DAILY STREAMFLOW FORECASTING BY ARTIFICIAL NEURAL NETWORK IN A LARGE-SCALE BASIN

Jingwen XU[1,2,3], *Xuemei ZHU*[1], *Wanchang ZHANG*[2], *Xiaoxun XU*[1], *Junren XIAN*[1]

1 College of Resources and Environment, Sichuan Agricultural University, Yaan 625014, P.R. China
2 Key Laboratory of Regional Climate-Environment Research for Temperature East Asia, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, P.R. China
3 Graduate School of the Chinese Academy of Sciences, Beijing 100039, P.R. China
Corresponding to： x.j.w@163.com

ABSTRACT

Artificial Neural Network (ANN) models have been successfully applied to daily stream flow forecasting in many basins. However, most of them are designed for small or meso-scale basins rather than large-scale basins. One of aims in this work is to develop an ANN model with an optimized combination of input variables and a more accurate architecture for daily stream flow forecasting. Another aim is to compare the performance of ANN models and a rainfall-runoff model—XXT, which is a new efficient hybrid model of Xinanjiang model and TOPMODEL, in one day in advance stream flow forecasting. Yingluoxia basin, with a drainage area of 10009 km^2, is chosen as a large-scale basin. The results show that the stream flow, precipitation and evaporation are all necessary to ANN modeling for this basin. The ANN model with an appropriate combination of stream flow, precipitation and evaporation as input vector performs much better than XXT in terms of Nash-Sutcliffe efficiency. Even if only using antecedent stream flow data as inputs ANN models are still better than XXT model for one day in advance flow forecasting.

Index Terms—ANN, XXT, Yingluoxia basin, stream flow, forecast

1. INTRODUCTION

Stream flow, an integrated process of atmospheric and topographic processes, is of prime importance to water resources planning [1]. Accurate prediction of stream flow is therefore of utmost importance. However, accurate predictions of rainfall-runoff and consequent stream flows from a large scale watershed are extremely difficult because of tremendous spatial and temporal variability of watershed characteristics and weather patterns and an incomplete understanding of complex underlying physical processes [2].

ANN models are particularly useful in real world applications involving complex, poorly understood systems with noisy observational data, where traditional mechanistic methods often perform poorly. Hence they are increasingly being applied in water resource management and planning including the modeling of monthly, daily and hourly rainfall-runoff process [2, 3]. However, the existent ANN-based stream flow forecasting models mostly aim to small or meso-scale basins rather than large-scale basins.

Process-based rainfall-runoff models such as Xinanjiang model and TOPMODEL are often used for daily stream flow forecasting[4, 5]. Few studies made comparisons between process-based models and ANN models. A. Sezin Tokar and Momcilo Markus made a comparison between the ANNs and Conceptual Models and the results indicate that ANN models can be powerful tools in modeling the precipitation-runoff process for various time scales, topography, and climate patterns [6]. Nevertheless, up to now comparisons between ANNs and process-based models are made in small or meso-scale basins. The aims of this paper therefore are to develop ANN-based daily stream flow forecasting models for a large-scale basin and made a comparison between the ANN and a new hydrological model XXT, which is based on popular TOPMODEL and Xinanjiang model.

2. STUDY AREA AND DATA

Yingluoxia watershed (10009km2), situated in the arid, semi-arid region of northwestern China (see Fig.1), in the upper stream of the Heihe River Basin, is selected as a large-scale basin for this study. It is a typical high altitude, cold forest-meadow ecological landscape. Daily precipitation and pan evaporation from Zhamashike, Qilian, Sunan, and Yinglouxia meteorological stations (their geographic position as showed in Fig.1b), and daily stream flow data observed at the Yinglouxia gauging station are used as input data of the ANN models and XXT. These data were recorded continuously for many years. In the present work 11 years data from 1991 to 2000 were used, out of which 5 years data (1991–1995) are used for calibration or training and 6 years data (1996–2000) are considered as the

978-1-4244-5076-3/09/$26.00 ©2009 IEEE

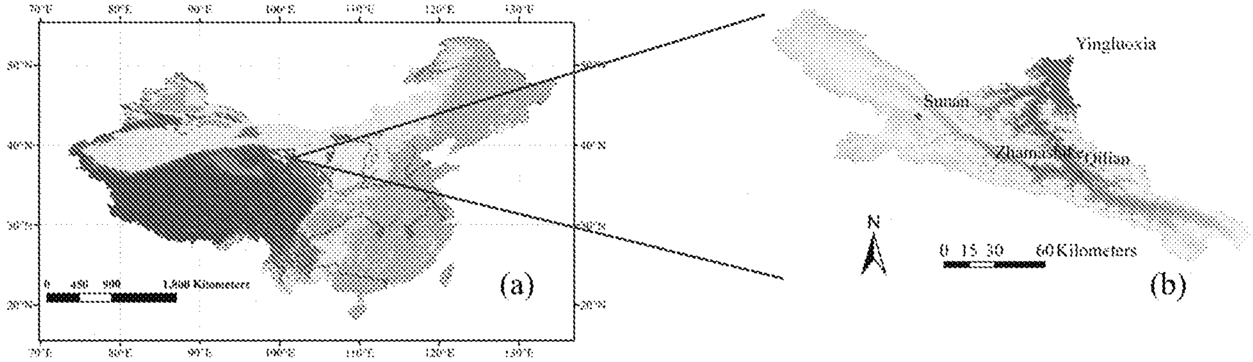

Fig.1. (a) Location of Heihe River basin in China. (b) Locations of the meteorological and hydrological stations in the Yingluoxia basin.

validation or testing set.

3. METHODOLOGY

3.1. Artificial neural networks

The ANN is an information processing system for emulating neuron information transfer in human systems. A typical feed-forward ANN has three main layers: input layer, which is used to present data to the network; hidden layer, which is used to act as a collection of feature detectors; and output layer, which is used to produce an appropriate response to the given input. Each layer may have multiple units interconnected completely with the adjacent layer and an adjusted weight is attached to each link in the system (Fig. 2). The three layer feed forward network, trained with a back-propagation learning algorithm have been found to provide the best performance with regard to input–output function approximation, such as forecasting, and therefore is used in this work.

3.2. A process-based hydrological model: XXT

Based on the soil moisture storage capacity distribution curve (SMSCC), the highlight of the Xinanjiang model, together with the simple model structure of TOPMODEL, Jingwen XU and Wanchang Zhang developed a new rainfall-runoff model named XXT. In XXT model, the water-table is incorporated into SMSCC and it connects the surface runoff production with base flow production, as can not only describe dynamically varying saturated contributing areas, but also physically depict underground water level.

3.3. Model Performance Criterion

The performance of XXT and ANN models in one day in advance stream flow forecasting is evaluated by the most commonly used criterion in assessing hydrologic modeling performance, the Nash-Sutcliffe efficiency coefficient (NE) [7], as expressed:

$$\mathrm{NE} = 1 - \frac{\sum_{i=1}^{n}\left(Q_{obs,i} - Q_{sim,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{obs,i} - \overline{Q}_{obs}\right)^2} \qquad (1)$$

where n is the total length of data; $Q_{obs,i}$ and $Q_{sim,i}$ denote the observed and simulated value at the ith time step , and $\overline{Q}_{obs}$ is the observed mean value. The higher NE is, the better a model performs.

4. RESULTS AND DISCUSSION

4.1. Artificial neural networks forecasting results

The stream flow data in m^3/s is first transformed into runoff depth in mm over Yingluexia basin. The precipitation, pan evaporation and stream flow data are then normalized linearly in the range of 0.1–0.9. This range is selected because of the logistic function (which varies from 0.0 to 1.0) used as the activation function. The normalization is done by utilizing equation (2):

$$Y=0.1+(0.9-0.1)\times(X-X\mathrm{min})/(X\mathrm{max}-X\mathrm{min}) \qquad (2)$$

where Xmin and Xmax are the minimum and maximum values in the data set, respectively; Y is the normalized value corresponding to X. A learning rate of 0.25 and a momentum term of 0.8 are used for ANNs training phase. Two stopping criteria, the cross validation and the fixed number of iterations, are adopted in the present study.

One of aims in this study is to develop the ANN models for stream flow forecasting for a large-scale basin and determine an appropriate combination of input vector and a more accurate architecture (i.e., number of neurons in hidden layer) in the design phase. A great number of ANN models with different combinations of input variables and number of neurons in hidden layer are investigated, and the

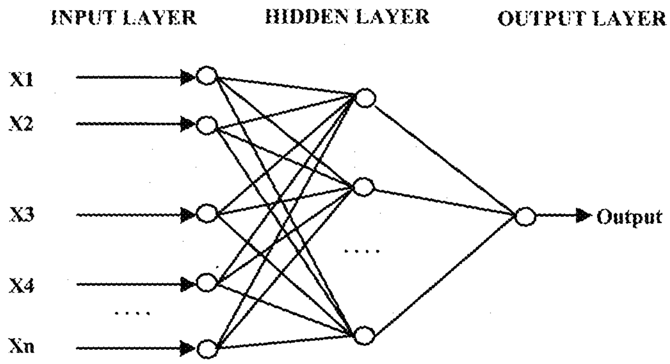

Fig.2. Conceptual diagram of three-layer feed-foreword ANN model.

Table 1: ANN models with optimized combinations of input variables and their performance*

Model NO.	Input combinations	Neurons in hidden layer	NE(Nash-Sutcliffe efficiency)	
			training period	testing period
1	Q(t-1), Q(t-2)	7	0.867	0.823
2	P(t),P(t-1),P(t-2),P(t-3),Q(t-1), Q(t-2), Q(t-3)	3	0.929	0.915
3	P(t-1),P(t-2),P(t-3),Q(t-1),Q(t-2), Q(t-3)	7	0.942	0.898
4	P(t),P(t-1),P(t-2), P(t-3),P(t-4)	7	0.635	0.644
5	W(t),W(t-1),W(t-2), W(t-3),W(t-4)	7	0.635	0.498
6	W(t),W(t-1),W(t-2),W(t-3),Q(t-1), Q(t-2), Q(t-3)	7	0.933	0.919

*Notations: Q = streamflow; P = precipitation, W = precipitation-evaporation.

Table 2: The calibrated parameters of XXT and the NE values.

Parameters						NE	
M(m)	SZQ (m/time step)	WM	B	SRmax(m)	Chevl(m/h)	calibration	validation
0.026	0.24	31.8	0.093	0.005	2680	0.595	0.732

Where M, the parameter of the exponential recession curve; SZQ, initial stream discharge; SRmax, the soil profile storage available for transpiration; WM and B are parameters with respect to the soil moisture storage capacity distribution curve; ChVel, an effective surface routing velocity for scaling the distance/area or network width function.

representative optimized combinations of input variables of these models is presented in Table 1. The performance of each of these optimized ANN models in terms of NE is also shown in Table 1. It can be seen that the ANN model with the combination of flow data + precipitation data + evaporation data as the input variables has the best performance in terms of NE, which means that for this large basin the stream flow, precipitation and evaporation data are all necessary to ANN modeling. The best performance was achieved for Model 6. This model resulted in a NE of 0.933 during training period and a NE of 0.919 during testing period. Hereafter this model is referred to as ANN1. Stream flows forecasted by ANN1 agree very well with the observed values in terms of NE since that the NE values are very high (higher than 0.9) in both training and test period. However, Fig.3 demonstrates that ANN1 underestimates the largest peak flow and overestimates the second largest peak flow in testing phase.

4.2. XXT model forecasting results

The rainfall-runoff simulations with XXT in Yingluoxia basin were carried out. The calibrated parameters and the NE values for this basin are shown in Table 2. The NE values for XXT simulation are 0.595 for calibration and 0.732 for validation period, respectively. Fig.3 shows that XXT underestimates the largest peak flow while captures the second and the third largest peak flows well in validation period. For one day in advance forecasting, ANN1 has a very good performance against the XXT in terms of NE(see

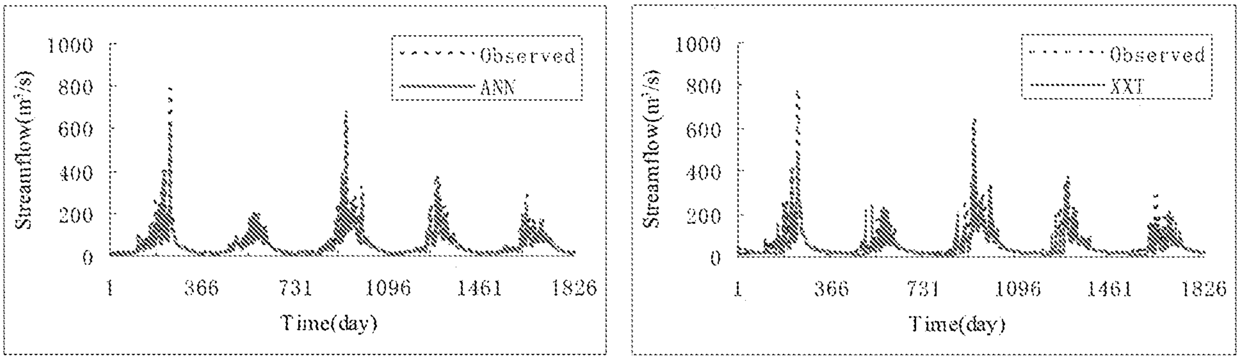

Fig.3. Observed and forecasted daily stream flow for the testing/validation period for Yingluexia basin (Day 1corresponds to January 1,1996).

Table 1 and Table 2). However, it has not advantages over XXT in capturing the main peak flows.

5. CONCLUSION

A great numbers of tests were performed in the present study to develop ANN-based daily stream flow forecasting models for the large-scale basin Yingluoxia, which is located the arid, semi-arid region of northwestern China. Furthermore, comparisons are made between the ANN models and a process-based rainfall-runoff model XXT, which is the hybrid of widely used Xinanjiang model and TOPMODEL. We find that The ANN model with an appropriate combination of stream flow, precipitation and evaporation data as input vector performs much better than XXT in terms of Nash-Sutcliffe efficiency. Even if only using antecedent stream flow data as inputs, ANN models are still better than the process-based hydrological model for one day in advance flow forecasting. However, both of them have drawbacks in capturing the main peak flows. ANN models are powerful tools for daily stream flow forecasting, especially in low stream flow phase.

Acknowledgements. This study is financially supported by the National Key Developing Program for Basic Sciences of China (Grant number 2006CB400502) and the Special Meteorology Project (GYHY200706001).

6. REFERENCES

[1] E. Kahya and J. A. Dracup, "U.S. Streamflow Patterns in Relation to the El Niño/Southern Oscillation," *Water Resour. Res.*, vol. 29, pp. 2491–2503, 1993.

[2] P. Srivastava, J. N. McVair, and T. E. Johnson, "Comparison of process-based and artificial neural network approaches for streamflow modeling in an agricultural watershed," *Journal of the American Water Resources Association,* vol. 42, pp. 545-563, Jun 2006.

[3] M. C. Demirel, A. Venancio, and E. Kahya, "Flow forecast by SWAT model and ANN in Pracana basin, Portugal," *Advances in Engineering Software,* vol. 40, pp. 467-473, Jul 2009.

[4] R. J. Zhao, "The Xinanjiang Model," in *Hydrological Forecasting Proceedings Oxford Symposium, IASH,* Oxford, 1980 pp. 351-356.

[5] K. J. Beven and M. J. Kirkby, "A physically based variable contributing area model of basin hydrology," *Hydrological Science Bulletin,* vol. 43, pp. 43-69, 1979.

[6] A. S. Tokar and M. Markus, "Precipitation-Runoff Modeling Using Artificial Neural Networks and Conceptual Models," *Journal of Hydrologic Engineering,* vol. 5, pp. 156-161, 2000.

[7] J. E. Nash and I. V. Sutcliffe, "River flow forecasting through conceptual models," *Journal of Hydrology,* vol. 273, pp. 282-290, 1970.