Short term river flood forecasting with neural networks

R.García-Bartual

Department of Hydraulic and Environmental Engineering,
Universidad Politécnica de Valencia, Camino de Vera s/n, 46071 - Valencia, Spain
(rgarciab@hma.upv.es)

Abstract: This paper reports results obtained using artificial neural networks (ANN) models for short-term river flow forecasting under heavy rain storms, in the upper Serpis river basin (460 km^2), with the outlet in Beniarrés reservoir (29 hm^3). The system is monitored by 6 raingauges, providing 5-min rainfall intensities, while reservoir inflows are derived from depth measurements in the reservoir every half hour and real-time data from controlled discharges in the spillway. In order to produce 1, 2 and 3 hours forecasts, the model makes use of the distributed rainfall information, together with observed discharges in the preceding hours. Several ANN topologies have been tested and compared, including linear and non linear schemes, being in all cases three-layer feedforward networks. Best results are obtained with different architectures for each forecasting horizon, basically due to the decreasing dependence of future inflows with respect to preceding values of the series as the time horizon is increased, while rainfall information increases its importance as a predictor. The ANN architectures finally proposed are achieved through pruning algorithms. Training is performed using the quasi-Newton approach.

A specific software was also develop to help in the real-time management of the dam during floods, which incorporates all the relevant information about the dam elements (Tainter-gates, depth-volume relationships,...), and is designed for real-time operation, accepting as inputs the rainfall measurements, reservoir levels and gates opening. Forecasts are made in real-time, including inflows to the dam and discharges under different assumptions for gates operations during the inmediate future hours.

***Keywords**:* Neural networks; Perceptron multilayer; Floods forecasting; Real-Time Operation

1. INTRODUCTION

Flash-floods are particularly complex and dangerous phenomena, which have produced in the past important economic losses and in some cases, life losses. A flood warning system is a technical way to effectively reduce such risks, supported by a real-time data collecting system and a running module producing future runoff estimations. If the basin or hydrological system under consideration includes a dam or several dams equipped with control gates, improved criteria for gates operation during the flood can be assessed if reliable forecasts of inflows to the reservoir are available.

There have been many recent papers and contributions proving the ability and potentials of artificial neural networks (ANN) modelling approaches in the field of rainfall-runoff modelling and time-series forecasting, both rainfall and runoff series [Lachtermacher, G. and Fuller, J. D.,1994; Boogard et al, 1998; Zealand et al, 1999; Luck et al, 2000; Coulibaly et al, 2000; Toth et al, 2000; Deo et al, 2000]. Although the ANN model building process is not yet consolidated in a systematic way, as outlined by Maier [Maier et al, 2000], there is no doubt about the practical benefits that can be derived from its use in different areas of applied hydrology and environmental sciences, whenever no physical considerations or knowledge is strictly required. On the other hand, results show that such techniques are effectively competing, and in many instances outperforming, other models based on linear recurrence relations or conceptual approaches [Hsu et al, 1995].

In the research reported herein, different ANN models are used to investigate and reproduce the unknown relationship existing between distributed rainfall over the basin and inflow discharges to Beniarrés Dam, located in the South East of Spain (Alicante's province). Such developments aim to the real-time prediction of floods from available rainfall data collected in the basin and specific data collected in the reservoir and dam, during the event. The sample data used for training and validation includes 2 historical events and 22 synthetic, stochastically generated rainfall patterns and runoff data using RAINGEN model [García-Bartual, 2001]. A software called PCTR-BENIARRÉS has been developed as a practical tool to help decision making during real-time operation of the dam, which incorporates the ANN models as prediction module, and is adequately inserted in a wider framework involving dam operation during the flood event.

2. STUDY AREA AND DATA SET

The upper Serpis basin covers an area of 460 km^2, with the outlet at Beniarrés reservoir (figure 1). The estimated concentration time is 10 hours, with isocrones depicted in figure 1. The dam has a storage capacity of 29 hm^3, and is equipped with three radial gates with spillways providing a maximum discharge capacity of 1050 m^3/s (fig 2).

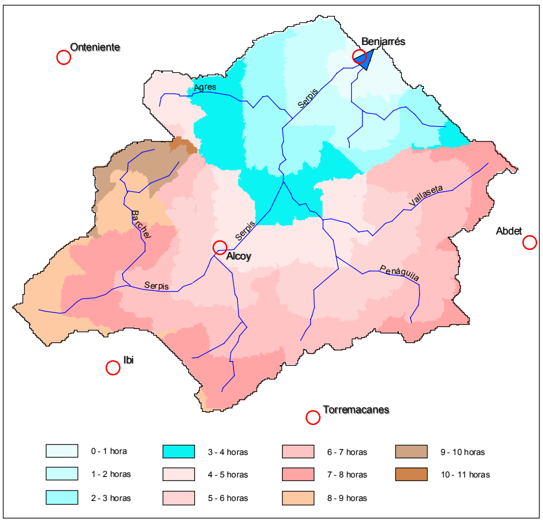

Figure 1. **Upper Serpis River Basin**

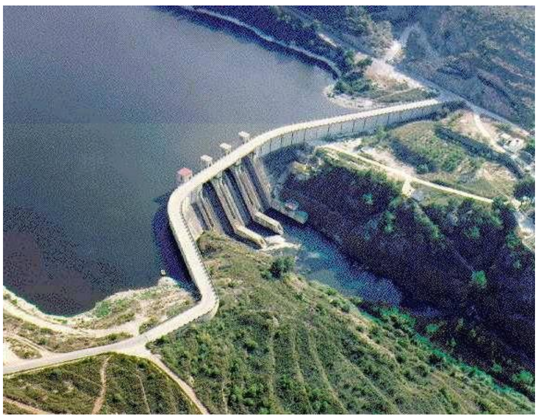

Figure 2. Beniarrés Dam

There are 6 automatic pluviometers measuring rainfall depths with 5-minutes time level of aggregation, with geographical location is shown in figure 1.

The estimated flood discharge for return period 100 years is 1000 m^3/s.

Only two historical events are available. Figure 3 shows the corresponding inflow hydrograph for one of them, occurred during February 1993.

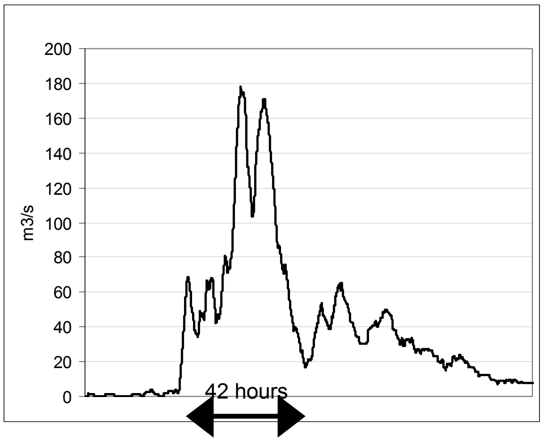

Figure 3. Inflow hydrograph (1-4 feb 1993)

Synthetic data is additionally used, enabling training with longer series, including hydrographs with peaks over 1000 m^3/s. We made use of total of 22 synthetic events generated with a stochastic rainfall generator previously calibrated with historical rainfall data. Such data were recorded over a wider geographical area, which includes the basin under consideration and has a climatic characteristics which are representative for the purpose of extreme rainfall modelling in the region. Figures 4 and 5 show the synthetic values or rainfall in Alcoy and runoff (inflow) in Beniarrés, for the synthetic event number C-10.

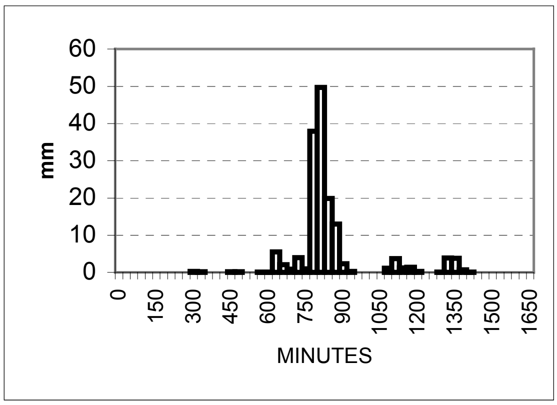

Figure 4. Rainfall in Alcoy – Event no. C-10

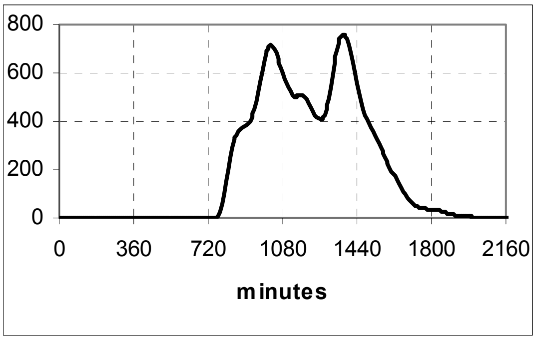

Figure 5. Inflow hydrograph in Beniarrés (C-10)

3. DIVISION OF DATA AND PRE-PROCESSING

The data are divided in three different groups. First, the calibration or training set, including one historical event and 12 synthetic events. A verification set is used for cross-validation, assessing the performance of the model at various stages of learning, including 4 synthetic events. Finally, a different independent test set including the remaining data is reserved for validation of the model, testing its generalisation ability achieved.

The pre-processing is applied to both, rainfall and discharge data, using equations 1 and 2.

$$p = \frac{P^{\alpha}}{P_m^{\alpha}} \qquad (1) \qquad\qquad q = \frac{Q^{\beta}}{Q_m^{\beta}} \qquad (2)$$

where p, q are the scaled values of rainfall and runoff, being P and Q the original values expressed in mm and m^3/s, respectively. The exponents α, β were adequately chosen to reduce the skewness of the samples, with respective values of 0.2 and 0.5. The values P_m and Q_m are introduced in the formulas for a convenient standarization, ensuring equal attention to all variables during training of the network. Q_m was taken equal to 1200, and P_m was given different values depending on the time interval chosen for rainfall aggregation. For Δt =1 hour, P_m= 100, and for Δt =1.5 hours, P_m= 120. These values were adopted largely by convenience, without particular attention to any rigorously established criteria. In fact, it has not been proved that functions (1) and (2) overperform other normalization techniques more usual in the literature, but it is also clear that they provide a suitable and straight-forward transformation of input data for simultaneous skewness reduction and scaling to uniform ranges. Nevertheless, unbounded transfer functions in the output layer were later used in all tested ANN topologies, and thus, scaling is not strictly needed, although recommended [Masters, 1993; Maier and Dandy, 2000].

4. NETWORK ARCHITECTURE

4.1 Input variables selection

Selection of the input variables set as predictor section of the exemplars involved basically three aspects. First, the number of previous runoff values. Second, which rainfall stations consider. And third, which time lag consider between each rain-gauge station data and runoff data corresponding to Beniarrés inflows. The lag is influenced by the distance of the rain gauge from the pluviometer, but also depends on the runoff generation dynamics and the state of the basin. For each of the pluviometers, we analized the problem in three ways:

a) Through correlation between rainfall and overall runoff (inflows to the reservoir). Figure 6 shows results for one of the events. The pluviometer is Alcoy, and lag steps are taken of ½ hour.

Figure 6. Correlation between Alcoy-rainfall and Beniarrés-runoff for different lags. Event number C-20.

b) Estimated lag from the isocrones analysis (figure 1).

c) Temporal pattern of the hydrologic response obtained after a short pulse of rainfall occurring in the pluviometer geographical location.

After the analysis, each pluviometer was asigned a characteristic lag, taking the average value after the three factors cited. The resulting lag for Alcoy pluviometer, which is the most representative of the rainfall occurring over the basin due to its geographical location, is τ = 5 hours.

Concerning selection of raingauge stations (out of the six available) , and number of previous runoff values, a pruning technique was applied during network training process, allowing for an optimal selection of input variables and dimensioning of the network.

4.2 Network architectures for 1-hour predictions

Six different multilayer feedforward networks were tested (number of nodes indicated for each model, in each of the layers):

M-1 (4-1-1): Linear model, ignoring precipitation information .
M-2 (4-1-1): Hidden nodes with logistic activation functions, also ignoring precipitation data.
M-3 (7-2-1): Hidden nodes with linear activation functions. Precipitation in 3 selected pluviometers, plus 4 previous runoff values. (figure 7)
M-4 (7-2-1): *Idem*., with logistic activation function. (figure 7).
M-5 (7-1-1): Hidden node with logistic activation functions. Same input values as M-3 and M-4.
M-6 (7-3-1): *Idem*., with 3 hidden nodes.

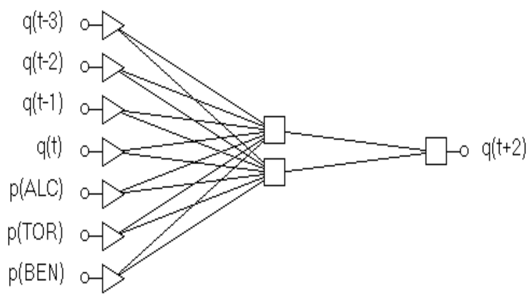

Figure 7. Network topology for models M-3 and M-4.

4.3 Network architecture for 2 and 3-hours predictions

After a trial and error procedure, adequately combined with pruning techniques, best dimensioning of the network was (9-5-2), being the output nodes always linear and producing in this case predictions for 2 and 3 hours ahead. Logistic function was used in the hidden nodes as activation function. Only 4 raingauge stations data are used, together with two previous values of the q's series.
Several topologies were also tested for longer term predictions, but when the time horizon was extended over 3 hours, the model's accuracy decreased significantly in all cases, being only acceptable during the falling limb of the flood, and thus, results were no longer useful for potential use in real time management.

5. RESULTS

Four common statistics for evaluation of models performance were used: r (correlation coefficient), R^2 (Nash-sutcliffe coefficient), RMSE (root mean squared error) and PFC (peak flow criteria).

$$PFC = \frac{\left[\sum_{j=1}^{j=n_p}\left(Q_j^{pr} - Q_j\right)^2 \cdot Q_j^2\right]^{1/4}}{\left[\sum_{j=1}^{j=n_p} Q_j^2\right]^{1/4}} \tag{3}$$

Table 1 resumes results for models M-1 to M-6, computed over the test (validation) set, with marked values corresponding with the best performance according to the criteria in each column.

Table 1. Evaluation of models for 1-h prediction

MODEL	r	R2	RMSE	PFC
M-1	0.9611	0.9224	78.4	**9.25**
M-2	0.9636	0.9274	75.8	9.58
M-3	0.9615	0.9216	78.8	11.35
M-4	**0.9749**	**0.9499**	**63.0**	9.30
M-5	0.9643	0.9282	75.4	10.34
M-6	0.9736	0.9469	64.9	10.56

Model M-4 was chosen, producing errors in the test-set which are shown in figure 8.

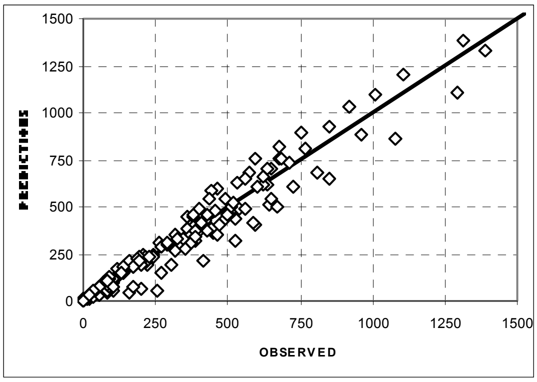

Figure 8. 1-hour prediction errors.

The R^2 coefficient is 0.95 for 1-h predictions, 0.84 for 2-h predictions, and 0.8 for 3-h predcitions.
Figures 9,10 and 11 show prediction results for the event of highest peak flow (Q_{max}= 1800 m^3/s). Predictions are made every ½ hour.

6. Software PCTR-BENIARRÉS

A software was developed incorporating the necessary features concerning dam gates and operation, together with the routines performing the ANN calculations for predictions. This software is called PCTR-BENIARRÉS (***P**redicción de **C**recidas en **T**iempo **R**eal* – **B**eniarrés). It is design to operate in real-time, and therefore, receives on-line data from raingauges and dam elements during the event, generating theree different types of outputs:

a) The predicted inflows to the reservoir for the inmediate future hours (1, 2 and 3 hours).

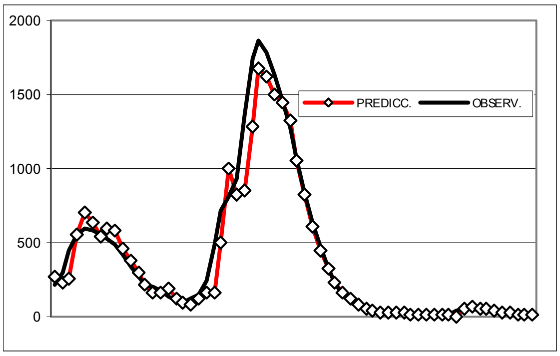

Figure 9. 1-hour predictions – event C-22

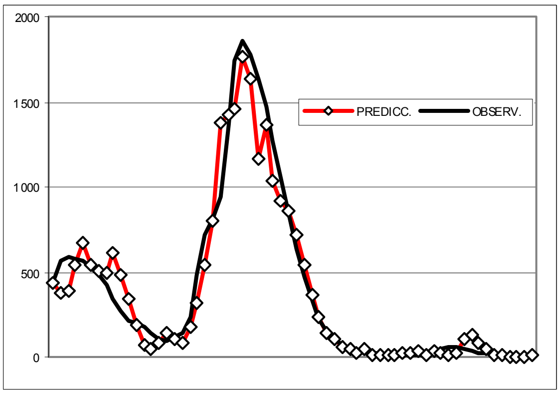

Figure10. 2-hour predictions – eventC-22

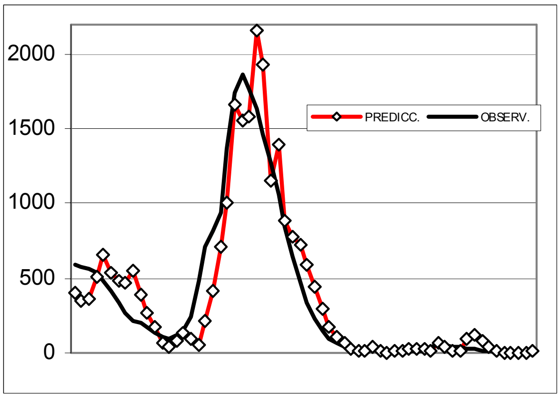

Figure 11. 3-hour predictions – event C-22

b) The recommended outflows or controlled discharges by the Water Authority, under two different risk assumptions concerning dam safety. These operations are functions of the present water level in the reservoir and the slope of the inflow hydrograph. If predictions are available, the can be applied also for the inmediate future intervals, and thus, an inference is made for the state of the system if the recommended operation rules in the gates were followed during the next hours.

c) The future situation in the inmediate next hours under any hypothesis of gates operation.

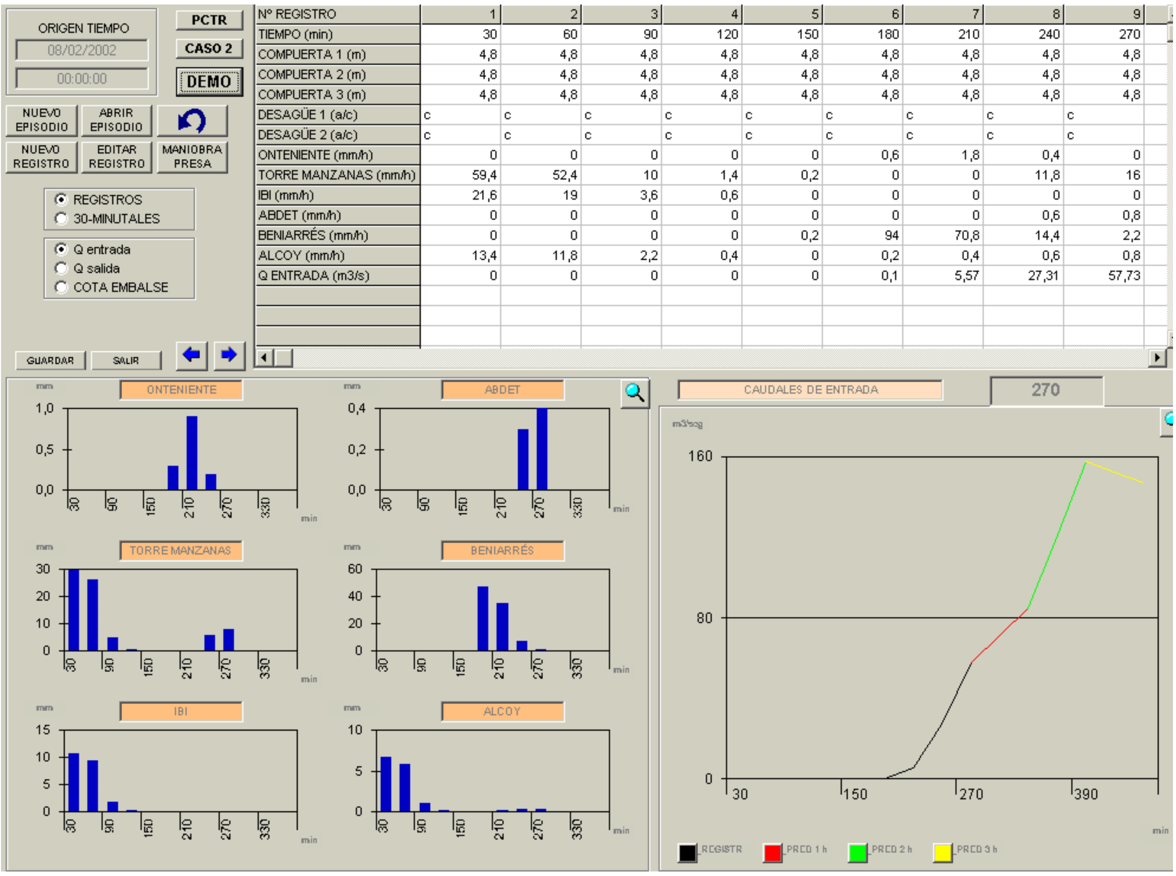

Figure 12. Main window of the PCTR-BENIARRÉS program

One of the main features of the program is that it accepts input data as they are usually collected by the system, i.e., including possiblie gaps in the series, or jumps in the time level of aggregation, or errors. The program includes all the needed filters to create in real-time an EXCEL datasheet properly arranged in constant intervals of ½ hour, with routins checking and completing series if needed. The program can also operate in batch mode, allowing for analyses of past events, prediction made, errors, operation of the dam and overall results obtained in the past.

Figure 12 shows the main window of the program, presenting precipitation measurements in each of the six automatic pluviometers, and the predictions made with the ANN networks. The black line in the hydrograph represents the observed inflows to the dam.

7. Conclusions

A case study is presented on the topic of real-time forecasting of floods using artificial neural networks (ANN) modelling approach, for a medium size basin with a dam at the outlet. Certain advantages from the ANN, as the computational speed and flexibility, are used to implement them in a practical application for flood warning and control purposes. The software PCTR-BENIARRÉS incorporates all the required features and databases due to gates specific discharge curves and volume of the reservoir as a function of depth, together with the routines derived from the trained ANN. The result is an operational tool that can be helpful as an efficient support during the real-time decision making process in flood events situations, and also for analysing goodness of the predictions and strategies for flood control.

8. REFERENCES

Boogard, H. F. P. van den, Gautam, D. K. and Mynett, A. E. *Auto-regressive neural networks for the modeling of time series*, Hydrodinamics98, Babovic & Larsen (eds), Balkema, Rotterdam. 741-748, 1998.

Coulibaly, P., Anctil, F. and Bobée, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology* 230, 244-257 (2000).

Deo, M. C. and Thirumalaiah, K. Real time forecasting using neural networks, *Artificial neural networks in Hydrology*, R. S. Govindaraju and A. Ramachandra Rao (eds), Kluwer Academic Publishers, Dordrecht, 53-71, 2000.

García Bartual, R. Identification of rain-cells from S.A.I.H. network measurements (Valencia): an analytical approach. *Proc. – 3^RD^ Plinius Conference – European Geophysical Society*. Italy, october 2001. Govindaraju, R. S. and Rao, A. R. (eds). *Artificial neural networks in Hydrology*, Kluwer Academic Publishers, Dordrecht, 2000.

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

Imrie, C.E., Durucan, S. and Korre, A. River flow prediction using artificial neural networks: generalisation beyond the calibration range. *Journal of Hydrology* 233, 138-153, 2000.

Lachtermacher, G. and Fuller, J. D. Backpropagation in hydrological time series forecasting, Stochastic and statistical methods in hydrology and environmental engineering, *Time series analysis in hydrology and environmental engineering*, K. W. Hipel et al. (eds), Kluwer Academic Publishers, Dordrecht, 229-242, 1994.

Luk, K.C, Ball, J.E. and Sharma, A. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *Journal of Hydrology*, 227, 56-65, 2000.

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

Masters, T. Practical Neural Networks Recipies in C++. Academic Press, San Diego, CA.

Raman, H. and Sunilkumar, N. Multivariate modelling of water resources time series using artificial neural networks, *Journal of Hydrological Sciences, 40*(2), 145-163, 1995.

Salas, J. D., Markus, M. and Tokar, A. S. Streamflow forecasting based on artificial neural networks, *Artificial neural networks in Hydrology*, R. S. Govindaraju and A. Ramachandra Rao (eds), Kluwer Academic Publishers, Dordrecht, 23-51, 2000.

Toth, E., Brath, A. and Montanari, A. Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of Hydrology* 239, 132-147, 2000.

Zealand, C. M., Burn, D. H. and Simonovic, S. P. Short term streamflow forecasting using artificial neural networks, *Journal of Hydrology 214*, 32-48, 1999.