

**A MATHEMATICAL ANALYSIS FOR THE FORECAST
RESEARCH ON TOURISM CARRYING CAPACITY TO
PROMOTE THE EFFECTIVE AND SUSTAINABLE
DEVELOPMENT OF TOURISM**

YANAN WANG, TAO XIE AND XIAOWEN JIE*

Business School, Sichuan University
Chengdu, Sichuan 610064, China

ABSTRACT. With the continuous and quick development of Chinese tourism industry over years, ecological environmental problems emerge consequently. The contradiction between the development of tourism economy and the protection of ecological environment has become the focus of scientific experts and Chinese government, and accordingly it is of vital importance to predict tourism carrying capacity accurately. In this paper, a new forecast approach is proposed for government staff and scenic spot management staff on tourist carrying capacity, which promotes the effective, healthy and sustainable development of the tourism country.

1. Introduction. With the continuous development of tourism, tourism carrying capacity forecast is becoming prominent. The accurate prediction of tourism carrying capacity can greatly improve the level of tourism economy, strengthen the pragmatism and foresight in drafting and implementing tourism industry development policies, and accordingly address challenges of tourist attractions (especially about the resultant damage to ecological environment) [15].

Since the 1960s, western scholars have begun to study on tourism demand forecast, reaching fruitful results in the 1980s by major means of tourist demand modelling or empirical analysis. Related research did not start until the 20th century end in China, targeted more at theoretical introduction and expansion of western research but less at empirical studies [28]. Currently, there are two types of tourism carrying capacity forecast: quantitative research and qualitative research, in which the former one is the most frequently-used way of research [23]. Quantitative predictive research is divided into four types: non-causal time series model, causality model, artificial intelligence model, and combinatorial model.

The time series model indicates that the state of motion obtained from historical data observations may reoccur in the future period of time, which provides us with the chance to forecast future with past data. In this way, the data collection costs are saved. The main time series models include auto-regression, exponential smoothing, and seasonal differential autoregressive moving average method (SARIMA). Among them, autoregressive integrated moving average model (ARIMA) is the most common one in tourism carrying capacity prediction. However, it is still

2010 *Mathematics Subject Classification.* Primary: 58F15, 58F17; Secondary: 53C35.

Key words and phrases. Tourism capacity, sustainable development, tourism industry.

* Corresponding author: Xiaowen Jie, jjxw@vip.163.com.

a controversial academic issue as to whether ARIMA is better than other models. For example, Cho (2001) argues that ARIMA is better than other time-series models [4, 7]; while through empirical analysis, Smeral and Wuger (2005) held that neither ARIMA nor SARIMA is as superior as the primary model (without change) [22]. In the time-series model, the generalized autoregressive conditional heteroskedasticity (GARCH) model is widely used in tourism demand forecasting conducted by Chan, Lim, McAleer (2005) [3] and Wang Fengbo (2009) [31], respectively.

Compared with the time-series model, the causal relationship model has the advantage of analysing the causal relationship between tourism carrying capacity and its influencing factors, so as to provide guidance for the development of relevant measures and policies for tourist attractions management institutions. To address the possible problem of spurious regression in traditional causal relationship model, models such as auto-regression distribution lag model (ADLM), error correction model (ECM), vector autoregressive model (VAR), and time-varying parameter model (TVP) were invented and widely used. For example, Kulendran and Wilson (2000) [11], Kulendran and Witt (2003b) [12], and Lim and McAleer (2001a) [17] all uses the error correction model to predict tourism demand; Shan and Wilson (2001) [21] and Song and Witt used the VAR model; Song, Witt and Jensen (2003) [25], and Song, Wong and Chon (2003) [26] used the autoregressive distribution lag model; this paper uses the time-varying parameter model, and the research results proves the superiority of this model over other causal relationship model.

With the emergence of artificial intelligence, artificial intelligence technology is applied to the prediction of tourism capacity, breaking the limitations of traditional linear prediction methods. In the middle and late of the 1990s, Rob Law (1999) introduced artificial neural networks into the field of prediction with respect to the demand for Japanese tourists in Hong Kong [13], after which the method was widely used.

Huang (1998) proposed a new self-adaptive method of empirical mode decomposition for nonlinear and nonstationary data [10]. After that, this method has been continuously improved and applied to various fields. Deng Yongjun et al. (2001) argue that this method is the best to extract trends or means of data sequences [5]. Liu Huiting et al. (2006) expatiate on the empirical mode decomposition method and its implementation [19].

Many of the single predictive models have achieved relatively high accuracy of the forecast results, but at the same time influenced by many other factors. In general cases, none of the single prediction models are comprehensive and effective enough to reveal travel demand information, which reduces the accuracy of the prediction results markedly [8]. To improve prediction accuracy, scholars focus on the development of combinatorial forecasting models by integrating the advantages of single models. For instance, Lei Kewei et al. (2007) first used ARIMA to predict the estimated value of China's domestic tourism, and then used artificial neural network to correct the error of the initial estimated value, thus improving the accuracy of the forecast [14]. Wang Yong et al. (2011) used the logistic model and the regression iterative model to combine the predictions, which improves the accuracy of the prediction results [30]. Li, Wong et al. (2006) combined the time-varying parametric model with the error correction model to prove that this combinatorial model is better than their parents [16]. In order to overcome the limitations of

quantitative research methods and improve prediction accuracy, some scholars combine quantitative research with qualitative research to form quantitative-qualitative combinatorial models. Such example include Tideswell et al. (2001) [29] and Jiao Song et al. (2013) who proposed a verification method based on empirical modal decomposition and grey relational analysis to solve the problem of simulation model validation of non-stationary fast changing data [27]. However, there are controversies over the large number of influencing factors of tourism carrying capacity; meanwhile, corresponding research results are affected by emergencies, which contributes to the non-linearity of inter-factor relationships. Therefore, the accuracy of combinatorial models is limited and waiting to be improved.

To this end, this paper uses the empirical modal decomposition method to decompose the time series of tourism carrying capacity into multiple modal functions to reflect its internal features, and then forecast tourism carrying capacity with the backpropagation artificial neural network.

2. Empirical mode decomposition - backpropagation artificial neural network prediction model.

2.1. Empirical modal decomposition - backpropagation artificial neural network combinatory forecasting ideas. Based on the empirical modal decomposition-backpropagation artificial neural network (ANN) model, the tourism carrying capacity of Mount Emei is forecasted in this paper. The basic idea is to use the empirical modal decomposition method to decompose the historical time-series data of Mount Emei tourism carrying capacity into multiple modal functions to reflect its internal features, and then forecast tourism carrying capacity with error backpropagation artificial neural network [32].

With the empirical modal decomposition – error backpropagation artificial neural network combinatory model, we implement the test procedures as shown in Figure 1:

① Empirical mode decomposition process

According to the above empirical mode decomposition, the tourism carrying capacity time series $x(t)(t = 1, 2, \dots, n)$ is decomposed into $IMF, c_i(t)(i = 1, 2, \dots, n)$ and the residual item $r_n(t)$.

② Grouped frequency

The obtained intrinsic mode functions are classified from high to low frequencies. The high frequency of intrinsic mode function means longer learning process; low frequency function means shorter learning process. Therefore, it is necessary to classify the time series in order to construct a network prediction model with short cycle and high precision.

③ Network prediction stage

According to the preparation phase of ① and ②, the artificial neural network model is constructed, trained and predicted in order. Finally, all the predicted results are linearly added and the final estimated value is obtained.

2.2. Empirical mode decomposition. In 1998, the American-born Chinese Huang initiated the idea of empirical mode decomposition which mainly uses Hilbert transform to select non-linear and non-stationary data sequences until the final data sequence is stationary. Each data sequence is called an intrinsic modal function (IMF). The intrinsic modal function is defined as follows:

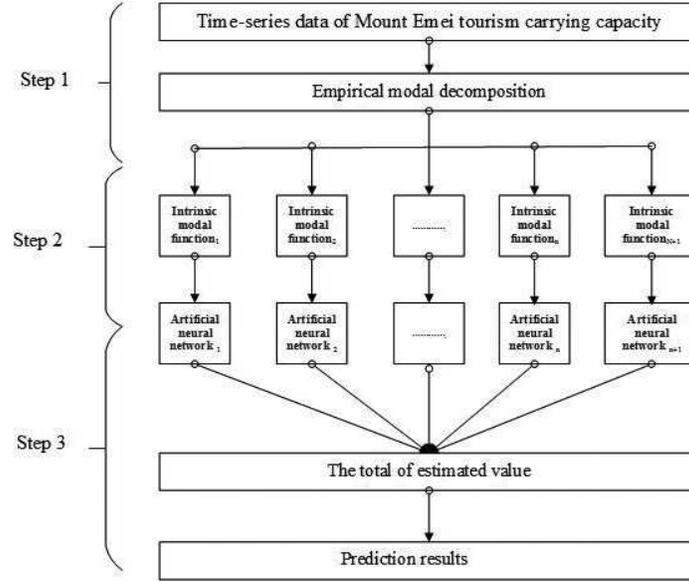


FIGURE 1. Schematic diagram of the combinatorial model based on empirical modal decomposition- error backpropagation artificial neural network

① In the whole time series, the difference between the extreme value (maximum and minimum) and the zero-crossing point is smaller than 1.

② In the arbitrary range of the time series, the average value of the upper and lower envelopes is 0 [9].

The procedure of empirical mode decomposition is as follows:

① find the original time series $x(t)$ and all the local minimum and maximum values

② Find the lower envelope $u_0(t)$ that contains all the minimum values and the upper envelope $v_0(t)$ with all the maximum values (the upper and lower envelopes should include all the data). Their mean value is represented by $m_1(t)$, and the difference between $x(t)$ and $m_1(t)$ is expressed by $h_1(t)$. Therefore,

$$m_1(t) = \frac{u_0(t) + v_0(t)}{2} \quad (1)$$

$$h_1(t) = x(t) - m_1(t) \quad (2)$$

③ Determine whether $h_1(t)$ meet the above definition of IMF. If it is satisfied, then $h_1(t)$ is *IMF*; otherwise, $h_1(t)$ is expressed by the next iteration symbol $x(t)$, so:

$$h_{11}(t) = h_1(t) - m_{11}(t) \quad (3)$$

④ Repeat the steps ① to ③ until the initial IMF is found the K th time, which is expressed by $c_1(t)$.

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t) \quad (4)$$

$$c_1(t) = h_{1k}(t) \quad (5)$$

⑤ The residual items $r_1(t) = x(t) - c_1(t)$ are marked as the new time series. Repeat steps ① to ④ until we find another IMF, represented by $c_2(t)$. When the residual term is $r_2(t) = r_1(t) - c_2(t)$, repeat ① to ⑤ until the number of residual items is lower than the threshold at the loop end. The final decomposed item are $nIMF$ and a residual item $r_n(t)$. The original sequence c is expressed as:

$$x(t) = \sum_{k=1}^n c_k(t) + r_n(t) \tag{6}$$

2.3. Error backpropagation artificial neural network. The error backpropagation artificial neural network is also known as the error backpropagation network, which is currently the most commonly used neural network [6]. Error backpropagation artificial neural networks include: input layer, output layer, and hidden layer (usually one or more layers). The error backpropagation of artificial neural network is shown in Figure 2:

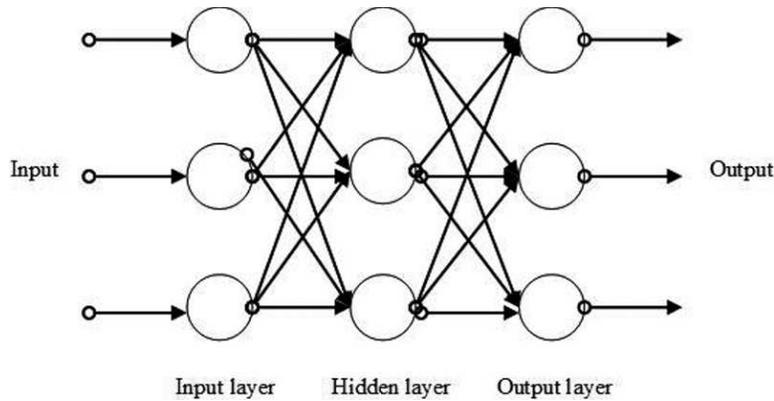


FIGURE 2. Structure map of error backpropagation of artificial neural network

(1) Error backpropagation algorithm

The structure map of error backpropagation of artificial neural network is shown in Figure 2, which consists of N neural cells in input layers, K neurons in hidden layers, and M neurons in output layers. Assume that O_{2pm} and O_{1pk} are the outputs of the output layer and the input layer, respectively, and that W_{2km} and W_{1nk} are the connection weight from the hidden layer to the output layer and from the input layer to the hidden layer, respectively. We let the learning sample be x_{pn} and the corresponding target output be t_{pm} .

The steps of the error backpropagation algorithm are:

① Initialize the weight. Assuming that the learning rate is α , the error is ε , the maximum cycle index is \max , and $i = 0$;

② Feedforward calculation: In the neural network model, we input $X_p(X_p = \{X_{p1}, \dots, X_{pn}\})$ and calculate O_{2pm} and O_{1pk} according to the following formula:

$$O_{1pk}(i) = f\left(\sum_{n+1}^N w_{1nk}(i)x_{pn}\right)$$

$$O2_{pm}(i) = f\left(\sum_{n+1}^N w2_{km}(i)O1_{pk}(i)\right) \quad (7)$$

Sigmoid function is used usually as the conversion function.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

③ Calculate the mean square error MSE. If $MSE \leq \varepsilon$, the loop ends. Otherwise continue to ④;

$$MSE = \frac{1}{M} \sum_{k=1}^M (t_{pm} - O2_{pm})^2 \quad (9)$$

④ Reverse calculation: Calculate the number of variable weights according to the following formula:

$$\Delta w1_{nk}(i+1) = \alpha \sum \delta_{pk}(i)x_{pn} \quad (10)$$

$$\Delta w2_{km}(i+1) = \alpha \sum \delta_{pm}(i)O1_{pk}(i) \quad (11)$$

Where:

$$\delta_{pm}(i) = (t_{pm} - O2_{pm}(i))(1 - O2_{pm}(i)) \quad (12)$$

$$\delta_{pk}(i) = O1_{pk}(i)(1 - O1_{pk}(i)) \sum_{m=1}^M \delta_{pm}(i)w_{km}(i) \quad (13)$$

Reset weight:

$$w1_{nk}(i+1) = w1_{nk}(i) + \Delta w1_{nk}(i+1) \quad (14)$$

$$w2_{km}(i+1) = w2_{km}(i) + \Delta w2_{km}(i+1) \quad (15)$$

⑤ Let $i = i + 1$, and return to ②.

(2) Improvement on the error backpropagation algorithm

Despite the wide use, the error backpropagation algorithm has some shortcomings [18] such as the large number of training samples, low learning efficiency, slow convergence, changeable parameter selection, and easiness to obtain local minimum value. The learning process of error backpropagation artificial neural network includes error backpropagation connection weight learning and network structure learning. Therefore, improvements should be implemented in these two aspects. The main improvement methods include simulated annealing algorithm, genetic algorithm, training sample normalization method, initial weight and threshold logic selection method, and network structure adjustment method.

The error backpropagation algorithm usually uses s-shape function to represent the transfer function with an interval of [0, 1]. The actual sample data may vary greatly, such that training samples are uncomparable. Therefore, for fear that small values are covered by big values, we normalize data samples within the interval of [0, 1]. In the calculation process, formula 19 is used for data pre-processing:

$$\bar{x}_i = \alpha + \beta \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (16)$$

In the formula, α, β are constants, x_{\min}, x_{\max} represent the minimum and maximum values in each sample set; x_i and \bar{x}_i represent the raw value and the processed value, respectively.

3. Forecast of tourism carrying capacity of Mount Emei.

3.1. Data source. In China, Mount Emei is a famous natural attraction visited by hundreds of thousands of tourists at home and abroad. The accurate prediction of daily tourist volume of Mount Emei can provide managers with base data, decision-makers with the way to distribute human resources comprehensively, restaurants and hotels with the optimal supply structure, and is helpful in formulating tourism traffic realignment schemes.

The author takes daily tourist volume of Mount Emei as the prediction object, receiving 365 data samples from the access control system in Mount Emei (2010.1.1–2010.12.31), as shown in Figure 3:

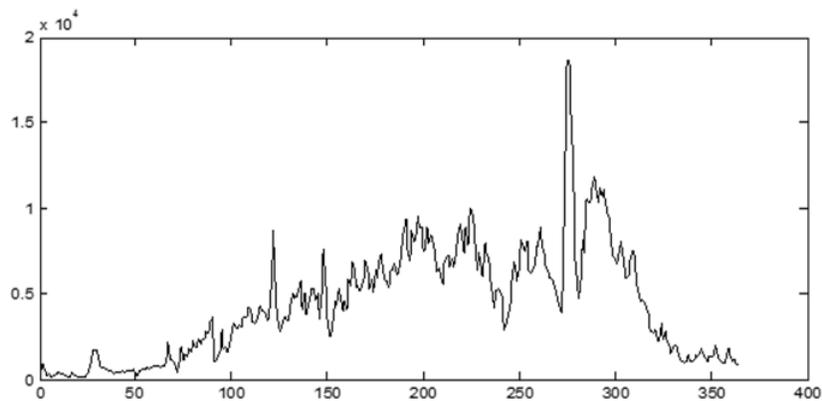


FIGURE 3. 2010 Mount Emei tourist area daily visitors capacity of the time series data

In order to ensure the effectiveness of the assessment model, this paper uses the error assessment method to study on tourism carrying capacity forecast, with the average deviation rate as the data for error assessment.

$$E = \frac{\sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{y_i}}{N} \quad (17)$$

In this formula, \hat{y}_i is the estimated value of I , y_i is the actual value, and N is the expected total sum.

3.2. Predictive results analysis. In this paper, the author first decomposed the empirical mode of original data, obtaining 11 intrinsic modal functions and 1 residual term. Then, a 3-layer artificial network is used to construct the forecast model of tourism carrying capacity of Mount Emei. In using the empirical modal decomposition-error backpropagation artificial neural network prediction model, since the number of training samples (365) is so large that it requires the number of neurons N_2 in hidden layers as many as possible, we let N_2 be 150. Actually, the number of neurons N_2 in hidden layers should be adjusted according to real training situations. The number of neurons N_1 in input layers is 1, and the number of neurons N_3 in input layers is 1, too. The matrix laboratory test is used in this paper, and the related parameters are: the conversion functions from the input layer to the middle layer and from the middle layer to the output layer are tangent

functions; the learning rate is 0.5; the training time is 5000; the training target error is 1×10^{-10} .

After the normalization process, in order to train neural network by matrix laboratory to obtain the output data, we should first restore data in the output layer. The corresponding formula can be derived from the normalization formula:

$$Y = 10^{(e \times 10)} \quad (18)$$

In this formula, Y is the output data of the restored neural network, y^i is the output data of the neural network.

Figure 4 is a comparison between the estimated value and the actual value of the tourism carrying capacity using the empirical modal decomposition-error backpropagation artificial neural network prediction model. The estimated curve is close to the actual curve in shape and, in fact, they almost coincide with each other.

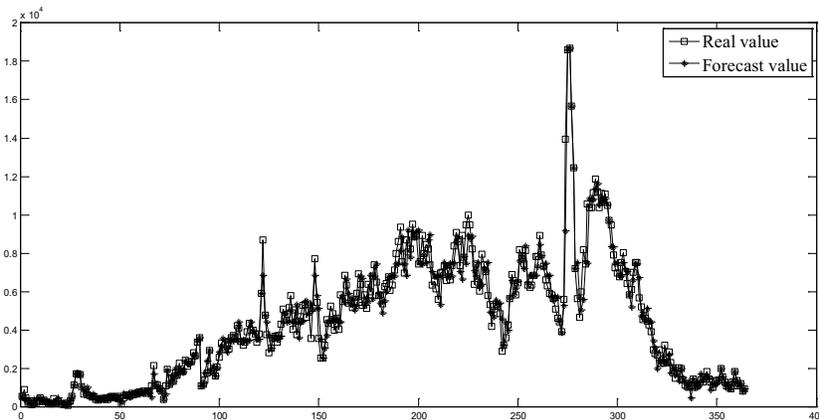


FIGURE 4. Comparison between the estimated value and the actual value of the tourism carrying capacity using the empirical modal decomposition-error backpropagation

In order to further confirm the effectiveness of the proposed method, this paper compares the empirical modal decomposition - error backpropagation artificial neural network prediction model with the single error backpropagation artificial neural network prediction model. Figure 5 is a comparison of the estimated errors of the two methods.

According to the predictions of the two models, the estimated value agrees with the actual value with respect to change trend. From Figure 4, we find that: 1, the actual annual tourism carrying capacity of Mount Emei shows the form of multi-modality; 2, the actual number of tourists and the estimated number of tourists are basically the same; 3, the comparison chart shows that the two expected values are basically correct. Similarly, for the empirical model decomposition - error backpropagation artificial neural network combination forecasting model (which uses empirical mode decomposition to process data and the error backpropagation artificial neural network for prediction), the predicted value of tourism carrying capacity and the actual value approach each other. Through calculation, we find that the corrected predictive value with neural network model is much less erroneous; and the corresponding average error rate is 8.8%, among which the values with errors

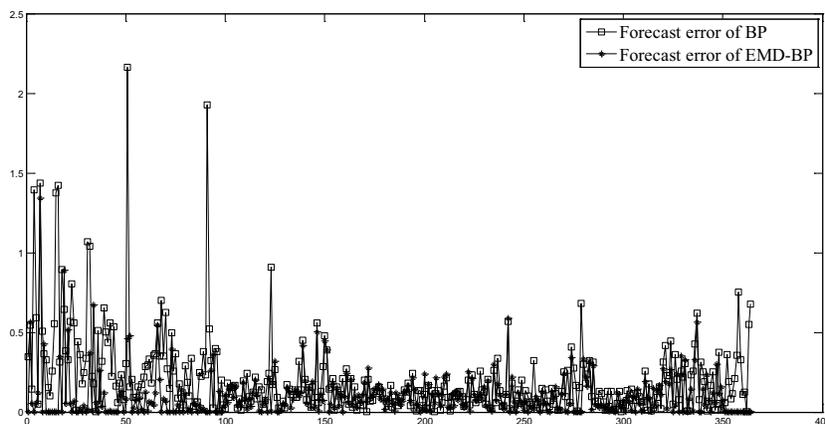


FIGURE 5. Estimation error comparison between the empirical modal decomposition - error backpropagation artificial neural network prediction model and single error backpropagation artificial neural network prediction model

less than 1% account for 31.3% of the total expected value, the values with errors less than 5% account for 41.7% of the total expected value, the values with errors higher than 10% have the percentage of 30.2% in off season and 69.8% in peak season. When the estimated value is smaller than the actual value, the large error mainly occurs in the busy season mostly as a result of more complicated factors influencing tourism in this period.

4. Conclusion. Domestic and foreign researches into tourism capacity prediction can be traced back to the 1960s. In recent years, there have been massive studies on tourism carrying capacity forecast, most of which are on an annual basis but rare ones are on a daily basis. In this paper, we construct the empirical mode decomposition and error backpropagation artificial neural network (ANN), and take the daily tourism carrying capacity of Mount Emei as the research object. Through empirical analysis, the author draws the following conclusions: compared to the single error backpropagation artificial neural, the empirical mode decomposition - error backpropagation artificial neural network combined prediction model achieves more accurate prediction results and raises convergence rate. However, there are some limitations in this paper, such as the exclusion of emergencies. In follow-up studies, the author will use textual data mining or other technologies to include the influence of emergent events on tourism carrying capacity to further improve the prediction accuracy.

Acknowledgments. Authors would appreciate the anonymous reviewers for their insightful and constructive comments and suggestions, which have been contribute to improve the manuscript and our future research. This work was supported by the National Social Science Foundation Key Project of China (No. 16AGL003), the National Natural Science Foundation of China (No. 71702118), the Soft science project of Chengdu science and technology bureau (No. 2016-RK00-00257-ZF), and the project of Systems Science and Enterprise Development Research Center (No. Xq18C01).

REFERENCES

- [1] G. E. P. Box and G. M. Jenkins, Time series analysis, forecasting and control, *San Francisco: Holden Day*, 1970.
- [2] A. Carreño, A. Vidal-Ferrández, D. Ginestar and G. Verdú, Multilevel method to compute the lambda modes of the neutron diffusion equation, *Applied Mathematics and Nonlinear Sciences*, **2** (2017), 225–236.
- [3] F. Chan, C. Lim and M. McAleer, Modelling multivariate international tourism demand and volatility, *Tourism Management*, **26** (2005), 459–471.
- [4] V. Cho, Tourism forecasting and its relationship with leading economic indicators, *Journal of Hospitality and Tourism Research*, **25** (2001), 399–420.
- [5] Y. Deng, et al, The processing of boundary problem in EMD method and Hilbert transform, *Chinese Science Bulletin*, **46** (2001), 257–263.
- [6] J. Gao, Principle of artificial neural network and simulation example, *Beijing: China Machine Press*, 2003, 1–2.
- [7] C. Goh and R. Law, Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention, *Tourism Management*, **23** (2002), 499–510.
- [8] W. Guo and J. Li, Tourism demand prediction based on improved optimized combined method, *Statistics and Decision*, **8** (2011), 75–77.
- [9] X. Guo and L. Wang, New algorithm and application of empirical mode decomposition, *Noise and Vibration Prediction*, **5** (2008), 70–71.
- [10] N. Huang, et al, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non - stationary time series analysis, *Proceedings of the Royal Society of London*, **454** (1998), 903–995.
- [11] N. Kulendran and K. Wilson, Modelling business travel, *Tourism Economics*, **6** (2000), 47–59.
- [12] N. Kulendran and S. F. & Witt, Leading indicator tourism forecasts, *Tourism Management*, **24** (2003b), 503–510.
- [13] R. Law, A neural network model to forecast Japanese demand for travel to Hong Kong, *Tourism Management*, **20** (1999), 89–97.
- [14] K. Lei and Y. Chen, Chinese inbound tourist capacity prediction based on bp neural network and arima combined model, *Tourism Tribune*, **22** (2007), 20–25.
- [15] K. Lei and Y. Chen, Prediction of chinese inbound tourist capacity based on bp neural network and ARIMA combined model, *Tourism Tribune*, **22** (2007), 20–25.
- [16] G. Li, K. F. Wong, H. Song and S. F. Witt, Tourism demand forecasting: A time varying parameter error correction model, *Journal of Travel Research*, **45** (2006), 175–185.
- [17] C. Lim and M. McAleer, Cointegration analysis of quarterly tourism demand by Hong Kong and Singapore for Australia, *Applied Economics*, **33** (2001a), 1599–1619.
- [18] X. Liu, *Improvement and Application of BP Algorithm*, Taiyuan: Taiyuan University of Technology, 2012.
- [19] H. Liu, et al, Empirical mode decomposition method and its implementation, *Computer Engineering and Applications*, **32** (2006), 44–47.
- [20] M. Rosa and M. L. Gandarias, Multiplier method and exact solutions for a density dependent reaction-diffusion equation, *Applied Mathematics and Nonlinear Sciences*, **1** (2016), 311–320.
- [21] J. Shan and K. Wilson, Causality between trade and tourism: Empirical evidence from China, *Applied Economics Letters*, **8** (2001), 279–283.
- [22] E. Smeral and M. Wuger, Does complexity matter? Methods for improving forecasting accuracy in tourism: The case of Australia D, *Journal of Travel Research*, **44** (2005), 100–110.
- [23] H. Song and G. Li, Tourism demand modelling and forecasting—A review of recent research, *Tourism Management*, **29** (2008), 203–220.
- [24] H. Song and S. F. Witt, Forecasting international tourist flows to Macau, *Tourism Management*, **27** (2006), 214–224.
- [25] H. Song, S. F. Witt and T. C. Jensen, Tourism forecasting: Accuracy of alternative econometric models, *International Journal of Forecasting*, **19** (2003), 123–141.
- [26] H. Song, K. K. F. Wong and K. K. S. Chon, Modelling and forecasting the demand for Hong Kong tourism, *International Journal of Hospitality Management*, **22** (2003), 435–451.
- [27] J. Song, et al, Simulation model verification method based on empirical mode decomposition and gray relational analysis, *Systems Engineering and Electronics*, **35** (2013), 2613–2618.
- [28] W. Tao and M. Ni, A comparative study of tourism demand prediction in china and the west: Theoretical basis and model, *Tourism Tribune*, **8** (2010), 12–16.

- [29] C. Tideswell, T. Mules and B. Faulkner, An integrative approach to tourism forecasting: A glance in the rearview mirror, *Journal of Travel Research*, **40** (2001), 162–171.
- [30] Y. Wang, F. Li and Y. Zhang, et al., Prediction of Long-term Tourist Capacity in Wulong Scenic Spot Based on Combined Model, *Mathematical Statistics and Management*, **30** (2011), 770–779.
- [31] F. Wang and W. Zheng, A study of chinese international inbound tourism demand fluctuation based on ARIMA and GARCH, *Industrial Economy*, **7** (2009), 53–57.
- [32] W. Yu and C. Mu, Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks, *Transportation Research*, 2012, 152–154.
- [33] X. Zhao, L. Wang and H. Zou, Overview of prediction methods of international tourism demand in tourist destination countries, *Tourism Tribune*, **6** (1996), 27–29.

Received September 2017; revised December 2017.

E-mail address: 603116839@qq.com

E-mail address: 2268701563@qq.com

E-mail address: jiexw@vip.163.com