Research on prediction of tourists’ quantity in Jiuzhaigou Valley scenic based on ABR@G integration model

Zhixue Liao, Maozhu Jin*, Yuyan Luo and Peiyu Ren
Business School, Sichuan University, Chengdu, 610064, China
E-mail: law1219@126.com
E-mail: jinmaozhu@scu.edu.cn
E-mail: luoyuyan_1999@yahoo.com.cn
E-mail: renpy.scu@163.com
*Corresponding author

Huafeng Gao
Business School, Sichuan University, Chengdu, 610065, China
and
Hubei University for Nationalities, Enshi, 445000, China
E-mail: gaohuaf@126.com

Abstract: As the uncertain changes of tourists’ quantity have challenged scenic management, which affects the environmental pollution, many researches confirm that forecasting, which is the foundation of the tourists’ management can provide guarantee of effective environment protection. Because the changes of tourists’ quantity with complex characteristics of the linear and non-linear are mutually integrated, prediction accuracy of a single model alone or traditional combination model using the simple linear combination method is poor. This paper proposes the AI techniques integration method (the special combination method) – GMDH which can improve forecasting accuracy of that kind of data. The ABR@G model which is applied to predict that kind of data is built through integrating the seasonal ARMA model, neural network model and revised ARIMA model with the AI techniques integration method. Finally, Jiuzhaigou Valley scenic is taken as the subject of the research to do empirical analysis which proves that the model in predicting the quantity of tourists is valid.

Keywords: ABR@G integration model; group method of data handling; GMDH; ARMA model; BP neural network; Jiuzhaigou Valley; tourists’ quantity.


Copyright © 2013 Inderscience Enterprises Ltd.
Biographical notes: Zhixue Liao received his Bachelor in Management Science at the Jiangxi University of Finance and Economics, China. He is currently a graduate student in Sichuan University. His five papers were received by academic journals, such as SCI, EI or CSSCI.

Maozhu Jin is an Instructor of Business School in Sichuan University, having presided over three ministerial and provincial projects, with one project supported by National Natural Science Foundation of China. He has published two books and over ten research papers in authoritative journals of high quality both at home and abroad, with ten of them retrieved by SCI and EI.

Yuyan Luo is a doctoral student in Sichuan University majoring in management science and engineering, and is a member of Information and Business Management Institute of Sichuan University. Her researches mainly relate to system engineering and modelling, system and information science, industrial engineering and evaluation technology, etc. In recent years, she has published about ten academic papers, some of which are retrieved by SCI, EI or CSSCI.

Peiyu Ren is a Professor and PhD Supervisor. He is currently acting as the Director of Information and Business Management Research Institute of Sichuan University. He has presided over and completed five surface projects of National Natural Science Foundation of China, being in charge of project research of Projects 863, 985 and 211, and having published 15 textbooks, monographs and more than 100 academic papers, including SCI, EI and CSSCI.

Huafen Gao is a doctoral student in Sichuan University majoring in business management, and is a member of Information and Business Management Institute of Sichuan University. His researches mainly relate to strategic management. In recent years, he has published about ten academic papers, some of which are retrieved by SCI, EI or CSSCI.

This paper is a revised and expanded version of a paper entitled ‘Research on prediction of tourists’ quantity in Jiuzhaigou Valley scenic based on ABR@G integration model’ presented at 2010 International Conference of Environment Materials and Environment Management (EMEM 2010) in Harbin, China was held on 24 July 2010.

1 Introduction

The tourism industry in China, which benefits the transportation, accommodation, catering, entertainment and retailing sectors, has bloomed in the past few decades and generated nearly US$2 trillion in economic activity (Chen, 2011; Song, 2011). However, as we all know that rapid development of the tourism industry not only makes major contributions to nation’s economic development, trade performance and prosperity, but also reveals lots of environmental destruction which may not recovery, for example, when the quantity of tourists exceeds the capacity of the scenic spot would make scenic spot work under the long-term overloaded situation which will destroy the ecological environment. In order to protect the ecological environment with the development of tourism industry, tourists’ quantity forecasting has been the research focus of tourism
managers and scholars. Accurate prediction can let scenic managers make reasonable planning for the development, scenic resource’s allocation plan, tourists tap and vehicle scheduling plan to avoid problems disused above according to forecasting results and then promote the sustainable development of scenic spot, obtain the ideal economic benefits. It is obvious that accurately forecasting tourists’ quantity in the future has become the important way to solve this problem. So the study about how to improve forecasts’ accuracy of the scenic spot, and sustainable development of scenic spot, tourist development of relevant industries can provide foundation for scenic manager’s decision.

The research on forecasting of tourist’s quantity began in the 1960s, and in recent years, scholars have done lots of efforts to improve forecasting accuracy of tourist’s quantity. Exponential model, ARIMA model and GARCH model were built to produce an accurate forecasting (Kim and Ngo, 2001; Chen, 2011; Law, 2000b; Qu and Zhang, 1996). Some scholars think that non-linear method can improve the forecasting accuracy because the data’s characteristics of the tourists’ quantity are not linear. Law (1999) built the first neural network model to forecast Japanese demand for travel to Hong Kong, showing that neural network model performed well in forecasting the tourism demand. Then neural network model had attracted lots of attention in forecasting such as Cho (2003), Law (2000a), Tugba and Casey (2005), and Palmer et al. (2006). As non-linear models like neural network model success in improving the forecasting accuracy, more and more kinds of non-linear models and revised non-linear models were introduced to forecasting tourists’ quantity, and Pai et al. (2006) used support vector machines to forecast tourist arrivals in Barbados (Chen and Wang, 2007), used genetic algorithms to improve support vector regression model showing this method can get more accurate result.

However, each single method has its own particular advantages and disadvantages. It is demonstrated that no single forecasting model can produce the best forecasts in all situations by empirical results (Makridakis, 1989). Furthermore, there is not a certain criterion used to choose certain forecasting method which can produce a best performance when a particular tourism demand forecasting task is performed. For example, Cho (2001) proposed that ARIMA model performed better than exponential model, however, (Smeral and Wuger (2005) confirmed that ARIMA model failed in performance compared with exponential model. For this reason, an important thought to combine different models is appearing which firmly believes that although individual model cannot identify the true process exactly, different models may play a complementary role in the approximation of the data generating process (Chen, 2011).

The idea of combining forecasts was firstly proposed by Bates and Granger (1969) to improve the forecasts’ accuracy. In their paper, two individual forecasts of a time series were given, and a linear combination method was enjoyed to combine the two individual forecast models showing that combined model can result in a better performance than the two original ones. The advantage of the combined forecasting model had attracted lots of scholars’ attention and further studies have shown that combined model of several models often results in a more accuracy than individual ones because that combined model can capture data characteristics from different aspects of different individual models (Makridakis, 1989; Makridakis and Winkler, 1983). In recent years, considerable efforts have been made to develop and improve the various forecast combination methods through empirical testing. (Makridakis, 1989; Palm and Zellner, 1992; Fang, 2003) shown that simple average method can generate reliable and accurate forecasting results. While different individual models cannot make the same contribution to the combined
model as simple average method did. Variance-covariance method was introduced to
determine the weights of individual models. Winkler and Makridakis (1983) compared a
simple combination method with five variants of the variance-covariance combination
method showing that variance-covariance combination method produced a better
forecasting result. After that, more sophisticated methods to determine the weights have
been developed such as regression method (Granger and Ramanathan, 1984), OLS
combination method (Chan et al., 1999), principal component method (Diebold and
Pauly, 1987), and Bayesian combination method (Min and Zellner, 1993).

The combined method discussed above is linear, but as we all know that the tourists’
quantity complex characteristics of the linear and non-linear are mutual integrated which
can be reflected by linear combined method. So the purpose of this paper is to propose a
forecasting model that can be used by the scenic spot managers to produce accurate
forecasts of the tourists’ quantity of the Jiuzhaigou Valley scenic. ABR@G integration
model is a special combined model which integrates the seasonal ARMA model, neural
network model and their traditional linear combined model with AI techniques
integration method which overcomes the defects of the linear combination method.

The plan of this paper is outlined as follows. In Section 2, the problem of forecasting
is analysed and Jiuzhaigou Valley scenic is introduced. Section 3 focuses on the
methodology, where three individual models are introduced and an ABR@G integration
model is developed. In Section 4, Jiuzhaigou Valley scenic is taken as the subject of the
research to do empirical analysis. And then the last section presents the conclusion
obtained from the compare analysis with the individual models and two common
combined models.

2 Problem analysis

2.1 Jiuzhaigou Valley

Jiuzhaigou Valley scenic located in the west of China, is a world famous natural scenic
spot. The tourism industry there has promoted economic development of the west of
China. So the tourism development of Jiuzhaigou Valley scenic relates to the local social
stability, the development of regional economy. Tourists’ quantity has the direct impact
on the local tourism economy development, so tourists’ quantity forecasting is the most
important work.

2.2 Data characteristic

The change of the tourists’ quantity shows volatility and irregularities revealed from time
series aspect (Yang et al., 2011; Liang and Bao, 2012). Yan et al. (2009) analysed the
characteristics of tourists’ quantity’s change in Jiuzhaigou Valley based on wavelet
analysis method, indicating that the rules of tourists’ quantity’s changes are different and
complex on the different time scale, and show uncertain and irregularities (its
characteristics sometimes is linear, sometimes is non-linear) in the different time range,
showing that it is difficult to determine that characteristics of tourists’ quantity’s changes
are linear or non-linear because the both characteristics exist, furthermore, it is hard to
distinguish where is linear characteristics and where is linear characteristics. An
important conclusion can be obtained that characteristics of tourists’ quantity’s change in
Jiuzhaigou Valley is volatility, complexity and irregularities, and its characteristics of linear and non-linear are interlaced, then the linear combined approach to combine the linear and non-linear forecasting models cannot capture the real inherent laws and get a accurate forecasting. Although the combined forecasting model reflects both linear and non-linear characteristics, the existing combined method is linear. So the efficient way to improve accuracy of forecasting is to bring in the AI techniques integration method to solve that tourists’ quantity’s characteristics of which linear and non-linear are interlaced.

3 Methodology

3.1 Modelling ideas

According to change characteristics of tourists’ quantity in Jiuzhaigou Valley, this paper builds ABR@G model as shown in Figure 1, where A refers to seasonal ARIMA model, B refers to the BP neural network model, R refers to the revised ARIMA model, G refers to integration technology (group method of data handling, GMDH). ABR@G model follows the thought of integration after decomposition, focusing on the advantage of non-linear combination method, reflecting the advantage of traditional statistical method, and advanced artificial intelligence technology. It pays more attention to non-linear integration method.

Figure 1 ABR@G model

Seasonal ARIMA model can predict the tourists’ quantity change characteristics of the linear and seasonal by linear way, the BP neural network model can predict the tourists’ quantity change characteristics of the non-linear by non-linear way. Revised ARIMA
Research on prediction of tourists’ quantity in Jiuzhaigou Valley scenic model can reflect the situation that the characteristic is linear while it may be affected by some non-linear factors. Due to the changes of the tourists’ quantity cannot only reflect a single characteristic (linear or non-linear) and cannot confirm how they interlaced, seasonal ARIMA model, the BP neural network model or revised ARIMA model only respectively reflect the change of linear or non-linear seasonal from one aspect, which definitely causes the great errors. Integrating single models above by the GMDH integration technology will produce forecasting results which have both characteristics of linear and non-linear caused by seasonal ARIMA model’s linear forecast, the BP neural network model’s non-linear prediction and revised ARIMA model’s revised linear forecast.

In this paper, the basic thought of the modelling is to use single models (seasonal ARIMA model, the BP neural network model and revised ARIMA model) to forecast from different aspects and then using GMDH integration technology to integrate single forecasting models to get an improved forecasting result which is better than utilising single models and traditional linear combined models.

3.2 Introduction of the main individual models

3.2.1 Seasonal ARIMA model

Seasonal ARIMA model is a transformation model of the ARIMA (Box-Jenkins) model, which is the common method used to handle time series existing cyclical changes. Firstly, use difference or seasonal difference method to eliminate the time series’ unstable of tourists’ quantity in Jiuzhaigou Valley, converting it into a steady time series, and then modelling according to the steady time series modelling method. Mathematical expression of \(ARIMA(p, d, q) \times (P, D, Q)\) as follows:

\[
\psi_p(B^d) = 1 - \psi_1 B - \psi_2 B^2 - \cdots - \psi_p B^p, \quad \theta_q(B^D) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q, \quad \Psi_p(B^s) = 1 - \Psi_1 B^s - \Psi_2 B^{2s} - \cdots - \Psi_p B^{ps}, \quad \Theta_q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_q B^{qs}.
\]

3.2.2 BP neural network model

Artificial neural network is proposed based on the modern neuroscience theory, the BP neural network is the network of multilayer feedforward its basic structure as shown in Figure 2, including three parts: input layer, and hidden unit layer and output layer.

Take Jiuzhaigou Valley as a example to describe the working principle of the BP neural network model, factors which affect the tourists’ quantity in Jiuzhaigou Valley get into the network from the input layer and then reach the hidden unit layer after translation by activation function, finally, get outputs reflecting the tourists’ quantity forecasting results in Jiuzhaigou Valley from the output layer through translation of hidden unit layer’s output by activation function.

Hidden unit layer’s output calculation: \(O_\kappa = f \left( \sum_{i=1}^n \omega_{\kappa i} X_i - \alpha_\kappa \right)\), \(\kappa = (1, 2, \ldots, \ell)\), where \(f(\cdot)\) is the activation function, \(n\) is the number of input layer nodes, \(\ell\) is the number of hidden unit layer nodes, \(\omega_{\kappa i}\) is the connection weight between the input layer node \(i\)
with the hidden unit layer node $\kappa$, $\alpha_\kappa$ is the threshold value of the hidden unit layer node $\kappa$.

**Figure 2**  BP neural network model

Output layer’s calculation: $h_j = f \left( \sum_{\kappa=1}^{\ell} \alpha_\kappa \omega_{\kappa j} - b_j \right)$, $(j = 1, 2, \ldots, m)$, $m$ is the number of output layer nodes, $\ell$ is the number of hidden unit layer nodes, $\omega_{\kappa j}$ is the connection weight between the hidden unit layer node $\kappa$ with the output layer node $j$, $b_j$ is the threshold value of the output layer node $\kappa$.

Calculating the forecasting errors $e_j = y_j - h_j$ and revising the weights and threshold value of each layers based on errors, finally, return to hidden unit layer to calculation until the errors of output layers reach the errors standard that can be accepted.

3.2.3 Revised seasonal ARIMA model

As we all know that the factors which affect tourists’ quantity are so much that seasonal ARIMA model cannot really reflect the characteristic of tourists’ quantity’s change. So this paper would use revised seasonal ARIMA model to forecast under situation that the characteristic is linear while it may be affected by some non-linear factors. The thought of this model is that using BP neural network to revise forecasting errors of seasonal ARIMA model.

3.3 Integration method

In order to adapt to the complexity characteristics, GMDH is introduced as a kind of AI techniques integration method.

GMDH is a kind of self-organising data mining method based on the external standards. The main thought of GMDH is that intersecting and combining a group of various simple initial inputs (the single forecasting models above) produce the first generation candidate models, and then produce the second generation candidate models by combining the several best models choosing from the first generation candidate models based on the certain external standards. Repeating such a generation, selection
and genetic evolution process makes model complex until the optimal complexity model is selected. The structure of GMDH is shown in Figure 3 as follows.

**Figure 3** GMDH network

![GMDH network diagram](image)

Every neuron’s input is produced by the output of last layer’s two neurons, firstly, begin with the $m$ input variables ($x_1, x_2, x_3, \ldots, x_m$), produce the first generation candidate models by non-linear fitting each pair input variables ($x_i, x_j$) ($C_m^2$ pairs in total) shown as follows:

$$ h_{ij} = \omega_0 + \omega_1 x_i + \omega_2 x_j + \omega_3 x_i x_j + \omega_4 x_i^2 + x_j^2 $$

where $q = 1, 2, \ldots, C_m^2$, $\omega_q = (\omega_0, \omega_1, \omega_2, \omega_3, \omega_4)$ is the coefficients of polynomials, $x_q = (1, x_i, x_i^2, x_i x_j, x_j^2)$ is the input variable, $h_{ij}$ is the neuron’s output of the first layer.

Calculating the neurons’ minimum bias selection criteria

$$ r_{ij} = \frac{\sum_i (y_i - h_{ij})^2}{y_i^2} $$

in the $j^{th}$ layer candidate model (where $n$ is the number of the samples, $Y$ presents the true tourist’s quantity, $k = 1, 2, 3, \ldots, \ell_j, \ell_j$ is the number of the $j^{th}$ layer’s neurons), select $R_j$ ($R_j$ is the certain number representing the number of the $j^{th}$ layer candidate model’s neurons) neurons of which the bias is minimum as the next layer candidate model’s input neurons, clear the other neurons which is not selected (shown as black circle in Figure 2), finally compare the minimum bias $\min(r_j)$ among the $r_j$ with the minimum bias $\min(r_{j-1})$ among the $r_{j-1}$, if $\min(r_j) < \min(r_{j-1})$, produce the next layer candidate models according to the first layer candidate model’ s produce, or means that GMDH model is complete and then input the test set data to the GMDH network to forecast the tourist’s quantity.

It is observed that GMDH is kind of non-linear integration method which intersects and combines a group of single forecasting model to capture the useful and valid information from several angles.
In practice, due to changes of tourists’ quantity showing the characteristics of volatility and irregularities in Jiuzhaigou Valley, integrate the single forecasting model by the GMDH integration methods and improve the forecasting accuracy comparing the individual ones by overcoming defects of single model or linear combined model because of the integration which gathers effective information of each single forecasting model.

### 3.4 Modelling steps

According to the analysis above, modelling steps can be summarised as shown below:

1. Build seasonal ARIMA model to predict the tourists’ quantity in Jiuzhaigou Valley, and get the results $\hat{y}_1$ which reflect the changes characteristics of the linear and seasonal.
2. Build the BP neural network model to predict the tourists’ quantity in Jiuzhaigou Valley and get the results $\hat{y}_2$ which reflect the changes characteristics of the non-linear.
3. Build the revised seasonal ARIMA model to predict the tourists’ quantity in Jiuzhaigou Valley and get the results $(\hat{y}_3 = \hat{y}_1 + \hat{e}_1)$ which reflect the changes characteristics of the linear revised by the non-linear.
4. Take the forecasting results $(\hat{y}_1, \hat{y}_2, \hat{y}_3)$ as the input data of ABR@G integration model to predict the tourists’ quantity in Jiuzhaigou Valley and get the results which reflect the changes characteristics of the linear and non-linear from the different aspects.

### 4 Empirical analysis

#### 4.1 Data and performance criteria

This article takes Jiuzhaigou Valley scenic spot as the research object, and collects the monthly data shown in Table 1 (from Jiuzhaigou Administration) of tourists’ quantity from January of 2001 to December of 2007 in Jiuzhaigou Valley. Take tourists’ quantity data collected from January of 2001 to December of 2006 as the training data to estimate the models’ parameters, and then take tourists’ quantity data collected from January to December in the year 2007 as the test data to test the forecasting accuracy of the model.

In order to evaluate the effective model, this paper sets the mean absolute percentage error as evaluation statistics,

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

where $\hat{y}_i$ presents forecasting result of the month $i$, $\hat{y}_i$ presents the truth tourists’ quantity of the month $i$, $N$ presents the number of months which is forecasted.
Table 1  Tourists’ quantity data from January of 2001 to December of 2007 in Jiuzhaigou Valley

<table>
<thead>
<tr>
<th>Year/month</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>1,020</td>
<td>4,108</td>
<td>2,678</td>
<td>18,021</td>
<td>6,010</td>
<td>12,843</td>
<td>15,973</td>
</tr>
<tr>
<td>February</td>
<td>4,679</td>
<td>10,074</td>
<td>13,806</td>
<td>16,055</td>
<td>15,488</td>
<td>21,808</td>
<td>26,954</td>
</tr>
<tr>
<td>March</td>
<td>24,911</td>
<td>24,124</td>
<td>28,702</td>
<td>52,493</td>
<td>37,050</td>
<td>70,867</td>
<td>69,691</td>
</tr>
<tr>
<td>April</td>
<td>69,511</td>
<td>85,702</td>
<td>58,069</td>
<td>134,833</td>
<td>108,419</td>
<td>144,926</td>
<td>155,033</td>
</tr>
<tr>
<td>May</td>
<td>154,171</td>
<td>146,275</td>
<td>4,576</td>
<td>202,199</td>
<td>207,710</td>
<td>195,792</td>
<td>252,742</td>
</tr>
<tr>
<td>June</td>
<td>94,957</td>
<td>84,114</td>
<td>20,852</td>
<td>181,652</td>
<td>108,334</td>
<td>187,964</td>
<td>263,726</td>
</tr>
<tr>
<td>July</td>
<td>225,009</td>
<td>1,899,884</td>
<td>185,087</td>
<td>353,990</td>
<td>395,649</td>
<td>438,358</td>
<td>430,469</td>
</tr>
<tr>
<td>August</td>
<td>238,067</td>
<td>228,232</td>
<td>259,246</td>
<td>344,705</td>
<td>281,600</td>
<td>442,441</td>
<td>480,902</td>
</tr>
<tr>
<td>September</td>
<td>161,712</td>
<td>179,703</td>
<td>219,223</td>
<td>296,339</td>
<td>310,319</td>
<td>380,567</td>
<td>381,816</td>
</tr>
<tr>
<td>October</td>
<td>180,393</td>
<td>231,054</td>
<td>292,916</td>
<td>337,507</td>
<td>381,896</td>
<td>413,386</td>
<td>447,165</td>
</tr>
<tr>
<td>November</td>
<td>38,115</td>
<td>36,309</td>
<td>86,448</td>
<td>82,688</td>
<td>101,523</td>
<td>124,440</td>
<td>131,965</td>
</tr>
<tr>
<td>December</td>
<td>7,499</td>
<td>8,534</td>
<td>16,217</td>
<td>16,118</td>
<td>24,197</td>
<td>31,700</td>
<td>41,942</td>
</tr>
</tbody>
</table>

4.2 Forecasting the tourists’ quantity in Jiuzhaigou Valley

4.2.1 Input data of ABR@G integration model

According to the analysis above, before forecasting the tourists’ quantity in Jiuzhaigou Valley, the input data of ABR@G integration model have to be obtained, namely the forecasting results \( \hat{y}_1, \hat{y}_2, \hat{y}_3 \) of seasonal ARIMA model, BP neural network model and revised ARIMA model. So seasonal ARIMA model, BP neural network model and revised ARIMA model should be built respectively to forecast tourists’ quantity in Jiuzhaigou Valley, and then forecasting errors would be calculated to compare with forecasting errors of ABR@G integration model according to equation (3).

1 Seasonal ARIMA model

The tourists’ quantity data in Jiuzhaigou Valley from January of 2001 to December of 2006 is input to the SPSS17.0 to be analysed, and then get the result as shown in Figure 4. The result shows that the time series is not stable, furthermore, there is obvious seasonal and upward trend. So before building seasonal ARIMA model, the time series should be treated. Figure 5 is the result of original time series which is treated by difference and seasonal difference expressing that the treated time series is stable (which fluctuates around zero).

Build seasonal ARIMA model after differential treatment, and then estimate the parameters of the seasonal ARIMA model based on the tourists’ quantity data from January of 2001 to December of 2006, finally, use the model to forecast the tourists’ quantity from January to December in the year 2007 to get the forecasting results \( \hat{y}_1 \).
and the forecasting errors according to the equation (3), the results are shown in Figure 6 and Table 2, respectively.

2 BP neural network model

This paper takes tourists’ quantity monthly data from 2001 to 2005 as the network’s input sample, and takes monthly data of 2006 as the network’s output sample. So set the neural network’s input layer units’ number for 5, output layer units’ number for 1, and hidden layer units’ number for 3 which is calculated. Train the BP neural network based on the tourists’ quantity data from 2001 to 2006, and use the trained BP neural network model to forecast the tourists’ quantity of year 2007 to get the forecasting result $\hat{y}_2$ and the forecasting errors according to the equation (3), the results are shown in Figure 7 and Table 2 respectively.
Research on prediction of tourists’ quantity in Jiuzhaigou Valley scenic

Figure 6  Forecasting results of Seasonal ARIMA model (see online version for colours)

![Figure 6](image)

Figure 7  Forecasting results of the BP neural network model (see online version for colours)

![Figure 7](image)

Table 2  Parameter value

<table>
<thead>
<tr>
<th>Names of models</th>
<th>ARIMA model</th>
<th>BP model</th>
<th>Revised model of ARIMA</th>
<th>Integrated model of ABR@G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction errors</td>
<td>0.1337</td>
<td>0.1183</td>
<td>0.1084</td>
<td>0.0321</td>
</tr>
</tbody>
</table>
3 Revised ARIMA model

This paper uses the errors of the last five months as the input of the network to revise the sixth month forecasting error of ARIMA model, namely taking the last five months’ forecasting errors as the input data sample, and the next month forecasting error is taken as the ideal output data sample (such as forecasting errors from January to May in the year 2006 as a network’s input training sample and predicting errors of June in the year 2006, as an ideal output training sample). So set the neural network’s input layer units’ number for 5, output layer units’ number for 1, and hidden layer units’ number for 3 which is calculated. Train the BP neural network based on the tourists’ quantity forecasting errors of ARIMA model from 2001 to 2006, and then use the trained BP neural network model to forecast the tourists’ quantity forecasting errors of ARIMA model of the year 2007 to get the forecasting errors $e_i$, then get the forecasting results of revised ARIMA model $\hat{y}_3 = \hat{y}_1 + e_i$ and the forecasting errors according to the equation (3), the results are shown in Figure 8 and Table 2, respectively.

Figure 8  Forecasting results of revised ARIMA model (see online version for colours)

4.2.2 ABR@G integration model’s forecasting

In this paper, GMDH is taken as the integration technology to integrate seasonal ARIMA model, BP neural network, and revised ARIMA model above. Take 12 months’ forecasting data of seasonal ARIMA model, BP neural network, and revised ARIMA model in the year 2006 as GMDH network’s input sample and take the actual value of tourists’ quantity in the year 2006 as the network’s ideal output, which forms the GMDH network’s training sample to train the network. Finally take 12 months’ forecasting data of seasonal ARIMA model, BP neural network, and revised ARIMA model in the year 2007 as GMDH network’s input to forecasting the tourists’ quantity in the year 2007, get
the forecasting results \( \hat{Y} \) and forecasting errors shown in Figure 9 and Table 2 respectively.

**Figure 9** Forecasting results of ABR@G integration model (see online version for colours)

4.3 **Compare analysis**

4.3.1 **Compare with the individual models**

Three single models’ and integrated model’s errors are shown in the Table 2, we can find that integrated model performs better than any other model, where revised model is better than other two single models and non-linear model (BP model) performs better than linear model (ARIMA model).

4.3.2 **Compare with other combined models**

Among the combined forecasting models, average combined model and optimal weights combined model are two of the most common models. While the similarity of those models is the combination method which is linear, this paper also uses average combined method and optimal weights combined method to combine the three forecasting models to show whether AI techniques integration method (GMDH) performs better than others or not.

**Table 3** Parameter value

<table>
<thead>
<tr>
<th>Names of models</th>
<th>Average combined model</th>
<th>Optimal weights combined model</th>
<th>Integrated model of ABR@G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction errors</td>
<td>0.0632</td>
<td>0.0573</td>
<td>0.0321</td>
</tr>
</tbody>
</table>
Three kinds of combined models forecasting errors are shown in Table 3, the conclusion can be obtained that integrated model of ABR@G performs better than other two models.

5 Conclusions

Based on the analysis of changes characteristic of tourists’ quantity in Jiuzhaigou Valley, the paper builds ABR@G integration model, and take Jiuzhaigou Valley as the research object to do empirical analysis from which, the following conclusions are obtained: firstly, single models’ (linear, non-linear, or revised model) performance is not ideal; secondly, linear combined model can improve the forecasting performance, but it is not the optimal forecasting model; thirdly, AI techniques integration method can effectively reduce the prediction errors and can be used as the effective forecasting tool. This paper also has some disadvantages, for examples, the influence of emergency such as road landslide has not be considered. So the future research will use text mining technology to consider the effects of emergency, which will improve forecasting performance further.

Acknowledgements

This work was supported by the Major International Joint Research Programme of the National Natural Science Foundation of China (Grant no. 71020107027), the National High Technology Research and Development Major Programme of China (863 Programme) (Grant no. 2008AA04A107), the National Natural Science Foundation of China (Grant no. 71001075), the China Postdoctoral Science Foundation (Grant no. 2012M521704), Doctoral Fund of Ministry of Education of China (20110181110034), 985, 211 projects of Sichuan University and Central University Fund of Sichuan University.

References


