International Congress on Environmental Modelling and Software

1st International Congress on Environmental Modelling and Software - Lugano, Switzerland - June 2002

Jul 1st, 12:00 AM

An Artificial Neural Networks Modeling Approach Applied to the Probabilistic Management of Water Resources Systems

J. C. Ochoa-Rivera

R. García-Bartual

J. Andreu

An Artificial Neural Networks Modeling Approach Applied to the Probabilistic Management of Water Resources Systems

J. C. Ochoa-Rivera, R.García-Bartual, and J. Andreu

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

Abstract: A multivariate non-linear model for synthetic generation of monthly discharge series is presented. The time series generator is built upon a multilayer feedforward neural network with an added multivariate random component normally distributed. The usual error backpropagation algorithm is used to train the network, with a sequential training scheme using shuffled patterns for a stochastic search in the weight space. The proposed model generates inputs to a wider methodology developed for simulation of the probabilistic managing of the upper Tagus River system in Spain. The model is validated in terms of its ability to reproduce some relevant statistics directly related to important features from the water resources system managing point of view, including droughts and storage derived statistics. The generator was inserted into a decision support system, and a variety of possible future hydrological scenarios in the river system were simulated, with particular consideration of demand failures probabilities under different assumptions and previous hydrological states of the system. For comparison purposes, similar experiments were also undertaken using synthetic series generated with a second-order autoregressive multivariate model, AR(2). The results obtained show higher percentages of demand failures when inputs from the artificial neural network generator (ANN) are used. The case study illustrates another practical application of ANN approaches, adequately combined with other frequently used tools in the context of water resources systems planning and management.

***Keywords**:* Neural networks; Perceptron multilayer; Hydrological scenario generation; Multivariate time-series analysis; Water resources systems management.

1. INTRODUCTION

In the past, water resources planners used to tackle the problem of the systems simulation only with historic hydrological records. Such approach introduces severe restrictions because typical uncertainties affecting hydrological processes are not taken into account, as different researchers reported [Loucks et al., 1981; Bras and Rodríguez-Iturbe, 1985]. This is the reason why several models for generation of future hydrological scenarios have been developed [Box and Jenkins, 1976; Hipel and Mc Leod, 1994]. Most of them are linear approaches, as simple and multiple linear regression, autoregressive models (AR), autoregressive moving average models (ARMA), AR and ARMA with periodic parameters, and temporal disaggregation and spatial disaggregation models, among others [Salas et al., 1980]. On the other hand, classical non-linear techniques typically require large amounts of exogenous data, which are not always available [Deo and Thirumalaiah, 2000]; some non-linear approaches outperform the linear techniques, as periodic gamma autoregressive processes (PGAR) for instance [Fernandez and Salas, 1986], but they can only be applied to a single site, i.e., they are limited to be univariate models.

During recent years, new technologies and algorithms have arisen as powerful tools for modeling several problems related to water sciences. *Artificial Neural Networks* (ANN) is one of them. ANN have been used to successfully solve many different kind of hydrological problems [ASCE, 2000]. Particularly, the ANN approaches applied to hydrologic time series modeling and forecasting have shown better performance than the classical techniques [Govindaraju and Rao, 2000]. Some of those works deal with univariate multi-step prediction of annual series [Lachtermacher and Fuller, 1994], bivariate monthly streamflow forecasting [Raman and Sunilkumar, 1995], short-term streamflow prediction [Zealand et al., 1999], hourly discharges modeling [Deo and Thirumalaiah, 2000], and single and multi-step forecasting of monthly streamflows [Salas et al., 2000].

The study reported herein presents a new methodology based on a stochastic multivariate ANN model that takes into account the strong uncertainty typically affecting multi-site monthly streamflow relationships. The deterministic component of the model consists of a three-layer feed-forward architecture which is trained with the popular *error backpropagation algorithm*, and a normally distributed random noise is added as the stochastic component. The developed model is used to generate multiple monthly hydrological scenarios spanning several months in the future and conditioned to present and past inflows. Then, such future scenarios are applied as hydrological inputs to a decision support system in order to estimate the risks of demand failures of a water resources system located in the upper Tagus River basin (Spain). The proposed model is validated by comparing some relevant droughts and storage derived statistics of generated streamflow series to those of historical ones. For comparison purposes, an AR(2) model is also used for time series synthetic generation which are then processed in a similar way.

In Section 2 of this paper, the theoretical development of the model is presented; in Section 3, the case study of a water resources system in the upper Tagus River basin (Spain) is described; and finally, Section 4 summarizes several conclusions derived from the study.

2. ARTIFICIAL NEURAL NETWORKS MODEL

The scheme used in this research consisted of a *multilayer perceptron artificial neural network* (MLP-ANN), trained with the well-known error-backpropagation learning algorithm [Rummelhart et al., 1986], which has been successfully used in a number of water resources systems applications as referred above.

2.1 Architecture

Only three-layer MLP-ANNs were tested, assuming that sufficient degree of freedom can always be provided by changing the number of nodes in the hidden layer [Hornik et al., 1989]. Let N_S be the number of streamflow sites (stations) in the water resources system, and N_T the number of past streamflow values used to predict the future values. Once original streamflow series are conveniently normalized, standardized [Salas et al., 1980], and scaled [Salas et al., 2000], the transformed values (Z) are allocated on the neural network in such a way that $\{Z\}_{\psi}$ (with $\{Z\}=\{Z_1, Z_2,...,Z_{NS}\}$ and $\psi=t$-1, t-2,...,t-N_T) are the input values of the input layer, and $\{Z\}_t$ are the output values. Therefore, the input layer (layer 0) has $m_0=N_S \times N_T$ nodes. In this study, input values to the network are denoted by $\{x^{(0)}_p\}$ (with p=1,2,...,m_0); likewise, the output layer (layer 2) has $m_2=N_S \times 1$ nodes, given that ANN is a multivariate model recursively applied on a single time-step (t) basis. The output values of the output layer are denoted as $\{y^{(2)}_r\}$ (with r=1,2,...,m_2). Figure 1 indicates the network topology.

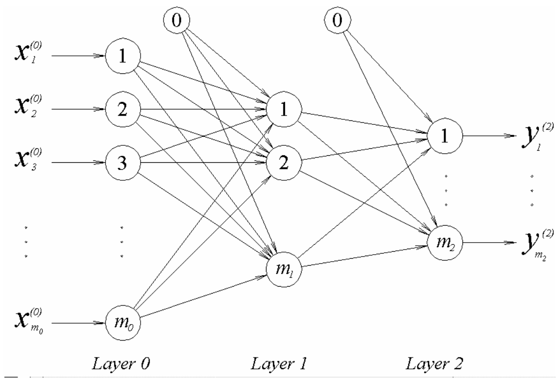

Figure 1. Artificial neural network topology.

In this way, the *patterns* to be used for training the ANN have a *predictor section* with the input values $\{x^{(0)}_p\}$, and a *criterion section* with the target values $\{y^{(2)}_r\}$.

As no general techniques nor rules have been established for achieving optimal network topology [Maier et al., 2000], the number of hidden nodes, m_1, must be determined applying a trial-error procedure. In this study, this process is started with a single hidden node, the network is trained saving the results, and then, new nodes are increasingly added repeating the process until no significant improvement in the network performance is obtained.

2.2 Multivariate Deterministic Component

The deterministic component is due to the overall operation of the ANN. Making use of *bipolar sigmoid function* as an activation function for each node of the ANN, due to its generally accepted good behaviour [Haykin, 1999], the nonlinear function relating inputs and outputs of the ANN can be written as

$$y_r^{(2)} = 2\{1+\exp[-\sum_{q=0}^{m_2} w_{qr}^{(2)}(2[1+\exp(-\sum_{p=0}^{m_0} w_{pq}^{(1)} x_p^{(0)})]^{-1}-1)]\}^{-1}-1; \quad r=1,2,...,m_2 \tag{1}$$

This equation shows the target values (vector $\{y_r^{(2)}\}$) as an explicit function of the ANN's input values (vector $\{x_p^{(0)}\}$), and of the synaptic weights ($w_{qr}^{(2)}$ and $w_{pq}^{(1)}$), whose values result from the training process. Eq. (1), in its matrix form, is written as

$$\{y_r^{(2)}\} = F_1(\{x_p^{(0)}\}) \tag{2}$$

where F_1 is a function that comprises the mathematical operations involved into the neural network. After the values x and y are de-scaled with the inverse of the scaling function, Eq. (2) becomes Eq. (3), which is the forecasting mode of the proposed model and is called *deterministic component*.

$$\{Y\}_t = F_2(\{X\}_\varphi); \quad \varphi = t-1, t-2, ..., t-N_T \tag{3}$$

where F_2 is a function that involves F_1 and the de-scaling function.

2.3 Training Process

The problem of estimating synaptic weights can be regarded as a non-linear optimization problem without constraints. A convenient objective function can be the *mean squared error function* (MSE), given by

$$MSE = \frac{1}{2N_p}\sum_{n=1}^{N_p}\sum_{j=1}^{m_2}[d_j(n) - y_j^{(2)}(n)]^2 \tag{4}$$

in which N_p is the number of patterns shown to the network; $d_j(n)$ and $y_j^{(2)}(n)$ are the target value and the predicted one for node j of the output layer, and corresponding to n^{th} pattern.

The error backpropagation algorithm is a sufficiently accurate and robust technique to train the neural network, and best results are obtained when it is applied in a sequential mode or *exemplar-mode training*. This allows the set of weights to be successively modified after each exemplar processing as

$$w_{jk}^{(l)}(n+1) = w_{jk}^{(l)}(n) + \Delta w_{jk}^{(l)}(n) \tag{5}$$

where $\Delta w_{jk}^{(l)}(n)$ is the corresponding weight change, given by

$$\Delta w_{jk}^{(l)}(n) = \alpha[\Delta w_{jk}^{(l)}(n-1)] + \eta\delta_k^{(l)}(n) y_j^{(l-1)}(n) \tag{6}$$

where α is the momentum constant; η is the learning rate; and $\delta_k^{(l)}(n)$ is the local gradient for node k in layer l. Details about the procedure can be found in Haykin (1999).

2.4 Multivariate Random Component

When Eq. (2) is applied with historical data, the difference between observed values and predicted ones leads to the residuals series, $\{\varepsilon\}_t$. A statistical analysis of $\{\varepsilon\}_t$ shows that they can be adequately modeled as normally distributed and cross-correlated series with means statistically equal to 0 and their respective variances. Therefore, they can be expressed as indicated in Eq. (7), which describes the *random component* of the ANN model.

$$\{\varepsilon\}_t = \mathbf{B}\{\xi\}_t \tag{7}$$

being

$$\mathbf{BB}^{\mathrm{T}} = \Sigma \tag{8}$$

where $\{\xi\}_t$ is a normally distributed and uncorrelated random signal with zero mean and variance statistically equal to 1; $\mathbf{\Sigma}$ is the covariance matrix of observed residual series. Since $\mathbf{\Sigma}$ is the Gramian matrix of matrix $\mathbf{B}$, this last one is unknown and must be obtained by solving the matrix equation (8). Bras and Rodriguez-Iturbe (1985) give details to solve it for a practical method.

2.5 Streamflow Generation Model

In order to build a multivariate model for synthetic generation of monthly streamflow series, $\{Q'\}_t$, both, the deterministic and the random component must be assembled to generate the synthetic values, given by

$$\{Q'\}_t = F(\{Y'\}_t + \{\varepsilon'\}_t) \tag{9}$$

where $\{Y'\}_t$ represents the normalized and standardized synthetic values obtained from Eq. (2); $\{\varepsilon'\}_t$ are synthetic residuals estimated from Eq. (7); and F includes the inverse functions of the normalization and standardization process.

3. CASE STUDY

The developed model was applied to Henares water resources system (HWRS), which is located in the upper Tagus River basin (Spain), the longest river of the Iberian Peninsula. HWRS comprises 7 streamflow stations, 2 agrarian demands, 2 hydraulic transfer demands, 1 urban demand, 4 reservoirs and several hydraulic structures (Figure 2). More details about the HWRS can be found in CHT (1999). The application consisted of three phases: calibration of the model, validation, and conditioned streamflow scenario generation inserted into a decision support system (DSS).

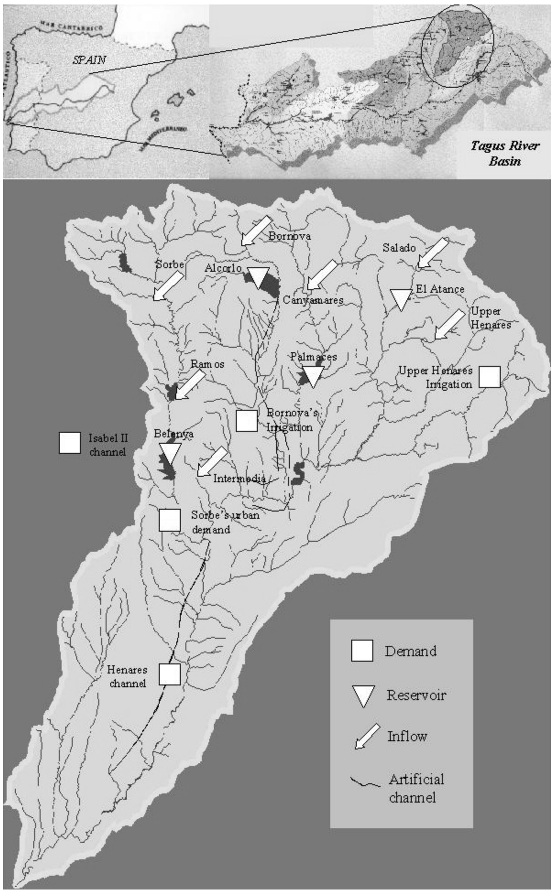

Figure 2. Henares water resources system.

3.1 Model Calibration

Monthly discharge data series of 53 years length were available for 7 different river stations of the HWRS. Those streamflow series were normalized with appropriate functions, and after standardized and scaled. Such values were allocated on the input and output layers of the ANN as indicated in Section 2.1, assembling in this way the training patterns. Then, the neural network topology was built, assuming that each discharge value was given as a function of the two past values from the all stations. Thus, the input and output layers were constrained to have 14 and 7 nodes respectively. Best training results were achieved for a topology with 6 hidden nodes (i.e., a topology 14-6-7), a learning rate value equal to 0.05, a momentum constant equal to zero. Next, residual series were calculated as described in Section 2.4, and finally, their covariance matrix, Σ, was calculated in order to estimate the parameter matrix of the model's random component, **B**.

3.2 Model Validation and Performance

The validation of the streamflow generation model was done through comparison of relevant statistics related to droughts and storage of the synthetic series to those of historical series [Stedinger and Taylor, 1982]. Consequently, the calibrated model was used for generating 200 synthetic series of monthly discharges in 7 rivers of the HWRS, each series of 53 years length, and then, some statistics were computed from such series. In addition to validation, performance of the ANN model was also analyzed by comparing its statistics to those of obtained from a second-order autoregressive multivariate model, AR(2). Obviously, same preprocessing of data series was used in both cases. It must be also remarked that parsimony indexes are not significantly different between them; while parsimony index of global modeling concerning to AR(2) approach is equal to 11, ANN entire modeling has a parsimony index equal to 10. Nevertheless, this number is not a suitable index for describing the generalisation properties of the ANN model, as Chakraborty et al. (1992, page 968) appointed.

Seven groups of statistics were calculated: monthly basic statistics, monthly average drought statistics (MAD), monthly maximum drought statistics (MMD), annual average drought statistics (AAD), annual maximum drought statistics (AMD), monthly storage statistics, and annual storage statistics. Mathematical formulation applied for such calculations is given in Salas et al. (1980). Next, relative root-mean-squared errors (RRMSE) of each one estimated statistics were computed for each of 7 stations of the HWRS, as indicated in Fernandez and Salas (1986), and then, such RRMSEs were averaged. They can be observed in Table 1, which shows that ANN model outperforms the AR(2) model.

3.3 Probabilistic Management of the HWRS

Finally, a probabilistic management of the HWRS was modeled in a DSS taking as hydrological inputs a number of future synthetic hydrological scenarios conditioned to a certain present hydrological state. This was performed with both, ANN and AR(2) models, in order to know the changes on the magnitude of demand failures due to a better preservation of the droughts and storage statistics. 2000 synthetic series for a time horizon equal to 24 months were generated with ANN model, conditioned to low flow discharges as initial hydrological state in each river station of the HWRS. Then, 2000 simulations of the entire HWRS were performed in SIMGES, which is a module of AQUATOOL package [Andreu et al., 1996], the DSS employed in this study.

Table 1. Averaged RMMSE of statistics.
ANN vs AR(2)

Statistics	ANN	AR(2)
Mean	0.041	0.019
Std. Dev.	0.092	0.119
Skew. Coef.	0.561	0.590
Correlogram	0.193	0.166
Frequency MAD	0.114	0.139
Length MAD	0.133	0.142
Intensity MAD	0.071	0.066
Magnitude MAD	0.090	0.095
Length MMD	0.244	0.245
Intensity MMD	0.038	0.029
Magnitude MMD	0.235	0.239
Frequency AAD	0.221	0.274
Length AAD	0.228	0.243
Intensity AAD	0.162	0.201
Magnitude AAD	0.226	0.277
Length AMD	0.356	0.372
Intensity AMD	0.070	0.095
Magnitude AMD	0.408	0.441
An. Reserv. Cap.	0.356	0.355
An. Hurst Coef.	0.068	0.075
Mn. Reserv. Cap.	0.383	0.385
Mn. Hurst Coef.	0.104	0.124

Low storage levels were assumed in 4 reservoir sites of the HWRS at the beginning of the simulations. As a result of these, the probabilities of demand failure along the time horizon were obtained for 5 demands referred above. Same process was carried out for the AR(2) model on identical basis, and then, its results were compared to those of the ANN model, as indicated in Figure 3. It can be observed that probabilities of demand failures are systematically higher when the ANN model is applied. It must be remarked that the management policies were exactly the same for both, ANN and AR(2) models; this was done in order to the differences of probabilities between two modelings were not influenced by management policies. Given that ANN model fits better than AR(2) model the historical statistics, as Table 1 shows, it can be said that the developed model leads to a more conservative decision, and hence a better preparedness for droughts.

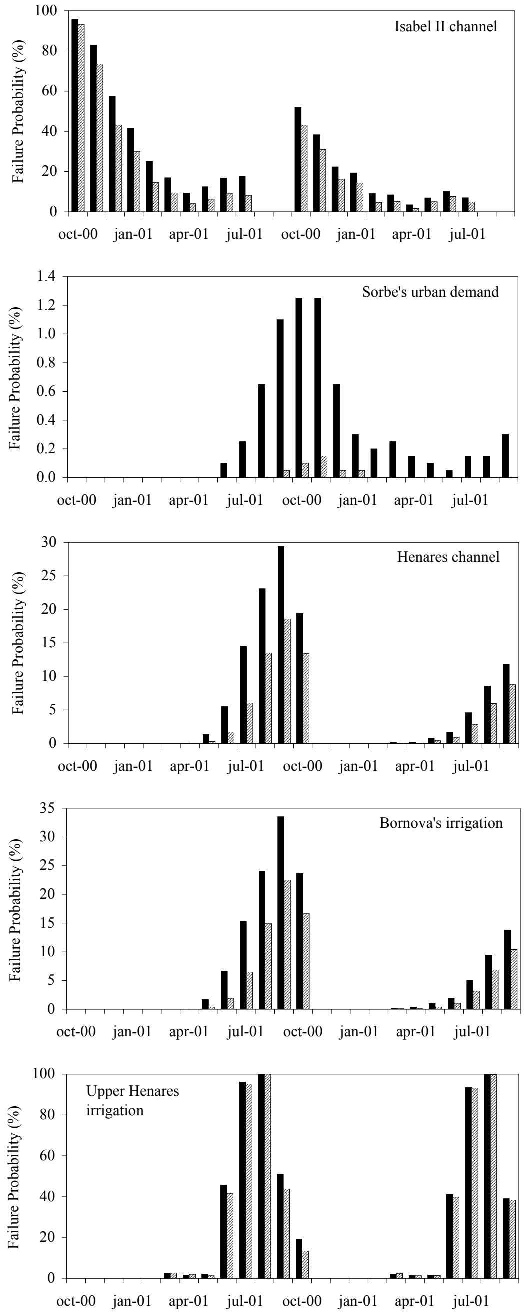

Figure 3. Probabilities of demand failures of the HWRS for ANN model (black) and AR(2) model (white).

4. CONCLUSIONS

A multivariate non-linear model for generating monthly streamflow series at several geographical sites has been developed. It consists of a deterministic core defined in terms of a three-layer feed-forward neural network with six hidden nodes, plus a stochastic white noise component multivariate normally distributed. Historical discharge records were used to train and test the potentials of the model for practical hydrological scenario generation.

The model was validated by comparing some relevant droughts and storage related statistics of the synthetic scenarios to those derived from historical series. The classical multivariate AR(2) model was also applied under identical basis and then the results were compared to those of the proposed model. Synthetic series from the models were inserted into a decision support system to simulate the probabilistic management of a water resources system located in the upper Tagus River basin (Spain). The outputs of the simulation show that the developed technique leads to a more realistic management than that of a simulation with hydrological inputs derived from an AR(2) model. This is basically due to the droughts statistics, which are better preserved with the artificial neural network model. Therefore, it can be concluded that the proposed model represents a viable alternative to be considered in future applications, competing with other classical techniques. The results presented herein provide a qualitative demonstration of its value as a candidate method for synthetic generation of monthly streamflow series in water resources system analysis. Preliminary results from an ongoing research show satisfactory behaviour of the proposed scheme adapted to higher autoregressive orders (>2). Other network architectures should be also explored, while different training techniques might be required for a more efficient operation, as the number of locations is increased and the network becomes more complex.

5. ACKNOWLEDGEMENTS

The research grants from the Spanish Ministry of Education and Culture (CICYT project HID99-0656) and European Commission (INCOMED "WAMME" project ICA3.199.00014) are acknowledged.

6. REFERENCES

Andreu, J., Capilla, J. and Sanchís, E., AQUATOOL, a generalized decision-support system for water-resources planning and operational management. *Journal of Hydrology, 177*, 269-291, 1996.

ASCE task committee on application of artificial neural networks in Hydrology (Rao Govindaraju), Artificial neural networks in Hydrology. II: Hydrologic applications, *Journal of Hydrologic Engineering 5*(2), 124-137, 2000.

Box, G. E. P. and Jenkins, G. M., *Time series analysis, forecasting and control*, Holden Day, San Francisco, 1976.

Bras, R. L. and Rodríguez-Iturbe, I., *Random functions and hydrology*, Addison-Wesley, Massachusetts, 1985.

Chakraborty, K., Mehrotra, K., Mohan, C. K. and Ranka, S., Forecasting the behavior of multivariate time series using neural networks, *Neural networks*, *5*, 961-970, 1992.

CHT - Confederación Hidrográfica del Tajo. *La cuenca del Tajo en cifras*, Oficina de Planificación Hidrológica, CHT, Ministerio de Medio Ambiente, Madrid, 1999.

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.

Fernandez, B. and Salas J. D., Periodic gamma autoregressive processes for operational hydrology, *Water Resources Research, 22*(10), 1385-1396, 1986.

Govindaraju, R. S. and Rao, A. R. (eds.). *Artificial neural networks in hydrology*, Kluwer Academic Publishers, Dordrecht, 2000.

Haykin, S., *Neural networks. A comprehensive foundation*, Prentice-Hall, New Jersey, 1999.

Hipel, K. W. and Mc Leod, A. I., *Time series modeling of water resources and environmental systems*, ELSEVIER, Amsterdam, 1994.

Hornik, K., Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators, *Neural Networks, 2*, 395-403, 1989.

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.

Loucks, D. P., Stedinger, J. R. and Haith, D. A. *Water resources system planning and analysis*, Prentice Hall, New Jersey, 1981.

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.

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.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J., Learning internal representations by error propagation, *Parallel distributed processing*, D. E. Rumelhart and J. L. McCleland (eds.), vol. 1, Chapter 8, Cambridge, MA:MIT Press, 1986.

Salas, J. D., Delleur, J. W., Yevjevich, V. and Lane, W. L., *Applied modeling of hydrologic time series*. Water Resources Publications. Littleton, Colorado, 1980.

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.

Stedinger, J. R. and Taylor, M. R., Synthetic streamflow generation 1. Model verification and validation, *Water Resources Research, 18*(4), 909-918, 1982.

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.