Available online at www.elixirpublishers.com (Elixir International Journal)

Aquaculture

Elixir Aquaculture 75 (2014) 27455-27464

ISSN: 2229-712X

Forecasting water quality time series with a hybrid methodology

Mohammad Reza Golabi, Ali Mohammad Akhondali and Feridon Radmanesh

Faculty of Water Engineering, Shahid Chamran University, Ahvaz, Iran.

ARTICLE INFO

Article history:

Received: 16 February 2013;

Received in revised form:

20 September 2014;

Accepted: 29 September 2014;

Keywords

Neural networks,

ARIMA,

Hybrid model,

Time series.

ABSTRACT

Typical time series prediction methods used in many real-world applications. The proposed approach consists of an ARIMA methodology and feed-forward, backpropagation network structure with an optimized conjugated training algorithm. The hybrid approach for time series prediction is tested using 144-month observations of water quality data, including water temperature, Total dissolved solids (TDS) and Sodium adsorption ratio (SAR), during 1997–2008 at Karun river, Iran. The correlation coefficients between the hybrid model predicted values and observed data for TDS, SAR and water temperature are 0.935, 0.939, and 0.892, respectively, which are satisfactory in common model applications. Predicted water quality data from the hybrid model are compared with those from the ARIMA methodology and neural network architecture using the accuracy measures.

Introduction

Water quality is a growing concern throughout the developing world. Deterioration of water quality has initiated serious management efforts in many countries. Most acceptable ecological and water related decisions are difficult to make without careful modeling, prediction and analysis of river water quality for typical development scenarios. The increasing water demands at the project development stage including agricultural networks, fish hatchery projects, and inter-basin water transfers, have caused a gloomy future for water quality of the Karun River. Computer science and statistics have improved modeling approaches for discovering patterns found in water resources time series data. Much effort has been devoted over the past several decades to the development and improvement of time series prediction models. In recent years, the artificial neural networks (ANN) have been applied to many areas of statistics. One of these areas is time series forecasting (Katijani et al, 2005). Since ANN can model both nonlinear and linear structures of time series, using neural networks in forecasting can give better results than the other methods. Zhang et al.(1998) review the literature of forecasting time series using ANN. Both theoretical and empirical findings in the literature show that combining different methods can be an affective and efficient way to improve forecasts. Therefore, hybrid ARIMA and ANNs methods have been used for modeling both linear and nonlinear patterns equally well. There are a number of studies in which neural networks are used to address water resources problems. Maier and Dandy (2000) reviewed recent papers dealing with the use of neural network models for the prediction and forecasting of water resources variables. Flood and Kartam (1994), Hassoun (1995) and Rojas (1996) have used feedforward networks with sigmoidal-type transfer functions for the prediction and forecasting of water resources variables. Chau (2006) has reviewed the development and current progress of the integration of artificial intelligence into water quality modeling. Hatzikos et al. (2005) utilized neural networks with active neurons as a modeling tool for the prediction of seawater quality indicators like water temperature, pH, dissolved oxygen (DO) and turbidity. Palani et al. (2008) demonstrated the application of ANNs to model the values of selected seawater quality variables, having the dynamic and complex processes hidden in the monitored data itself. Most of the studies reported above were simple applications of using traditional time series approaches and ANNs. Many of the real-life time series are extremely complex to be modeled using simple approaches especially when high accuracy is required. There have been several studies suggesting hybrid models, combining the ARIMA model and neural networks. Durdu (2009) proposed a hybrid approach for time series prediction of water quality data, that including water temperature, boron and dissolved oxygen, during 1996–2004 at Buyuk Menderes river, Turkey. Jain and Kumar (2006) proposed a hybrid approach for time series forecasting using monthly stream flow data at Colorado river. They indicated that the approach of combining the strengths of the conventional and ANN techniques provides a robust modeling framework capable of capturing the nonlinear nature of the complex time series and thus producing more accurate forecasts. Zhang (2003) tested a hybrid ARIMA and ANN model over three kinds of time series, and concluded that the linear and non-linear time series patterns in the combined model improved forecasting more than either of the models used independently. In the present paper, a hybrid approach, combining seasonal ARIMA model and neural network backpropagation model, is developed to predict water quality time series data. The motivation behind this hybrid approach is largely due to the fact that a water quality problem is often complex in nature and any individual model may not be able to capture different patterns equally well. The objectives of the present study are to: (1) develop a hybrid model, an ANN and an ARIMA model, to predict water quality time series data, (2) assess the performance of each modeling approach using observed data versus predicted data and (3) evaluate the predictive performance of hybrid model in comparison to ANN architecture and ARIMA model using accuracy measures.

Tele:

E-mail addresses: hamid_golabi65@yahoo.com

Materials and methods

Study area and water quality data

The Karun River is the largest river by discharge in Iran, the Karun River's watershed covers 65,230 square kilometres (25,190 sq mi) in parts of two Iranian provinces. The river is around 950 kilometres (590 mi) long and has an average discharge of 575 cubic meters per second (20,300 cu ft/s). and discharges into the Persian Gulf. The Karun River begins high in the Zagros Mountains 75 km southwest of Isfahan and flows in a general westerly direction through a descending series of anticlinal ridges and synclinal valleys before emerging on to the Khuzestan Plain at Gotvand, 400 km downstream. The headwaters of the Karun River are located in very high mountain terrain, almost unpopulated, where agriculture is seldom practical. Its main upstream tributaries are the Khersan, the Vanak and the Bazuft. The Karun River passes near the city of Ahwaz, the principal inland city of the Khuzestan plain, and then continues southward for 185 km to discharge into the Arvand-rud at Khoramshahr. The largest city on the river is Ahvaz, with over 1.3 million inhabitants. There are 83 hydrometric stations in the branches of the rivers in the great Karoon River basin. In these stations, discharge, sediment concentration and qualitative parameters are measured. The greatest number of data in this basin is related to Ahvaz Station from where data is available since 1950. We are used of these information of Ahvaz hydrometery station.

The dataset of one measurement location, comprising three water quality parameters monitored over 12 years (1997–2008), was measured by the Ahvaz hydrometery station The selected water quality parameters include water temperature, total dissolved solids (TDS) and sodium adsorption ratio (SAR). The data for these parameters available for the analysis are on amonthly basis for the period of 12 years from 1997 to 2008. In this study, the first 96 months of water quality data for water temperature, TDS and SAR were used for model training. The remaining 48 months of water quality data were used for verification of the model prediction results. The statistical roperties of water quality time series data are demonstrated in Table1. The minimum, maximum, mean, standard deviation (SD), skewness and kurtosis can describe variability of a water quality parameter. As described in Table1, SAR and TDS are the water quality variables that have low skewness coefficients. water temperature has a large skewness coefficient and this indicates that the mean and median values of water temperature have large differences. Probably the mean of water temperature data set is heavily influenced by the presence of afew extreme values.

Table 1- Statistical properties of the water quality parameters

Kurtosis	Skewness	Std .Dev.	Mean	Min	Max	Parameters
3.51	-0.97	0.55	3.8	0.64	4.37	Water temperature (°C)
-0.28	0.6	3.54	7.81	2.9	17.8	TDS
1.57	0.83	1.06	3.95	2.23	8.5	SAR

ARIMA modeling approach

The ARIMA model, sometimes called Box-Jenkins (1976) models is an important forecasting tool, and is the basis of many fundamental ideas in time-series analysis. An autoregressive model of order p is conventionally classified as AR (p) and a moving average model with q terms is known as MA (q). A combined model that contains p autoregressive terms and q moving average terms is called ARMA (p, q). If the object series is differenced d times to achieve stationarity, the model is classified as ARIMA (p, d, q). Thus, an ARIMA model is a combination of an autoregressive (AR) process and a moving average (MA) process applied to a non-stationary data series. So the general non-seasonal ARIMA (p, d, q) model is defined by; AR: p = order of the autoregressive part, I: d = degree of differencing involved and MA: q = order of the moving average part.

The simplest ARIMA (p,d,q) model is given in equation (1)

$$\mathbf{Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \; .... + \emptyset_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \square - \theta_p e_{t-p} \; ... . (1)}$$

or in backshift notation

$$\mathbf{\left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) Y_t = c + \left(1 - \theta_1 B - \theta_2 B^2 - \square - \theta_q B^q\right) e_t \; ..... (2)}$$

Where, c = constant term, $\boldsymbol{\emptyset_i = i^{th}}$ autoregressive parameter, $\theta_j = j^{th}$ moving average parameter, e_t = the error term at time t and $B^k = k^{th}$ order backward shift operator.

In addition to the non-seasonal ARIMA (p, d, q) model, seasonal ARIMA (P, D, Q) parameters for the used data can be identified. These parameters are: Seasonal autoregressive (P), seasonal Differencing (D) and seasonal moving average (Q). The general form of the Seasonal ARIMA (p, d, q) (P, D, Q)S model using backshift notation is as in equation (3).

$$\mathbf{\emptyset_{AR}(B)\emptyset_{SAR}(B^s)(1 - B)^d (1 - B^s)^D Y_t = \theta_{MA}(B)\theta_{SMA}(B^s) e_t \; ... ... (3)}$$

Where, s = number of periods per season, $\boldsymbol{\emptyset_{AR}} =$ non-seasonal autoregressive parameter, $\theta_{MA} =$ non-seasonal moving average parameter, $\theta_{SMA} =$ seasonal moving average parameter.

The ARIMA modeling approach involves the following three steps: model identification, parameter estimation, diagnostic checking. Identification of the general form of a model includes two stages: (1) if it is necessary, appropriate differencing of the series is performed to achieve stationary and normality; (2) the temporal correlation structure of the transformed data is identified by examining its autocorrelation (ACF) and partial autocorrelation (PACF) functions (Mishra and Desai, 2005). The ACF is a useful statistical tool that measures if earlier values in the series have some relation to later values. PACF is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all low order lags. Considering the ACF and PACF graphs of water quality concentration series, different ARIMA models are identified to model selection.

The model which has the minimum AIC value is the model of interest. The mathematical formulation for the AIC is developed as

$$AIC=n(\ln(2\pi \mathbf{RSS})/\mathbf{n})+\mathbf{1})+\mathbf{2m} \quad (4)$$

where m=(p+q+P+Q) is the number of terms estimated in the model and RSS denotes the sum of squared residuals.

After the functions of the ARIMA model have been specified, the parameters of these functions must be estimated. Once an appropriate model is chosen and its parameters are estimated, the Box–Jenkins methodology requires examining the residuals of the model to verify that the model is an adequate one for the series. Several tests are employed for diagnostic check to determine whether the residuals of the selected ARIMA models from the ACF and PACF graphs are independent, homoscedastic and normally distributed. If the homoscedasticity and normality assumptions are not provided, the observations are transformed by a Box–Cox transformation (Wei, 1990). For a good forecasting model, the residuals, left over after fitting the model, must satisfy the requirements of a white noise process (uncorrelated and normally distributed around a zero mean). In order to determine whether water quality time series are independent, the residual auto- correlation (RACF) function of the series is studied. There are several useful tests related to RACF for the independence of residuals. The first one is the correlograms drawn by plotting the residual ACF function against lag number. If the ARIMA model is correct, the estimated autocorrelations of the residuals are uncorrelated and distributed approximately normally about zero. The second one is Ljung–Box–Pierce statistics. In order to test the null hypothesis that acurrent set of autocorrelations is white noise, test statistics are calculated for different total numbers of successive lagged autocorrelations using the Ljung–Box–Pierce statistics (Q(r) test) to test the adequacy of the model. Q(r) values are compared to a critical test value(χ^2) distribution with respective degree of freedom at a 5% significant level. The third one is the cumulative periodogram, employed to diagnose the residuals for a white noise sequence.(El-Din and Smith, 2002).

Structure of neural network model

In this study, an empirical neural network algorithm was applied to estimate water quality parameters (SAR, TDS, water temperatur). ANN models are highly flexible function-approximators that have shown their utility in a broad range of water resources applications. Feed forward, backpropagation networks have previously been identified as the most common type of ANN models used in water resources applications. Therefore, such networks were used in the current study. Networks constructed in the current study comprised of three layers: an input layer, a hidden layer and an output layer. An example of a network topology is shown in Fig. 1.

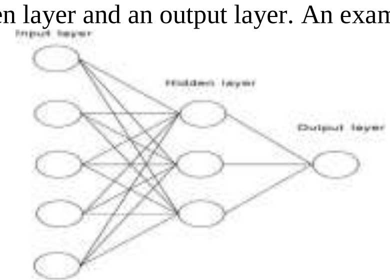


Fig. 1. An example of an artificial neural network topology with one input layer, one hidden layer and one output layer

A neural network must be trained to determine the values of the weights that will produce the correct outputs. In a training step, a set of input data is used for training and presented to the network many times. The performance of the network is tested after the training step is stopped. The backpropagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient). This is the direction in which the performance function is decreasing most rapidly. It turns out that although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. Therefore, the basic gradient descent training algorithm is inefficient owing to its slow convergent speed and at times the poor accuracy in model predictions (Huang et al., 2004). From an optimization point of view, training a neural network can be considered as equivalent to minimizing a multivariable global error function of the network weights. There are several optimized training algorithms, as described by Haykin (1999), such as resilient backpropagation, Levenberg–Marquardt and conjugated gradient backpropagation . On of the optimized methods developed by Moller (1993) is the scaled conjugate gradient (SCG) algorithm. The SCG training algorithm was developed to avoid the time-consuming line search. In the conjugate gradient algorithm a search is performed along conjugate directions, which produces faster convergence than steepest descent directions . The standard backpropa gation algorithm, traditionally employed in neural network learning, evaluates the gradient of the global error function with respect to the weights, f (W^k), at each iteration k, and updates the weights according to

$$\mathbf{W^{k+1}=W^k-\alpha^k\nabla_f(W^k)} \quad (5)$$

The step size $\alpha^k > 0$ is a user-selected learning rate parameter, which affects the performance of the learning algorithm to a great extent. In all cases, the backpropagation algorithm may follow a zigzag path to the minimum, typical for a steepest gradient descent method (Falas and Stafylopatis, 2005). A conjugate gradient algorithm avoids the zigzag approach to the minimum point by incorporating a special relationship between the direction and gradient vector at each iteration. If D^k represents the direction vector at iteration k of the algorithm, then the weight vector is updated according to the rule

$$\mathbf{W^{k+1}=W^k+\alpha^k D^k} \quad (6)$$

Given values of W^k and D^k, a particular values of α^k that reduce the objective function as much as possible needs to be found. After a small number of iterations, the search along the line direction to find the optimum step size for the actual minimum should stop. Estimating the optimum step size with scaled conjugate gradient (SCG) training algorithm increases the learning speed and eliminates the dependence on critical user-selected parameters. The main idea behind the algorithm is the use of a factor ρ which is raised or lowered with in each iteration during the execution of the algorithm, looking at the sign of the quantity δ^k, which reveals if the Hessian matrix is not positive definite (Falas and Stafylopatis, 2005).

In this study, before the training of the network both input and output variables were normalized within the range 0.1– 0.9 as follows:

$$\mathbf{x} = 0.8\,\frac{(\mathbf{x} - \mathbf{x}_{\min})}{(\mathbf{x}_{\max} - \mathbf{x}_{\min})} + 0.1 \tag{7}$$

is the normalized value of a certain parameter, x is the measured value for this parameter, x_{min} and x_{max} are the minimum and maximum values in the database for this parameter, respectively (Dogan et al., 2009).

Hybrid model

A hybrid model having both linear and nonlinear modeling abilities could be a good alternative for predicting water quality data. By combining different models, different aspects of the underlying patterns may be captured. Following the hybrid model structure of Zhang (2003), a water quality time series could be composed of a linear autocorrelation structure and a nonlinear component. That is,

$$\mathbf{y_t = L_t + S_t} \tag{8}$$

where L_t represents the linear component and S_t represents the nonlinear component. Both of these two parameters have to be estimated from the time series data. First ARIMA model is used to capture the linear component, then the residuals from the linear model will contain only the nonlinear relationship. The residuals e_t at time t from the linear model is defined by

$$\mathbf{e_t = y_t - \hat{L}_t} \tag{9}$$

where $\widehat{L_t}$ is the predicted value of the ARIMA model at time t. The diagnostic check of the residuals is important to determine the adequacy of the ARIMA models. An ARIMA model is not sufficient is not adequate if there are still linear correlation structures left in the residuals. However, diagnostic check of the residuals is not able to detect any nonlinear patterns in the times series data. For this reason, even if the residuals pass the diagnostic check and the model is an adequate one, the model may still not be sufficient in that nonlinear relationships have not been appropriately modeled. Therefore, the residuals can be modeled by using ANNs to discover nonlinear relationships. With N input nodes, the ANN model for the residuals will be

$$\mathbf{e_t = f(e_{t-1}, e_{t-2}, \ldots\ldots, e_{t-N}) + \varepsilon_t} \tag{10}$$

Where f is a nonlinear function determined by the neural network and e_t is the random error. Finally the combined prediction will be

$$\mathbf{\widehat{Y_t} = \widehat{L_t} + \widehat{S_t}} \tag{11}$$

where $\widehat{S_t}$ represents the prediction from Eq.(10).

Model verification and comparison methods

Three different forecast consistency measures are used in order to compare the performances of obtained ARIMA and artificial neural network (ANN): root mean square error (RMSE), the mean absolute percentage error (MAPE) and the correlation coefficient (r).

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - Q_i)^2} \tag{12}$$

$$\mathbf{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{P_i - Q_i}{P_i}\right| 100 \tag{13}$$

$$\mathbf{r} = \sqrt{1 - \frac{\sum_{i=1}^{n}(P_i - Q_i)^2}{\sum_{i=1}^{n} Q_i^2 - \frac{\sum_{i=1}^{n} P_i^2}{n}}} \tag{14}$$

Here Pi and Oi are the predicted and observed values, respectively. And n is the total number of observations.

Results and discussion

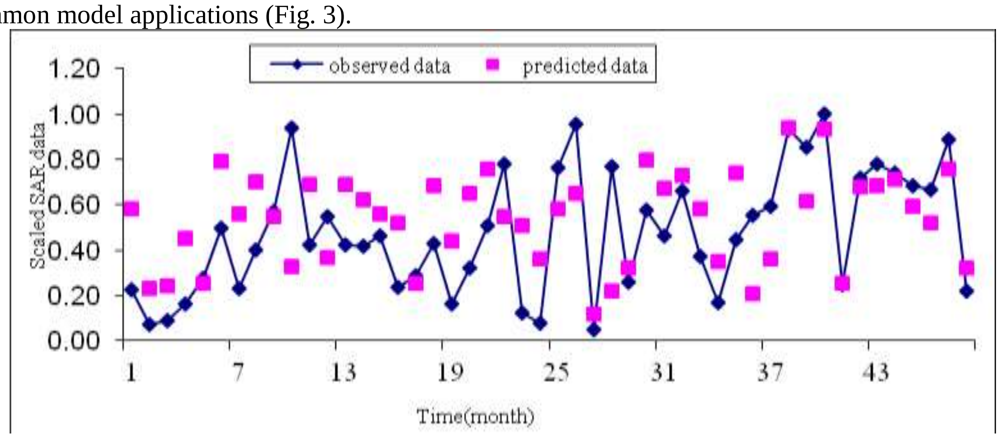
ARIMA modeling

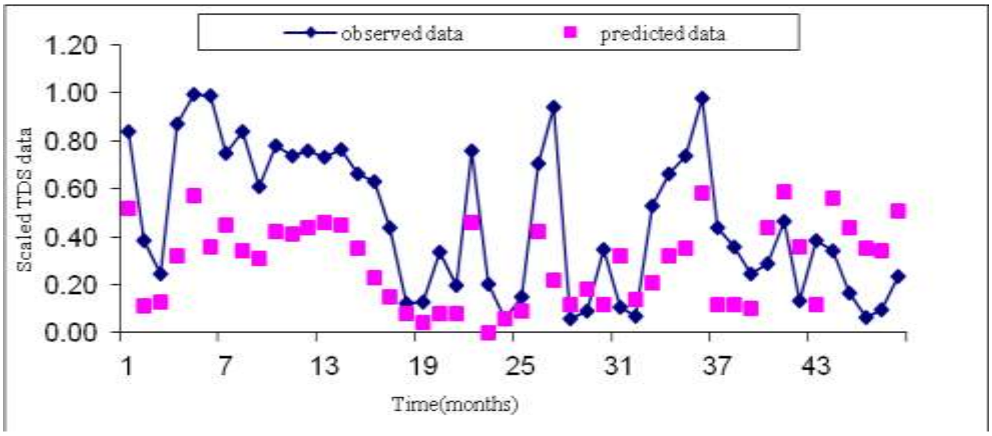
The ARIMA models were used to predict monthly water quality time series over the period between 1997 and 2008. The water quality data for the period between 1997 and 2004 were used for model calibration and to obtain the best model fit for each water quality parameter. The data for the period between 2005 and 2008 were used for model verification and comparisons for prediction purposes. To fit ARIMA model to the available water quality time series data, three-stage procedure of model identification, estimation of model parameters and diagnostic checking of the estimated parameters was employed. In the identification stage, to determine the possible persistence structure in the time series data, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) were used. Using Akaike Information Criteria (AIC), the best fitted model has been identified out of the various competing models. As demonstrated in Table2. In the estimation of model parameters stage, the best fit ARIMA model statistical parameters were estimated. The computational method outlined by Box and Jenkins (1976) was employed to estimate model parameters. In the diagnostic checking of the estimated parameters stage, diagnostic checks were done to insure that the best fit model was selected by checking that assumptions of ARIMA modeling such as independence, homoscedastic (constant variance) and normality of the residual a_t were satisfied. In order to check the independence of residuals, the residual autocorrelation function (RACF), Ljung–Box–Pierce statistics and cumulative periodograms were used. The values of residual autocorrelation functions (RACF) were well settled within confidence limits except very few individual correlations appear large compared to the confidence limits, which were acceptable among 30 lags. The results exhibited no significant correlation between the residuals of the each water quality parameter. The values of Ljung–Box–Pierce Q(r) statistic is shown in Table2. The values of Q(r) were compared to a critical test value (χ^2) distribution with respective degree of freedom at a 5% significant level. It was obvious that the computed values were less than the actual (χ^2) values, which indicated that the residuals from the best models were white noise (Table2). Cumulative periodograms confirmed that no significant periodicity was available in the residual series at 95% confidence level and indicated that the points were clustering closely about the theoretical line and there was no evidence of periodic characteristics buried in the residual series. In order to check the normality of residuals, the histograms and normal probability plot of residuals were investigated and they clearly supported the assumption of normality. In order to investigate homoscedasticity of the residuals, a plot of residuals versus fitted values were examined and the plots showed a random scatter around zero. In other words, the residuals were evenly distributed around mean, which explains the models were adequate.

Table 2- Summary of the statistical parameters of the best fitted multiplicative ARIMA models fitted to water quality parameters

AIC	λ	Q	θ_{SMA}	$Ø_{SAR}$	θ_{MA}	$Ø_{AR}$	Model	Parameters
100.07	33.9	27.5	-	-0.81	-	0.021	(1,0,0)(1,1,0)12	Water temperature (°C)
46.073	33.9	21.6	0.67	-	0.606	-	(0,1,1)(0,1,1)12	TDS
163.77	33.9	25.5	-	-0.31	-0.44	-	(0,0,1)(1,1,0)12	SAR

The trained ARIMA model was then tested using water quality data set for the period of 48 months. As shown in Fig. 2. The correlation coefficient values between models predicted values and observed data for TDS, SAR and water temperature are 0.549, 0.524 and 0.721, respectively, which are not satisfactory in common model applications (Fig. 3).





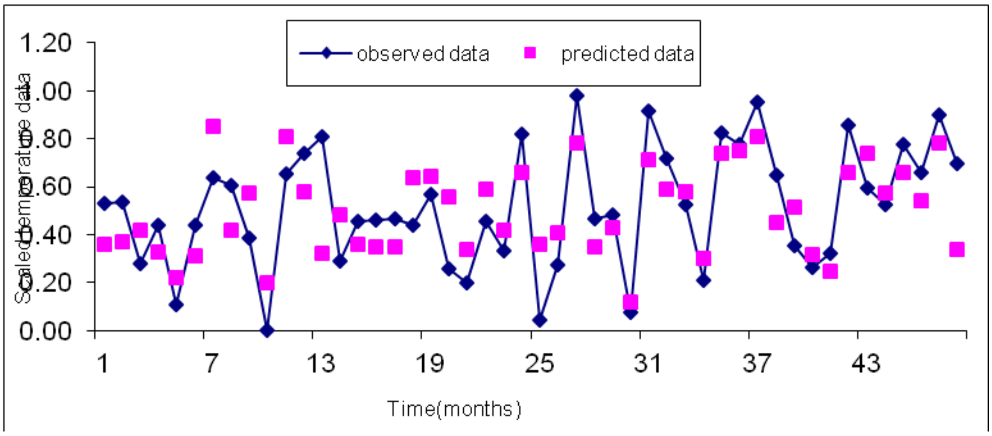


Fig. 2. ARIMA model verification for each water quality constituent.

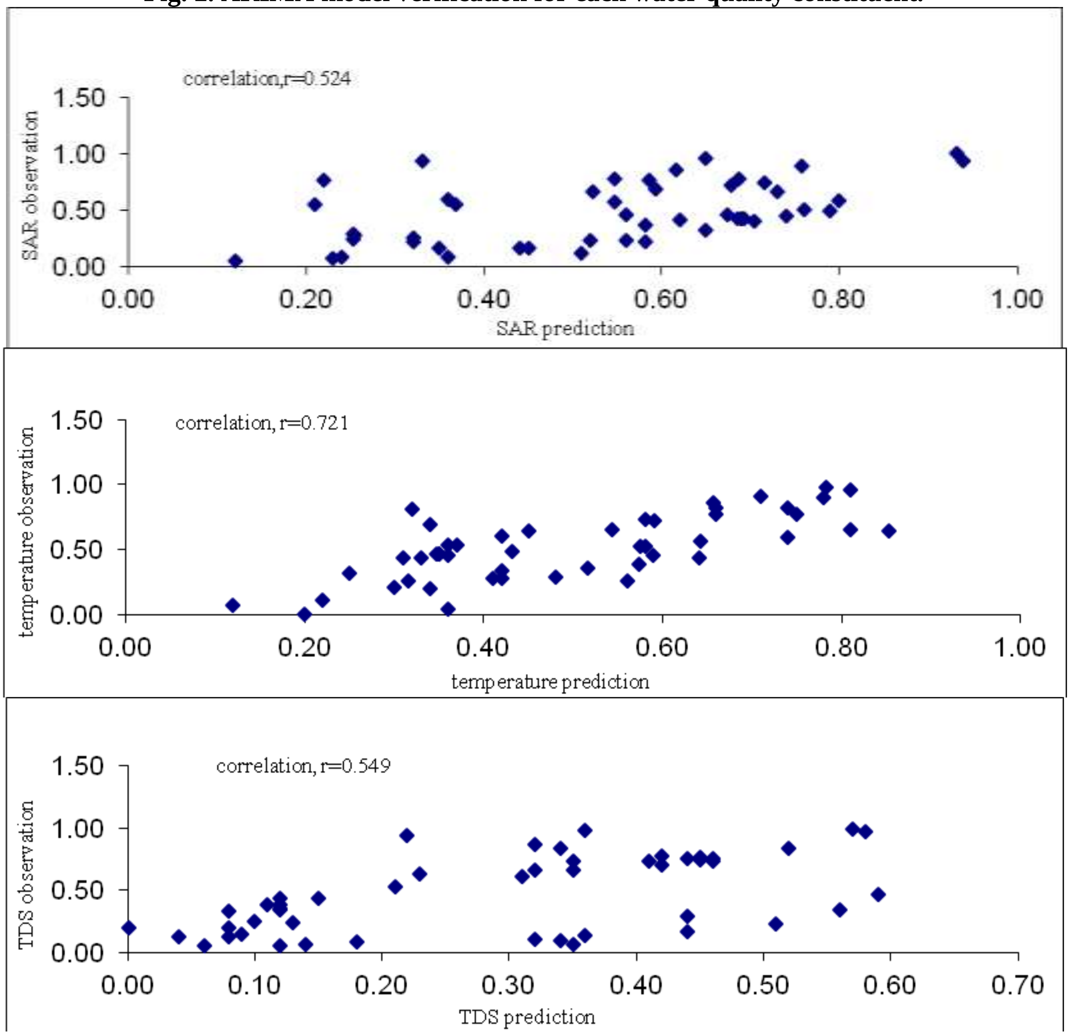


Fig. 3. Observed versus ARIMA predicted data for each water quality constituent

Neural network modeling

In the present study, the scaled conjugated gradient algorithm was selected as the optimized training method. In the following part, artificial neural network model performances were validated for flow prediction under monthly time-step condition. The data for the period between1997 and 2008 were available for the modeling purposes. TDS, SAR and water temperature time series data were divided into two independent data sets. The first data set was used for model training, and the second data set was used for model verification purposes. In the ANN modeling process. In the model verification phase, the trained network was used to predict the monthly water quality parameters. Fig. 4 compares the model predictions for water quality parameters with the observations. The verifications stage indicate that the model prediction results reasonably match the observed water quality parameters. The correlation coefficient between the ANN model predicted values and observed data for TDS, SAR and water temperature are 0.920, 0.926 and 0.875, respectively, which are satisfactory in common model applications (Fig. 5). These results indicate that the neural network model is able to recognize the pattern of the water quality parameters to provide good predictions of the monthly variations of water quality data of the Karoon river.

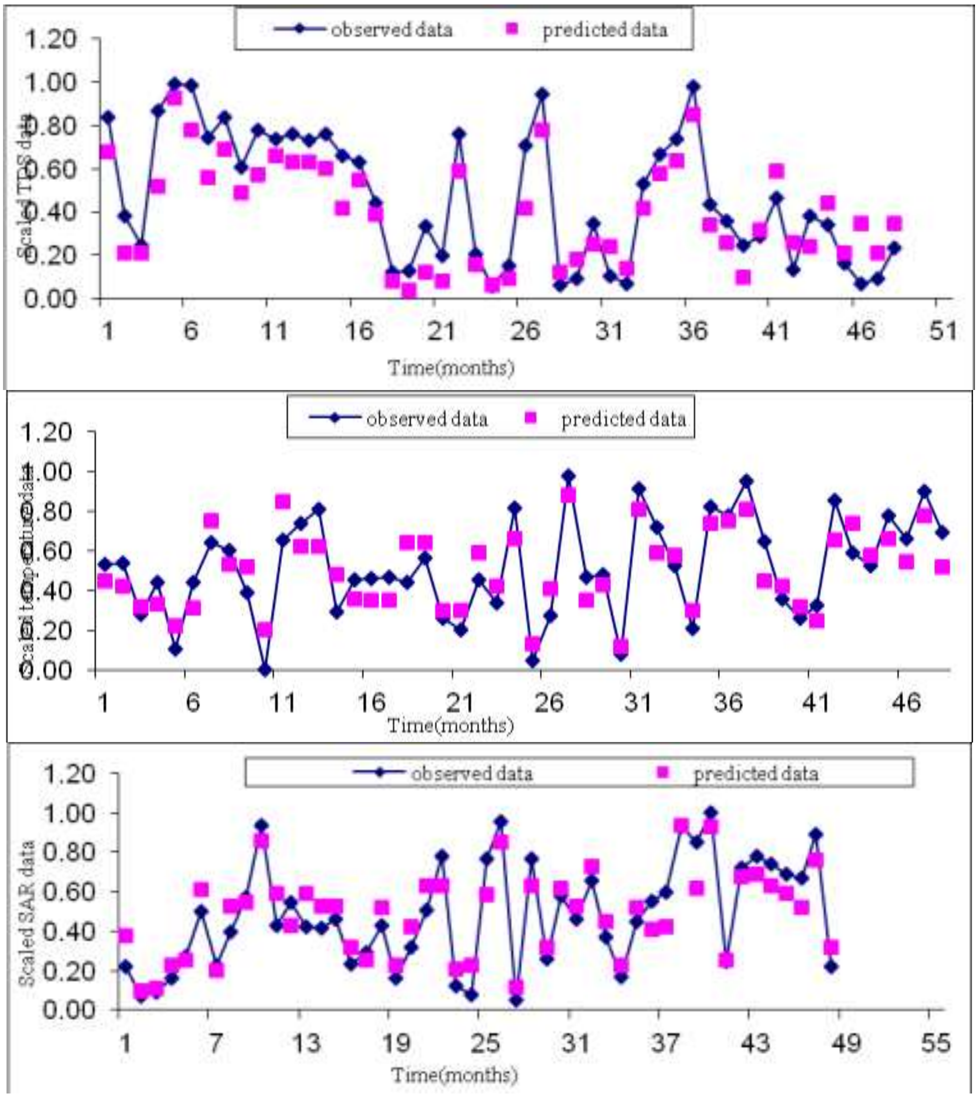
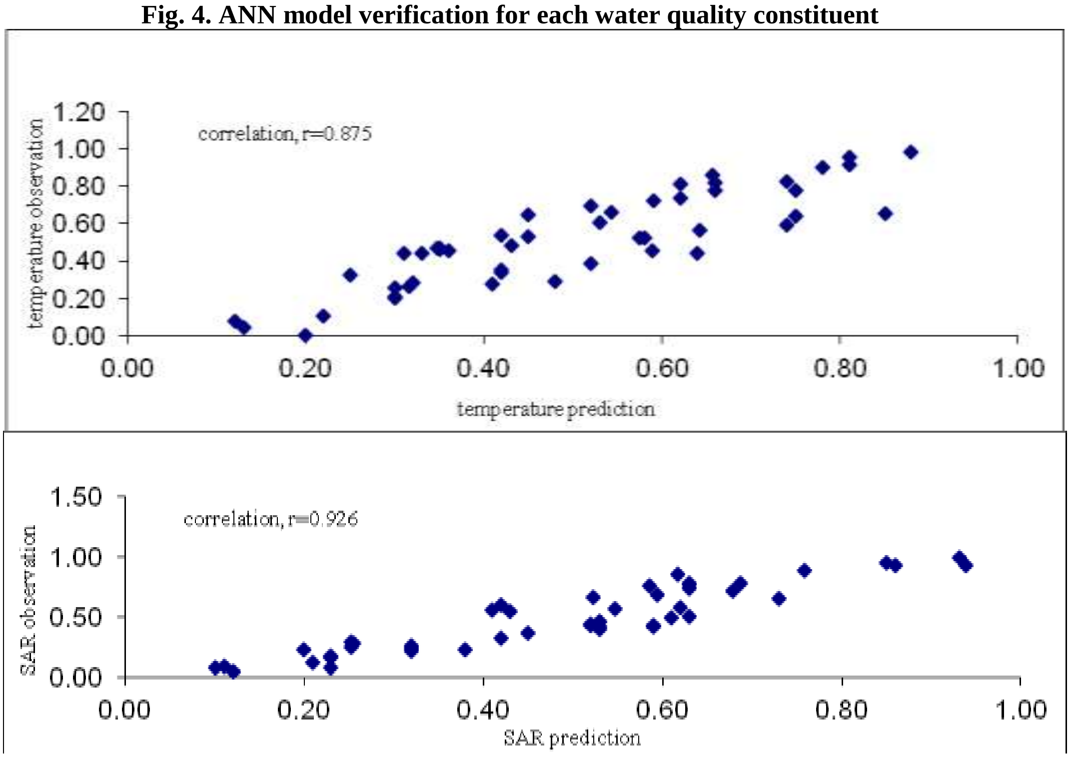


Fig. 4. ANN model verification for each water quality constituent



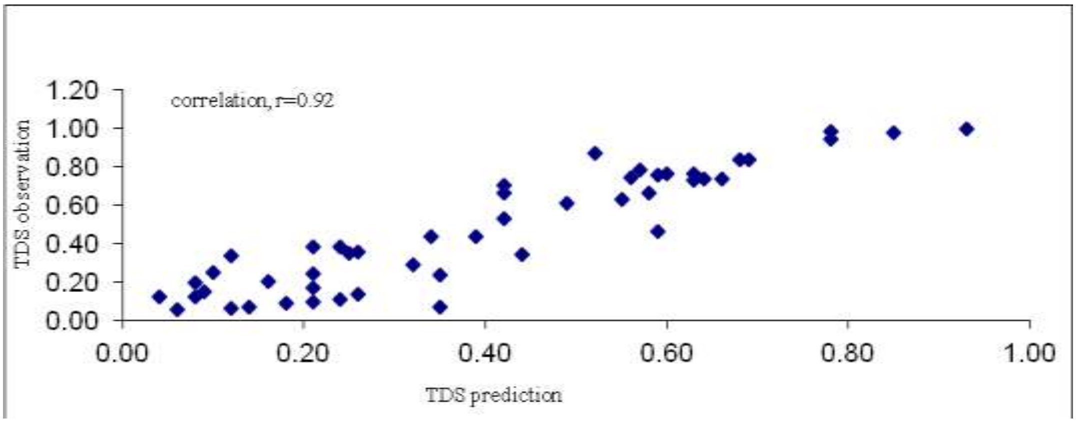


Fig. 5. Observed versus ANN predicted data for each water quality constituent.

Hybrid modeling

In the modeling process, the hybrid model was trained to adjust the model so that the model predicted water quality parameters match well with observed data. Fig. 6 compares the model predictions for water quality parameters with the observations. The verifications results indicate that the model prediction results reasonably match the observed water quality parameters. The correlation coefficient between the hybrid model predicted values and observed data for TDS, SAR and water temperature are 0.935, 0.939 and 0.892, respectively, which are satisfactory in common model applications (Fig. 7).

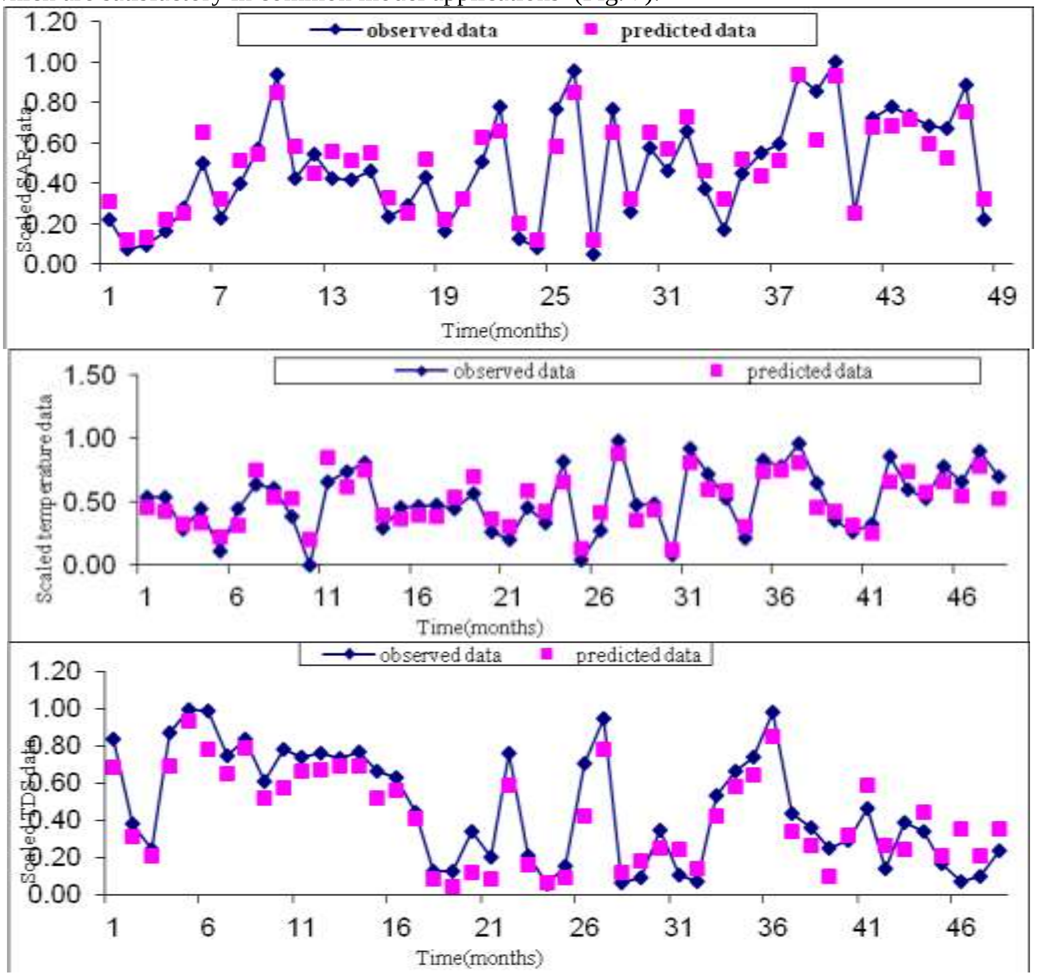


Fig. 6. Hybrid model verification for each water quality constituent

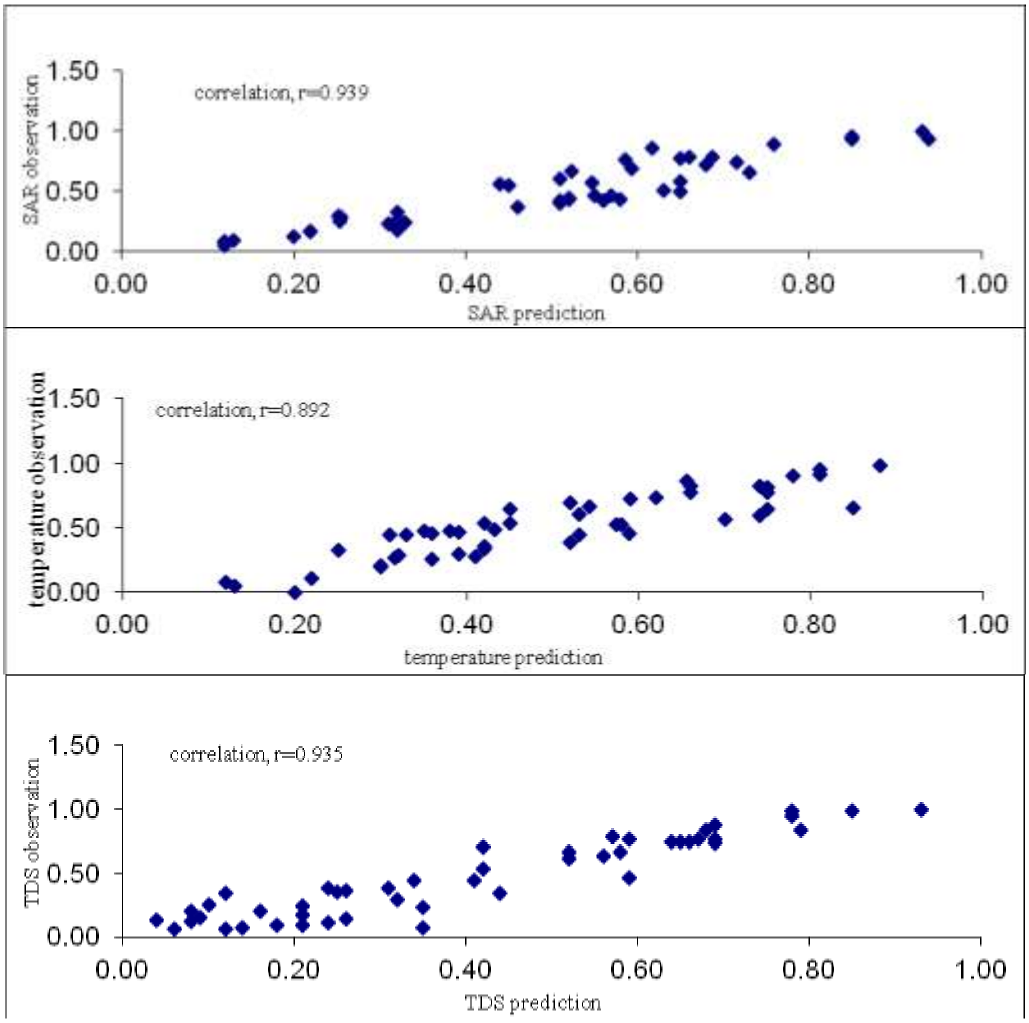


Fig. 7. Observed versus hybrid model predicted data for each water quality constituent

Comparison of model performances

Employing accuracy measures (RMSE, MAPE and R), the predicted data and observed data for the period of 48-months from the hybrid, ANN and ARIMA models were compared to determine the best performed model. The predicted water quality parameters using the ARIMA models were not found to be in reasonable agreement with the observed data. However, the hybrid and ANN approaches provided reasonable precision for all water quality parameters. Table3 gives the error estimates of the two different approaches used in the study for predicting water quality parameters. The RMSEs between observed and predicted data were calculated in ARIMA models as 0.127 ^{0}C, 0.192 and 0.141 for water temperature, TDS and SAR, respectively. In the case of ANN modeling approach, the RMSEs between observed and predicted data were computed as 0.07 ^{0}C, 0.098 and 0.086 for water temperature, TDS and SAR, respectively. Applying the hybrid method, there were a decrease of 15.71%, 9.18% and 8.139% in the RMSE values of ANN for water temperature, TDS and SAR, respectively. Furthermore, the MAPEs between observed and predicted data for water temperature, TDS and SAR were appeared to be slightly lower for the ANN modeling approach. Prediction error statistics for the ANN approach produced MAPEs of 26.22%, 41.44% and 33.85% for water temperature, TDS and SAR, respectively. In case of the MAPE values, the improvement of the hybrid model over the ANN model were 8.89%, 6.51% and 6.28% for water temperature, TDS and SAR, respectively. These results indicated that the hybrid model performed well for adequate predicting of water temperature, TDS and SAR. Therefore, it can be concluded that the hybrid modeling approach can give more reliable predictions of water temperature, TDS and SAR time series of a river than the ANN and ARIMA modeling approach.

Table 3- Statistical comparison of observed and predicted data from the hybrid, ARIMA and artificial neural network(ANN) modeling approaches.

		MAPE				RMSE		Parameters
Hybrid	ANN	ARIMA		Hybrid	ANN	ARIMA		
23.887	26.219	46.181		0.059	0.07	0.127		Water temperature (°C)
38.746	41.446	61.145		0.089	0.098	0.192		TDS
31.723	33.85	50.341		0.079	0.086	0.141		SAR

Conclusions

An empirical comparative evaluation of the performance of hybrid model to the ANN and ARIMA modeling approach was presented for river water quality predictions. The proposed modeling framework gradually receives the data filtered using the ARIMA models and then the residuals from the ARIMA approach were analyzed by ANNs to capture the nonlinearity in the time series involved. Investigations were conducted to examine the hybrid model performance for predicting river water quality in monthly time steps. The results from the ARIMA models poorly represented the pattern of water quality data for TDS and SAR, but the model produced acceptable results for water temperature. The results from the ANN model were capable of providing accurate predictions of water quality parameters at the proposed time step. In the proposed hybrid model, an ARIMA model was used to analyze the linear part of the problem and then the residuals from the ARIMA model were modeled by using a neural network model. The results from the hybrid model indicated that the modeling approach gave more reliable predictions of water temperature, TDS and SAR time series data. The predictions from hybrid model were compared with those obtained from the ANN and ARIMA traditional time series approaches. Owing to its ability in recognizing time series patterns and nonlinear characteristics, the accuracy measures RMSE and MAPE demonstrated that the hybrid model provided much better accuracy over the ANNs and ARIMA methods for water quality predictions. Therefore, the proposed hybrid algorithm can be used for the Karun river and other hydrometerologically similar rivers for predicting water quality data of monthly time step to detect water quality severity with respect to water temperature, TDS and SAR future. The hybrid model developed for the Karun river can be employed for the development of a water quality emergency management plan so as to ensure sustainable water resources management in the basin.

References

1- Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis Forecasting And Control. Holden-Day, San Francisco, USA.

2- Chau, K.W., 2006. A review on integration of artificial intelligence into water quality modeling. Marine Pollution Bulletin 52, 726–733.

3-Dogan E, Sengorur B, Koklu R.,2009. Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. Journal of Environmental Management 90: 1229–1235.

4- Durdu, O.F., 2010. A hybrid neural network and ARIMA model for water quality time series prediction. Engineering Applications of Artificial Intelligence 23, 586–594.

5- El-Din, A.G., Smith, D.W., 2002. A combined transfer-function noise model to predict dynamic behavior of a full-scale primary sedimentation tank. Water Research 36, 3747–3764.

6- Falas, T., Stafylopatis, A., 2005. Implementing temporal difference learning with the scaled conjugated gradient algorithm. Neural Processing Letters 22, 361–375.

7- Flood, I., Kartam, N., 1994. Neural networks in civil engineering I: principles and understanding. Journal of Computing in Civil Engineering 8, 131–148.

8-Hassoun, M.H., 1995. Fundamentals of Artificial Neural Networks. MIT Press, Cambridge.

9- Hatzikos, E., Anastasakis, L., Bassiliades, N., Vlahavas, I., 2005. Simultaneous prediction of multiple chemical parameters of river water quality with tide. In: Proceedings of the Second International Scientific Conference on Computer Science, IEEE Computer Society, Bulgarian Section.

10- Haykin, S., 1999. Neural Networks, a Comprehensive Foundation. Prentice Hall, Upper Saddle River, NJ, USA.

11-Huang, W., Bing, X., Chan-Hilton, A., 2004. Forecasting flows in Apalachicola River using neural neworks. Hydrological Process 18, 2545–2564.

12- Jain, A., Kumar, A., 2006. An evaluation of artificial neural network technique for the determination of infiltration model parameters. Applied Soft Computing 6, 272–282.

13-Katijani,Y., Hipel, W.K., Mcleod, A.I., 2005. Forecasting nonlinear time series with feedforward neural networks: A case study of Canadian lynx data, Journal of Forecasting 24 , 105-117.

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

15- Mishra, A.K., Desai, V.R., 2005. Drought forecasting using stochastic models. Stochastic Environmental Research Risk Assessment 19, 326–339.

16- Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks 6, 525–533.

17- Palani, S., Liong, S.Y., Tkalich, P., 2008. An ANN application for water quality forecasting. Marine Pollution Bulletin 56, 1586–1597.

18-Rojas, R., 1996. Neural Networks: A Systematic Introduction. Springer-Verlag, Berlin.

19-Wei, W.W.S., 1990. Time series analysis. Addison-Wesley Publishing Company, Inc., New York.

20- Zhang, G., Patuwo, B.E., Hu,Y.M., 1998. Forecasting with artificial neural networks: The state of the art, International Journal of Forecasting 14 , 35-62.