Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 315 (2005) 236–251
Journal of Hydrology
www.elsevier.com/locate/jhydrol

An artificial neural network model for generating hydrograph from hydro-meteorological parameters

Sajjad Ahmad[a,1], Slobodan P. Simonovic[b,*]

[a]Department of Civil, Architectural and Environmental Engineering, University of Miami, Coral Gables, FL 33146-0630, USA
[b]Department of Civil and Environmental Engineering and Institute for Catastrophic Loss Reduction, University of Western Ontario, London, Ont., Canada N6A 5B9

Received 22 September 2003; revised 7 March 2005; accepted 31 March 2005

Abstract

Conceptual models are considered to be the best choice for describing the runoff process in a watershed. However, enormous requirements for topographic, hydrologic and meteorological data and extensive time commitment for calibration of conceptual models (especially for distributed models) are often prohibitive factors in their practical applications. Artificial neural networks (ANN) can be an efficient way of modeling the runoff process in situations where explicit knowledge of the internal hydrologic processes is not available. An ANN is a flexible mathematical structure that is capable of identifying complex nonlinear relationships between input and output data sets. This paper presents the use of ANN for predicting the peak flow, timing and shape of runoff hydrograph, based on causal meteorological parameters. Antecedent precipitation index, melt index, winter precipitation, spring precipitation, and timing are the five input parameters used to develop runoff hydrograph for the Red River in Manitoba, Canada. A feed-forward artificial neural network is trained by using back-percolation algorithm. Peak flow, time of peak, width of hydrograph at 75 and 50% of peak, base flow, and timing of rising and falling sides of hydrograph are the output parameters obtained from the neural network model to describe a runoff hydrograph. The ANN generated results are evaluated using statistical parameters: percentage error and correlation. For six flood events for which forecasts are made the average absolute error in peak flow and time of peak is 6% and 4 days, respectively. Correlation between observed and simulated values of peak flow and time of peak is 0.99 and 0.88, respectively.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Artificial neural networks; Hydrograph estimation; Meteorological parameters; Red River

1. Introduction

Runoff prediction has been an active area of research in surface water hydrology and will remain so in the foreseeable future because of the uncertainties associated with both the meteorological and hydrological parameters causing extreme flood events. A reasonable prediction of runoff not only

* Corresponding author. Tel.: +1 519 661 4075; fax: +1 519 661 3779.
E-mail addresses: sajjad@miami.edu (S. Ahmad), simonovic@uwo.ca (S.P. Simonovic).
[1] Tel.: +1 305 284 3457; fax: +1 305 284 3492.

0022-1694/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2005.03.032

provides useful information for management of water resources, but also reduces losses to life and property caused by extreme events. Predicted runoff is also vital for the economic analysis of flood management alternatives. With increasing population and economic activity in floodplains and along major rivers the importance of accurate runoff prediction is increasing.

Several techniques have been developed to predict runoff ranging from empirical or statistical relationships to detailed mathematical models. While empirical or statistical relationships can provide magnitude and frequency of floods, they are not capable of generating a runoff hydrograph with complete information on timing of peak, volume of flood and shape of hydrograph, especially slope of rising and falling sides of hydrograph. Though frequency analysis of past flood peaks can provide information on the risk, it is limited by its lack of consideration of the forcing factors producing floods. Mathematical models, based on the consideration of physical processes, have been divided into two categories in the literature, i.e. conceptual models, or black box models. Conceptual models, based on characteristics of model parameter and variables, can be further divided into two categories, i.e. distributed and lumped models. Both distributed, e.g. MIKE SHE (Singh et al., 1999) and lumped, e.g. HEC-1 (US Army Corps of Engineers, 1998), models are designed to approximate the general processes and physical mechanisms, which govern the hydrologic cycle. In terms of data requirement, simple black box models and highly sophisticated distributed conceptual models fall on two extreme ends of a spectrum, while lumped conceptual models with moderate data requirement are somewhere in the middle. Due to the realistic representation of watershed topography and ability to capture surface and ground water interaction, the most desirable method to predict runoff hydrograph is a distributed conceptual dynamic hydrologic model. However, extensive topographic, meteorological, and hydrologic data required to describe the runoff process and time needed to calibrate conceptual models (especially distributed models) are important factors to be considered in their practical applications. Implementation and calibration of conceptual models (especially distributed models) can typically present various difficulties, requiring sophisticated mathematical modeling tools, significant amount of calibration data, and some degree of expertise and experience with the model (Hsu et al., 1995).

While conceptual models are of importance in the understanding of hydrologic processes, there are many practical situations, such as runoff prediction, where the main concern is with making accurate predictions at specific locations along a river. In such a situation, a hydrologist may prefer not to expend the time and effort to develop and implement a conceptual model and instead may choose to implement a simpler black box model. In the black box approach, difference/differential equation models are used to identify a direct mapping between the inputs and outputs without detailed consideration of the internal structure of the physical processes. The linear time series models, such as ARMAX (auto-regressive moving average with exogenous inputs) model developed by Box and Jenkins (1976), have been most commonly used in such situations because they are relatively easy to develop and implement. ARMAX type model has been found to provide satisfactory predictions in many applications where input–output characteristics are approximately linear, e.g. forecasting water level or discharge at a point along the river based on a water level or discharge value at some location upstream of that point. However, such models do not attempt to represent the nonlinear dynamics inherent in the transformation of snow/rainfall to runoff and therefore may not always perform well. Based on the success of ANN in the field of water resources, and working with an assumption that simplest model that can satisfactorily describe the system for the given input data should be used, artificial neural networks (ANN), belonging to the class of black box models, are explored in this paper for the development of runoff hydrograph.

This paper outlines a general framework for developing a runoff hydrograph using artificial neural network approach. While such a model is not intended as a substitute for a conceptual model, it can provide a viable alternative when the hydrologic application requires that an accurate forecast of stream flow behavior be provided using only the available time series data, and with relatively moderate conceptual understanding of the hydrologic dynamics of the particular watershed under investigation. A feed-forward artificial neural network with back-percolation algorithm is used to estimate the shape of

the runoff hydrograph for major floods on a river. Based on the available historic data on flood events, five causal parameters, which play an important role in flood generation, have been identified. These five meteorological parameters that also have physical base are used as input to the neural network model. The network has been trained to predict eight output parameters that are used to describe a complete runoff hydrograph. ANN model has been trained using historic data and verified on a different data set never seen by the network. The proposed approach is implemented for the Red River in Manitoba, Canada. The neuro-computing approach for developing runoff hydrograph is presented along with details on input and output parameters of neural network model. The paper concludes with a discussion of results.

2. Application of ANN in hydrology

The application of ANN in hydrology started in the early 1990s. A state-of-the-art review of ANN applications in hydrology can be found in the ASCE task committee report (2000b). Some applications of ANN in water resources include: precipitation–runoff modeling (Rajurkar et al., 2004; Elshorbagy and Simonovic, 2000; Tokar and Markus, 2000; Zealand et al., 1999; Fernando and Jayawardena, 1998; Hsu et al., 1995); stream flow forecasting (Moradkhani et al., 2004; Anctil et al., 2004; Özgür, 2004) and river stage forecasting (Liong et al., 2000; Thirumalaiah and Deo, 1998; Karunanithi et al., 1994). Attempts have been made to develop runoff hydrographs using different input parameters. Muttiah et al. (1997) used information on the drainage basin, elevation, average slope, and average annual precipitation to predict 2-year peak discharge from a watershed. Carriere et al. (1996) used ANN with a recurrent back-propagation algorithm to generate a runoff hydrograph using a virtual runoff hydrograph system. They used rainfall intensity, duration, catchment slope, and catchment cover to estimate runoff hydrographs. Smith and Eli (1995) used a back-propagation ANN to predict the peak discharge and the time of peak resulting from a single rainfall event. They used a synthetic watershed to generate runoff from stochastically generated rainfall patterns. Markus et al. (1995) used ANNs with back-propagation algorithms to predict monthly stream flows from snow water equivalent and temperature data. Zhu et al. (1994) predicted upper and lower bounds on the flood hydrograph based on rainfall and previous flood data. The ANN approach presented in this research uses five causal parameters, i.e. antecedent precipitation index; melt index; winter precipitation; spring precipitation; and timing to develop a runoff hydrograph (Ahmad and Simonovic, 2001).

3. Issues in implementation of artificial neural networks (ANN)

Details on the theoretical aspects of ANN can be found elsewhere (ASCE Task Committee, 2000a; Anderson, 1995; Maren et al., 1990). There are two important issues concerning the implementation of artificial neural networks. The first issue involves a specification of the network size (the number of layers in the network and the number of nodes in each layer). This task involves decision on the number of nodes required in the hidden layer. Generally, the more complex the mapping the larger the number of hidden nodes required. However, too many hidden nodes can help network memorize the training set and behave like a look-up-table at the expense of any useful generalization that will result in poor performance on unseen data. The second issue involves finding the optimal values for the connection weights. Starting with a small number of nodes and gradually increasing the network size until the desired accuracy is achieved address the first problem. However, this approach heavily depends on the ability to find the optimal weights. The objective function (typically mean square error) used to train the ANN has many local optima and extensive regions of poor sensitivity to variations in weights (Hsu et al., 1995). Back-propagation algorithms (BPA), commonly used to train the network weights, display poor performance under these conditions. The back-percolation (Jurik, 1990) method has been used in this research to improve the convergence speed of BPA. Back propagation is a learning algorithm for a multi-layered neural network in which the weights are modified via propagation of an error gradient backward from the output to the input. A back-percolation algorithm modifies the error propagation method of

the back-propagation algorithm and considers the errors in the hidden nodes separately from the error in the output node. The model of the runoff hydrograph prediction presented in this research, using artificial neural networks, uses back-percolation algorithm to solve a feed-forward artificial neural network. The runoff hydrograph prediction model and its application to a case study are presented in Section 4.

4. Case study application

The proposed artificial neural network approach for predicting runoff hydrograph has been applied to the Red River in Manitoba, Canada (Fig. 1). The discharge measurement station near St Agathe has been selected as a site to estimate runoff hydrograph as this is the only station on the Red River, close to the city of Winnipeg, where natural flow (not affected by flood control structures) measurements are available. The Red River originates in the north-central United States in Minnesota and flows north. It forms the boundary between North Dakota and Minnesota and enters Canada at Emerson, Manitoba. It continues northward to Lake Winnipeg. From origin to its outlet in Lake Winnipeg, the river is 563 km long. The drainage area of the Red River basin at St Agathe is 116,500 km^2 of which nearly 103,600 km^2 are in US. The remaining 13,000 km^2 are in Canada. The basin is remarkably flat. The slope of the river averages about 0.05 m/km. The basin is about 100 km across at its widest. During major floods entire valley becomes the floodplain.

The Red River basin has a sub-humid to humid continental climate with moderately warm summers, cold winters, and rapid changes in daily weather patterns. Extreme temperature variations are the norm. On average, the Red River basin mean monthly temperature ranges from −15 to +20 °C. About three-forth of the basin's approximately 50 cm of annual precipitation occurs during April through September, with almost two-thirds falling during May, June and July. November through February are the driest months (IJC, 1997). These hydrological, meteorological and topographic factors are very important in understanding the flooding in the basin.

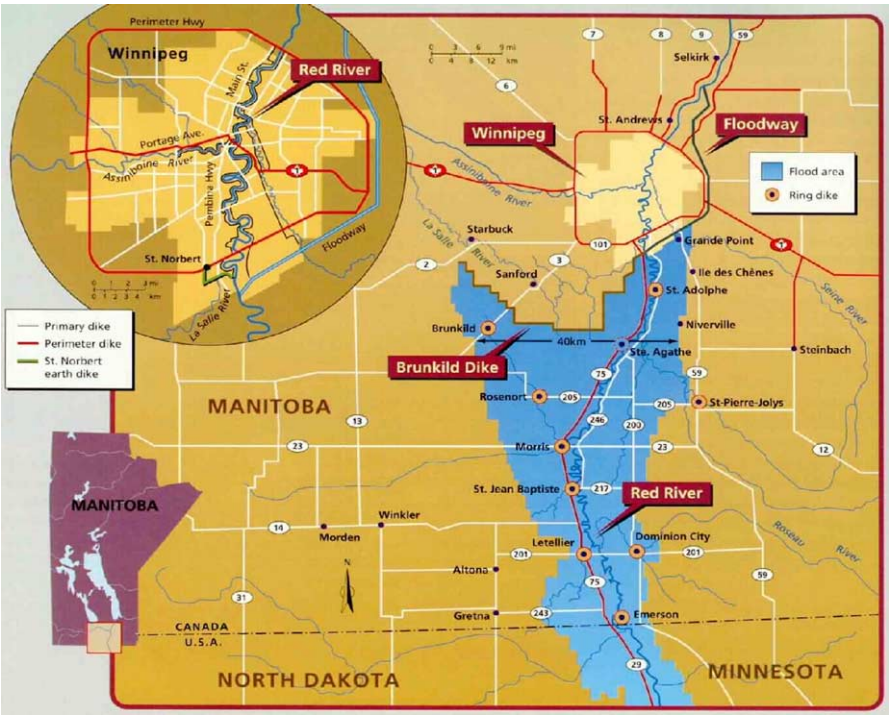


Fig. 1. Red River Basin, Manitoba, Canada (source Winnipeg Free Press).

The low absorptive capacity of the basin's clay soil is also a contributing factor towards floods.

5. Modeling approach

The modeling approach used to develop runoff hydrograph along with details on input and output parameters is presented in this section. The selection of appropriate input parameters that will allow an ANN to successfully produce the desired output is a complex task. Good understanding of the hydrologic system under consideration is an important prerequisite for successful application of ANNs. Physical understanding of the process being studied leads to better choice of input variables.

A careful study of historic flood data in the Canadian portion of the Red River basin suggests a causal pattern followed by large floods. The climatic conditions, which lead to the likelihood of an extreme flood, may be summarized as follows (Royal Commission, 1958):

- a wet summer or fall in the preceding year which thoroughly saturates the ground before freeze-up;
- cold weather with little snow during the early winter, allowing a deep penetration of frost;
- a cold winter with heavy snowfall over the entire drainage basin;
- a late spring followed by a sudden rise in temperature, producing a rapid runoff; and heavy rain during the runoff period.

Based on this information the following five parameters, i.e. antecedent precipitation index (API), melt index (MI), winter precipitation (WP), spring precipitation (SP) and timing (T) are identified as input to the neural network model for predicting a runoff hydrograph. These five parameters are collected in the form of processed data from the Water Resources Branch, Manitoba Department of Conservation. The parameters are defined in the following section. Details on calculation of these parameters can be found in Warkentin (1999).

Antecedent precipitation index (API). This is the index of soil moisture at freeze-up the previous autumn, based on weighted basin precipitation from May to October.

Melt index (MI). This is expressed in average degree-days/day at Grand Forks during the active melt period (°F).

Winter precipitation (WP). Total basin precipitation from November 1 of previous year to the start of active melt during the flood year, measured in inches.

Spring precipitation (SP). Total basin precipitation from the start of active snowmelt to the start of the spring crest, measured in inches.

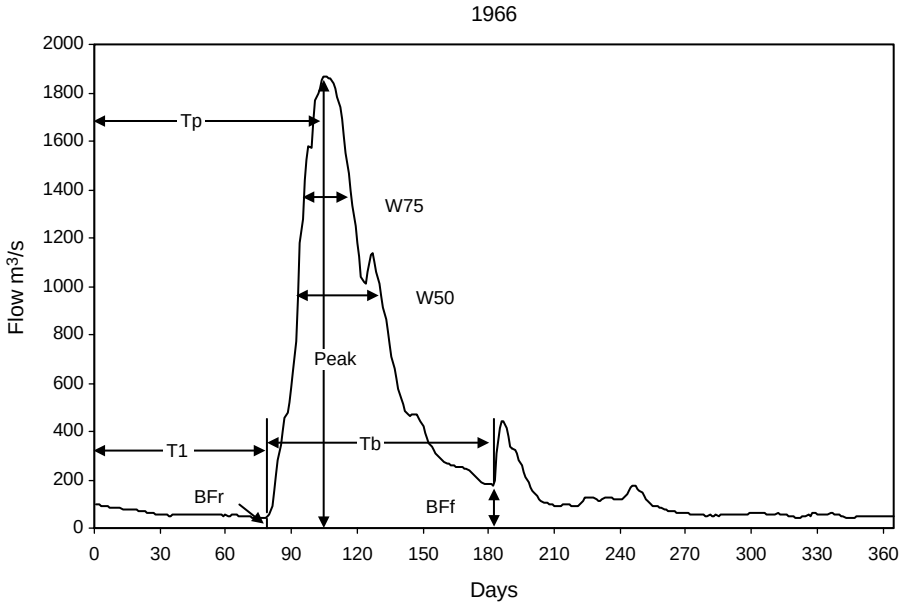
Timing factor (T). An index of the south–north time phasing of the runoff based on the percentage of tributary peaks experienced on the date of the mainstream peak. This is basically a percentage of the worst possible south–north progression of melt and rain. The value varies between 0 and 100, where 100 refers to worst. For 1997 flood T was 62.

A correlation test, for two parameters at a time, is done to determine whether any dependency exists between the causal parameters. The inter-correlation matrix is shown in Table 1. The results of the correlation test (lower values of correlation coefficient) suggest that causal parameters are statistically independent of each other. The strongest correlation between winter precipitation and the melt index is 0.33.

Eight network output parameters are identified that are required to estimate a runoff hydrograph. These output parameters are peak flow (Peak), timing of peak (T_p), timing of the rising (T_1) and recession sides (T_b) of the hydrograph, base flow both at the start of the rising side (BF_r) and end of the recession side (BF_f) of the hydrograph, width of the hydrograph at 50% of peak (W_{50}) and the width of hydrograph at 75% of the peak (W_{75}). The eight output parameters are defined in Fig. 2. To provide a data set for training

Table 1
Correlation between input parameters

	API	MI	WP	SP	T
API	–	0.02	0.18	0.01	0.32
MI	0.02	–	0.33	0.19	0.19
WP	0.18	0.33	–	−0.04	0.17
SP	0.01	0.19	−0.04	–	0.23
T	0.32	0.19	0.17	0.23	–



Peak: Peak value of hydrograph (m3/s)

Tp : Time of peak (days)

T1 : Time when rising side of hydrograph starts (days)

Tb : Total time between rising and recession sides of hydrograph (days)

BFr : Base flow at time T1, i.e., when hydrograph starts rising (m^3/s)

BFf : Base flow at time T1+Tb (m^3/s)

W50: Width of hydrograph at 50 % of peak (days)

W75: Width of hydrograph at 75 % of peak (days)

Fig. 2. Definition of hydrograph parameters.

of the neural network these parameters are extracted from hydrographs of observed floods.

6. Neuro-computing procedure

Daily discharge data is available for 40 years (1958–1997) from a gauging station near St Agathe on the Red River. The peak annual discharge recorded at this station range between 321 and 3230 m^3/s. Because the objective of this work is to predict high flow/floods, 7 years of data with very low discharge, i.e. peak less than 400 m^3/s, are removed from the data set. These years are: 1958; 1961; 1968; 1977; 1981; 1990; and 1991. The remaining 33 years of data are divided into three sets, i.e. training data (calibration), test while training data (verification), and forecasting data. Six years, i.e. 1976, 1978, 1979, 1989, 1996, and 1997 are separated from the data set for forecasting purposes. Within the 40 years of

Table 2
ANN model input parameters for selected floods

	Flood 1976	Flood 1978	Flood 1979	Flood 1989	Flood 1996	Flood 1997	Mean (training data)	Max (training data)
API	2.29	2.95	2.03	1.81	2.89	2.49	2.4	3.24
MI	11.4	7	18	3.5	10	11.6	7.4	14.9
WP	3.62	4.29	4.88	4.96	4.88	8.6	3.6	6.18
SP	0.12	0.94	2.89	0.70	0.12	0.1	0.7	2.9
T	50	53	47	43	75	62	45.6	81

the data used the largest flood occurred in 1997. Keeping in mind that the objective is an estimation of a hydrograph for large floods, three major floods are included in the forecasting data set. From the remaining 27 years, 7 years are separated by random selection to be used for testing while training the network. These years are 1959, 1963, 1970, 1972, 1980, 1984, and 1985. The remaining 20 years of data is used for training the network.

The method for generating the runoff hydrograph, based on hydro-meteorological parameters, using ANN approach is summarized as follows:

- Important causal parameters responsible for generating large floods are identified. These parameters are: antecedent precipitation index (API); melt index (MI); winter precipitation (WP); spring precipitation (SP); and timing factor (*T*). These input parameters, for the six selected floods for which forecasts are made, are given in Table 2. Mean and maximum values of input parameters for the training and test while training datasets are also provided.
- Important characteristics are identified that can be used to develop a runoff hydrograph. These hydrograph characteristics are: peak flow; time of peak; base flow at both the rising and recession sides of the hydrograph; timing of the rising and recession sides of the hydrograph; and width of the hydrograph at 75 and 50% of the peak. The hydrograph characteristics are extracted from a historic runoff time series to train the ANN.
- An ANN model is developed by relating input parameters to flood hydrograph characteristics. Two networks are developed with five input and four output parameters in each network. One hidden layer with four nodes is used in each network.
- The performance of the network is evaluated during 'training' and 'test while training' stages. In addition to considering the agreement between observed and simulated hydrographs, statistical performance measures of percentage error and correlation are used to evaluate the performance of the network. After comparing the observed flood hydrograph to that generated with ANN, adjustments in the controlling parameters of the ANN, such as number of nodes and weights, are made.
- Once network training is complete, it is used to predict the hydrograph parameters using the input data never seen by the network (forecasting data set).
- The output parameters obtained from the neural network model are used to develop a runoff hydrograph.

In short, the modeling process starts with defining the purpose/goal of the model. Next step involves identification of key input and output variables in the system. These system variables are mapped in the ANN. Then structure of ANN is adjusted by trial and error making decisions on number of nodes in the hidden layer, etc. Neural network is then run to test the behavior of the system. Network is evaluated by comparing the predicted and observed values. Based on the evaluation, improvements are made to the network structure. When the network performance is satisfactory it is ready for forecasting.

7. Performance evaluation statistics

The statistical evaluation methods are more objective than the visual inspection of the agreement between simulated and observed discharge hydrographs (ASCE Task Committee, 1993). In order to

evaluate the performance of the neural network, percentage error and correlation were calculated for each output parameter

Percentage Error

$$= \frac{\text{Predicted Value} - \text{Observed Value}}{\text{Observed Value}} \times 100 \quad (1)$$

A positive or negative value of percentage error represents an over prediction or under prediction by the model, respectively.

The correlation statistics measures the correlation between observed and forecasted values; the optimal value is 1.0. The correlation calculation returns the covariance of two data sets divided by the product of their standard deviations.

For two variables x and y, the correlation is defined by

$$\mathrm{Cor}(x, y) = \frac{\mathrm{Cov}(x, y)}{\sigma_x \sigma_y} \quad (2)$$

where σ_x, σ_y is standard deviation, Cov (x,y) is the covariance and x and y represent observed and predicted values, respectively

$$\mathrm{Cov}(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \quad (3)$$

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2} \quad (4)$$

$$\sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (5)$$

where $\bar{x}$ and $\bar{y}$ represent the mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (6)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad (7)$$

Correlation can be used to determine whether two ranges of data move together; that is, whether large values of one set are associated with large values of the other (positive correlation), whether small values of one set are associated with large values of the other (negative correlation), or whether values in both sets are unrelated (correlation near zero). The relationship between the peak flow and percentage error in each parameter of the predicted hydrograph is studied. This relationship helps in identifying a trend (if there is any) in model performance for flood events of different magnitudes.

Average absolute error is used to estimate overall average error in all predictions for a certain parameter

Average Absolute Error

$$= \frac{\text{Error in Prediction } 1 + \cdots + \text{Error in Prediction } n}{\text{Total number of Predictions}} \quad (8)$$

8. Results

The comparison of observed and predicted hydrographs for all 6 years of unseen data is shown in Fig. 3a–f. Error in individual hydrograph parameters, for all 6 years for which forecasts are made, is given in Table 3. The average absolute error for all 6 years for peak, T_p, T_1, T_b, BF_r, BF_f, W_{50} and W_{75} is 6%, 4.17, 5.5 days, 8, 25, 32, 19, and 20%, respectively. The relationship between the peak flow and error in predicted hydrograph parameters is shown in Fig. 4a–h. The correlation between observed and predicted values of hydrograph output parameters is shown in Fig. 5a–h. The hydrometeorological characteristics of floods are summarized in Table 2. In the following sections, first the hydrometeorological characteristics of floods are described, and then a discussion of results with respect to each hydrograph parameter predicted by the network is presented.

Lowest values for three of the five input parameters, i.e. API, MI and T belong to 1989 flood, ranked number 5 (in terms of flood peak) among six flood events for which forecasts are made. The lowest value of WP corresponds to 1976 flood that is ranked number 6. Surprisingly, the lowest spring precipitation value belongs to 1997 flood, the largest among six flood events. Highest value of MI and SP belong to 1979 flood, second largest flood among the six flood events. The highest winter precipitation was in 1997, the largest flood event. The highest API and T correspond to 1978 and 1996 flood events, ranked

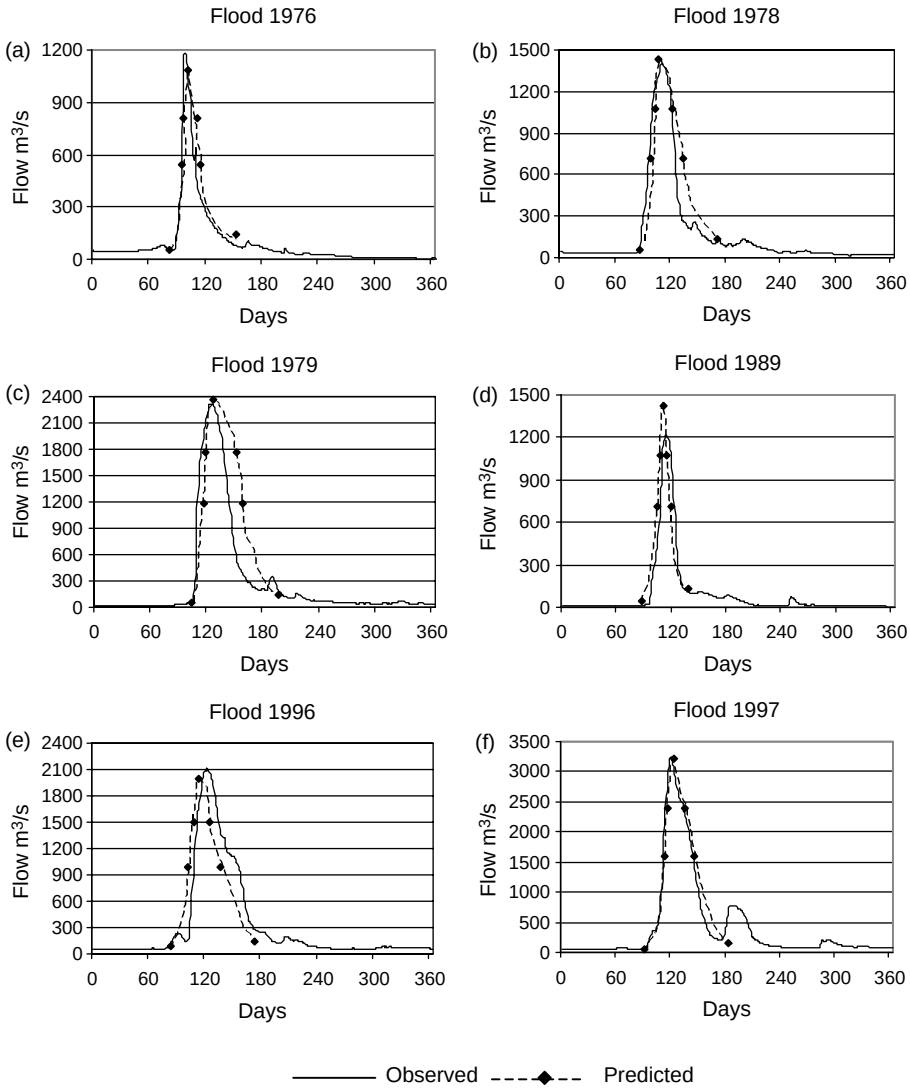


Fig. 3. (a–f) Comparison of observed and ANN generated runoff hydrographs.

numbers 3 and 4, respectively. It is clear that the largest flood event does not result from the largest value for each input parameter; it can result from any combination of input parameters.

8.1. Peak flow (peak)

The network produced an excellent prediction of peak flows for 1978, 1979, and 1997 (Fig. 3b,c,f, respectively); however, it over estimated the peak in 1989 (Fig. 3d) and under estimated it in 1976 (Fig. 3a) and 1996 (Fig. 3e). The error in estimating peak ranges from −8 to 17%. The average absolute error for all 6 years is 6%. The relationship between peak flow and percentage error in peak (Fig. 4a) and error values shown in Table 3 indicate that peak flow forecast is better for wet years. The network tends to fit the higher flows quite well. The error in the predicted peak for the three largest floods (in descending order) is found to be −1, 2 and −6%,

Table 3
Error in observed (O) and predicted (P) hydrograph parameters for selected floods

	1976			1978			1979			1989			1996			1997			Average absolute error
	O	P	Error	O	P	Error	O	P	Error	O	P	Error	O	P	Error	O	P	Error	
Peak	1180	1081	−8.40%	1400	1429	2.10%	2320	2356	1.60%	1220	1428	17.00%	2120	1994	−5.90%	3230	3207	−0.70%	5.95%
T_{peak}	98	102	4 days	112	108	−4 days	125	128	3 days	115	112	−3 days	122	115	−7 days	121	125	4 days	4.17 days
T_1	88	82	−6 days	85	88	3 days	105	105	0days	96	89	−7 days	102	85	−17 days	92	92	0 days	5.5 days
T_b	74	72	−2.70%	84	85	1.20%	78	93	19.20%	47	50	6.40%	92	89	−3.30%	82	93	13.40%	7.70%
BF_f	55	54	−1.80%	37	53	43.20%	66	56	−15.20%	21	33	57.10%	141	97	−31.20%	59	59	0.00%	24.75%
BF_f	85	138	62.40%	105	138	31.40%	200	141	−29.50%	105	133	26.70%	165	142	−13.90%	200	145	−27.50%	31.90%
W_{50}	14	18	28.60%	29	35	20.70%	34	41	20.60%	17	14	−17.60%	46	35	−23.90%	32	32	0.00%	18.57%
W_{75}	8	10	25.00%	24	0	−20.80%	24	29	20.80%	12	8	−33.30%	22	17	−22.70%	19	19	0.00%	20.43%

respectively. The overall performance of network in estimating the peak is excellent; it is also supported by the correlation value of 0.99 between the observed and forecasted peak flow values (Fig. 5a).

8.2. Time of peak (T_{peak})

The error in estimating the time of peak is in the range of −7 to 4 days. The average absolute error is 4.17 days for all 6 years for which forecasts are made. There is no detectable trend in the relationship between peak flow and percentage error in time of peak (Fig. 4b). Network performance is very good in estimating the time of peak. The correlation value between observed and forecasted time of peak is 0.88.

8.3. Time at the start of rising side of hydrograph (T_1)

Over all, the network captured the timing of the start of the rising side of the hydrograph quite well. The percentage error in estimating the time T_1 is in the range of −7 to 3 days. The average absolute error is 5.5 days for all 6 years for which forecasts are made. The relationship between the peak flow and percentage error in T_1 shown in Fig. 4c and error values shown in Table 3 indicate no detectable trend. Network performance is very good in estimating the time at the start of the rising side of hydrograph. Although correlation value between observed and predicted T_1 is 0.59, it should be noted that the presence of a large error in a single record in a time series can significantly affect the correlation value. Because of a 17% error in estimating T_1 in 1996, the correlation coefficient is low.

8.4. Time of base (T_b)

The percentage error in estimating the time of base (time between the rising and recession side of the hydrograph) is in the range of −3 to 19%. The average absolute error is 8% for all 6 years for which forecasts are made. The relationship between the peak flow and the percentage error in T_1 shown in Fig. 4d and error values shown in Table 3 indicate that there is a trend (though not decisive) that the percentage error increases with an increase in the peak flow. The network performance is very good in estimating

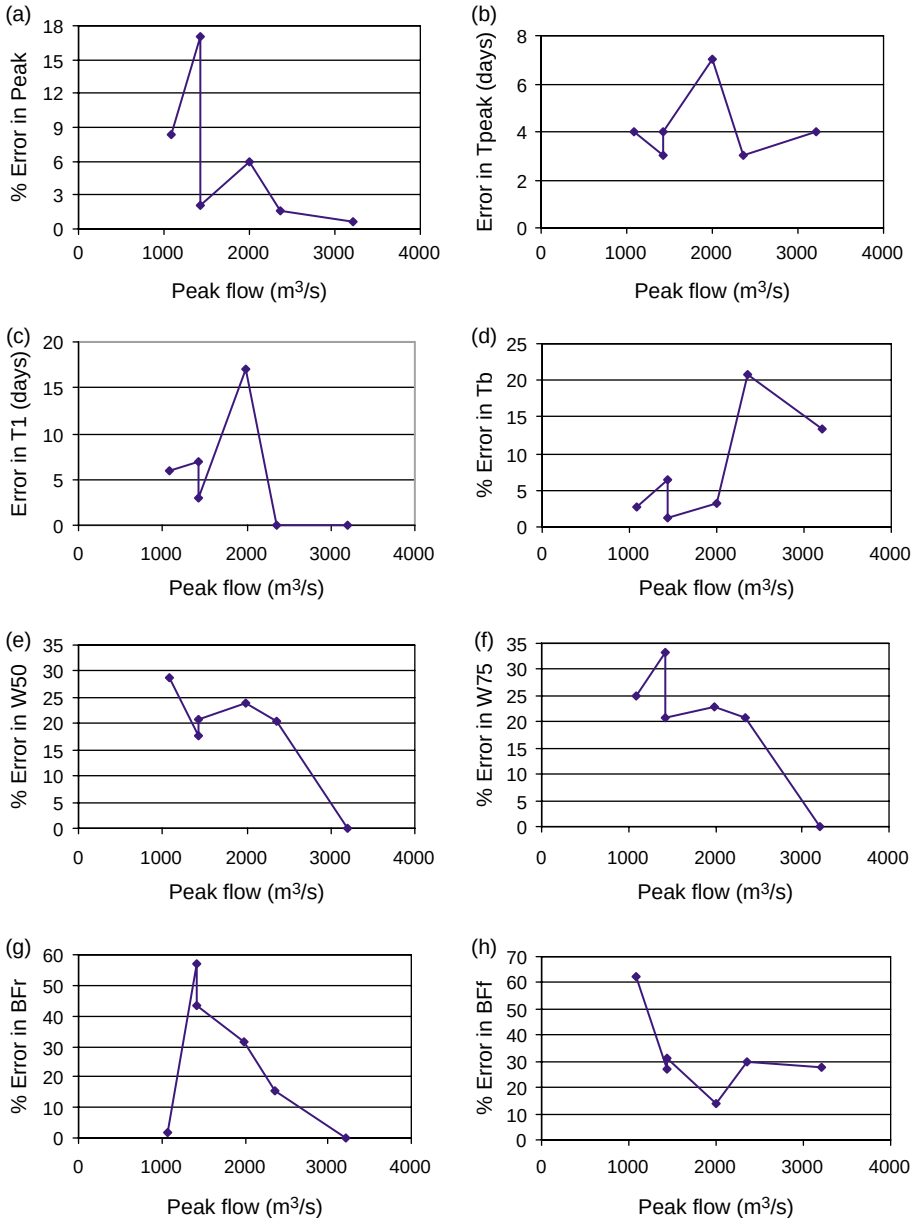


Fig. 4. (a–h) Relationship between peak flow and %error for hydrograph parameter.

the time of base. The correlation value between observed and forecasted T_b is 0.91.

8.5. Width of hydrograph at 50% of the peak (W_{50})

The network produced an excellent forecast of W_{50} for 1997 (Fig. 3f). The overall error in estimating the W_{50} is in the range of −24 to 28%. The average absolute error is 19% for all 6 years for which forecasts are made. The relationship between the peak flow and the error in W_{50} shown in Fig. 4e and error values in Table 3 indicate that model performance is better for wet years. The percentage error appears to decrease with increase in peak flow. However, if we

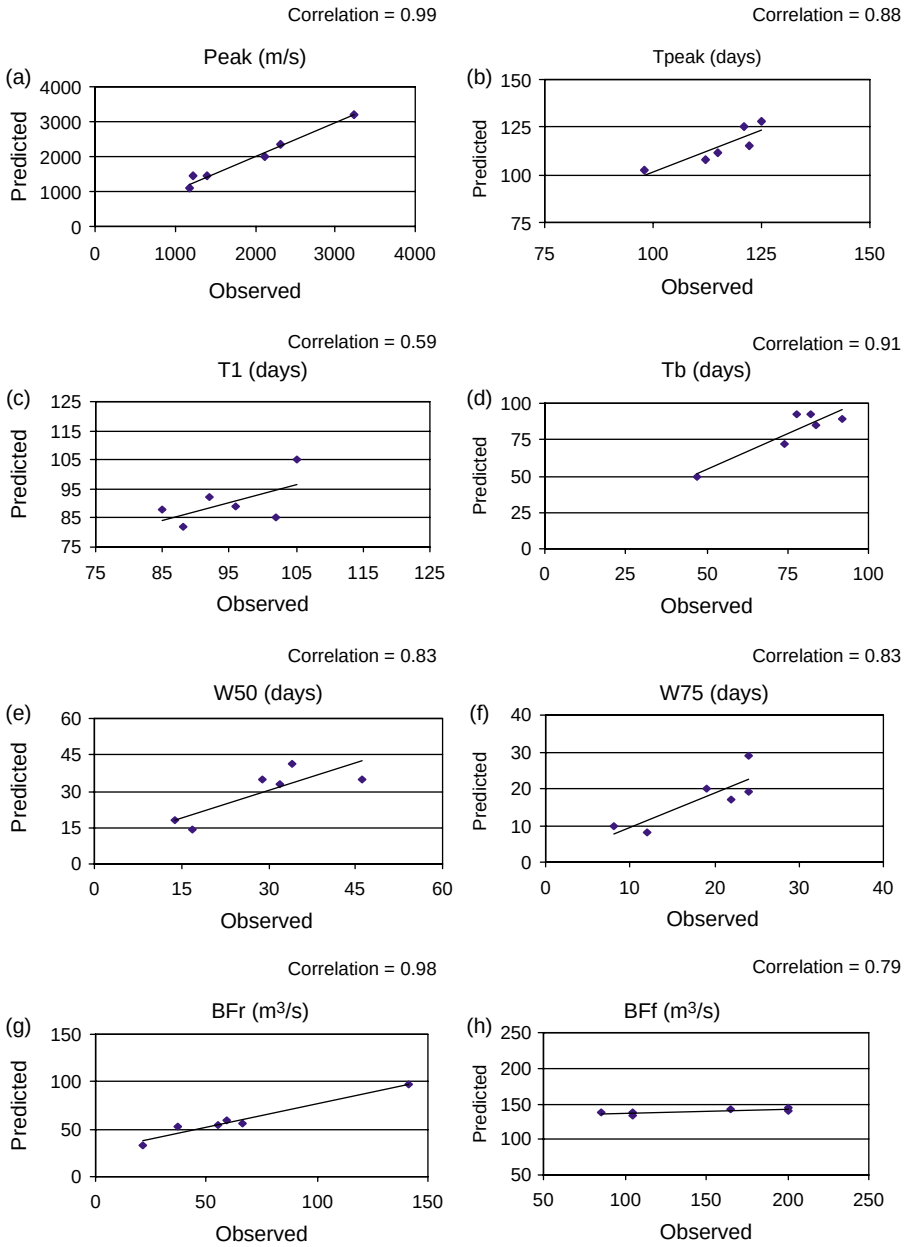


Fig. 5. (a–h) Correlation between observed and predicted value of hydrograph parameters.

ignore the results of 1997, there is no significant trend in the results. The overall network performance is average in estimating the width of hydrograph at 50% of the peak; the correlation value between observed and forecasted W_{50} is 0.83.

8.6. Width of hydrograph at 75% of the peak (W_{75})

The network estimate of W_{75} for 1997 (Fig. 3f) is excellent. The percentage error in estimating the W_{75} is in the range of −33 to 25%. The average error is

20% for all 6 years for which forecasts are made. The relationship between the peak flow and the percentage error in W_{75} is shown in Fig. 4f and error values are given in Table 3. The network performance is average in estimating the width of the hydrograph at 75% of the peak. The correlation value between observed and predicted W_{75} is 0.84.

8.7. Base flow at the rising side of hydrograph (BF_r)

The average absolute percentage error in estimating BF_r is 25% for all 6 years for which forecasts are made. The error ranges from −31 to 57% for different years. This relatively high error value might be due to the lumped values of parameters used as input for predicting runoff. The relationship between peak flow and %error in BF_r (Fig. 4g) shows a trend that error reduces for wet years. The correlation value between observed and forecasted BF_r is 0.98.

8.8. Base flow at the recession side of hydrograph (BF_f)

Network performance in estimating the BF_f is very much similar to its performance in estimating BF_r. The percentage error in estimating the BF_f is in the range of −29–62%. The average absolute error is 32% for all 6 years for which forecasts are made. This high error value might be due to the lumped values of parameters used as input to the network or it may suggest that the model is not successful in adequately representing the process by which precipitation is converted to runoff, especially at the end of rainfall event. Perhaps, some other watershed characteristics, not currently used as input into the model, affect the recession curve of the hydrograph. The relationship between the peak flow and the percentage error in BF_f (Fig. 4h) shows a trend where error is reduced for wet/high flow years. The correlation value between observed and forecasted BF_f is 0.79.

To explore the relative importance of input variables and their impact on final results, a comparison is made. The input parameters for the network are ranked in order of their effect on the output of network. This ranking is based on weights associated with each input node and being transferred to the hidden layer. The final weights associated with input parameters are calculated from several training runs and using both training and test while training data sets. The results are presented in Fig. 6 in the form of a bar chart with range of value between 0 and 1. Each bar corresponds to a different input to the network. The higher the bar, the more significant influence a particular input has on the output. Results in Fig. 6 show that summer precipitation (SP), winter precipitation (WP) and antecedent precipitation index (API) are the most important input parameters (in descending order). Melt index and timing are relatively less important factors.

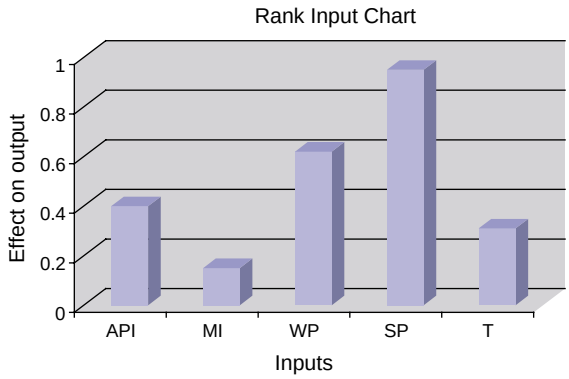


Fig. 6. Effect of input parameters on output of ANN model.

Water Resources Branch, Manitoba Department of Conservation has developed a formula (Eq. (9)) based on nonlinear regression analysis of data from past flood years to estimate peak flows at James avenue, a station located downstream of St Agathe

$$\text{Peak (cfs)} = 149\,(\text{API})^{1.5}(\text{WP}+\text{SP})^{1.5}(\text{MI})^{0.3}(T)^{0.4} \tag{9}$$

This relationship works reasonably well for forecasting peak discharge. However, the ANN model, presented here, can provide a complete description of the hydrograph including, peak flow, timing, base flow, shape of hydrograph and flood volume.

9. Summary of results

Artificial neural network proved to be time efficient and a reasonably accurate approach for the development of runoff hydrograph based on

meteorological parameters. The network is able to predict peak flow, timing of peak, timing of start of the rising side of hydrograph, and the time base of hydrograph very well with 6%, 4.17, 5.5 days and 8% average absolute error, respectively, for 6 years of prediction. Model performance is good in predicting the width of hydrograph at 50 and 75% of the peak with an average absolute error of 15–20%. The average absolute error in predicting the base flow at start and end of hydrograph is in the range of 25–30%.

The model performance in predicting the peak flow and time of peak is very good; this suggests that these hydrograph characteristics can be estimated well using the current input parameters. However, the inability of the simulated hydrograph to properly match the shape of the recession side of the hydrograph and the tendency to be out of phase may be either due to the lumped nature of input parameters used or it is an indication that additional input parameters may be required.

In general, the network performance is good during the high flows. For eight output parameters predicted by network, the highest prediction error is for the base flow at the start and end of the hydrograph. Fortunately, these are not the most important parameters in developing a runoff hydrograph, as base flow is typically 5–6% of peak flow. Considering the lumped values of parameters used for this study, the network performance is very good.

10. Conclusions and discussion

The research presented in this paper focuses on the estimation of the shape of flood hydrograph using the artificial neural network approach. The flood hydrograph on the Red River in Manitoba, Canada is estimated as a case study to demonstrate the application of the proposed technique. A feed-forward artificial neural network is trained on historic data using back-percolation algorithm. The five input parameters used are, antecedent precipitation index (API), melt index (MI), winter precipitation (WP), spring precipitation (SP), and timing factor (T). As output, the network is trained to generate information on peak flow, time of peak, base flow and timing for rising and falling sides of hydrograph and width of hydrograph at 50 and 75% of peak flow. These parameters are used to develop a complete runoff hydrograph. Evaluation statistics of percentage error and correlation are employed to estimate the accuracy of predicted hydrograph.

The ANN approach presented here does not take into account the physical processes involved in runoff generation; it is by no means a substitute for conceptual watershed modeling. However, the proposed ANN-based hydrograph estimation technique is a valuable alternative to conceptual watershed simulation techniques, where time is a constraint, limited topographic data is available, and where a complete understanding of the physical processes of watershed is not available.

One limitation of the neural network approach presented in this research, with current input parameters, is its inability to predict a two-peak hydrograph. The second peak usually occurs in the later part of the year (in July) due to rainfall only; this occurred in 1964 and 1975. The reason for this limitation lies in the nature of the input parameters used. Using lumped parameters as input (a single value for entire year), it is not possible for the network to generate two peaks. One way of dealing with this limitation is to use higher resolution data (instead of using a single input parameter value for 12 months, parameters can be calculated for say every 3 months). Further exploration is required.

Lead-time is an important issue in forecasting of hydrologic events. Parameter values for API, WP, and MI can be calculated well in advance of an anticipated flood. However, estimates of SP and T become available just before the actual flood event. SP is measured as the total basin precipitation from the start of active spring melt to the date of spring crest at Emerson (station upstream of St Agathe located at the Canadian–US border). Once flood arrives at Emerson, approximate lead-time of 3–4 days is available until flood peak reaches St Agathe. This lead-time may be sufficient for emergency response such as evacuation or operation of floodway. Any attempt to increase the lead-time will involve forecasting the values for SP and T.

For the six flood events for which forecasts were made in this paper, the input parameter values for API, SP and T were within the range of the data used for training. Input parameter values for MI in 1979

and WP in 1997 were outside the historic range. For output parameters, all values were within the normal range of data used for training with exception of peak flow in 1997. In spite of this, the network produced good forecasts for all six flood events.

This approach for estimating a flood hydrograph is generic in nature and can be applied to other locations in the Red River basin or to other rivers having similar flood characteristics. For example, transferring the model to any other location in the Red River will require training the network for output parameters extracted from hydrograph at that location, while using same input parameters. For watersheds that are similar in characteristics (e.g. heavy snow, flat topography), same architecture of ANN, in terms of input and output parameters, can be used, but data is watershed specific. There are simulation scenarios, in addition to what has been demonstrated in this study that can be tested using the existing framework. Training the model in reverse, i.e. using output parameters as input, a scenario of hydrologic and meteorological conditions can be identified that may lead to a catastrophic flood.

Acknowledgements

The authors would like to thank Mr Alf Warkentin from the Water Resource Branch, Manitoba Department of Conservation, for providing the necessary data.

References

Ahmad, S., Siminovic, S.P., 2001. In: Phelps, D., Shalke, G. (Eds.), Developing Runoff Hydrograph using Artificial Neural Networks. Proceedings of the World Water and Environmental Resources Congress: Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges, Orlando, FL.

Anctil, F., Claude, M., Charles, P., Vazken, A., 2004. A soil moisture index as an auxiliary ANN input for stream flow forecasting. Journal of Hydrology 286 (1–4), 155–167.

Anderson, J.A., 1995. An Introduction to Neural Networks. MIT Press, Cambridge.

ASCE Task Committee, 1993. Definition of criteria for evaluation of watershed models. Journal of Irrigation and Drainage Engineering 119 (3), 429–442.

ASCE Task Committee, 2000a. Artificial neural networks in hydrology I: preliminary concepts. Journal of Hydrologic Engineering 5 (2), 115–123.

ASCE Task Committee, 2000b. Artificial neural networks in hydrology II: hydrologic applications. Journal of Hydrologic Engineering 5 (2), 124–137.

Box, G.E., Jenkins, G.M., 1976. Time Series Analysis: Forecasting and Control. Holden-Day, Oakland, CA.

Carriere, P., Mohaghegh, S., Gaskari, R., 1996. Performance of a virtual runoff hydrograph system. Journal of Water Resources Planning and Management, ASCE 122 (6), 421–427.

Elshorbagy, A., Simonovic, S.P., 2000. Performance evaluation of artificial neural networks for runoff prediction. Journal of Hydrologic Engineering, ASCE 5 (4), 424–427.

Fernando, D.A., Jayawardena, A.W., 1998. Runoff forecasting using RBF networks with OLS algorithm. Journal of Hydrologic Engineering, ASCE 3 (3), 203–209.

Hsu, K.-L., Gupta, H.V., Sorooshian, S., 1995. Artificial neural network modeling of the rainfall–runoff process. Water Resources Research 31 (10), 2517–2530.

IJC, 1997. Red River Flooding: Short-Term Measures, Interim Report of the International Red River Basin Task Force, Ottawa, Washington.

Jurik, M., 1990. Back-percolation: assigning local error in feed-forward perceptron networks. Report by Jurik Research & Consulting P.O. 2379, Aptos, California, USA 95001. Web reference: http://www.jurikres.com/down/backperc.pdf.

Karunanithi, N., Grenney, W.J., Whitley, D., Bovee, K., 1994. Neural networks for river flow prediction. Journal of Computing in Civil Engineering, ASCE 8 (2), 201–220.

Liong, S.-Y., Lim, W.-H., Paudyal, G.N., 2000. River stage forecasting in Bangladesh: neural network approach. Journal of Computing in Civil Engineering 14 (1), 1–8.

Maren, A., Harston, C., Pap, R., 1990. Handbook of Neural Computing Applications. Academic Press, San Diego.

Markus, M., Salas, J.D., Shin, H.-K., 1995. Predicting streamflows based on neural networks. Proceedings, First International Conference on Water Resource Engineering, ASCE, 1641–1646.

Moradkhani, Hamid, Kuo-lin Hsu, Hoshin, Gupta, V., Sorooshian, S., 2004. Improved streamflow forecasting using self-organizing radial basis function artificial neural networks. Journal of Hydrology 295 (1–4), 246–262.

Muttiah, R.S., Srinivasan, R., Allen, P.M., 1997. Prediction of two-year peak stream discharges using neural networks. Journal of American Water Resources Assocation 33 (3), 625–630.

Özgür, K., 2004. River flow modeling using artificial neural networks. Journal of Hydrologic Engineering 9 (1), 60–63.

Rajurkar, M.P., Kothyari, U.C., Chaube, U.C., 2004. Modeling of the daily rainfall–runoff relationship with artificial neural network. Journal of Hydrology 285 (1-4), 96–113.

Royal Commission, 1958. Report of the royal commission on flood cost benefit, Winnipeg, Manitoba.

Singh, R., Subramanian, K., Refsgaard, J.C., 1999. Hydrological modeling of a small watershed using MIKE SHE for irrigation planning. Agricultural Water Management 41, 149–166.

Smith, J., Eli, R.N., 1995. Neural-network models of rainfall–runoff process. Journal of Water Resources Planning and Management, ASCE 121 (6), 499–508.

Thirumalaiah, K., Deo, M.C., 1998. River stage forecasting using artificial neural networks. Journal of Hydrologic Engineering, ASCE 3 (1), 26–32.

Tokar, A.S., Markus, M., 2000. Precipitation–runoff modeling using artificial neural networks and conceptual models. Journal of Hydrologic Engineering 5 (2), 156–161.

US Army Corps of Engineers, 1998. HEC-1: Flood Hydrograph Package, User Manual, CPD-1A, Hydrologic Engineering Center, Davis, California.

Warkentin, A.A., 1999. Red River at Winnipeg: Hydro-meteorological Parameter Generated Floods for Design Purposes, CWRA Symposium on Red River Flooding—Decreasing our Risks, Winnipeg, Manitoba 1999.

Zealand, C.M., Burn, D.H., Simonovic, S.P., 1999. Short term stream flow forecasting using artificial neural networks. Journal of Hydrology, Amsterdam 214, 32–48.

Zhu, M., Fujita, M., Hashimoto, N., 1994. Application of neural networks to runoff prediction. In: Hipel, K.W. et al. (Ed.), Stochastic and Statistical Method in Hydrology and Environmental Engineering, vol. 3. Kluwer, Dordrecht, The Netherlands, pp. 205–216.