Using artificial neural network approach for modelling rainfall–runoff due to typhoon

S M Chen, Y M Wang* and I Tsou

Department of Civil Engineering, National Pingtung University of Science and Technology, Neipu Hsiang, Pingtung 91201, Taiwan (ROC).
**Corresponding author. e-mail: wangym@mail.npust.edu.tw*

In Taiwan, owing to the nonuniform temporal and spatial distribution of rainfall and high mountains all over the country, hydrologic systems are very complex. Therefore, preventing and controlling flood disasters is imperative. Nevertheless, water level and flow records are essential in hydrological analysis for designing related water works of flood management. Due to the complexity of the hydrological process, reliable runoff is hardly predicted by applying linear and non-linear regression methods. Therefore, in this study, a model for estimating runoff by using rainfall data from a river basin is developed and a neural network technique is employed to recover missing data. For achieving the objectives, hourly rainfall and flow data from Nanhe, Taiwu, and Laii rainfall stations and Sinpi flow station in the Linbien basin are used. The data records were of 27 typhoons between the years 2005 and 2009. The feed forward back propagation network (FFBP) and conventional regression analysis (CRA) were employed to study their performances. From the statistical evaluation, it has been found that the performance of FFBP exceeded that of regression analysis as reflected by the determination coefficients R^2, which were 0.969 and 0.284 for FFBP and CRA, respectively.

1. Introduction

Due to tremendous spatial and temporal variability in precipitation, rainfall–runoff relationship becomes one of the most complex hydrologic phenomena. In Taiwan, because of its location, rainfall is more prevalent during the typhoon season which usually brings about two-thirds to three-quarters of the annual rainfall amount. Taiwan is surrounded by sea, bringing a lot of water vapour that is elicited by typhoon; with natural physiographic characteristics often inducing flood disasters. For example, 2009 typhoon Morakot in Pingtung County of southern Taiwan caused serious river disasters and embankment failure. A quick and accurate flood forecasting is required particularly in flood prone regions, like Taiwan, for the issuance of warnings in order to allow ample time for the evacuation of population endangered. Wu *et al.* (2008) developed a distributed support vector regression model (D-SVR) equipped with genetic algorithm based Artificial Neural Network (ANN-GA) as part of flood control measures and outlined the significance of flood forecasting.

Vos and Rientjes (2005) reported that one of the main research challenges in hydrology is the development of computational models that are able to simulate accurately a catchment's response to rainfall. Such models are capable of forecasting future river discharge under different rainstorm return period, which are needed for hydrologic and hydraulic engineering design and water

Keywords. Feed forward back propagation network; river flow; typhoon; rainfall; runoff.

management purposes. Chiang *et al.* (2004) indicate that the complex watershed rainfall–runoff process is non-linear and dynamic in nature.

For decades, many rainfall runoff models have been developed for forecasting stream flow from two different aspects and they are; knowledge-driven and data-driven models. However, simulating the real-world relationship using these rainfall–runoff (R–R) models is far from a trivial task because to transform rainfall into discharge is very complex and highly variable. The best examples of knowledge driven models are; physically based model approaches, which generally use a mathematical framework based on mass, momentum and energy conservation equations in a spatially distributed model domain, and parameter values that are directly related to catchment characteristics. Furthermore, for proper application of these models, complex parameters for respective model need to be determined. The predictive performances of the physically based models are usually subjective and are dependent on the watershed characteristics and the ability of the model user.

Solomatine and Dulal (2003) mentioned that data-driven models treat the studied watershed as a black box and try to find a relationship between historical inputs (e.g., rainfall, temperature, etc.) and outputs (e.g., runoff). Traditionally, data-driven models borrow techniques developed in such areas, like statistics, soft computing, computational intelligence, machine learning and data mining. Among many data-driven techniques, the artificial neural network (ANN) is the most widely applied.

The American Society of Civil Engineers (ASCE) Task Committee (2000) summarised applications of ANN for the solution of many hydrologic problems. Muttil and Chau (2006) discovered that through analysis of ANN and genetic programming (GP) scenarios, long term trends in algal biomass can be obtained only by chlorophyll. Cheng *et al.* (2008) proved the effectiveness of using chaos genetic algorithm (CGA) in optimal operations of complex reservoir systems through a series of monthly inflows for 38 years. Dawson and Wilby (1998) discussed the application of ANN to flow forecasting in two flood-prone catchments in England using hourly hydrometric data. Wang *et al.* (2012) employed a hybrid genetic algorithm to optimize parameter values for Xinanjiang model for flood forecasting in Shuangpai reservoir. ANN models were established to predict flow with 6 hours lead time for Amber and Mole rivers. Rajurkar *et al.* (2002) applied ANN for modelling daily flows during monsoon flood events for a catchment in India using daily rainfall data as input vector of the network model. Tokar and Johnson (1999) constructed ANN model to forecast daily runoff as a function of daily precipitation, temperature, and snowmelt for a watershed in Maryland. Tokar and Markus (2000) applied ANN to predict monthly stream flow for the Fraser River Watershed in Colorado. Also, ANN has been applied in sediment transport (Tayfur 2002; Wang *et al.* 2009); stream flow forecasting (Anctil and Rat 2005; Wu *et al.* 2005); rainfall forecasting (Ramírez *et al.* 2005); seepage (Tayfur *et al.* 2005); dispersion (Tayfur and Singh 2006); and evapotranspiration (Wang *et al.* 2009). Most of the previous studies employed only daily or monthly hydrometric data in the input vectors of neural networks to make long-term predictions for various watersheds and tested ANNs against the regression-based, simple conceptual black box, or stochastic models. The complex nonlinearity process of rainfall–runoff in southern Taiwan provides an impetus for evaluating other alternatives like the FFBP. This could enable monitoring episodic river flows at a short time step during storm events. However, in Taiwan, very few studies reported the use of ANN on river flow modelling by introducing the available hourly rainfall data of the basin. River flow forecasting by using local available hourly rainfall records of the watershed is essential and crucial to provide basic information on a wide range of problems related to the water system design and the river management while the gauge station is damaged especially during extreme flow periods.

Hence, the purpose of this paper is to establish a model for estimating runoff by using hourly rainfall data from the river basin and to recover missing data by ANN technique.

2. Materials and methods

2.1 *Study area*

The study area was Linbien River (22°20′∼22°30′N; 120°28′∼120°40′E), a basin located in the low-lying areas of Pingtung County in the southern part of Taiwan. Because of its location, Linbien is more prone to natural disasters. For example, the 2005 Haitang typhoon, 2006 Bilis typhoon and 2009 Morakot typhoon caused serious embankment failures in the river. Because of abundant rainfall during the summer season in Linbien area, the total accumulated rainfall were around 2114.5 mm for the year 2009 with abundant rainfall occurring in the wet season from May to October, conversely in the dry season from November to April. The river is 42 km long with a drainage area of 336.30 km^2, a catchment area of 127.01 km^2 and a catchment average slope of 0.5122 m/m. Three major rainfall stations in the basins were selected for this study as shown in figure 1 and

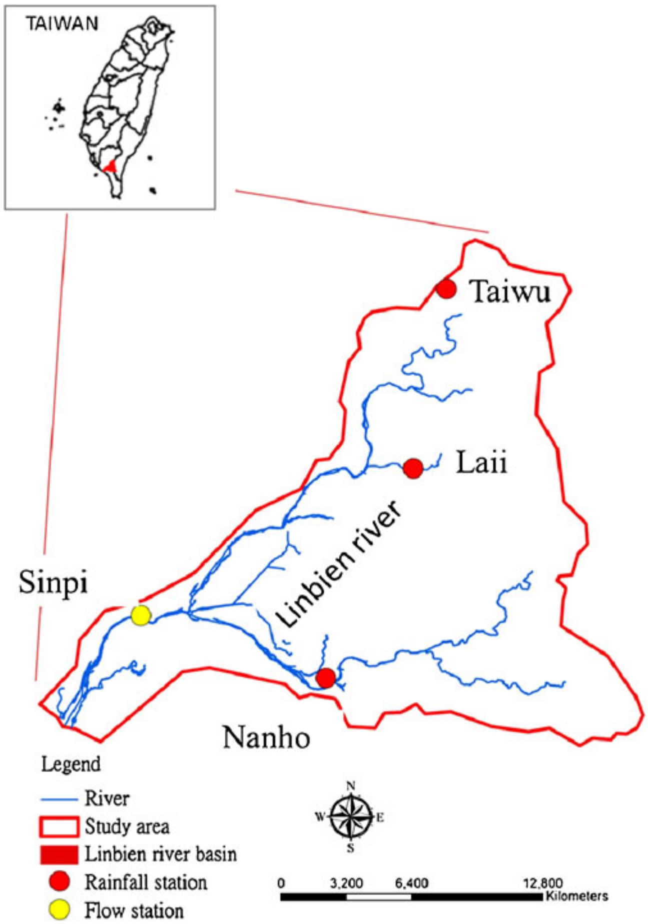

Figure 1. Location of the study area.

they are; Nanhe (22°26′4.49″N, 120°38′18.605″E), Taiwu (22°36′24.439″N, 120°41′43.75″E) and Laiyi (22°31′38.462″N, 120°40′47.316″E). Besides, the flow station selected was Sinpi located at 22°27′43.177″N, 120°33′1.039″E in the basin.

2.2 *Artificial neural network model*

An ANN is an information-processing paradigm inspired by biological nervous systems such as our brain (Lin *et al.* 2006). Although there are many types of ANNs, the most popular is the feed forward neural network, which is organised as layers of computing elements connected via weights between layers. Generally, there is an input layer (receiving inputs from the environment), one or more intermediate layers and an output layer (producing the network's response).

This study used the NeuroSolution 5.0 software presented by the NeuroDimension, Inc. Intelligence software solution. Feed forward back propagation (FFBP) was selected for forecasting river flow because it is the most commonly used ANN approach in hydrological predictions and in approximating nonlinear functions (Traore *et al.* 2010). The FFBP is a supervised learning technique used for training ANN. Basically, it is a gradient descent technique to minimize some error criteria because of the method used in its training. Training is a process of adjusting the connection weights in the network so that the network's response best matches the desired response (Muttil and Chau 2006). Although this can be treated as an optimization method, the FFBP avoids this costly exercise by using an approximation to a gradient descent method (Muttil and Chau 2006).

The neural network structure in this study had a three-layer learning network consisting of an input layer, a hidden layer and an output layer as illustrated in figure 2. In the figure, hourly rainfall data of Taiwu, Laii and Nanho stations were used as input for building the model to estimate hourly flow for Sinpi station (output) in the river basin studied herein.

The data was divided into three parts for the purpose of training (70%), validation (20%) and testing (10%) as recommended by Wang and Traore (2009). First, training data (from August 2005 to July 2008) are used to train the network by minimizing error data. Secondly, validation data (from August 2008 to July 2009) are used to find the network performance by monitoring the training and guard against overtraining. Lastly, the testing data (from August 2009 to October 2009) are used for checking the overall performance of trained and validated network. Table 1 shows the conditions of the training performance variables for FFBP.

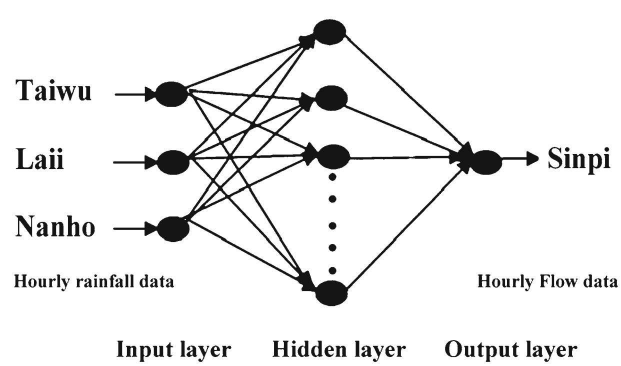

Figure 2. Structure of typical FFBP neural network model in the study.

Table 1. *Conditions of the training performance variables for FFBP.*

Training variables	Assigned value
Step size	0.1
Momentum	0.7
Maximum iterations	5000
Training threshold	0.001

2.3 *Data processing*

It is essential to scale the data before applying ANN to predictions. First, it helps to avoid attributes in greater numeric ranges dominating those in smaller numeric changes. Also, it is essential to avoid numerical difficulties during the calculations. In this study, TanhAxon (figure 3) is used as a transfer function since it is one of the most widely used. The TanhAxon is normally employed as hidden and output layers in multilayer perceptron topologies. The TanhAxon applies a bias and hyperbolic tangent (tanh) function to each neuron in the layer, thus outputting values within the range −1 to +1 for each neuron. This hyperbolic tangent (tanh) function is expressed as:

$$\mathrm{F(net)} = \tanh(\alpha\ \mathrm{net}). \tag{1}$$

The initial weights were started at random values. The use of varied random starting weights on each run could generate different outcomes; therefore, five independent runs were made on each topological model in order to get the best result.

2.4 *Model performance evaluation*

The performance of the predictions resulting from training, validation and testing is evaluated by the following measures for goodness-of-fit: root mean square error (RMSE) and coefficient of determination (r), and they are expressed as following:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - y_i'\right)^2}{N}} \tag{2}$$

$$r = \frac{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)\left(y_i' - \overline{y}'\right)}{\sqrt{\sum_{i=1}^{N}\left(y_i - \overline{y}\right)^2 \sum_{i=1}^{N}\left(y_i - \overline{y}'\right)^2}} \tag{3}$$

where y_i represents the observed flow record, y_i' is the alternative methods estimated flow values; $\overline{y}$ and $\overline{y}'$ represent the averages values of the corresponding variable; and N represents the number of data considered. Additionally, a linear regression $y = \alpha_1 x + \alpha_0$ is applied for evaluating the models' performance statistically, where y is the dependent variable (alternative methods); x the independent variable (observed); α_1 the slope and α_0 the intercept.

3. Results and discussion

3.1 *Conventional regression analysis*

Conventional regression is the simplest and well developed representation of a casual, time invariant relationship between an input function model and the corresponding output function. Hence, it is developed as a benchmark for comparison in flood forecasting (Chau *et al.* 2005). In the regression study, the 1745 datasets collected were divided into 1221 sets for linear regression and 524 sets were for regression model testing. Poor performance is observed as it is found that the determination coefficient R^2 performed by the regression model is only 0.284 as shown in figure 4.

3.2 *Processing elements determination*

This study adopted a hidden layer since it is well known that one hidden layer is enough as suggested by Traore *et al.* (2010). The determination of the optimum processing element (PE) in the hidden layer providing the best testing results was the initial process of the learning procedure. Determination of the optimal PE is an important factor, which affects the performance of the trained network (Muttil and Chau 2006). According to Wang *et al.* (2011), one of the important issues in the

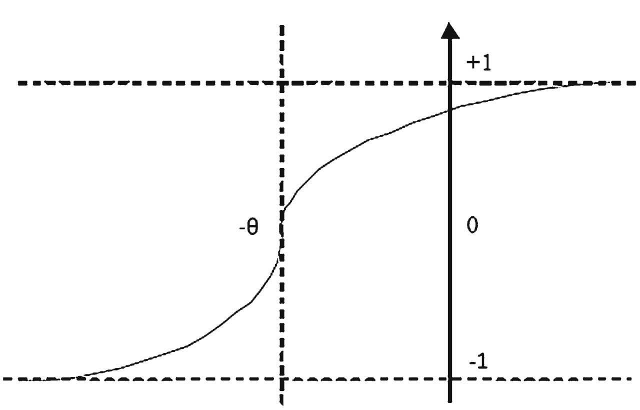

Figure 3. A TanhAxon-with-bias type of transfer characteristic.

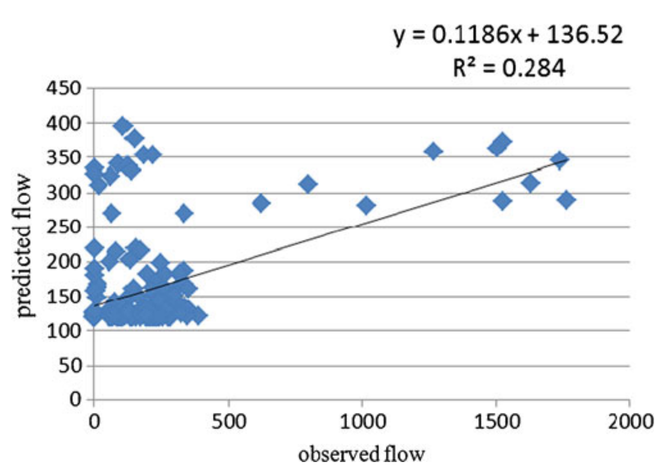

Figure 4. Scatter plot for regression model.

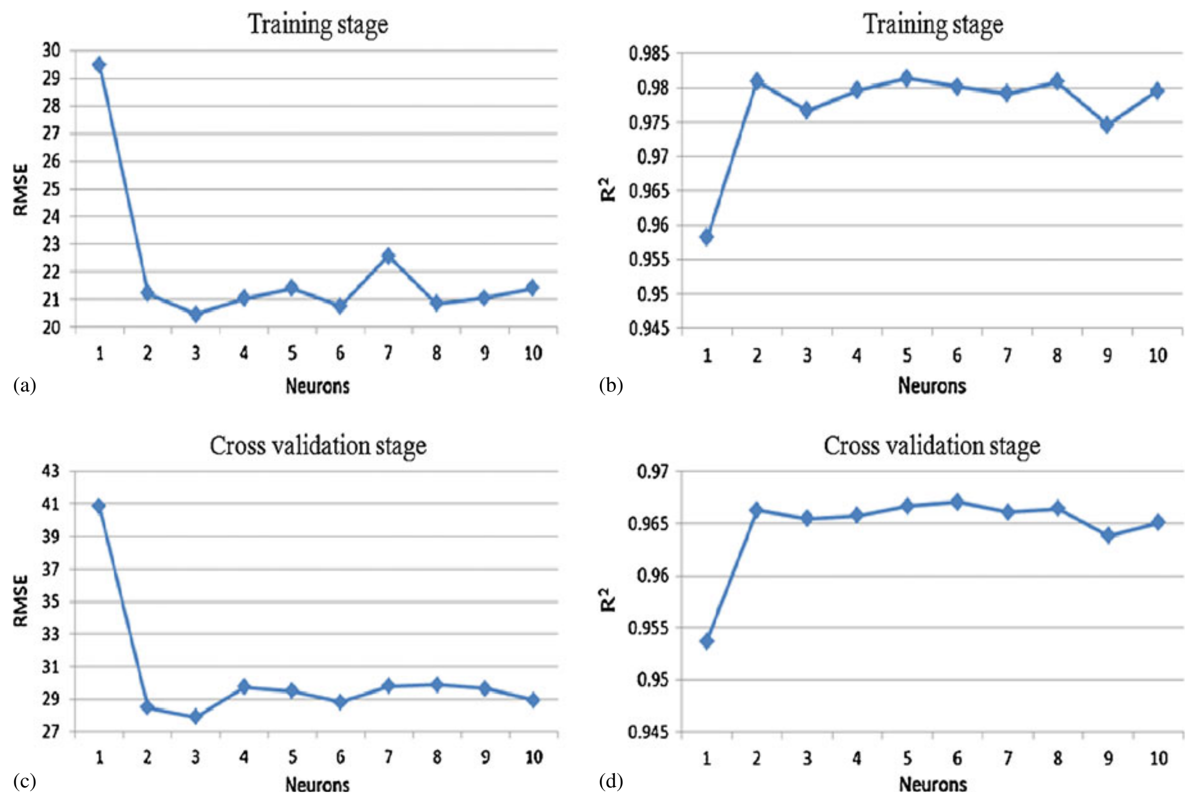

Figure 5. FFBP neural network accuracy for different number of nodes during the training (**a**, **b**) and cross validation stage (**c**, **d**).

development of neural networks model is the determination of the optimal number of processing elements that can satisfactorily capture the nonlinear relationship existing between the input and the output variables. The number of neurons in the hidden layer was varied between 1 and 10 and were trained based on the trial and error procedure. In general, networks with fewer hidden neurons are preferred, because they have better generalization capability and less overfitting problems. The dataset aside for the testing period was used to find the optimal number of processing elements. The optimum number of processing elements was found to be three and eight based on the minimum RMSE and maximum coefficient of determination (r) as illustrated in figure 5.

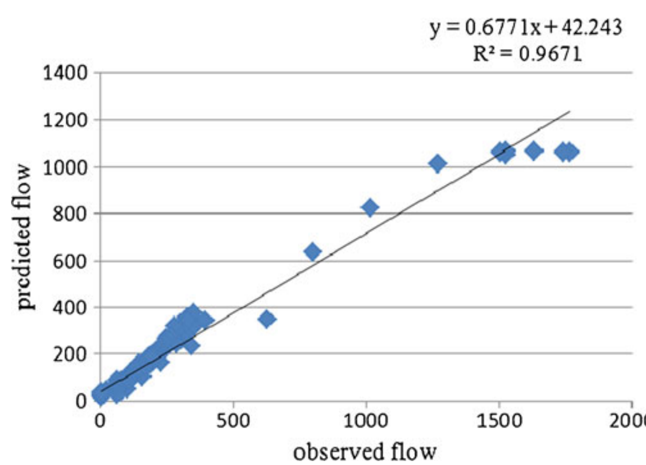

Figure 6. Scatter plot for three neurons in the model.

3.3 *Comparison among models and discussion*

After the determination of the best architecture for FFBP, we applied the same data to the FFBP neural network to obtain the predicted results. Figures 6 and 7 presented the scatter plot of observed and predicted data of testing dataset in this study.

From the results listed in figures 6 and 7, one can find the best result when the data ratio is 70:20:10 with eight neurons, by comparing with the other nine ratios. The RMSE of the 70:20:10 is 20.858 and R^2 is 0.969, which is superior to the results obtained in the regression model ($R^2 = 0.284$) shown in figure 4. Moreover, one can easily conclude that the ANN approach performed better than the conventional statistical method.

Figures 8–10 presented the desired and actual network output for training, validation, and testing stages respectively. Observing the figures, trends of the estimated outputs are similar to the observed

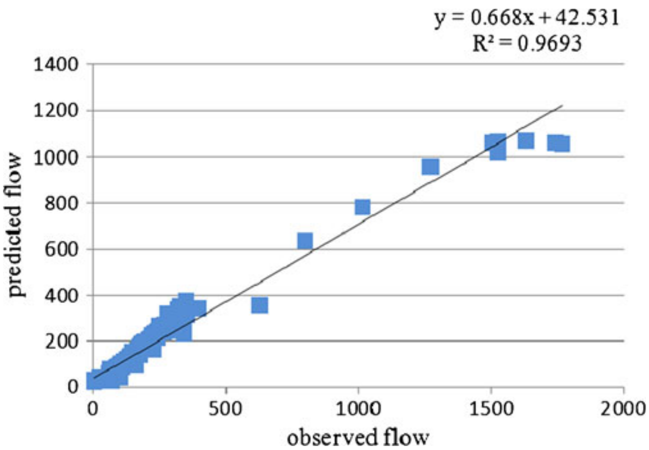

Figure 7. Scatter plot for eight neurons in the model.

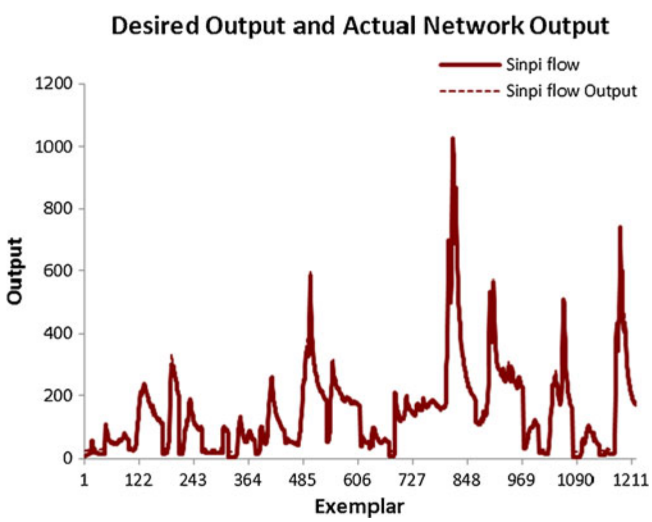

Figure 8. Desired and actual network output at training stage.

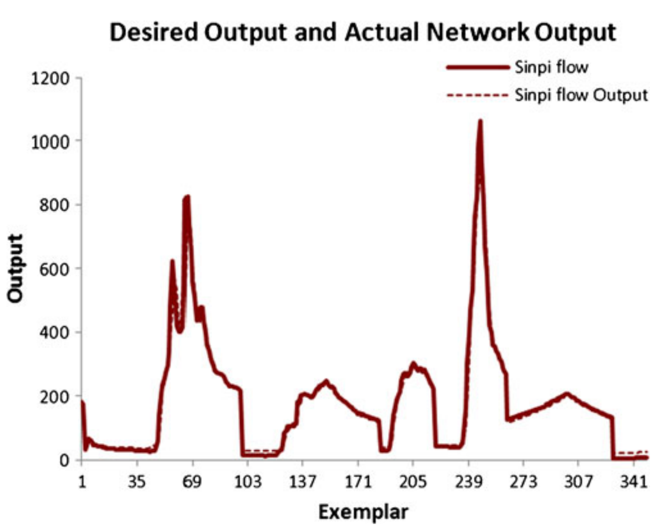

Figure 9. Desired and actual network output at validation stage.

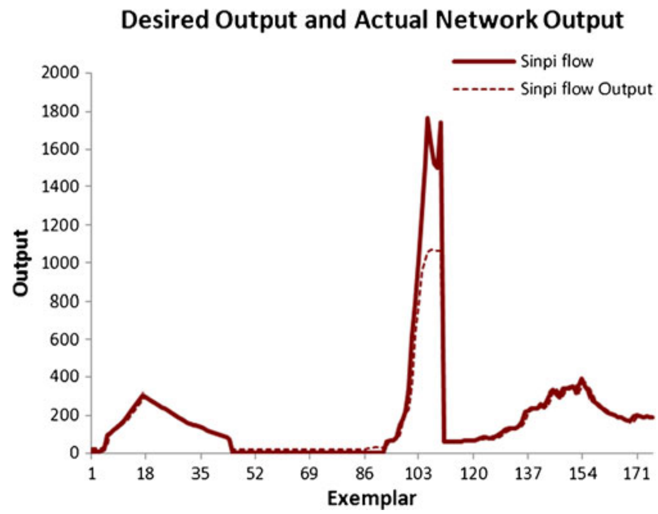

Figure 10. Desired and actual network output at testing stage.

data, although in the testing stage some obvious differences can be distinguished.

Finally, the study applied both built models to class predictive accuracy test. The superiority of ANNs over the conventional method in the reviewed prediction study can be attributed to their capability to capture the nonlinear dynamics and generalize the structure of the whole dataset. Besides, using the FFBP model is more reliable than conventional regression method for estimating runoff in the area studied herein.

4. Conclusions

In this study, the applicability of artificial intelligence techniques was investigated in the Linbien River in southern Taiwan. The performances of the models and observations were compared and evaluated based on their performance in training and testing sets. Feed forward back propagation (FFBP) perform better than the conventional regression analysis method.

The regression analysis is suitable for trend of small variation whereas artificial neural network is suitable for trend of large variation. In addition, the constructed model does not use complicated mathematics and hydrological processes; therefore, it will be easier to use than the traditional hydrological models. The ability of FFBP network to reconstruct hydrological models by learning the experience makes the application of the model favourable. FFBP model can predict the flood discharge more accurately than the conventional regression analysis. Justified by the results, we recommend the use of FFBP for rainfall–runoff modelling, flood forecasting, and to explore the model in other basins.

Acknowledgement

The authors gratefully acknowledge the financial support from NSC Taiwan under the grant of NSC101-2625-M-020-003.

References

Anctil F and Rat A 2005 Evaluation of neural network stream flow forecasting on 47 watersheds; *J. Hydrol. Eng.* **10** 85–88.

ASCE Task Committee 2000 Artificial Neural Networks in Hydrology II: Hydrologic applications; *J. Hydrol. Eng.* **5** 115–123.

Chau K W, Wu C L and Li Y S 2005 Comparison of several flood forecasting models in Yangtze River; *J. Hydrol. Eng.* **10** 485–491.

Cheng C T, Wang W C, Xu D M and Chau K W 2008 Optimizing hydropower reservoir operation using hybrid genetic algorithm and chaos; *Water Res. Manag.* **7** 895–909.

Chiang Y M, Chang L C and Chang F J 2004 Comparison of static-feedforward and dynamic-feedback neural networks for rainfall–runoff modelling; *J. Hydrol.* **290** 297–311.

Dawson C W and Wilby R 1998 An artificial neural network approach to rainfall–runoff modelling; *J. Hydrol. Sci.* **43** 47–66.

Lin J Y, Cheng C T and Chau K W 2006 Using support vector machines for long term discharge prediction; *Hydrol. Sci. J.* **22** 895–909.

Muttil N and Chau K W 2006 Neural network and genetic programming for modelling coastal agal blooms; *Int. J. Environ. Pollut.* **28** 223–238.

Rajurkar M P, Kothyari C and Chaube U C 2002 Artificial neural networks for daily rainfall–runoff modelling; *J. Hydrol. Sci.* **47** 865–877.

Ramírez M C V, Velho H F C and Ferreira N J 2005 Artificial neural network technique for rainfall forecasting applied to the São Paulo region; *J. Hydrol.* **301** 146–162.

Solomatine D P and Dulal K N 2003 Model trees as an alternative to neural networks in rainfall–runoff modelling; *J. Hydrol. Sci.* **48** 399–411.

Tayfur G 2002 Artificial Neural Networks for sheet sediment transport; *J. Hydrol. Sci.* **47** 879–892.

Tayfur G and Singh V P 2006 ANN and fuzzy logic models for simulating event-based rainfall–runoff; *J. Hydraul. Eng.* **132** 1321–1330.

Tayfur G, Swiatek D, Wita A and Singh V P 2005 Case study: Finite element method and artificial neural network models for flow through Jeziorsko earth fill dam in Poland; *J. Hydraul. Eng.* **131** 431–440.

Tokar A S and Johnson P A 1999 Rainfall–runoff modelling using artificial neural networks; *J. Hydrol. Eng.* **4** 232–239.

Tokar A S and Markus M 2000 Precipitation-runoff modelling using artificial neural networks and conceptual models; *J. Hydrol. Eng.* **5** 156–161.

Traore S, Wang Y M and Kerh T 2010 Artificial neural network for modelling reference evapotranspiration complex process in Sudano-Sahelian zone; *Agric. Water Manag.* **97** 707–714.

Vos N J and Rientjes T H M 2005 Constraints of artificial neural networks for rainfall–runoff modelling: Trade-offs in hydrological state representation and model evaluation; *Hydrol. Earth Syst. Sci.* **9** 111–126.

Wang W C, Cheng C T, Chau K W and Li Y S 2012 Calibration of Xinanjian model parameters using hybrid genetic algorithm based fuzzy optimal model; *J. Hydroinform.* **14** 784–799.

Wang Y M, Kerh T and Traore S 2009 Neural network approach for modelling river suspended sediment concentration due to tropical storms; *J. Global Nest* **11** 457–466.

Wang Y M and Traore S 2009 Time-lagged recurrent network for forecasting episodic event suspended sediment load in typhoon prone area; *Int. J. Phys. Sci.* **4** 519–528.

Wang Y M, Traore S, Kerh T and Leu J M 2011 Modelling reference evapotranspiration using feed forward back propagation algorithm in arid regions of Africa; *Irrigation Drainage* **60** 404–417.

Wu C L, Chau K W and Li Y S 2008 River stage prediction based on a distributed support vector regression; *J. Hydrol.* **358** 96–111.

Wu J S, Han J, Annambhotla S and Bryant S 2005 Artificial neural networks for forecasting watershed runoff and stream flows; *J. Hydrol. Eng.* **10** 216–222.

MS received 9 August 2012; revised 31 October 2012; accepted 2 November 2012