Application of GA-BP neural network model for small watershed flood forecasting in Chun'an county, China

D J Huang1,2,5, C C Tian3 and J Y Jiang4

1Key Laboratory for Technology in Rural Water Management of Zhejiang Province, Hangzhou 310018, China
2College of Water Resources and Environmental Engineering, Zhejiang University of Water Resources and Electric Power, Hangzhou 310018, China
3Zhejiang Design Institute of Water Conservancy and Hydroelectric Power, Hangzhou 310002, China
4Shandong Survey and Design Institute of Water Conservancy, Jinan 250013, China
5E-mail: huangdj@zjweu.edu.cn

Abstract: Flood forecasting for small basins in hilly areas is often plagued by poor performance of hydrological models due to lack of observed data, meanwhile, the traditional Back Propagation (BP) neural network is easy to fall into the local minimum. This paper put forward an approach combined Genetic Algorithms (GA) with BP neural network and established a GA-BP neural network model to promote the flood forecasting. The flood hygrograph of Fenglingang small watershed, in Chun'an county, simulated by GA-BP model indicates that the deviation of runoff volumes is controlled within 10%, the deviation of peak discharge is kept below 20%, and absolute error of time to peak is less than 2h. Additionally, the correlation coefficient of simulation result of GA-BP model for each rainstorm event is above 0.75, which is smaller than that of traditional BP model. Consequently, it is demonstrated that the GA-BP model has a higher simulation precision and can provide reference for local forecasting in the future.

Key words: Genetic Algorithm; BP Neural Network; Flood Forecasting; Hydrologic model

1. Introduction

In hilly areas of China, there are a huge number of small watersheds with well-developed water system, where flood disaster happens frequently and is difficult to forecast because of the regional rainstorm typically with the features of short duration, great intensity and sudden breakout, especially during plum rain season [1]. What is more, it can often trigger a series of geological disasters such as debris flow and landslide that are very likely to devastate the ecological environment. Therefore, to strive to make more accurate flood predictions is of great significance for small watersheds.

There have been lots of in-depth studies on small watershed flood forecasting at home and abroad, such as the application study of XAJ model in flood forecasting in mountainous areas [2], the research of flood forecasting of upper Lijiang river based on TOPMODEL model [3], the study of hydrological response of small watershed in north China based on the distributed hydrological model MIKESHE [4], and the application of HBV in a flash flood case in Slovenia [5], etc. As we can see the hydrological models have been widely used in flood forecasting. However, due to the difficulties in survey stations construction in mountainous area, topographic data and long series of measured hydrological data are

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd.
normally not available for hydrological modeling, which can account for the poor accuracy of flood forecasting. How to adopt the BP neural network, which is able to describe the complex relationship between the inputs and outputs of the system as long as the relevant data can be obtained, to predict floods and simulate rainfall-runoff process has been a prominent topic of research in recent years [6]. Considering the conventional BP algorithm is easy to be trapped in slow convergence and local minimum, a number of studies has attempted to apply Genetic Algorithm (GA) to make up for the disadvantages, which can keep searching the whole solution space to get the global optimal solution [7]. Consequently, this paper built up a GA-BP neural network flood forecasting model integrated the GA with BP neural network, and then applied it to the Fenglingang small watershed, in Chun'an county, Hangzhou city, Zhejiang province, aimed at achieving a more efficient and accurate flood forecasting.

2. Study area
Fenglingang small watershed is located in Fengshuling Town, in the southwest of Chun'an county, hilly areas of western Zhejiang province, China. It belongs to the Xin'an river system attached to the Qiantang river basin, with a drainage area of about 227 km². The rainfall in Chun'an county displays an uneven characteristic both in time and space, according to the statistics from the precipitation stations, and it is the Fengshuling town in which the study catchment lies has the highest mean annual precipitation, reaching 1907 mm [8].

The river system in the study area composed of dozens of rivers is well-developed, mainly including Fenglingang river and Baimaxi river, etc. Besides, there are a few precipitation stations in and around the study area, such as Dayuan, Tongsan, Baima, Yanchang, Fengshuling station, etc., the observed data of which is the basic essential for our research. Figure 1 shows the study area with the river system and the distribution of precipitation stations.

3. Methods

3.1. BP neural network
BP neural network is one of the most typical feedforward networks, which will adjust the weight matrix and threshold value through the forward transfer of network structure and reverse correction of training function until the error between the predicted result and the actual one is smaller than allowed [9]. In addition, the neurons in adjacent layers are connected with each other, but there is no any connection between the neurons in the same layer. Figure 2 shows the structure of BP neural network.
Here are how the BP neural network trains the network:

1. Normalize the sample data.
2. Determine the number of nodes \((n, l, m)\), connection weights \((w_{ij}, w_{jk})\), threshold values \((a, b)\) and excitation functions of each layer.
3. Calculate the output value of each layer: in hidden layer, where \(H_j = f\left(\sum_{i=1}^{n} w_{ij}X_i - a_i\right)\); in the output layer, where \(O_k = f\left(\sum_{j=1}^{l} w_{jk}H_j - b_k\right)\).
4. Compute the output error: \(E_k = Y_k - O_k\).
5. Adjust the weights and thresholds according to the error precision and the number of iterations until the generated network can satisfy the request, and then finish running.

3.2. Genetic Algorithms

Genetic Algorithms (GA) is an adaptive and probabilistic global searching method [10], based on Darwinian evolution and the principle of "survival of the fittest" [11]. In GA, the initial population is a random sampling, each member of which is a data set and represents a candidate solution. Through fitness function, the better members of the initial population can survive and transmit to the next generation following some certain rules. Then, the next generation generates new members through crossover and mutation, and so repeatedly, the optimal solution will eventually emerge.

Afterwards, this study proposed a method coupled the GA with BP neural network, which could absorb the merits of the both. That GA provided a solution space enabled BP neural network to find out the optimal solution, which help to avoid a local minimum or maximum.

4. Calculation and analysis

This paper set up a neural network model combined GA and BP (called GA-BP model) for small watershed flood forecasting, then applied it to Fenglingang basin in Chun'an county, Hangzhou city, and aimed to improve the local flood prediction accuracy.

4.1. Select the sample data

Sample data includes input part and output part, but small watersheds in mountainous areas are usually lack of monitoring data in practice, such as rainstorm, underlying surface, temperature, evaporation, etc., which are all the most important factors that can affect the formation of flood. And just as expected there were only rain stations within the study area, so we got the series rainfall data from 2011 to 2016 from the 5 rain stations and employed Thiessen Polygons method to calculate the weighted average rainfall, that could be regarded as the series mean areal rainfall of Fenglingang basin. Then, 11 rainstorm events were selected as the input data. Moreover, we obtained the outflow series of Fengshuling reservoir from 2011 to 2016, as well as the measured water level, so the inflow series were figured out through reversely push. In result, the corresponding flows of the 11 rainstorms were selected as our output data.
Finally, there were 11 groups of sample data in total, and each of them included 50 input data and 50 output data, and the time period was 1h (Table 1).

| Number | Flood No. | Peak discharge(m³·s⁻¹) | Max rainfall depth(mm) |
|--------|-----------|-------------------------|------------------------|
| 1      | 20110614  | 680.83                  | 22.76                  |
| 2      | 20110618  | 456.00                  | 16.03                  |
| 3      | 20120424  | 294.00                  | 17.06                  |
| 4      | 20120810  | 353.00                  | 11.12                  |
| 5      | 20130607  | 159.00                  | 10.28                  |
| 6      | 20140514  | 158.00                  | 19.45                  |
| 7      | 20140620  | 288.00                  | 24.27                  |
| 8      | 20140724  | 197.00                  | 11.17                  |
| 9      | 20150608  | 472.00                  | 13.78                  |
| 10     | 20150701  | 329.00                  | 41.83                  |
| 11     | 20160628  | 447.35                  | 28.85                  |

4.2. Normalize the sample data
In order to improve the convergence rate and precision of the model, it is quite essential to normalize the sample data by Formula 1 before training the neural network, that is able to simplify the calculation through converting a dimensional expression to a dimensionless one. Ultimately, the data is mapped to a range of 0 to 1 and processed as a scalar.

\[
\bar{X}_i = \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{1}
\]

Where \( \bar{X}_i \) is the normalized value of the i'th factor; \( X_i \) is the actual value of the i'th factor; \( X_{\text{max}} \) is the maximum value of the i'th factor; \( X_{\text{min}} \) is the minimum value of the i'th factor.

4.3. Set the parameters
After that, we set the parameters of both the GA and BP neural network as follows: in GA, population size was 100, hereditary algebra was 80, crossover probability was 0.7, and mutation probability was 0.01; in BP neural network, the max training times was 1000, target error was 0.0001, learning rate ranged from 0.01 to 0.3, and exciting function of hidden layer was “sigmoid”.

4.4. Establish network structure
Among the 11 groups of sample data, 7 were randomly selected for network training, 2 groups for verification, and remaining 2 for testing. According to the sample data, it was obvious that there were 50 nodes both in input layer and the output layer, and only the number of nodes in the hidden layer was needed to determine through cut and try method by the empirical Formula 2. Running the code in MATLAB, the model became capable of optimizing and adjusting the weight of each layer and threshold of each neuron after training the 7 groups of sample data. Then, the verification of the 2 groups was carried out, and the number of neurons in hidden layer was finally determined to be 21. Hence, the three-layer network structure that would be adopted in Fenglingang basin was 50-21-50.

\[
l < n - 1 \quad \text{or} \quad l < \sqrt{n + m} + a, \quad a = 1 - 10 \tag{2}
\]

Where \( l \) is the number of hidden layer nodes; \( n \) is the number of units in the input layer; \( m \) is the number of output nodes; \( a \) is constant.

4.5. Results of network tests
We utilized the remaining two rainstorms (20110618 and 20140724) to simulate the flood processes in conventional BP model and GA-BP model respectively. The simulation results were shown in Table 2 and Table 3. Figure 3 illustrated the comparison of the simulation results of the 2 models and measured value.

![Comparison of the simulation results of conventional BP model and GA-BP model with observed value.](image)

**Figure 3.** Comparison of the simulation results of conventional BP model and GA-BP model with observed value.

It can be seen from Figure 3 that the simulated flood hydrograph of GA-BP model was more consistent with the measured process compared to that of conventional BP model from overall trend. Furthermore, three evaluation criteria including the deviation of runoff volumes, the deviation of peak discharge, as well as absolute error of time to peak were selected to assess the accuracy of flood forecasting in accordance with the related provisions in "forecasting norm for hydrology intelligence" (GB/T 22482-2008). (Hereinafter referred to as the national standard)

| Period | 20110618 | BP model | GA-BP model |
|--------|----------|----------|-------------|
|        | Observed value (m³.s⁻¹) | Simulated value (m³.s⁻¹) | Deviation (m³.s⁻¹) | Simulated value (m³.s⁻¹) | Deviation (m³.s⁻¹) |
| Max    | 456.0    | 298.5    | -157.5      | 388.0      | 545.5      |
| Sum    | 7429.2   | 5858.1   | -1571.1     | 7934.8     | 505.6      |
| The deviation of runoff volumes /% | -21.1 | 6.8 |
| The deviation of peak discharge /% | -34.5 | -14.9 |
| Absolute error of time to peak /h | -3 | -2 |
IOP Conf. Series: Earth and Environmental Science 612 (2020) 012066
doi:10.1088/1755-1315/612/1/012066

Table 3. Comparison of simulation results of conventional BP model and GA-BP model for 20140724 rainstorm event.

| Period | Observed value (m$^3$·s$^{-1}$) | Simulated value (m$^3$·s$^{-1}$) | Deviation (m$^3$·s$^{-1}$) | Simulated value (m$^3$·s$^{-1}$) | Deviation (m$^3$·s$^{-1}$) |
|--------|---------------------------------|---------------------------------|---------------------------|---------------------------------|---------------------------|
| Max    | 197.0                           | 258.5                           | 61.5                      | 231.0                           | 169.5                     |
| Sum    | 5217.7                          | 5367.9                          | 150.2                     | 5134.7                          | -83.0                     |
| The deviation of runoff volumes /% | 2.9                             | -1.6                            |                           |                                |                           |
| The deviation of peak discharge /% | 31.2                            | 18.7                            |                           |                                |                           |
| Absolute error of time to peak /h | 6                               | -2                              |                           |                                |                           |

The evaluation results of the three indicators given in Table 2 and Table 3 indicated that: ① The deviations of runoff volumes of GA-BP model were both under 10%, that was 6.8% in 20110618 rainstorm and -1.6% in 20140724 rainstorm, far less than the permissible error 20% in national standard. Meanwhile, the deviations of traditional BP model fluctuated greatly, that was 2.9% in 20140724 rainstorm but reached up to -21.1% in 20110618 rainstorm, exceeding the requirements of national standard; ② The deviations of peak discharge of GA-BP model were -14.9% in 20110618 rainstorm and 18.7% in 20140724 rainstorm, both controlled within the acceptable error requirement of 20% in national standard. However, in traditional BP model, they were -34.5% in 20110618 rainstorm and 31.2% in 20140724 rainstorm, which were seriously out of allowable rate; ③ The requirement of absolute error of time to peak was less than 3h in national standard for time to peak, while the simulation results of GA-BP model for two rainstorms both only appeared 2h in advance. Besides, it was clear that the traditional BP model was not satisfying, because the simulated peak of 20110618 rainstorm reached 3h early, but 20140724 rainstorm delayed 6h.

To arrive at a more reasonable conclusion, some other indexes, such as root mean square error and correlation coefficient, were adopted to assess the applicability and feasibility of the model. A smaller root mean square error indicates a better performance and a higher accuracy of the model, as well as a lower dispersion degree between the predicted value and the simulated one. For correlation coefficient, if the value is close to 1, it reveals a strong correlation between the predicted value and the simulated one; conversely, if close to 0, a weak correlation.

Table 4. Comparison of evaluation indexes of simulation results.

| Flood No. | Runoff volume (m$^3$·s$^{-1}$) | Root mean square error | Correlation coefficient |
|-----------|-------------------------------|-----------------------|------------------------|
|           | BP model | GA-BP model | BP model | GA-BP model |
| 20110618  | 7429.2   | 83.68       | 71.00     | 0.67        | 0.75        |
| 20140724  | 5217.7   | 57.10       | 28.10     | 0.67        | 0.90        |

From Table 4, the findings can be summarized as follows: ① In a single model, the root mean square error increased with the rise of runoff volume; For the same rainstorm event, the simulation errors of GA-BP model were apparently smaller than that of traditional BP model; ② For correlation coefficient, no matter which rainstorm it was, the simulation error of GA-BP model was above 0.75, which pointed out a stronger correlation than BP model. Generally, it became evident that the GA-BP model has made a great advance in the simulation accuracy.
5. Conclusion

This paper aimed to promote the accuracy of flood forecasting in Fenglingang small watershed in Chun'an county, so set up a GA-BP neural network flood prediction model combined GA with BP neural network. We employed random seven rainstorm events to train both the conventional BP neural network and GA-BP neural network, and two rainstorm events to verify, and another two to test, then selected the deviation of runoff volumes, the deviation of peak discharge, absolute error of time to peak, root mean square error and correlation coefficient as the evaluation indexes to compare the simulation results of the two models. It was indicated that the GA-BP model showed a better consistency between the simulated and measured flood hydrographs, which definitely represented a higher prediction accuracy. Owing to the good performance of GA-BP model in the study area, it is believed that it can provide scientific reference for flood control in Chun'an county and make a great contribution to reduce the damage caused by rainstorm in a sense in the future.

Acknowledgments

The paper is supported by Open Fond of Key Laboratory for Technology in Rural Water Management of Zhejiang Province (Grant No. ZJWEU-RWM-20200202B).

References

[1] Li X S 2017 Research and application of flood prediction method for small watershed Agricultural Science-Technology and Information 24 35-6
[2] Li P Y 2017 The application study of the XAJ model in flood forecasting at hilly areas (Zhengshou: North China University of Water Resources and Electric Power)
[3] Wei C J 2015 The flood forecasting research of Lijiang river upstream based on TOPMODEL (Guangxi: Guangxi University)
[4] Xiao J Q 2006 The application of distributed hydrologic model MIKESHE to simulate hydrologic responses of small watershed in north China (Beijing: Beijing Forestry University)
[5] Grillakis M G, Tsanis I K and Koutroulis A G 2010 Application of the HBV hydrological model in a flash flood case in Slovenia Natural Hazards Earth System Sciences 10 2713–25
[6] Hu J W, Zhou Y L and Jin J L 2015 Flood Forecasting Model on BP Neural Networks and Its Application in Flood Forecasting Systems J. China Hydrol. 35 20-5
[7] Li S X 2004 Studies on application of combination of Genetic Algorithm and artificial neural network in city water environment ultimate carrying capacity (Ningxia: Ningxia University)
[8] Feng H J 2016 Rainfall-triggered landslide development regularity analysis and Hazard assessment in Chun'an, west Zhejiang (Beijing: China University of Geosciences)
[9] Liu K, Guo W Y, Shen X L and Tan Z F 2012 Research on the Forecast Model of Electricity Power Industry Loan Based on GA-BP Neural Network Energy Procedia 14 1918-24
[10] Zhang S 2016 Research on methods for single-objective optimization of hydrological model and multi-objective decision making of multi-reservoir operation (Wuhan: Huazhong University of Science and Technology)
[11] Jennifer A, Rian B and Jesson Hu 2019 Genetic algorithm for nuclear data evaluation applied to subcritical neutron multiplication inference benchmark experiments Annals of Nuclear Energy 133 853-62