Daily Reservoir Inflow Forecasting Using Time Delay Artificial Neural Network Models

M. J. DIAMANTOPOULOU*, P.E. GEORGIOU** & D.M. PAPAMICHAIL**
* Faculty of Forestry and Natural Environment, Aristotle University of Thessaloniki, 54124 Thessaloniki, GREECE
** Faculty of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, GREECE

Abstract: - Accurate real-time reservoir inflow forecasting is an important requirement for operation, scheduling and planning conjunctive use in any basin. In this study, Time Delay Artificial Neural Network (TDANN) models, which are time lagged feed-formatted networks with delayed memory processing elements at the input layer, are applied to forecast the daily inflow into a planned Reservoir (Almopeos River basin) in Northern Greece. The network topology is using multiple inputs, which include the one time lagged daily reservoir inflow values and the time lagged daily precipitation values from three meteorological stations which are inside the Almopeos river basin and a single output, which are the daily reservoir inflow values. The choice of the precipitation input variables introduced to the input layer was based on the cross-correlation. In the forecasting part of this study, predictions of one day ahead were investigated. The training of ANNs suitable for the current application is the cascade correlation algorithm. Kalman's learning rule was used to modify the artificial neural network weights. The networks are designed by putting weights between neurons, by using the hyperbolic-tangent function for training. The results show a good performance of the TDANN approach and demonstrate its adequacy and potential for forecasting daily reservoir inflow.

Key-Words: - Reservoir inflow, Real-time forecasting, Time delay artificial neural networks

1 Introduction

Forecasting the reservoir inflow is an important requirement for planning conjunctive use in any basin. The purpose of forecasting is to reduce the risk in decision making. Information regarding reservoir inflow, is necessary in the analysis and design of several water resources projects such as dam construction, reservoir operation, flood control and wastewater disposal. There are different types of inflow forecasting models, and they can be classified into three types: empirical (or black box) models, lumped conceptual models and distributed physically based models. Black box models may also be divided into sub groups according to their origin, namely empirical hydrological methods (such as unit hydrograph model), statistically based models (such as ARMA, ARMAX, SARIMA [25], gauge to gauge correlation models etc.) and artificial intelligence based models (such as artificial neural networks).

In recent years, ANN models have become extremely popular for prediction and forecasting in a number of areas, including finance, power generation, medicine, water resources and environmental science [22]. A number of researchers have investigated the adaptability of ANN models to the field of hydrology, water resources and hydrologic time series [1, 2, 4, 10, 12, 16, 17, 18, 19, 20, 21, 23, 28, 29, 30, 31, 33, 34].

In this paper, three layer cascade correlation Time Delay Artificial Neural Network (TDANN) models were developed to forecast the one day ahead daily inflow into a planned Reservoir (Almopeos River basin) in Northern Greece by using multiple inputs. These inputs include the one time lagged daily reservoir inflow values and the time lagged daily precipitation values from three meteorological stations which are inside the Almopeos river basin.

2 Artificial neural networks methodology

Artificial Neural Network is an information processing system that tries to replicate the behavior of a human brain by emulating the operations and connectivity of biological neurons [9]. The basic structure of an ANN model, usually, consists of three distinctive layers, the input layer, where the data are introduced to the ANN, the hidden layer or layers, where data are processed, and the output layer, where the results of ANN are produced. The structure and operation of ANNs is discussed by a number of authors [6, 7, 9, 14, 15, 27].

The ANNs are designed by putting weights between neurons, by using a transfer function that controls the generation of the output in a neuron, and using adjustable laws that define the relative importance of weights for input to a neuron. In the training, the ANN defines the importance of the weights and adjusts them through an iterative procedure.

The training of ANNs suitable for the current application is the cascade correlation algorithm [6,8], which produces the cascade correlation Time Delay Artificial Neural Network (TDANN) that belongs to the feedforward type, which is a supervised algorithm in the multilayer feed-forward ANNs. The Cascade part refers to the architecture and its mode of construction entails adding hidden units once at a time and always connecting all the previous units to the current unit. The Correlation part refers to the way hidden units were trained by trying to maximize the correlation between output of the hidden unit and the desired output of the network across the training data. The training procedure of TDANNs is composed of a forward pass. The information is processed in the forward direction from the input layer to the hidden layer or layers to the output layer. Kalman's learning rule [3, 5, 13, 24] was used to modify the TDANN weights. Such, a network has the ability to approximate any continuous function. As it was mentioned the input nodes receive the data values and pass them on to the hidden layer nodes. Each one of them collects the input from all inputs nodes after multiplying each input value by a weight, attaches a bias to this sum and passes on the result through a nonlinear transformation, the hyperbolic-tangent function [9].

The objective of the training algorithm needed by the network for training, is to reduce the global error [12] by adjusting the weights and biases. In each training step, a new hidden neuron is added and its weights are adjusted to maximize the magnitude of the correlation. Each hidden neuron is trained just once and then its weights are frozen.

The error between the output of the TDANN and the target value of the output was computed, as well. In order to achieve an estimation of the one day ahead daily reservoir inflow, the one day lagged daily reservoir inflow values and the time lagged daily precipitation values from three meteorological stations, which are inside the Almopeos river basin, are introduced as inputs into TDANNs. In this sense, the input layer of TDANNs consists of a number input neurons and one output neuron, which is the daily reservoir inflow.

The choice of the precipitation input variables introduced to the input layer based on the cross-correlation. The use of cross-correlation between the ith input series and the output provides a short cut to the problem of the delayed memory determination [26]. The cross-correlation coefficient of the ith input series and the output records on a span of N times intervals has been given by Yevjevich [32]. As the output increases after the occurrence of the ith input series and then decreases gradually towards to its original level, the cross-correlation coefficient is expected to decrease gradually with increase of the time lag, k. The first minimum positive value of the correlogram approximately indicates the delayed memory. Therefore, personal judgement must be exercised in interpreting the correlogram.

During the training of TDANNs in the calibration period, the simulated daily reservoir inflow values are compared with the corresponding observed daily inflow values to identify the simulation errors. The geometry of TDANNs, which determines the number of connection weights and how these are arranged, depends on the number of hidden layers and the number of the hidden nodes in these layers. In the developed TDANNs, one hidden layer is used and the number of the hidden nodes is optimized by maximizing the correlation between output of the hidden unit and the desired output of the network across the training data. However, the final network architecture and geometry are tested to avoid over-fitting as suggested by Maier and Dandy [21].

3 Application and results

The study area is the Almopeos river basin, Northern Greece (Fig. 1), between $40^o49'$ S to $41^o09'$ N and $21^o47'$ W to $22^o19'$ E. The basin is 1021 km^2 in size. In this paper, TDANN models were developed to forecast the daily inflow into a planned Reservoir (Almopeos river basin) by using multiple inputs, which include the one day lagged daily reservoir inflow values and the time lagged daily precipitation values from three meteorological stations (Exaplatanos, Promachoi and Theodoraki) (Fig. 1), which are inside the Almopeos river basin and a single output, which are the daily reservoir inflow values [11]. Thus, there are four separate independent input functions of time and a single output function of time.

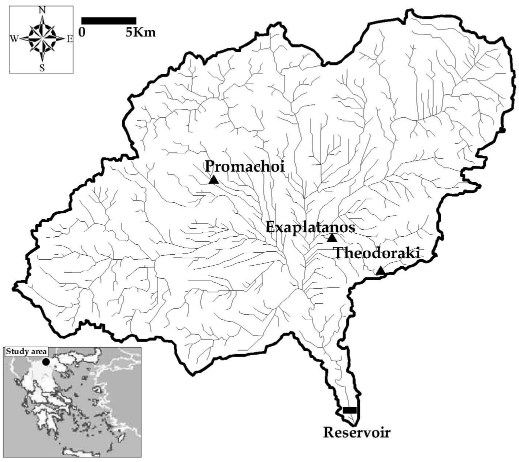

Fig. 1. Map showing the Almopeos river basin, the reservoir and the meteorological stations

The TDANN models were developed by using the daily data from the year 1987 as the calibration period and from the year 1995 as the validation period. Related information for the reservoir inflow and the three precipitation stations (Fig. 1) and statistical parameters of their time series of daily values for the years 1987 and 1995, are given in Table 1.

For TDANN models construction, daily inflow data from the year 1987 randomly partitioned into training (90% of all data) and test (the remaining 10% of all data) data sets, were used. The one day lagged daily reservoir inflow values and the time lagged daily precipitation values from the three meteorological stations, were used as inputs. The delayed memory corresponding to the ith input series was determined by using the cross-correlation between the ith input series and the output. This procedure was applied to the Almopeos river system (Fig. 1), using the correlograms, between the daily reservoir inflow values and the daily precipitation values from the three precipitation stations (Exaplatanos, Promachoi and Theodoraki) during the calibration year. The results indicate that the delayed memory of the system corresponding to the ith input series appears to be equal to 10 days.

Numerous TDANN structures tested in order to determine the optimum number of hidden layers and the number of nodes in each. The architecture of the best TDANN model for forecasting reservoir inflow is composed of one input layer with thirty one input variables, one hidden layer with eight nodes and one output layer with one output variable. The input $Q_{R(t-1)}$ represents the one day lagged daily reservoir inflow value and the inputs P_{Et}, $P_{E(t-1)}$, $P_{E(t-2)}$, … $P_{E(t-9)}$, P_{Pt}, $P_{P(t-1)}$, $P_{P(t-2)}$, … $P_{P(t-9)}$, and P_{Tt}, $P_{T(t-1)}$, $P_{T(t-2)}$, … $P_{T(t-9)}$ represent the delayed daily precipitation values recorded at Exaplatanos, Promachoi and Theodoraki, respectively. The output Q_R represents the one day ahead forecasted daily reservoir inflows.

Table 1. Related information for the Almopeos reservoir inflow and the three precipitation stations and statistical parameters of their time series of daily values for the calibration year 1987 and the validation year 1995

Almopeos reservoir inflow				
Year	Mean (m^3/s)	Min (m^3/s)	Max (m^3/s)	st dv (m^3/s)
Calibration (1987)	12.03	2.25	152.28	14.83
Validation (1995)	6.06	3.21	45.94	4.48
Exaplatanos meteorological station (132.9m a.s.l.)				
Year	Sum (mm/year)		Max (mm/day)	
Calibration (1987)	815		110	
Validation (1995)	585		35	
Promachoi meteorological station (260 m a.s.l.)				
Year	Sum (mm/year)		Max (mm/day)	
Calibration (1987)	954		125	
Validation (1995)	789		90	
Theodoraki meteorological station (424 m a.s.l.)				
Year	Sum (mm/year)		Max (mm/day)	
Calibration (1987)	877		54	
Validation (1995)	745		42	

The best TDANN model, the correlation coefficient (R), the mean absolute error (MAE), the root mean square error (RMSE), the RMSE (%) of the mean, between the output of the hidden unit and the desired output of the TDANN model, for the Almopeos reservoir daily inflow, for the calibration, the training, the test and the validation data sets, are given in Table 2. The notation (Q_R / TDANN: 31-8-1/0.9957) (Table 2) means that the best architecture of the specific TDANN model is composed of one input layer with thirty one input variables, one hidden layer with eight nodes and one output layer with one output variable, with value of correlation coefficient equals to 0.9957.

Table 2. TDANN model, Correlation coefficient (R), Mean absolute error (MAE), Root mean square error (RMSE) and the (%) of the mean of the TDANN model, for the daily inflow values into Almopeos reservoirfor the calibration, the training, the test and the validation data sets

Q_R / TDANN: 31-8-1/0.9957			
Data	R	MAE	RMSE
Calibration (365) (1987)	0.9957	0.7156	1.4984 (12.45%)
Train (328)	0.9958	0.6717	1.5087 (12.51%)
Test (37)	0.9946	1.1046	1.4035 (11.89%)
Validation (365) (1995)	0.9688	1.1478	1.1478 (18.95%)

According to the results of Table 2 we can see that the difference in the R, MAE and RMSE obtained using the test data set is not markedly different than that obtained using the training data, meaning there is no overfitting. Also, the results of Table 2 show a good performance of the chosen TDANN model for forecasting the Almopeos reservoir daily inflow.

Table 3 depicts the percentage error in daily peak flow estimates for the TDANN model during the calibration and validation years. The low percentage error in daily peak flow estimates imply that the TDANN model is able to forecast the peak flows with reasonable accuracy. Also, the times to peak are well estimated.

Table 3. Percentage error in daily peak estimation for the TDANN model during the calibration and validation years

Calibration year			
Date	Peak flow (m^3/sec)		Error
	Historical	TDANN	(%)
15 Feb 87	152.28	140.72	-7.59
21 Feb 87	64.43	61.18	-5.04
23 Mar 87	133.65	122.92	-8.02
31 Mar 87	81.29	72.02	-11.41
27 Nov 87	78.56	76.73	-2.33
Validation year			
26 Apr 95	34.74	37.31	7.40
2 Dec 95	30.20	28.59	-5.33
30 Dec 95	45.94	43.38	-5.57

An analysis to assess the potential of the chosen TDANN model to preserve the statistical properties of the historic inflow series reveals that the inflow series forecasted by the TDANN model reproduces the first three statistical moments (i.e. mean, standard deviation and skewness) for the calibration and validation years.

The comparisons were also made by using the paired t-test with the two-sided tabular value (a=0.05) and the 45–degree line test. The computed t-values and the slopes of the chosen TDANN model, for the Almopeos reservoir daily inflow values, for the calibration, the training, the test and the validation data sets, are given in Table 4.

Table 4. t-value, two-sided tabular value (a=0.05) and slope of the TDANN model, for the daily inflow values into Almopeos reservoir, for the calibration, training, test and validation data sets

Q_R / TDANN: 31-8-1/0.9957			
Sample size	t-value	Two-sided tabular value (a=0.05)	Slope (^{o})
Calibration (365) (1987)	0.7564	1.9665	45.98
Train (328)	1.1721	1.9673	45.99
Test (37)	1.2280	2.028	45.94
Validation (365) (1995)	1.0258	1.9665	44.24

The computed t-values of the chosen TDANN model were less than two-sided tabular t-values, for the calibration, the training, the test and the validation data sets (Table 4). These imply that there were no significant differences between the observed and the forecasted values. Also, the observed values and the forecasted values yielded slopes close to 45 degrees, for the calibration, the training, the test and the validation data sets (Table 4). It can be observed that the TDANN model tended to make an angle of 45 degrees with the axes, meaning there is no significant difference between the observed and the forecasted values. Since the data in the test and validation data sets were never seen by the chosen TDANN model, the good predictions on these data sets (Tables 2, 3 and 4) demonstrated the adequacy and the potential of the chosen TDANN model for forecasting daily reservoir inflow. Tables 2, 3 and 4 clearly demonstrate the ability of the chosen TDANN model to forecast very well daily inflow values, into Almopeos reservoir. Consequently, the TDANN models seem promising to be applicable for forecasting daily reservoir inflow.

4 Conclusions

In this paper, Time Delay Artificial Neural Network (TDANN) models were developed for forecasting daily inflow values into Almopeos river reservoir. The training of the TDANNs was achieved by the cascade correlation algorithm which is a feed-forward and supervised algorithm with dynamic expansion. Kalman's learning rule was used to modify the artificial neural network weights. The networks are designed by putting weights between neurons, by using the hyperbolic-tangent function for training. The number of nodes in the hidden layer was determined based on the maximum value of the coefficient of correlation. In the training process, the test data were not used with no way neither using them as part of the training procedure or as part of the decision when to stop training. No fixed number of iterations used as the stopping criterion of the procedure. The choice of the input variables introduced to the input layer based on the cross-correlation. The use of cross-correlation between the ith input series and the output during the calibration period provides a short cut to the problem of the delayed memory determination.

The results, for the training, the test, the calibration and the validation data sets clearly demonstrate the ability of the TDANN models for forecasting daily reservoir inflow. The TDANN models introduced in this study have the ability to forecast the peak reservoir inflows with reasonable accuracy, to develop a generalised solution as there is no overfitting and to overcome the problems in data of daily reservoir inflows such as outliers and noise in the data. Since the proposed methodology is based on the information contained in the data series itself, the Time Delay Artificial Neural Network approach becomes more explicit and can be adopted for any reservoir daily inflow forecasting.

Acknowledgements

The authors are grateful to the Ministry of National Education and Religious Affairs and European Union for the financial support received through the research project "Optimal Reservoir Operation for Irrigation Purposes" (No 80869) which is granted through the Managing Authority of Operational Programme "Education and Initial Vocational Training" – Environment – PYTHAGORAS II.

References

[1] Agarwal, J. & Singh, R.D., Runoff modelling through back propagation artificial neural networks with variable rainfall-runoff data, *Water Resour. Manage.,* Vol. 18, 2004, pp. 285-300.

[2] Anmala, J., Zhang, B. & Govindaraju, R.S., Comparison of ANNs and empirical approaches for predicting watershed runoff, *J. Water Resour. Planning Manage.,* Vol. 126, No. 3, 2000, pp. 156-166.

[3] Brown, R.G. & Hwang, P.Y.C., *Introduction to random signals and applied Kalman filtering,* 2nd Ed., Wiley, New York, 1992.

[4] Coulibaly, P., Anctil, F. & Bobee, B., Daily reservoir inflow forecasting using artificial neural networks with stopped training approach, *J. Hydrol.,* Vol. 230, 2000, pp. 244-257.

[5] Demuth, H. & Beale, M., *Neural network toolbox. For Use with Matlab®*, User's Guide, V4., 2001.

[6] Diamantopoulou, M., Artificial neural networks as an alternative tool in pine bark volume estimation, *Comput. Electron. Agric.,* Vol. 48, 2005, pp. 235-244..

[7] Dowla, U.F. & Rogers, L. *Solving problems in environmental engineering and geosciences with artificial neural networks,* Massachusetts, USA., 1995.

[8] Fahlman, S.E. & Lebiere, C., *The cascade correlation learning architecture,* in D.S. Touretsky (ed.), Advances in Neural Information Processing Systems 2, San Mateo, CA, Morgan Kaufmann, pp. 524-532, 1990.

[9] Fausett, L., *Fundamentals of neural networks architectures. Algorithms and applications,* Prentice Hall, USA, 1994.

[10] French, M.N., Krajeswski, W.F. & Cuykendall, R.R., Rainfall forecasting in space and time using a neural network, *J. Hydrol.,* 137, 1992, pp. 1-31.

[11] Georgiou, P.E. *Reservoir sizing by using Rippl and Transition Probability Method in historical and synthetic inflow time series,* M.Sc., School of Agriculture A.U.TH., 1996 (in Greek).

[12] Govindaraju, R.S. & Rao, A.R. (eds) *Artificial neural networks in hydrology,* Kluwer Academic Publishers, Amsterdam, 2000.

[13] Grewal, M.S. & Andrews, A.P., *Kalman filtering theory and practice,* Upper Saddle River, Prentice Hall, NJ, USA, 1993.

[14] Gurney, K., *An introduction to neural networks,* UCL Press, UK, 1999.

[15] Haykin, S., *Neural networks: A comprehensive foundation,* Macmillan, New York, USA, 1994.

[16] Islam, S. & Kothari, R., Artificial neural networks in remote sensing of hydrologic processes, *J. Hydrol. Eng.,* Vol. 5, No. 2, 2000, pp. 138-144.

[17] Kumar, D.N, Raju, K.S. & Sathish, T., River flow forecasting using recurrent neural networks, *Water Resour. Manage.,* Vol. 18, 2004, pp. 143-161.

[18] Karunanithi, N., Grenmey, W.J., Whitley, D. & Bovee, K., Neural networks for river flow forecasting, *J. Comp. Civil Eng.,* Vol. 8, No. 2, 1994, pp. 201-218.

[19] Lorrai, M. & Sechi, G.M., Neural nets for modelling rainfall-runoff transformations, *Water Resour. Manage.,* 9, 1995, pp. 299-313.

[20] Loukas, A., Vasiliades, L. & Dalezios, N., A hydrologic modeling procedure using a watershed model and artificial neural networks, *Proc. of Inter. Conf., Water Resources Management in the ERA of Transition,* European Water Resources Association, Athens, 2002, pp. 219-228.

[21] Maier, H.R. & Dandy, G.C., The use of artificial neural networks for the prediction of water quality parameters, *Water Resour. Res.,* Vol. 32, No. 4, 1996, pp. 1013-1022.

[22] Maier, H.R. & Dandy, G.C., Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications, *Environ. Model. Soft.;* Vol. 15, 2000, pp. 101-124.

[23] Mantoglou, A. & Kourakos, G., Modeling the rainfall-runoff process using neural networks with emphasis on optimization of network configuration using genetic algorithms, *Proc. of Inter. Conf., Water Resources Management in the ERA of Transition,* European Water Resources Association, Athens, 2002, pp. 229-242.

[24] Masters, T., *Practical neural networks recipes in C^{++}*, Academic Press, San Diego, CA, 1993.

[25] Papamichail, D.M. & Georgiou, P.E., Seasonal ARIMA inflow models for reservoir sizing, *J. Am. Water Resour. Assoc.,* Vol. 37, No. 4, 2001, pp. 877-885.

[26] Papamichail, D.M. & Papazafiriou, Z. G., Multiple input-single output linear functional models for river flow routing, *J. Hydrol.,* Vol. 133, 1992, pp. 365-377.

[27] Patterson D. *Artificial neural networks,* Prentice Hall, Singapure, 1996.

[28] Raman, H. & Sunilkumar, N., Multivariate modelling of water resources time series using artificial neural networks, *Hydrol. Sci. J.,* Vol. 40, No. 1, 1995, pp. 145-163.

[29] Shamsheldin, A.Y., O'Connor, K.M. & Liang, G.C., Methods for combining the outputs of different rainfall-runoff models, *J. Hydrol.,* Vol. 197, 1994, pp. 203-229.

[30] Thirumalaiash, K. & Deo, M.C., River stage forecasting using artificial neural networks, *J. Hydrol. Eng.,* Vol. 3, No. 1, 1998, pp. 26-32.

[31] Tokar, A.S. & Markus, M., Precipitation-runoff modelling using artificial neural networks and conceptual models, *J. Hydrol. Eng.,,* Vol. 5, No. 1, 2000, pp. 156-161.

[32] Yevjevich, V., *Stochastic processes in hydrology,* Water Resources Publications, Fort Collins, CO, 1972

[33] Zealand, C.M., Burn, D.H. & Simonovic, S.P., Short tern streamflow forecasting using artificial neural networks, *J. Hydrol.,* Vol. 214, 1999, pp. 32-48.

[34] Zhang, Q. & Stanley, S.J., Forecasting raw-water quality parameters for the North Saskatchewan river by neural network modelling, *Water Res.,* Vol. 31, No. 9, 1997, pp. 2340-2350.