ISBN: 978-981-18-5852-9
WCSE 2022 Spring Event: 2022 9th International Conference on Industrial Engineering and Applications
doi: 10.18178/wcse.2022.04.087

Forecasting Traffic land Demand in Guangdong-Hong Kong-Macao Greater Bay Area Based on Gray-BP Neural Network Model

Shuhui Lyu [1 +], Chun-Hsien Chen [1]

[1] A School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore

Abstract. In addition to promoting rapid regional socio-economic development, traffic land will occupy a large amount of land resources, leading to conflicts between different land types. Therefore, it is necessary to forecast the traffic land demand with scientific methods. This paper takes the Guangdong-Hong Kong-Macao Greater Bay Area as a case study to forecast its traffic land demand for the next 8 years. The data were obtained from the regional yearbook from 2008 to 2020 and the shared application service platform of land survey results of the Ministry of Natural Resources. A reasonable index system of impact factors was established according to its development characteristics, and a gray correlation model was used to rank the importance of impact factors, and finally a coupled gray-BP neural network model was constructed to forecast the traffic land demand. The forecasting error of traffic land demand of 11 cities in Greater Bay Area from 2021-2028 obtained under this method is about 0.1%.

Keywords: BP neural network, gray relational analysis, traffic land, forecasting

1. Introduction

Nowadays, with the rapid progress of urbanization and the expansion of motor vehicle traffic in China, the problem of traffic land has become more and more serious [1]. The expansion of traffic land has a stimulating effect on regional social and economic development. At the same time, the development of traffic land will also bring conflicts between different types of land [2].The Greater Bay Area is one of the economic development regions with greatest openness in China. The Greater Bay Area include Guangzhou, Shenzhen, Zhuhai, Foshan, Huizhou, Dongguan, Zhongshan, Jiangmen, Zhaoqing, Hong Kong and Macau. Due to differences in spatial location and historical processes, the Greater Bay Area still has many differences in terms of industrial development, economic construction, and urban development in urban agglomerations. People's demand for transportation continues to increase, placing higher and higher demands on daily travel. However, considering the ecological construction of urban development and the imbalance between supply and demand of urban roads, there are still more serious problems. Reasonable expansion will also promote the rapid interflow of people and logistics, and promote the rapid growth of regional production capacity. Therefore, The realization of urban planning, land use, and regional land resource conservation and sustainable development requires scientific methods to reasonably predict the demand for traffic land and determine the scale of land use [2].

The Back Propagation Neural Network (BP neural network) is the most classic ANN training method. BP neural networks have been applied to many fields, especially in classification and prediction, due to their strong learning and generalization abilities [3-4]. Wang et al. developed a tax forecasting model using BP neural networks to demonstrate that it is feasible to analyze highly nonlinear forecasts through neural networks, and it has fast convergence, high accuracy and good generalization capability. Wei, M. et al [5] constructed a BP neural network model to forecast agricultural prices, which circumvents the previous forecasting methods that emphasize linear correlation between prices, which has obvious limitations and leads to low prediction accuracy. In analyzing the effect of hidden layer on stock price prediction, D. Zhang et al. [6] found that the accuracy of prediction of BP neural network is superior to the prediction model of deep learning fuzzy algorithm. In the field of predicting land use, the reason for the poor fit between the results and the actual value is that the data selection system is unreasonable and incomplete [7]. In the process of urbanization, people's demand for traffic land has phase characteristics. The complete reference to historical data and planning goals will conceal the risk and uncertainty characteristics of social development, and make the prediction results of traffic land distorted. On the other hand, there is no correlation analysis on the basic data before forecasting, which results in factors with low or no correlation being mixed into the impact factors. Therefore, we used gray relational analysis in this paper to sort the impact factors, and constructs a gray-BP neural network coupling model to predict the traffic land in the Greater Bay Area from 2021 to 2028.

2. Analysis of The Major Impact Factors on Traffic Land in The Greater Bay Area

2.1. Data Sources

We selected five basic data on the current status of land use, population size, industrial development level and economic and social accounting data of relevant departments and spatial data of administrative divisions of single city in the Greater Bay Area. The data were obtained from 50 statistical yearbooks from 2008-2020, including Region Statistical Yearbook, China Traffic Statistical Yearbook, as well as the shared application service platform for land survey results of the Ministry of Natural Resources, and the data are authentic and reliable.

2.2. Structure of Evaluation Index System for Impact Factors of Traffic Land Demand

The factors affecting traffic land use involve physical geography, socio-economics, institutions and policies, etc. Various production and consumption activities in cities also increase the demand for traffic land. Therefore, the selection of traffic land demand factors should be based on the criteria of harmonizing with the development of socio-economic productivity. In this paper, eight evaluation dimensions were selected to influence the demand for traffic land in the Greater Bay Area, including urbanization level, labor resource endowment, economic development level, industrial structure heightening, regional financial capacity, industrial development level, regional investment level and regional consumption capacity. The evaluation layer is refined into a factor layer under the evaluation layer, including nine indicators on the scale of urban land, urbanization rate, total population, regional GDP, general budget revenue of local finance, total social fixed asset investment, per capita consumption expenditure, total industrial output value, and the number of employees in secondary and tertiary industries. The model for constructing the evaluation index system for the impact factors of traffic land demand is shown in Table 1.

Table 1: impact factors of traffic land demand

Evaluation layer	**Factor layer**	**Unit**	**Variable Settings**
Urbanization level	Urban land scale	Km^2	X_1
Labor resource endowment	Urbanization rate	%	X_2
The level of economic development	Total population	10 thousand	X_3
Regional financial capacity	Regional GDP	100 million yuan	X_4
Regional investment level	General budget revenue of local finance	100 million yuan	X_5
Regional spending power	Total investment in fixed assets of the whole society	100 million yuan	X_6
Industrial development level	Consumption expenditure per capita	Yuan	X_7
Degree of advanced industrial structure	Industrial output	100 million yuan	X_8

2.3. Gray Relational Analysis

Gray theory is a system theory founded by Deng Julong in 1982. It can weaken the randomness of the original data and obtains valuable information from it. Find out the law in the original data, apply the differential fitting method to transform the time series into differential equations, and then establish the model [7]. The calculation steps of gray relational analysis are as follows:

A. Identify reference and comparison sequences

The traffic land size of the city group in the Greater Bay Area from 2008 to 2020 is used as the dependent variable group (Y) as in (1).

$$Y = \{Y(k)|k = 1,2, \dots, n\} \tag{1}$$

Set the scale of urban land use, urbanization rate, total population, regional GDP, general budget income of local finance, total fixed asset investment of the whole society, consumption expenditure per capita, total industrial output value, and the number of employees in secondary and tertiary industries as independent variable groups. Set 7 sets of comparison sequences, as in (2).

$$X_i = \{X_i(k)|k = 1,2, \dots, n\}，\ i = 1,2, \dots, m \tag{2}$$

B. Dimensionless processing of indicators

Due to the different dimensions of the original data, the data cannot be analyzed and compared. Therefore, the non-dimensional processing is required before the data is associated to eliminate the dimensional differences between different data. The dimensionless methods commonly used in gray relational analysis include initial value conversion method, mean value conversion method and standardized conversion method. Here, use the initial value conversion method to obtain the original value image of the reference sequence-based raw data. The calculation formula is as in (3).

$$X_i = \frac{X_i}{x_i(1)} = (x_i'(1), x_i'(2), \dots, x_i'(n))，\ i = 0,1,2, \dots, m \tag{3}$$

C. Calculate the relationship coefficient between each factor and the traffic land demand, and the formula is as in (4).

$$\varepsilon_i(k) = \frac{min\, i\, min\, k\, |y(k) - x_i(k)| + \rho\, max\, i\, max\, k\, |y(k) - x_i(k)|}{|y(k) - x_i(k)| + \rho\, max\, i\, max\, k\, |y(k) - x_i(k)|} \tag{4}$$

Where $y(k)$ is the corresponding value in the reference sequence, and $x_i(k)$ is the corresponding value of the comparison sequence after initial value conversion, and ρ is the resolution coefficient, which is between [0,1].

D. Calculate the correlation between each factor and the demand for traffic land, and the formula is as in (5).

$$r_i = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_i(k)，\ k = 1,2, \dots, n \tag{5}$$

The above calculation process is implemented in the $MATLAB\ 2016a$ operating platform, and set resolution coefficient ρ as 0.5. The results are shown in Table 2.

Table 2: Gray relational degree

Factors classification	**Gray relational degree**
Urbanization level X_1	0.7186
Labor resource endowment X_2	0.6811

Factors classification	Gray relational degree
The level of economic development X_3	0.8226
Regional financial capacity X_4	0.6363
Regional investment level X_5	0.8608
Regional spending power X_6	0.8282
Industrial development level X_7	0.8060
Degree of advanced industrial structure X_8	0.8132

The gray correlation degree between the dependent variable and the independent variable is sorted as follows: $X_5 > X_6 > X_3 > X_8 > X_7 > X_9 > X_1 > X_2 > X_4$. That is, the relevance between the general budget income of the local finance, the total fixed asset investment of the whole society, the total population, the total industrial output value, the number of employees in the secondary and tertiary industries, the scale of urban land, the urbanization rate, the regional GDP and the demand for traffic land drops by times.

The degree of correlation between the independent and dependent variables is above 0.8, and the correlation is considered to be great; 0.5~0.8 is considered to have a certain correlation; but when the correlation is less than 0.5, the two are considered unrelated, and should Eliminate this factor. According to the results, the correlation between the selected 9 independent variables and the dependent variables are all greater than 0.5, indicating that the selection of factors affecting the demand for traffic land is convincing. Among them, the total population, the general budgetary income of the local finance, the total fixed asset investment of the whole society, the per capita consumption expenditure, and the total industrial output value are greater than 0.8, indicating the endowment of labor resources, the regional financial capacity, the level of industrial development, the level of regional investment and the region Consumption capacity is the main influence layer of the needs for traffic land in the Greater Bay Area.

3. BP Neural Network Prediction Model

3.1. Model Construction

Nowadays, BP neural networks have broad applications in the field of pattern recognition, image processing, analysis and control [8]. The network is composed of three neuron levels: input layer, hidden layer and output layer [7]. The BP neural network is based on gradient descent to search the error surface for the point with the smallest error [8]. Each layer neuron is only influenced by the neuron in previous layer. The model is shown in Fig. 1. The input layer $X_1, X_2, \dots, X_j$ is transmitted in from the instance feature vector of the training set, passed into the next layer after the weights of the connected nodes, and then the output is transformed according to the nonlinear equation, and the output of the previous layer is the input of the next layer. If the desired result cannot be reached at the output layer, then the error signal is propagated backwards along the original path. In this iteration, the weight of each neuron is modified to minimize the error signal. The coupling of gray systems with BP neural networks can better predict social and economic problems with complex influencing factors. The forecast of traffic land demand is a good example.

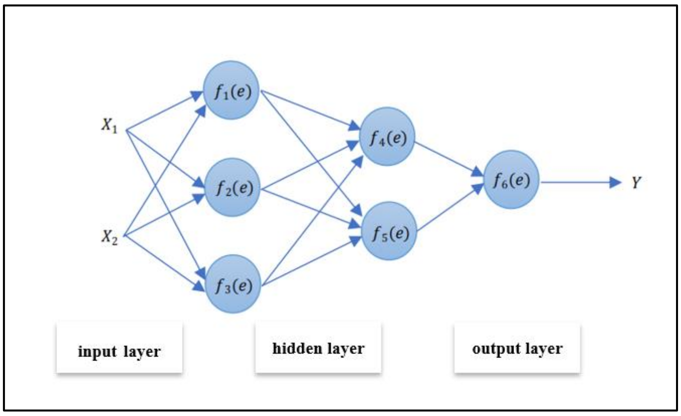


Fig. 1: BP neural network model

In predicting the demand for traffic land, the operation construction of gray-BP neural network prediction model is constructed as shown in Fig. 2. Firstly, the gray relational analysis method is used to determine the primary factors affecting the change in traffic land demand among various social and economic factors, and then the main influencing factors are used as the input layer unit, and the traffic land demand as the output layer unit. At the same time, the relevant historical data is used to influence the system. Carry out simulation training, and constantly adjust the parameters such as the hidden layer's neurons to reach the best simulation output.

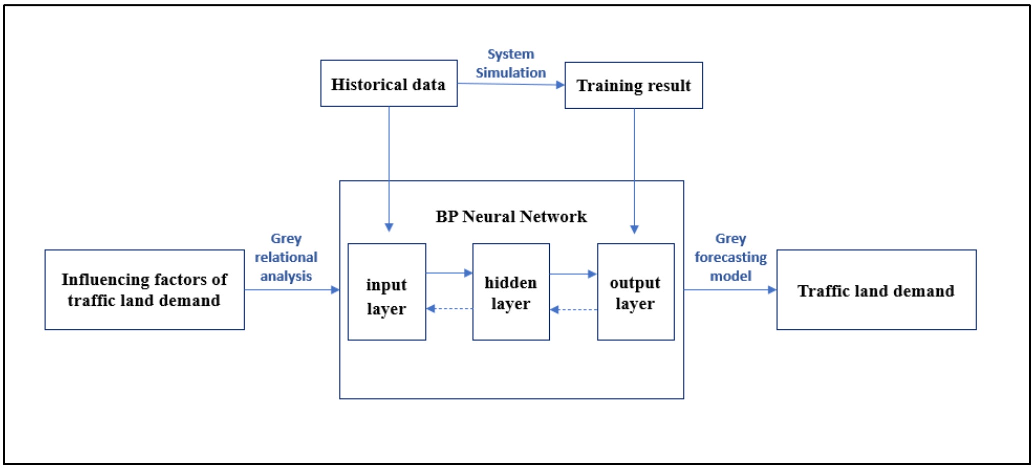


Fig. 2: Gray-BP neural network prediction model

Use empirical formula as in (6) to calculate the number of hidden layers and nodes [9], which is shown in Fig. 3.

$$L = \sqrt{m+n} + a \tag{6}$$

where m is input layer's number of neurons, n is output layer's neurons, and a is a constant, which is between1~10.

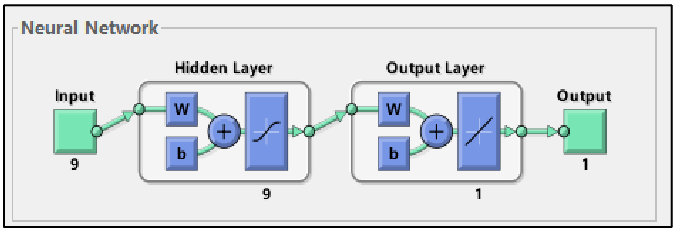


Fig. 3: The number of hidden layers and nodes

The main process of BP neural network model and gray prediction model is as follows:

A. Network initialization: A random number between$(-\mathbf{1}, \mathbf{1})$ is assigned to each connection weight and an error function $\boldsymbol{e}$ is set, and then the computational accuracy ε and the maximum learning number $\boldsymbol{M}$ are given. The formula is as in (7).

$$e = \frac{1}{2}\sum_{j=1}^{n}\left(y_i - y_j'\right)^2 \tag{7}$$

B. Calculate the output in each unit in hidden layer and output layer according to the input sample value and output expected value, and modify the weights of every neuron input node according to the gradient direction. The correction value of each weight is $\boldsymbol{\Delta\omega}$, and the formula is as in (8).

$$\Delta\omega_{ij} = -\eta * \frac{\partial\varepsilon}{\partial\omega_{ij}} = -\eta * \frac{\partial\varepsilon}{\partial I_j} * \frac{\partial I_j}{\partial\omega_{ij}} \tag{8}$$

ω_{ij} is the connected weights from node i to node j in the output layer; η is the learning efficiency value, which is the transfer function of the jth hidden layer. The output layer to the hidden layer are $tansig$ functions, while the hidden layer to the output layer are $trainlm$ functions.

C. Iteration: Select the next input mode and return to the second step. If the output layer cannot obtain the desired result, it will return an error signal along the original route of the connection. During iteration, each neuron weight are modified to minimize the error signal, and it stops until the output error reaches the accuracy requirement.

3.2. Forecast Results

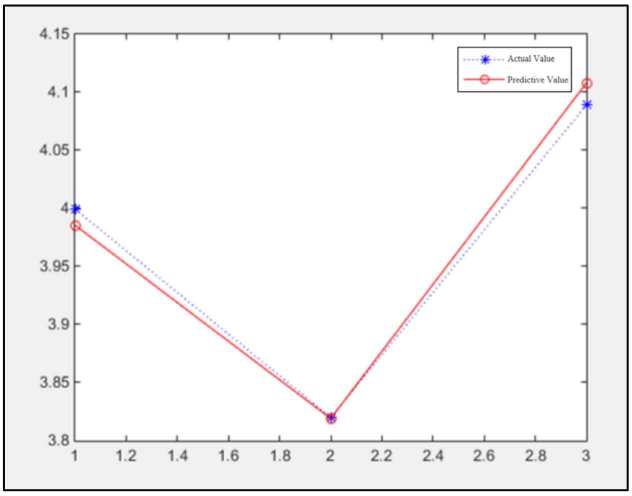
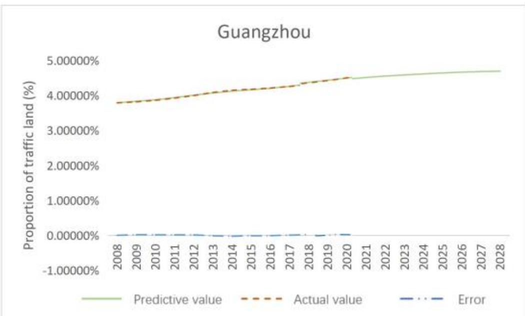
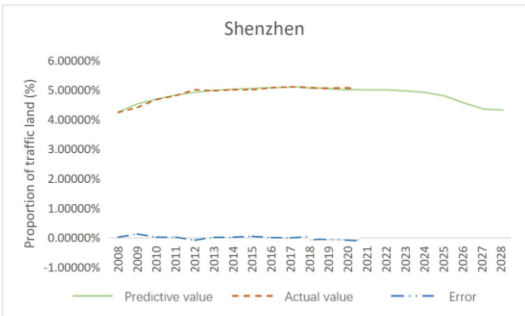


Fig. 4: Forecasting results in MATLAB 2016a

The forecasting results of 11 cities are shown in Fig. 4-5. The prediction results of a single city have a high degree of fit with historical data, and the errors are all around 0.1%. The demand for traffic land in Guangzhou, Dongguan, and Zhuhai in the urban agglomeration is showing a steady and slow upward trend. In the next 8 years, the traffic land in Guangzhou and Dongguan will continue to expand. The demand for traffic land in Hong Kong, Macau, Jiangmen, Zhuhai, and Zhaoqing began to stabilize around 2016 and will maintain a stable demand for the next 8 years. The demand for traffic land in Shenzhen, Zhongshan, and Foshan showed a downward trend after peaking in 2014, 2020, and 2021. As land resources become more scarce and urban construction becomes more and more complete, the demand for traffic land decreases. The situation confirms the reliability of our predictions.





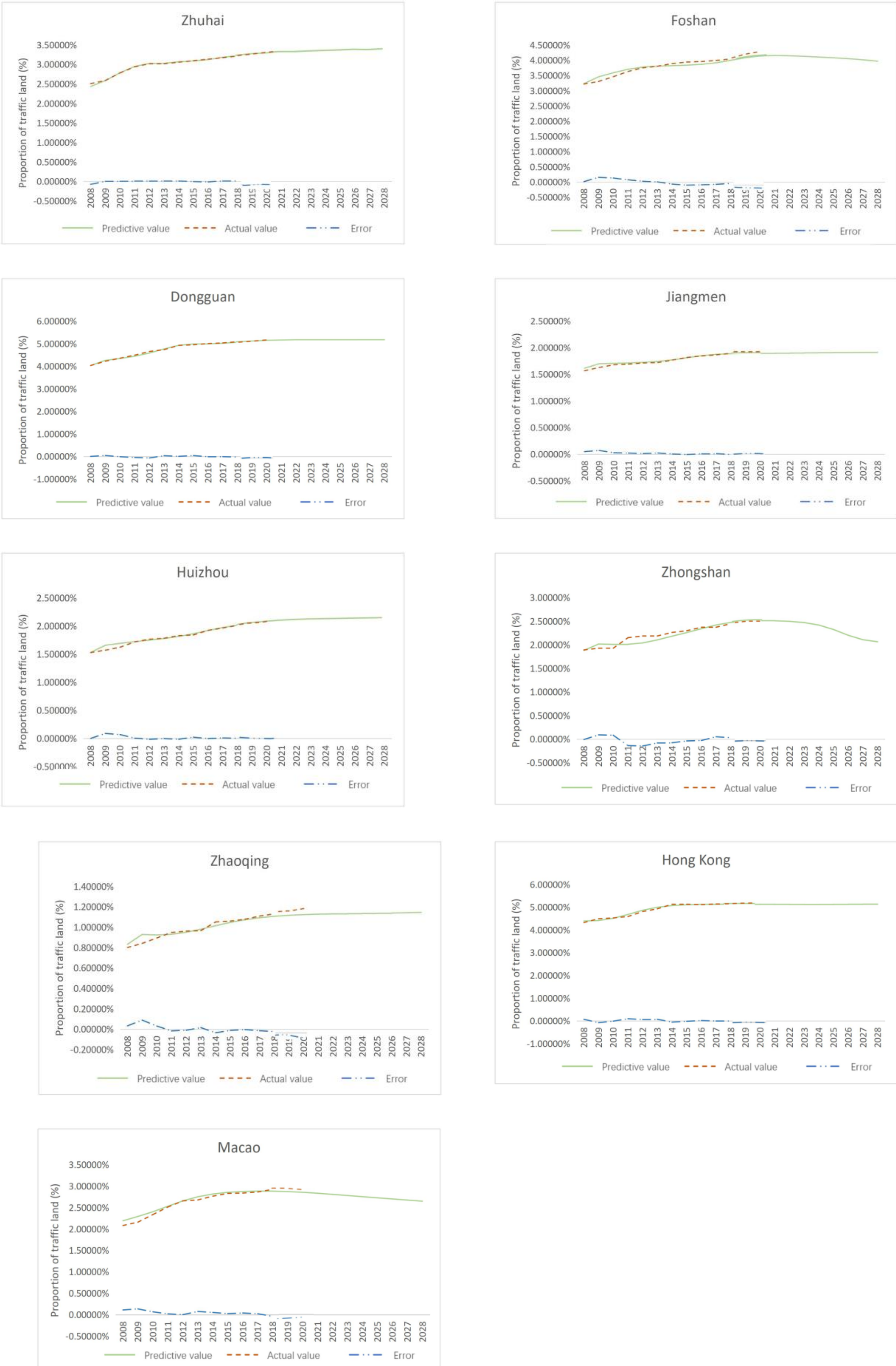


Fig. 5: Forecasting results of 11 cities

4. Conclusion

We used the gray-BP neural network in this paper to forecast the traffic land demand of the Greater Bay Area city group. A reasonable demand factor evaluation system is constructed before the prediction. The gray relational model is used to calculate and screen the factors with strong correlation, providing true and credible data support for prediction. According to the BP neural network forecasting model, we have obtained the traffic land demand in the next 8 years with an error of 0.1%. The forecast result can provide a certain basis for the traffic land plan of the Greater Bay Area. Traffic land can be further subdivided into five secondary land types: railways, highways, rural roads, civil airports, and ports. To further predict the demand for secondary traffic land, a new index system needs to be constructed according to the development characteristics of different land types. The suitability of the gray-BP neural network model in the second-level traffic land demand forecasting needs to be researched and verified. Refining the traffic land demand forecasting will become the goal of subsequent research.

5. References

[1] W. Meng, T. You, Y. Duan, and J. He, "Planning Practice of Coordinated Development of Urban Traffic and Land Use," in IOP conference series. Earth and environmental science, 2021, vol. 632, no. 2, p. 22051–. doi: 10.1088/1755-1315/632/2/022051.

[2] A. Azad and X. Wang, "Land Use Change Ontology and Traffic Prediction through Recurrent Neural Networks: A Case Study in Calgary, Canada," ISPRS international journal of geo-information, vol. 10, no. 6, p. 358–, 2021, doi: 10.3390/ijgi10060358.

[3] Huize Chen, "Construction and Application of Forecast Model for Eco-Tourism Market Demand Based on BP Neural Network[J]", The Frontiers of Society Science and Technology, vol. 1, no. 6, pp. 87-95, 2019.

[4] L. Zhang, J. Luo, and S. Yang, "Forecasting box office revenue of movies with BP neural network," Expert systems with applications, vol. 36, no. 3, pp. 6580–6587, 2009, doi: 10.1016/j.eswa.2008.07.064.

[5] M. Wei, Q, Zhou, Z. Yang, and J. Zheng, "Prediction model of agricultural product's price based on the improved BP neural network," in 2012 7th International Conference on Computer Science & Education (ICCSE), 2012, pp. 613–617. doi: 10.1109/ICCSE.2012.6295150.

[6] D. Zhang and S. Lou, "The application research of neural network and BP algorithm in stock price pattern classification and prediction," Future generation computer systems, vol. 115, pp. 872–879, 2021, doi: 10.1016/j.future.2020.10.009.

[7] Y. Shi, M. Li and Y. Fu, "Multi-scenario traffic land demand forecasting based on gray system-BP neural network model:A case study of urban agglomeration in the middle reaches of the Yangtze River," Journal of China Agricultrural University, vol. 25, no. 6, pp. 142-153, 2020, doi: 10.11841/j.issn.1007- 4333.2020.06.15.

[8] L. Yang, "The prediction of tmmel border displacement by BP neural network," Value Engineering, vol. 31, no. 25, pp. 125–127, 2012.

[9] D. Dai and J. Chen, "Research on Mathematical Model of Box Office Forecast through BP Neural Network and Big Data Technology," Journal of physics. Conference series, vol. 1952, no. 4, p. 42118–, 2021, doi: 10.1088/1742-6596/1952/4/042118.