International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

Sequential Neural Network with Error Updating for Improved Higher Lead Time Flood Forecasts

Om Prakash[1], K.P. Sudheer[2] and K. Srinivasan[3]
Department of Civil Engineering, Indian Institute of Technology Madras
Chennai - 36, INDIA
Email: [1]omprakash@iitm.ac.in; [2]sudheer@iitm.ac.in; [3]ksrini@iitm.ac.in,

ABSTRACT: A novel ANN architecture is proposed for forecasting river flows at higher lead times with greater accuracy. The paper predominantly demonstrates the potential in computing paradigm, through 'sequential ANN (SANN)', to extend the lead time of forecast. In SANN, a series of ANNs are connected sequentially,, each of them taking forecast value from an immediate preceding network as input. The output of each network modifies itself by adding an expected value of error so that residual variance of the forecast series is minimized. The efficacy of the developed model has been tested through a real case study for the data on hourly river flow forecasting for Kolar River, India. The binary-coded genetic algorithm is used to establish the weights among the neurons because of the dynamic nature of input layer in SANN model. The main objective function of the proposed model is to minimize the root mean square error. Our results demonstrate that the SANN is capable of providing accurate forecasts up to 8 hours ahead. The SANN model tends to preserve the performance at higher lead times compared to both ANN1 and ANN2 models.

INTRODUCTION

Flood forecasting essentially aim at ensuring the structural safety of the hydraulic structures like dams; providing optimum protection against inundation of urbanized area; minimizing the adverse impact on the commercialized activities like hydropower generation; navigation and the living habitat in the downstream of the river along with retaining sufficient storage to fulfill the ever increasing demand of the multipurpose projects associated with the adjoining river or reservoir. A good flood forecasting system should address technical issues that would make it possible to provide an accurate forecast with sufficient lead time. Also, an improvement in the accuracy of the higher lead time flood forecast enables better mitigation of anticipated flood damage at the control points through improved operation of the reservoirs.

HISTORICAL PERSPECTIVE

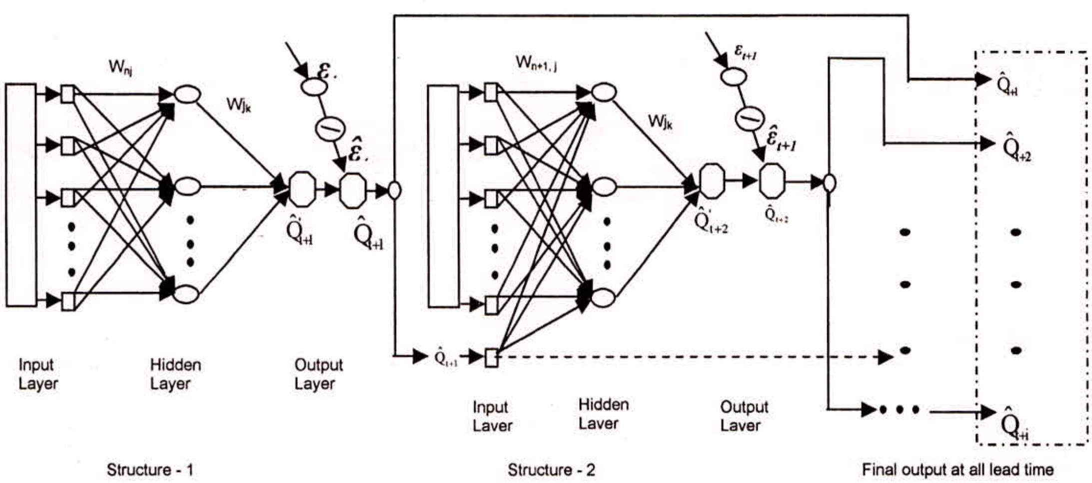
Traditionally, flood forecasting systems have been developed by coalescing conceptual hydrological models for the land-phase with suitable hydraulic routing models to simulate flood propagation throughout the drainage network (Arduino *et al.*, 2005). Such models are highly complex usually compromised by linearity, analogous to nonlinear dynamics, and calibration of such models is not trivial (Duan *et al.*, 1992; Hsu *et al.*, 1995). The inherent limitation of traditional conceptual models restricts its application in representing the real world system dynamics. Owing to the complexities associated with physically based hydrologic and hydrodynamic simulation models, nonlinear system theoretic techniques, called Artificial Neural Networks (ANN) was developed, to provide flood forecasting (Hsu *et al.*, 1995). The development of ANN technique has resulted in a plethora of application in hydrology, most of them pertaining to rainfall-runoff modeling in flood forecasting (ASCE Task Committee, 2000a, b; Maier and Dandy, 2000; Dawson and Wilby, 2001; Dawson *et al.*, 2006). Despite its successful functioning in rainfall-runoff modeling, ANNs have not been deployed in operational flood warning systems. This has been attributed to various inherent practical limitations and inaccurate flood forecasting at higher lead times (Dawson *et al.*, 2006; Bruen and Yang, 2005; Birkundavyi *et al.*, 2002). On the contrary, Campolo *et al.* (1999) noted that the capacity of a basin to respond to the perturbation is more accurate when recent input information is used. This emphasizes the fact that, an improved forecast at higher lead time is possible only when updated information about the basin saturation is provided to the network.

Continued endeavors on updating the output variables dates back to early 1990's. Data assimilation procedure implemented in the forecast system improves the estimate of initial state of the system and reduces the simulation errors in the forecast period (Madsen and Skotner, 2005; WMO, 1992). In addition, an error correction mechanism is also been used successfully, to update the output variables for better forecast at higher lead time (Refsgaard, 1997; Madsen, 2000). Recently, Shamseldin and O'Conner (2001) applied an error correction mechanism in ANNs for daily flood forecasting (till 4 days) to improve the forecast accuracy. Since error correction forecast model is superimposed on the simulation model (ANN), it can neither modify the model parameters nor the model internal storage contents.

These restrictions of the existing models, entail the need for a more sophisticated model for updating the output variables. As a result, a novel ANN architecture model is proposed in this paper, to forecast the river flows at higher lead times with greater accuracy. This paper predominantly demonstrates the computing potential of the proposed model called 'Sequential ANN (SANN), in extending the lead time of forecast. The 'sequential ANN (SANN)',which is a series of ANNs are connected sequentially to extend the lead time of forecast, each of them taking a forecast value from an immediate preceding network as input. The output of each network is modified by adding an expected value of error so that residual variance of the forecast series is minimized. The applicability of the model in flood forecasting is illustrated through a real world case study. The paper also intends to evaluate the relative performance of SANN with that of two traditional ANN models (called as ANN1 and ANN2 considering single and multi-output neurons in the output layer respectively), developed for the same case study.

ARTIFICIAL NEURAL NETWORKS AND SANN FOR FLOOD FORECASTING

Three-layer feed-forward neural networks have been widely used for hydrological modeling. Earlier works reveal that three layers are sufficient to simulate the dynamic and nonlinear properties of the rainfall-runoff transformation (Lippmann, 1987). A neural network consists of a set of neurons, logically arranged into two or more layers (Takahashi, 1993), namely, the input layer, the hidden layer and the output layer. Input layer consists of input neurons that receive the external stimuli but do not carry out any signal processing. There is one hidden layer between the input and the output layers. The hidden layer is connected with the input layer and the output layer using adjustable weights as measures of correlation between each layer of neurons. The output neurons, on the other hand, are used for outputting the processed signals. Neural networks are trained using a set of observed input and output data pairs (called patterns) (Camargo and Yoneyama, 2001), which is repeatedly processed to calibrate ANN. During calibration, the network weights gradually converge to values such that each input vector produces output values that are in close proximity to the desired target output vector. The activation function used in the hidden layer as well as output layer is non-linear sigmoid function. Moreover, in case of SANN that integrates the series of ANN models, each taking input from a previous ANN for extending the lead time of forecasts, and an error updating procedure together in a single framework (Figure 1).



n = number of inputs neurons in the input layer; j = number of hidden neurons in the hidden layer; k = number of neurons in the output layer, in this case k = 1; i = lead time; t = time; W = connection weight between the neurons; $\hat{Q}'$ = network output; $\hat{Q}$ = forecast value (network output after correction); ε = forecast error; and $\hat{\varepsilon}$ = output from the error neurons which employs an error model.

Fig. 1: Sequential Artificial Neural Network architecture

The training of ANN1, ANN2 and SANN structures entails a training algorithm or a learning algorithm to establish the weights among the neurons. It is to be noted that the SANN model parameters gets added at every lead time of forecast because number of input increases with forecast lead time. Hence, the network parameter optimization by traditional methods becomes tedious or infeasible. However, the parameter optimization of the SANN being a nonlinear optimization problem, any nonlinear optimization algorithm can be employed for SANN model identification. In the contemporary work, binary-coded genetic algorithms (Holland, 1975) have been deployed to train the neural networks model. The objective function of the parameter estimation is to minimize the total forecasting error defined in terms of root mean square error (RMSE). The Genetic Algorithm Toolbox available with the MATLAB has been used for parameter estimation (Mathworks, 2004).

MODEL APPLICATION

The computing potential of the proposed SANN model has been illustrated by applying it to a real case example (Kolar River basin, India), to predict the river flow up to 8 hours in advance. The rainfall and runoff data on an hourly interval for Kolar basin in India during the monsoon season (July, August, and September) for three years (1987–1989) are used for the study which is depicted in Figure 2. The Kolar River is a tributary of the river Narmada that drains an area of about 1350 sq km before its confluence with Narmada near Neelkanth. In the present study the catchment area up to the Satrana gauging site is considered, which constitutes an area of about 903.87 sq km. The 75.3 km long river course lies in between north latitude of 21° 09′ N to 23° 17′ N and east longitude of 77° 01′ E to 77° 29′ E. The rainfall data available were in the form of real average values in the basin. The total available data set was divided into two equal sets in terms of number of patterns employed for calibration and validation of the model.

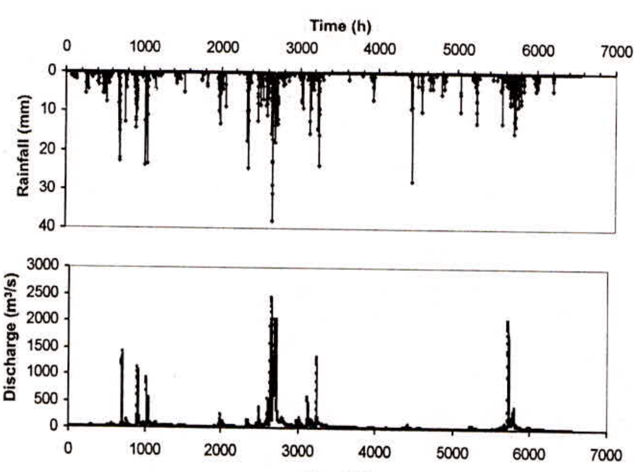


Fig. 2: Rainfall and runoff series of the Kolar basin

DETECTION OF INPUT PATTERNS

The current study employs a statistical approach suggested by Sudheer *et al.* (2002) to identify the appropriate input vector. The method is based on the heuristic that potential influencing variables corresponding to different time lags can be identified through statistical analysis of the data series. The procedure uses cross-, auto-, and partial auto-correlations among the variables in question at 95% confidence interval. The input vector identified for modeling the river flow, include a total of 5 variables and the functional form of the model is,

$$\left[Q(t+1),Q(t+2),...,Q(t+8)\right] = f\left[Q(t-1),Q(t-2),R(t-7), R(t-8),R(t-9)\right] \quad ...(1)$$

where $R(t)$ is the precipitation at any given hour t, $Q(t)$ is the river flow at any time t. Note that the current study explores forecasting the time series up to 8 steps ahead mainly for the demonstration of increased lead time forecasts.

MODEL ASSESSMENT

The purpose of assessing the performance of a hydrological model is to provide a quantitative appraisal of the model ability to reproduce historic and predict future behavior. Often, evaluations of model performance utilize a number of statistics and techniques, usually referred to as "goodness of fit" statistics. Many of the principal measurements that have been used in the hydrological literature is critically reviewed by Legates and McCabe (1999); Hsu *et al.* (1995). The effectiveness of any model could be ascertained by some of the statistical error measurements like correlation coefficient, coefficient of efficiency, relative error, etc. Due to the non availability of single evaluation measure (Sudheer and Jain, 2003), a multi criteria assessment was performed in the current study with different goodness of fit statistics. These measures could be grouped into two types: relative and absolute. Relative goodness of fit measure are nondimensional indices, which provide a relative comparison of the performance of one model against another. The relative statistics indices which are considered in this study include Coefficient of Correlation (CC), the Coefficient of Efficiency (CE) (Nash and Sutcliffe, 1970) and the Relative Error Peak Flow (REPF). In contrast, absolute goodness of fit statistics is measured by the units of flow measurement. The criterion for evaluation that is employed is the Root-Mean-Square Error (RMSE) between the observed and forecasted values, the Standard Error of Estimate (SEE) and the noise to signal ratio (Nayak *et al.*, 2005; Dawson and Wilby, 2001; Kneale *et al.*, 2001).

RESULTS AND DISCUSSIONS

Forecasts at 1-Lead Time

The statistical performance indices of the identified models for the 1 hour ahead forecast for Kolar basin is summarized in Table 1. The correlation statistics, which evaluate the linear correlation between the observed and computed runoff (Hsu *et al.*, 1985), is consistent (> 0.97) for all models during calibration as well as for the validation period. The model efficiency that evaluates the capability of the model in predicting runoff values away from the mean are found to be more than 93% during the calibration and validation periods for all models, which proves to be very satisfactory in accordance to the works of Shamseldin (1997). The RMSE statistic, which indicates a quantitative measure of the model error in units of the variable, is also satisfactory for all the models during both calibration and validation periods. The mean bias error that measure the efficiency of the model, is found to be very-very less in SANN when compared to other models. The Volume Error (VE) statistics, which measures the error in volume (bias) between the observed and computed runoff hydrographs, was negligible and showed consistently good performance for SANN model. In general, for a 1-step ahead forecast, although the performance of all the models are compatible performance of SANN model is proved best as measured by statistical indices.

Forecasts at Higher Lead Times (> 1 Step)

The variation of RMSE with different lead times is presented in Figure 3. From the Figure 3, it can be observed that the models ANN1 and ANN2 shows an increasing trend in RMSE with lead time, while SANN model has no significant increase of RMSE with lead time. Also , it is evident that the slope of the RMSE vs. prediction time horizon is the least in case of SANN model during both calibration and validation period. The SANN model forecasted the flows with a RMSE of values of 76.69 m^3/s, whereas ANN1 and ANN2 models forecasts an values of 130.80 m^3/s and 137.45 m^3/s respectively during validation period at a lead time of 8 hours. Note that SANN has a slightly higher value of RMSE compared to other two models at 1 hour ahead forecast, but SANN showed little deterioration at higher lead times of prediction. On average, the model SANN perform best as measured by this statistics for higher lead times.

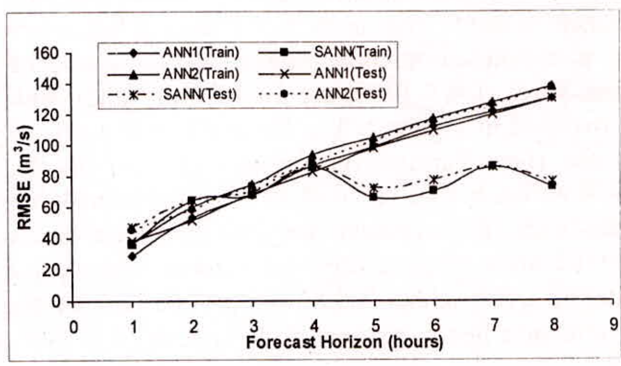


Fig. 3: Variation of RMSE along the forecast time horizon for Kolar basin

Table 1: Performance Statistics of All Models at 1-Hour Lead Forecasting for Kolar Basin

	Calibration			*Validation*		
Performance Index	*ANN1*	*SANN*	*ANN2*	*ANN1*	*SANN*	*ANN2*
CC	0.9892	0.9830	0.9830	0.9801	0.9711	0.9735
CE	0.9781	0.9650	0.9648	0.9602	0.9405	0.9464
RMSE	28.7442	36.3379	36.4674	38.6877	47.3295	44.9176
VE	−0.0716	0.0000	−0.1203	−0.0667	0.0000	−0.1208
SEE	29.0854	42.8219	37.3490	39.1470	55.7747	46.0035
NS	0.1496	0.2203	0.1922	0.2018	0.2875	0.2371
REPF	−0.0490	−0.0124	−0.0331	−0.0619	−0.0301	−0.0463

(–ve) values indicate underestimation; (+ve) values indicate overestimation.

The noise to signal ratio statistics is shown in Figure 4 at different lead times. From the Figure 4, it can be illustrated that although all the models have comparable value of this performance index at 1st lead time, the performance of models ANN1 and ANN2 is found to deteriorate at higher lead times, while SANN model has trivial increment with lead time. It is also noticed that SANN model has lowest noise to signal ratio gradient with lead time. However, it is worth mentioning that the value of the noise to signal ratio is less than 0.5 for the model SANN even up to 8th lead times, while the models ANN1 and ANN2 are also not exceeded the limiting value but more than the SANN at higher lead times. The model SANN performs best as measured by this statistics.

The error statistics discussed above provide relevant information on overall performance of the models, but do not provide specific information about model performance at peak flow, which is of critical importance in any flood forecasting context. Hence two additional event-specific evaluation measures were considered: percentage relative error peak flow and time difference to peak. The relative error peak flow is computed as the ratio of peak residual and observed values expressed as a percentage, whereas, time difference to peak is computed between observed and predicted value. The forecast error on a few typical peaks flow during the period of the analysis is presented in Tables 2 during the validation period. From the results, it can be concluded that the value of the percent error in peak flow prediction, which is a useful index in simulating floods events, is within acceptable limits by model SANN. Note that these peak flows were observed at different periods of time, and do not correspond to the same flood event. Furthermore, the model SANN was able to forecast most of the peaks with reasonable accuracy even up to 8 hours in advance, while the error increased drastically with an increase in prediction lead time for models ANN1 and ANN2. On the other hand, it is worth mentioning that the model SANN was able to forecast the peak flows with minimum error, irrespective of the magnitude of the peak flow.

Table 2: Comparison of Model Estimated Hydrograph Characteristics at Different Forecast Lead Times during Validation Period for Kolar Basin

		Percent Error in Forecasted Peak Flows			*Time Difference to Peak (hours)*		
Forecast Lead Time	*Observed Peak Flow*	*ANN1*	*SANN*	*ANN2*	*ANN1*	*SANN*	*ANN2*
1 hour	1392.1	5.064	–26.945	11.443	0	0	0
1 hour	2427.7	3.188	3.011	4.613	–1	–1	0
1 hour	2029.0	5.352	7.935	4.071	1	0	1
2 hours	1392.1	7.155	–7.931	14.367	0	0	0
2 hours	2427.7	9.495	1.495	9.804	0	–1	0
2 hours	2029.0	9.803	2.489	7.447	1	–1	1
3 hours	1392.1	6.113	–4.526	17.140	–1	–1	–1
3 hours	2427.7	13.276	7.031	10.557	1	–1	1
3 hours	2029.0	14.993	7.156	10.907	0	0	0
4 hours	1392.1	9.288	10.639	17.887	–1	–1	–1
4 hours	2427.7	18.161	2.130	15.995	1	–1	1
4 hours	2029.0	17.122	3.499	18.847	2	–1	0
5 hours	1392.1	20.473	–14.733	25.867	–2	0	–2
5 hours	2427.7	11.352	–7.798	23.224	0	0	0
5 hours	2029.0	9.783	12.400	26.925	1	–1	0
6 hours	1392.1	31.436	–30.659	30.864	–2	–1	–2
6 hours	2427.7	12.716	–23.339	29.180	1	2	1
6 hours	2029.0	9.783	–34.825	8.167	3	1	3
7 hours	1392.1	30.423	6.724	34.178	–2	–1	–3
7 hours	2427.7	25.695	1.594	35.342	–1	–1	0
7 hours	2029.0	15.392	2.440	18.216	0	–1	1
8 hours	1392.1	39.530	–26.227	35.824	–2	–1	0
8 hours	2427.7	32.084	–8.209	39.239	0	0	–1
8 hours	2029.0	16.284	–2.109	23.968	1	–1	1

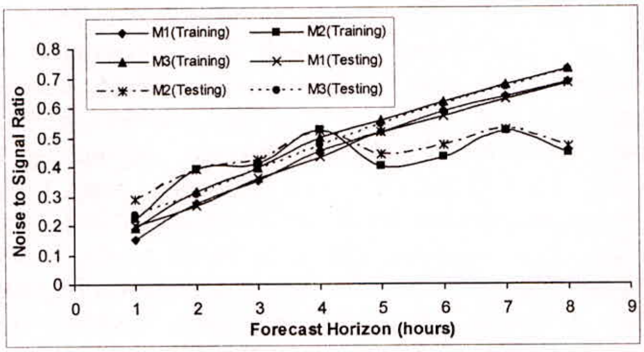


Fig. 4: Variation of noise to signal ratio along the forecast time horizon for Kolar basin

SUMMARY AND CONCLUSIONS

Despite innumerable methods available for flood forecasting, none of them are effective in providing accurate forecasts at higher lead times. The accuracy of forecasts gets deteriorated as the forecast lead time increases. This paper presents a novel ANN architecture model that can be employed for forecasting river flows at higher lead times with greater accuracy. The focal point of the paper is to demonstrate the potential of the proposed computing paradigm in extending the lead time of forecast. The proposed model, called 'sequential ANN (SANN)', is a series of ANNs connected sequentially to extend the lead time of forecast, each of them taking a forecast value from an immediate preceding network as input. The output of each network is modified by adding an expected value of error so that residual variance of the forecast series is minimized. The developed model is tested for Kolar Basin, India for a hourly river flow forecast. Since, SANN model parameters gets added at every lead time of forecast , a binary-coded genetic algorithms is used to train the SANN model parameters with the objective of minimizing the total forecasting error, in terms of root mean square error (RMSE). The results of the study shows that the model SANN is implicitly doing a better job than the contemporary models ANN1 and ANN2 in modeling the rainfall-runoff process. The performance of these models was acceptable at 1st lead time, but only the model SANN excels in preserving its performance at higher lead times as compared to the ANN1 and ANN2 models. This data put forth the arena for a new and promising research area in the field of flood forecasting.

REFERENCES

Arduino, G., Reggiani, P. and Todini , E. (2005). Recent advances in flood forecasting and flood risk assessment, Hydrology and Earth System Sciences, 9(4), 280–284.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a). "Artificial neural networks in hydrology I: Preliminary concepts", *J. Hydrol. Engg.*, 5(2), 115–123.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b). "Artificial neural networks in hydrology II: Hydrologic applications", *J. Hydrol. Engg.*, 5(2), 124–137.

Birikundavyi, S., Labib, R., Trung, H.T. and Rousselle, J. (2002). "Performance of neural networks in daily streamflow forecasting", *Journal of Hydrologic Engineering*, Vol. 7, No. 5, pp. 392–398.

Bruen, M. and Yang , J. (2005). Functional networks in real-time flood forecasting- novel application, Advances in Water Resour., 28, 899–909.

Camargo, L.S. and Yoneyama, T. (2001). Specification of Training Sets and the Number of Hidden Neurons for Multilayer Perceptrons. Neural Computation, 13, 2673–2680.

Campolo, M., Andreussi, P. and Soldati, A. (1999). "River flood forecasting with neural network model", *Water Resour. Res.*, 35(4), 1191–1197.

Dawson, C.W. and Wilby, R.L. (2001). Hydrological modelling using artificial neural networks, Progress in Physical Geography, 25(1), 80–108.

Dawson, C.W., See, L.M., Abrahart, R.J. and Heppenstall, A. (2006). Symbiotic adaptive neuro-evolution applied to rainfall–runoff modelling in northern England, Neural Networks, 19, 236–247

Duan, Q., Sorooshian, S. and Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res. 28(4), 1015–1031.

Holland, J.H. (1975). Adaptation in Natural and Artificial System. Mass. Inst. of Technol. Cambridge.

Hsu, K., Gupta, V.H. and Sorroshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process, Water Resour. Res., 31(10), 2517–2530.

Kneale, P.E., See, L. and Smith, A. (2001). "Towards defining evaluation measures for neural network forecasting models." *Proceedings of the Sixth International Conference on GeoComputation,* University of Queensland, Australia, available at http://www.geocomputation.org/2001/papers/kneal.pdf (accessed on March 2, 2007).

Legates, D.R. and G.J. McCabe, Jr. (1999). "Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation", *Water Resour. Res.,* 35(1), 233–241.

Lippmann, R.P. (1987). An introduction to computing with neural nets. IEEE ASSP Magazine, 4–22.

Madsen, H. (2000). "Automatic calibration of a conceptual rainfall–runoff model using multiple objectives", *J. Hydrol.*, 235, 276–288.

Madsen, H. and Skotner, C. (2005). "Adaptive state updating in real-time flow forecasting—a combined filtering and error forecasting procedure", *J. Hydrol.*, 308, 302–312.

Maier, H.R. and Dandy, G.C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, Environmental Modelling & Software, 15, 101–124.

Mathworks (2004). Genetic algorithm and direct search tool box for use in MATLAB, The MathWorks, Inc., Natick, MA.

Nash, J.E. and Sutcliffe, J.V. (1970). "River flow forecasting through conceptual models." 1. *A discussion of principles, J. Hydrol.*, 10, 282–290.

Nayak, P.C., Sudheer, K.P., Rangan, D.M. and Ramasastri, K.S. (2005). Short-term flood forecasting with a neurofuzzy model, Water Resour. Res., 41(4), W04004.

Refsgaard, J.C. (1997). Validation and intercomparison of different updating procedures for real-time forecasting, Nordic Hydrol., 28, 65–84.

Shamseldin, A.Y. (1997). Application of a neural network technique to rainfall runoff modeling, J. Hydrol., 199, 272–294.

Shamseldin, A.Y. and O'Connor, K.M. (2001). A non-linear neural network technique for updating of river flow forecasts, Hydrology and Earth System Sciences, 5(4): 577–597.

Sudheer, K.P., Gosain, A.K. and Ramasastri, K.S. (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models, Hydrol. Processes, 16, 1325–1330.

Sudheer, K.P. and Jain, S.K. (2003). "Radial basis function neural networks for modeling stage discharge relationship," *J. Hydrol. Engg.*, 8(3), 161–164.

Takahashi, Y. (1993). Generalization and approximation capabilities of multilayer networks. Neural Computation, 5, 132–139.

WMO (1992). "Simulated real-time intercomparison of hydrological models." *Operational Hydrology Report No. 38,* World Meteorological Organization, Ganeva.