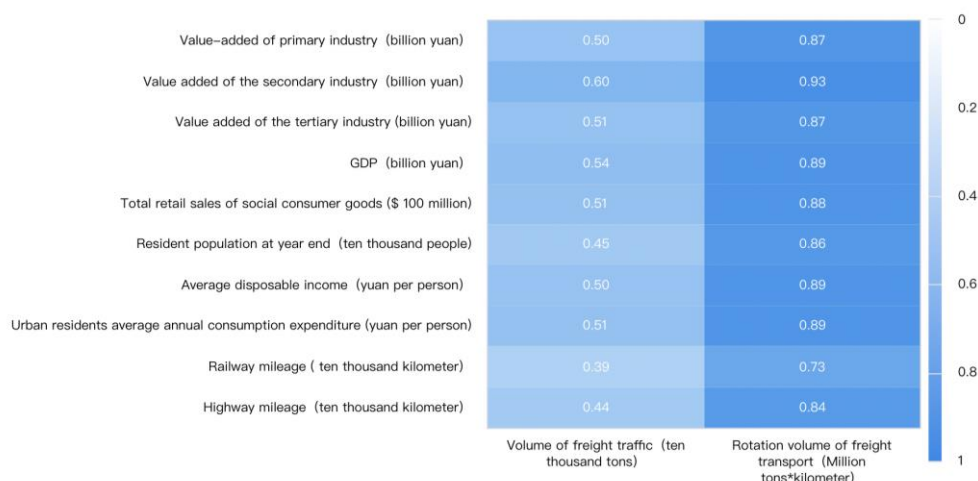


Fig.3 Correlation between all dependent variables and independent variables
(Photo credit: Original)

Then is to choose independent variables. First, this paper selects the value-added data of primary, secondary and tertiary industries. Since GDP is the synthesis of the value added of primary, secondary and tertiary industries, it is redundant to exist simultaneously with these variables. In addition, the correlation between GDP and total retail sales of consumer goods is very large, at 0.993, so I left total retail sales of consumer goods and deleted GDP. Moreover, the population size is highly correlated with the total retail sales of consumer goods, and the correlation degree is 0.993. Therefore, the population size is also abandoned in this paper.

The average per capita annual consumption expenditure of urban residents is highly correlated with the average disposable income of residents, which is 0.993. In this paper, the average disposable income of residents is left, and the average per capita annual consumption expenditure of urban residents is deleted. In addition, this paper chooses highway mileage and railway mileage as independent variables.

To sum up, considering economic indicators and infrastructure indicators, this paper finally selects Value-added of primary, secondary and tertiary industry (billion yuan), Total retail sales of social consumer goods (\$ 100 million), Average disposable income (yuan per person), Railway mileage (ten thousand kilometer), Highway mileage (ten thousand kilometer) as independent variables.

3.2 Linear Regression

3.2.1. Multiple Linear Regression

In this paper, SPSS Pro software is used for regression analysis and the main parameters of the regression model are shown in Table 3.

Table 3. Linear regression analysis result table

Multiple Linear Regression n=11								
	Unstandardized coefficient		Standardized coefficient	t	p	VIF	R ²	Adjusted R ²
	B	SE	Beta					
Constant	-5297.832	2178.881	-	-2.431	0.093*	-		
Value-added of primary industry(billion yuan)	-0.699	0.705	-1.941	-0.992	0.394	388.855		
Value added of the secondary industry(billion yuan)	0.314	0.104	2.748	3.018	0.057*	84.2		
Total retail sales of social consumer goods (\$ 100 million)	0.373	0.314	5.718	1.188	0.320	2351.524	0.97	0.902
Value added of the tertiary industry (billion yuan)	0.057	0.061	1.151	0.945	0.414	150.518		
Average disposable income(yuan per person)	-0.58	0.376	-9.785	-1.543	0.220	4082.018		
Railway mileage (ten thousand kilometer)	1230.871	2689.858	0.194	0.458	0.678	18.306		
Highway mileage(ten thousand kilometer)	302.57	148.947	3.085	2.031	0.135	234.276		
Dependent Variable: Rotation volume of freight transport(Million tons*kilometer)								

From the results of the above regression parameters, it can be concluded that the regression equation is : $y = -5297.832 + (-0.699) \times \text{Value-added of primary industry(billion yuan)} + 0.314 \times \text{Value added of the secondary industry(billion yuan)} + 0.373 \times \text{Total retail sales of social consumer goods}$

(\$ 100 million)+0.057*Value added of the tertiary industry (billion yuan)+(-0.58)*Average disposable income(yuan per person)+1230.871*Railway mileage (ten thousand kilometer)+302.57*Highway mileage(ten thousand kilometer).

3.2.2. Model test

3.2.2.1. The F test of the regression equation

The results of the F test shows that P value is $0.026 \sim * < 0.1$, which is significant at level $\alpha = 0.1$. So, the original hypothesis that the regression coefficient is 0 should be rejected. Then reach the conclusion that the model satisfies variable collinearity, and the model shows a significant linear relationship.

3.2.2.2. The goodness-of-fit test

The above result indicates that $R^2 = 0.97$, adjusted $R^2 = 0.902$. Both of the values are close to 1. Therefore, the regression line fits well with this linear model, as shown in Fig.4.

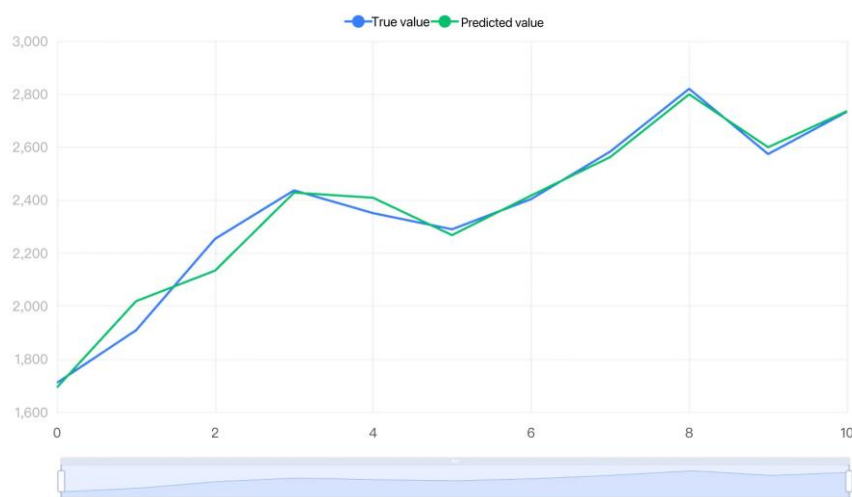


Fig.4 The fitting effect of multiple linear regression
(Photo credit: Original)

3.2.2.3. t test of the regression coefficient

Given the significance level $\alpha = 0.1$, by looking up the t distribution table, it is found that the critical value of $n-k-1$ is $t_{\alpha/2}(n-k-1) = 2.353$. According to the t values of all the variables, it is shown that only the t value of the added value of the secondary industry is larger than the critical value of t, $3.018 > 2.353$, and its p value is $0.057 < 0.1$. Therefore, the added value of the secondary industry has a significant impact on the model. Since the t values of other variables are smaller than the critical value, their impact on the model is not significant.

3.2.2.4. Test for collinearity of variables

VIF is a value that represent the severity of multicollinearity. It is used to test whether the model is collinear or if a high correlation exists between several independent variables. A model without multicollinearity should have VIF value less than 10 or 5. If “inf” appears in VIF, the VIF value is infinite.

The result of SPSS Pro shows that the VIF value of the added value of the primary , second and tertiary industry (100 million yuan), the average disposable income of residents (yuan / person), the total retail sales of consumer goods, railway mileage (10,000 km), road mileage (10,000 km) are all greater than 10. So, there exists a strong collinear relationship between these variables and a stepwise regression should be used to remove variables' collinearity.

3.2.3. Stepwise Regression

A stepwise regression is down using SPSS. The added value of the primary industry, secondary industry and tertiary industry (100 million yuan), the total retail sales of consumer goods, the average disposable income of residents (yuan/person), the length of railways (ten thousand kilometers), and the length of highways (ten thousand kilometers) were taken as independent variables, freight turnover (100 million tons · km) is taken as dependent variable. It is shown in Table 4 that after stepwise regression, the only independent variable left in the model is value added of the secondary industry.

Table 4. Stepwise regression result table

Stepwise Regression n=11								
	Unstandardized coefficient		Standardized coefficient	t	p	VIF	R ²	Adjusted R ²
	B	SE	Beta					
Constant	950.242	193.956	0	4.899	0.001***	-		
Value added of the secondary industry	0.106	0.014	0.928	7.473	0.000***	1	0.861	0.846
Dependent Variable: Rotation volume of freight transport								

The model passed F-test ($F=55.840$, $p=0.000<0.05$), so the model is valid and has significant linear relationship. The value of R square is 0.861, which means that the added value of the secondary industry can explain 86.1% of the change of freight turnover, so the equation has a good fitting effect of the model, shown in Figure 5.



Fig.5 The fitting effect of stepwise regression
(Photo credit: Original)

In addition, multicollinearity test is used to test the collinearity problem. From the SPSS result, it is found that the VIF value of all the left independent variables was less than 5. Moreover, the D-W value is around 2. Both of the results shows that no correlation exists among the sample data, so there was no collinearity problem in this model.

Given the significance level $\alpha = 0.1$, by looking up the t distribution table, it is found that the critical value of $n-k-1$ is $t_{\frac{\alpha}{2}}(n-k-1) = 2.353$. Since the t value of the added value of the secondary industry is 7.473, which is larger than the critical value of $t=2.353$, and its p value is $0.000^* < 0.1$. Therefore, the only left variable, value-added of the secondary industry, is significant.

Since the model passes all the tests above, it can be concluded that the regression coefficient of added value of the secondary industry was 0.106, which means that the value-added of the secondary industry will have a significant positive influence on the freight turnover. The regression equation is: $y=950.242+0.106 \times \text{Value added of the secondary industry (billion yuan)}$.

3.2.4. Forecasting accuracy

Using the equation of the stepwise regression, this research checks the fitness of the data from 2010-2020. As it is shown from Table 5, the model's prediction of 2010-2020 freight turnover is pretty accurate, with a maximum error of 7.7%, and an average error of 4.41% (less than 5%). Therefore, when the data of value-added of the secondary industry is available, the model can be used to predict the freight turnover of Sichuan Province in the coming years. The forecasting accuracy can roughly reach more than 95%.

Table 5. The forecasting error rate of stepwise regression

Year	Value added of the secondary industry	Real rotation volume of freight transport	Forecast rotation volume of freight transport	Forecasting error	Error rate
2010	8283.21	1711	1828.26226	117.2623	6.85%
2011	10014.39	1909	2011.76734	102.7673	5.38%
2012	11231.06	2254	2140.73436	113.2656	5.03%
2013	12418.94	2437	2266.64964	170.3504	6.99%
2014	13082.69	2351	2337.00714	13.99286	0.60%
2015	13192.45	2290	2348.6417	58.6417	2.56%
2016	13450.13	2404	2375.95578	28.04422	1.17%
2017	14569.17	2583	2494.57402	88.42598	3.42%
2018	16056.94	2820	2652.27764	167.7224	5.95%
2019	17187.92	2574	2772.16152	198.1615	7.70%
2020	17571.11	2734	2812.77966	78.77966	2.88%

3.3 Back Propagation neural network

3.3.1. Data Normalization

It is shown from Figure 1-3 that the correlation between variables is very high, so neural network model can be built using MATLAB. Selected independent variables are chosen to be the input variables of neurons and rotation volume of freight transport as output variables of neurons. This research compares the learning efficiency and prediction accuracy of two neural networks models divided by different training and testing sets. In the first model, data from 2010-2018 form the training set and 2019-2020 as testing set. In the second model, data from 2010-2019 form the training set and 2020 are testing set.

Since the original units and the magnitude of the input data is quite different, the data should be converted into uniform magnitude and dimensionless expression so as to accelerate the training of the network convergence, preventing neurons output saturation phenomenon and improving the accuracy of prediction. Since Mapminmax Function can only operate on each row, the matrix p , t should be transverse first. Then mapminmax function is used to normalize each row of the above data to the range $[0,1]$.

3.3.2. Building the network

The number of neurons in the middle layer was verified one by one according to Kolmogorov theorem. When the value is 8, the convergence speed and prediction accuracy are the best, the neural network is constructed accordingly. Then, weights are initialized, and target parameters are set.

In the first model, the capability of the training result table manifests that the MSE of the training set is 7.37×10^{-9} , which is much smaller than the target error this paper set. Besides, the overall R is 0.99266 which is pretty close to 1, indicating that the regression of neural network training fits the data very well. Thus, using BP neural network to forecast is very accurate.

The second model shows that the MSE of the training set is 2.46×10^{-7} , which is smaller than the target error but bigger than the first model. Besides, the overall R value of this regression is 0.92, which is relatively high indicate good fitness but smaller than 0.99266 of the first model. So, both MSE and R value prove that the first model is better.

3.3.3. Examining the network

Then, both models are examined. Data of 2019 and 2020 is used to check the effectiveness of the first model. After transverse and normalize the input variables, the output value of network goes through inversely normalization by which the forecast rotation volume of freight transport is obtained. Both the results and the prediction error of 2019 and 2020 are shown in the Table 6.

Table 6. Result and prediction error under the first model

	Real value	Prediction value	Error rate
2019	2574	2307	10.373%
2020	2734	2495	8.742%

Data of 2020 is used to check the fitness of the second model. Results and the prediction error of 2020 are shown in the Table 7. The error rate of the first model is smaller than the second model, indicating that when forecasting more years, the model has higher accuracy in terms of the last year.

Table 7. Result and prediction error under the second model

	Real value	Prediction value	Error rate
2020	2734	3061	11.961%

3.3.4. Forecasting using the network

When predicting the rotation volume of freight transport for the next few years, the value of the seven input variables should be known first. The value of these variables can be estimated using the average growing trend of the previous years. After inputting these estimated variables into the network, the prediction result of rotation volume of freight transport can be gained.

3.4 Grey prediction models

3.4.1. Level ratio test

The level ratio for every two years can be calculated. If it is within the range $(e^{\frac{-2}{n+1}}, e^{\frac{2}{n+1}})$, the data is good for model building. Since the ratio is from 0.847 to 1.096, which is within the range (0.846, 1.181), so it is suitable for building Grey Prediction model.

3.4.2. Model Building

The model is built by SPSS PRO. Since the posterior error ratio is 0.145, which is smaller than 0.35, this model has high prediction precision.

3.4.3. The fitness of the model

Table 8. Relative prediction error of year 2010-2020 using grey prediction model

Year	Actual value	Prediction value	Error	Relative error(%)
2010	1711	1711	0	0
2011	1909	2116.347	-207.347	10.862
2012	2254	2181.647	72.353	3.21
2013	2437	2248.962	188.038	7.716
2014	2351	2318.354	32.646	1.389
2015	2290	2389.886	-99.886	4.362
2016	2404	2463.626	-59.626	2.48
2017	2583	2539.642	43.358	1.679
2018	2820	2618.003	201.997	7.163
2019	2574	2698.781	-124.781	4.848
2020	2734	2782.052	-48.052	1.758

Table 8 manifests that the biggest relative error is 10.86%, and the average relative error is 4.133%, so this model is accurate. Besides, Figure 6 shows that this model fits well.

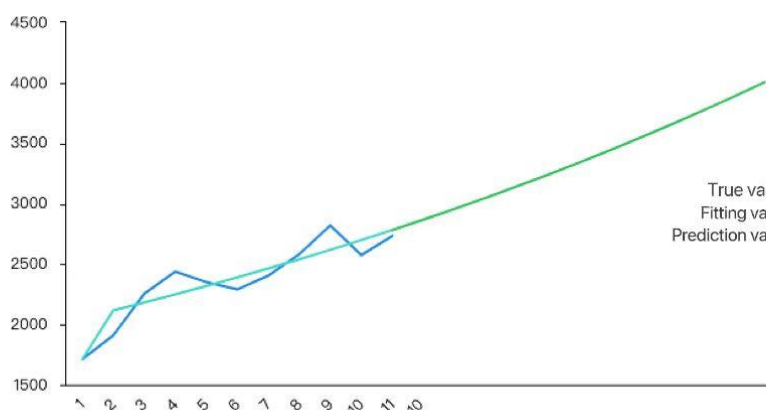


Fig.6 The fitness effect of grey prediction model
(Photo credit: Original)

3.4.5. Forecasting

Using grey prediction method, the forecasting value of the freight turnover in the coming three years are 2868, 2956 and 3048 million tons*kilometer.

4. Combined prediction model and limitation

4.1 Best models

Four kinds of criteria are used to evaluate and compare the prediction error of these three models. Results are obtained by MATLAB using data from 2010-2020. According to Table 9, Stepwise Regression has the smallest MSE and RMSE. So, if MSE and RMSE are used as the evaluation criterion of these three models, Stepwise Regression has the best forecasting ability. BP neural network has the smallest MAE and grey prediction model has the smallest MAPE. So, if MAE is used as the evaluation criterion, BP neural network is the best predicting model. In the same way, if MAPE is used, grey prediction model should be chosen.

Table 9. Prediction error of these three models under four kinds of criterion

	Forecasting Using Stepwise Regression	Forecasting Using BP	Forecasting Using Grey Prediction
MSE	13819	22698	14431
RMS E	117.5528	150.66	120.1311
MAE	103.4013	96.1952	98.0076
MAP E	4.41%	4.21%	4.13%

4.2 Combined model

These three models all show good prediction accuracy from different indexes and have their own advantages and disadvantages (not explain due to word limitation), so they should be combined to analyze the logistics demand of Sichuan Province in order to avoid the technical limitation of a single model and improve the overall prediction accuracy.

Using the mean average percentage error shown in Table 10, the mean average percentage accuracy of each model can be calculated. Since model with higher accuracy show better performance, it should be assigned a higher weight ratio to the combination model. Note that the total weight ratio of these three models add up to 100%.

Table 10. Weight ratio of these three models in the combined prediction model

	MAPA	Weight Ratio
Stepwise Regression	95.59%	33.28%
BP Neural Network	95.79%	33.35%
Grey Prediction Model	95.87%	33.37%

According to the weight calculation results in table (10), linear addition method can be adopted. Then logistics demand prediction formula based on multi-model combination is:

$$y_{overall}(n) = 0.3328y_S(n) + 0.3335y_{BP}(n) + 0.3337y_G(n). \quad (2)$$

Where, n is the period that this paper forecast. $y_{overall}(n)$ represents the freight turnover prediction result of the combined model in the n^{th} period. $y_S(n), y_{BP}(n), y_G(n)$ denotes the prediction result using stepwise regression, BP Neural Network, and Grey Prediction Model in the n^{th} period respectively.

According to the freight turnover in the past eleven years, the average error rate of Stepwise Regression, BP Neural Network and Grey Prediction Model are 4.41%, 4.21% and 4.13% respectively. So, the average prediction error rate of the multi-model combination is 4.25%.

4.3 Limitations and potential future research directions

This paper only considers several quantitative factors. But the change of freight turnover in Sichuan Province may also be affected by several unquantifiable factors, for example government policies, industry regulations and macro environment, which should be considered in later research.

5. Conclusion

As the “third source of profit”, the modern logistics industry has encountered a good opportunity for development. However, the formulation of logistics improvement and development policies and the layout of logistics infrastructure construction all lack of quantitative basis. Therefore, this paper establishes a scientific and reasonable forecasting model to reveal the relevant mechanism and forecast the logistics demand of Sichuan province. Multiple regression model, BP neural network model and grey forecasting model are applied to study and analyze the logistics demand forecast of

Sichuan Province. These three models all show good prediction accuracy from different error criterion. Stepwise Regression has the smallest MSE and RMSE. BP neural network has the smallest MAE and grey prediction model has the smallest MAPE. But these three models have their own disadvantages, so this paper give each of them a weight ratio and use a combination model to analyze the logistics demand of Sichuan Province in order to avoid the technical limitation of a single model and improve the overall prediction accuracy. Using the logistics demand prediction formula based on multi-model combination, the freight turnover of Sichuan Province in the next few years can be forecasted with accuracy of 95.75%, which provide a credible reference when planning the logistics system and constructing logistics infrastructure in the future.

References

- [1] Lin Xiaoping, Yuan Jie. Logistics forecasting of Chengdu Shuangliu Airport based on Grey Model. *Journal of Wuhan University of Technology (Traffic Science and Engineering Edition)*, 2007, (03): 457,459.
- [2] Dong Suqin, Li Xiaoyan. Forecast of logistics demand of agricultural products in Sichuan Province under direct broadcast mode. *Northern economy and Trade*, 2022, (05): 50-52.
- [3] Wang Yang. Combination Forecast of Highway passenger Volume in Sichuan Province. *Guide to Economic Research*, 2019, (12): 56-57.
- [4] Tian Fangming, Liu Qingfu. Study on Railway Freight Volume Forecast in Sichuan Province. *Technology Information*, 2011, (30): 40-41.
- [5] Wang Chenglin, Wang Meng, Luo Yun. Demand Forecast of Agricultural products Logistics based on DNGM (1,1) Model. *Practice and understanding of Mathematics*, 2021, Piaget 51 (11): 113-119.
- [6] Wang, Yingchun, Liu, Wenjuan. "Based on the Grey Prediction Methods Forecasting the Logistics Development of FuJian Province." *World-Association-of-Science-Engineering Global Congress on Science Engineering (GCSE 2010)*, 2011
- [7] Minling Zeng, Runmin Liu, Ming Gao, Yizhang Jiang, "Demand Forecasting for Rural E-Commerce Logistics: A Gray Prediction Model Based on Weakening Buffer Operator", *Mobile Information Systems*, vol. 2022, Article ID 3395757, 8 pages, 2022.
- [8] Cui Yuxin. Research on Logistics demand Forecast of Hebei Province based on Lasso-BP Neural Network. MS thesis. Hebei University of Economics and Economics, 2019.
- [9] Gao Xuexue. Logistics demand forecasting of Hainan Province based on BP Neural Network. *Logistics Technology*, 2018. 37 (01): 87-90.
- [10] Statistical Yearbook, Sichuan Provincial Bureau of Statistics, 2011-2021, postal code 610041, <http://tjj.sc.gov.cn/scstjj/c105855/nj.shtml>
- [11] Adam Hayes, "Multiple Linear Regression (MLR) Definition, Formula, and Example", June 23 2022, September 19 2022, <https://www.investopedia.com/terms/m/mlr.asp>
- [12] Great Learning Team, "An introduction to R-Square", July 22 2020, September 19 2022, <https://www.mygreatlearning.com/blog/r-square/>
- [13] S. S. Haykin, *Neural Networks and Learning Machines*, vol. 3, Pearson Education, Upper Saddle River, NJ, USA, 2009.
- [14] Anas AI-Masri, "How Does Back-Propagation in Artificial Neural Networks Work?", January 30 2019, September 19 2022, <https://towardsdatascience.com/how-does-back-propagation-in-artificial-neural-networks-work-c7cad873ea7>
- [15] Chen M (2013) MATLAB neural network principles and examples of fine solution. Tsinghua University Press, Beijing, pp. 156–158.