

River Water Prediction Modeling Using Neural Networks, Fuzzy and Wavelet Coupled Model

Kulwinder Singh Parmar · Rashmi Bhardwaj

Received: 23 March 2014 / Accepted: 10 October 2014 /

Published online: 21 November 2014

© Springer Science+Business Media Dordrecht 2014

Abstract In this paper, new prediction model introduced by coupling of neural networks model, fuzzy model and wavelet model for the water resources management. Artificial neural network (ANN), fuzzy, wavelet and adaptive neuro-fuzzy inference system (ANFIS) are found to be a sturdy tool to model many non-linear hydrological processes. Wavelet transformation will improve the ability of a prediction model by capturing valuable information on different resolution levels. The target of this research is to compare our model with other famous data-driven models for monthly forecasting of water quality parameter chemical oxygen demand (COD) level monitored at Nizamuddin station, New Delhi, India of river Yamuna based on the past history. The data has been decomposed into wavelet domain constitutive sub series using Daubechies wavelet at level 8 (Db_8). Statistical behavior of wavelet domain constitutive series has been studied. The foretelling performance of the wavelet coupled model has been compared with classical neuro fuzzy, artificial neural network and regression models. The result shows that the wavelet coupled model produces considerably higher leads to comparison to neuro fuzzy, neural network, regression models.

Keywords Water management · Hydrological model · Water quality prediction · Neural network · Fuzzy logic · Daubechies wavelet

1 Introduction

In nature, several hydrological processes occur to sustain the water cycle. Some of them include rainfall-runoff process, groundwater process, river water systems etc. These processes need to be modelled by making use of events that occurred in the past to predict. Adequate supply of pure water is essential for maintaining health and sanitary conditions (Bhardwaj and Parmar 2013a). River water is the most important resource for drinking water, ground water and it also affects the climate of that area. Yamuna river is the largest tributary of Ganga in the northern republic of India. It originates from the Yamunotri glacier ($38^{\circ} 59' N 78^{\circ} 27' E$) on the south western slopes

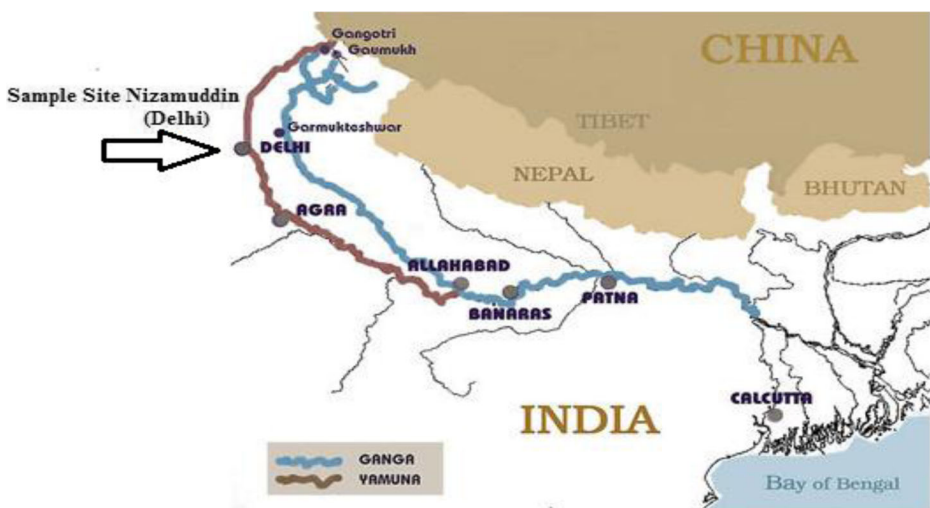
K. S. Parmar · R. Bhardwaj (✉)

Non-Linear Dynamics Research Lab, Department of Mathematics, University School of Basic and Applied Sciences, Guru Gobind Singh Indraprastha University, Dwarka, Delhi 110078, India
e-mail: rashmib22@gmail.com

K. S. Parmar
e-mail: Kulmaths@gmail.com

of Banderpooch peaks within the lower range of mountains in Uttarakhand. Map of the sample location is shown in Map 1. It crosses many states, Uttarakhand, Haryana, Himachal Pradesh, Delhi, Uttar Pradesh by travel a complete length of 1,376 kilometers. It has a mixing of drainage system of 366,233 km² before merging with Ganga at Allahabad, a total of 40.2 % of the entire Ganga basin (CPCB and Water Quality Status of Yamuna River 2006). Nizamuddin (sample site) is approximately 14 km downstream from the Wazirabad barrage at Delhi (capital of India) and 410 km from Yamunotri. Pollution in river water is continuously increasing due to urbanization, industrialization, population growth etc. The water quality at Nizamuddin (Delhi) has the impact of industrialization, sewerage, domestic discharge from Haryana and Delhi. Many rivers are dying due to pollution, which is an alarming signal (Parmar et al. 2009).

There are many different methods of mathematical modelling, which have been developed to forecast long-term precipitation (Rangarajan and Ding 2000; Kahya and Kalayci 2004; Shukla et al. 2008; Doyle and Barros 2011; Dökmen and Aslan 2013). In these methods, a statistical approach, such as time series modelling, is a conventional method that has been widely used (Zhang et al. 2011; Sachindra et al. 2012; Underwood 2012; Diodato et al. 2014; Parmar and Bhardwaj 2014). Statistical modelling has many advantages over mathematical models. But the shortcomings of the statistical approach include handling nonlinear characteristics of data because the statistical models are usually based on the linear correlations of the data can be expressed with a correlation coefficient. To overcome the shortcomings of the statistical methods, many other models that address the nonlinearity of data were developed, including an artificial neural network (ANN), fuzzy, wavelet are capable methods (Nayak et al. 2004; Partal and Kisi 2007; Yeniguna and Ecer 2012). ANN is a pattern matching technique that uses a pair of input and output sets and has a learning algorithm for improving the weights between the nodes (Hsu et al. 1995; Jeong et al. 2012). FIS is another data-driven model that is similar to ANN and would be a proper method for representing linguistic fuzzy if-then rules that are difficult to formulate through a model with crisp parameters (Yeon et al. 2009). A fuzzy logic model was developed to estimate flow for poorly gauged mountainous basins, using Mamdani approach. The entire data were distributed into two parts, namely training and testing. The results of the model were compared with calculated data. This fuzzy model provides more precise and consistent results (Toprak et al.



Map 1 Sample site description

2004; Aksoy et al. 2004; Hung et al. 2009; Toprak et al. 2009; Chen and Chang 2010). Wavelet analysis in the last decade has become an important mathematical tool for research. Wavelet analysis has both good time and frequency multi resolution, which can use effectively to diagnose signal. Wavelet can typically be visualized as a “brief oscillation” like the signal that might be seen from a seismograph or a heart monitor (Grapes 1995; Bhardwaj et al. 2011; Bhardwaj and Parmar 2013b). Wavelet analysis is a better tool than the Fourier transform. Wavelets used for time-scale demonstration of the time series and their interaction to resolve time series with non-stationarities (Adamowski and Chan 2011).

In this study, Neuro, fuzzy and wavelet coupled model is introduced to predict the monthly COD level of Yamuna river considering past history data of 10 years. The study has an importance as the quality of water impacts the 70 % supply of Delhi water from Yamuna river and more than 57 million people depend on river water for their daily usage.

2 Methodology

2.1 Sample Site

The monthly mean value of water quality parameter COD (Chemical Oxygen Demand) monitored at Nizamuddin of river Yamuna in Delhi (India) of the last 10 years have been considered sample site in the present study. Out of 120 samples from 1999 to 2009, 108 have been used to develop the prediction model for COD level time series and 9 data have been used for the test model. First three data points are used as the input values. Sample site is selected according as the utilization of river water and population ratio. It shows in Map 1 below.

2.2 Regression Analysis

It is a technique used for modeling and analyzing several variables. Regression analysis helps in understanding the variation in value of the dependent variable as independent variables is varied, while the other independent variables are held fixed. Regression line of Y (dependent variable) on X (independent variable) (Parmar and Bhardwaj 2014) defined as

$$Y = b_{yx}X + C$$

where C is a constant of integration.

$$b_{yx} = \text{regression coefficient} = r \times \frac{\sigma_y}{\sigma_x}.$$

$$r = \text{Correlation coefficient} = \frac{E(XY) - E(X)E(Y)}{\sqrt{(E(X^2) - E(X)^2)(E(Y^2) - E(Y)^2)}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

σ_Y, σ_X are Standard Deviation of variables Y and X respectively.

2.3 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) is commanding non-linear modelling approach. Which is based upon human brain functioning. It identifies and learns the correlated patterns between input values and objective values. It networks with nodes or neurons, which are interconnected to each other. A general three-layered neural network, consists of several elements nodes. These systems hold an input layer comprising of

hubs speaking to diverse input variables, the hidden layer comprising of numerous concealed hubs and a output layer comprising of output variables. ANN as a forecasting method, is broadly connected in hydrology and water asset studies. Feed-forward neural networks are successfully applied in many different studies (Chang and Chang 2006; Chaturvedi, et al. 2004). Partially recurrent systems begin with a completely repetitive net and include a feed-forward association that detours the repeat, adequately treating the recurrent part as a state memory. Recurrent networks are mostly used in nonlinear time series prediction, pattern analysis and recognition system.

A three layered FFNN based on backpropagation has been selected. There are three input nodes (depending upon input variable), three hidden nodes with tan sigmoid transfer function and one output node with linear transfer function. So for predicting the output, past three inputs have to be estimated. The momentum and learning rate has kept to 0.9 and 0.1 and it trained for 2,000 epochs.

2.4 Wavelet Transforms

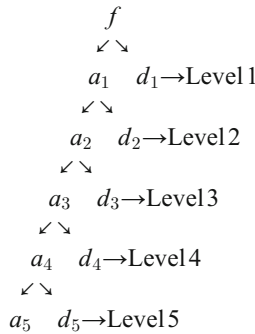
Wavelet could be depicted as a pulse of short time with finite energy that integrates to zero. Wavelets are located both in time and space. Wavelet method crosses the boundaries of the Fourier technique by utilizing capacities that hold a convenient tradeoff between time location and frequency detail. Wavelet transform can be used to locate the information when and where a particular event occurs (Daubechies 1992). Thus, it is very useful in matching the trends in a given time series of data. Wavelet analysis employs a prototype function called mother wavelet $g(t)$. This mother wavelet function has zero average and piercingly drops in an oscillatory way (Can et al. 2005; Labat 2008). Data is described in two versions, superposition of scaled and translated of the pre specified mother wavelet. Continuous wavelet transform (CWT) of a given signal $x(t)$ with respect to $g(t)$ is defined in (1) where a, b are scale and translation factor respectively (Daubechies 1992).

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) g\left(\frac{t-b}{a}\right) dt \quad (1)$$

A continuous wavelet transforms (a, b) coefficient, depicts how accurately signal $x(t)$ and the mother wavelet matches at a particular scale and translation. Thus, the set of all wavelet coefficient CWT (a, b) , associated with a particular signal $x(t)$, is the wavelet representation of the signal with respect to the mother wavelet $g(t)$.

The first step of wavelet transforms corresponds to the mapping signal f to its wavelet coefficients. Using this process two components are received i.e. a smooth version called approximation and a second component that corresponds to the deviations or details of the signal. A decomposition of the signal f into a low frequency part a_1 and a high frequency part d_1 at level 1 is represented by $f = a_1 + d_1$. Similarly the same procedure is performed on a_1 in order to obtain a decomposition in finer scales: $a_1 = a_2 + d_2$. Recursive decomposition for the low frequency parts follows the directions. The resulting low frequency parts a_1, a_2, \dots, a_n are approximations of f and

the high frequency parts d_1, d_2, \dots, d_n contain the details of f . It also classifies as follows (Bhardwaj and Parmar 2013b):



Let $f=(f_1, f_2, f_3, \dots, f_N)$, N is an even integer. $a_m = \frac{f_{2m-1} + f_{2m}}{\sqrt{2}}$, $m = 1, 2, 3, \dots, N/2$

$a^1=(a_1, a_2, \dots, a_{N/2})$. The first trend sub-signal (approximation)

$d_m = \frac{f_{2m-1} - f_{2m}}{\sqrt{2}}$, $m = 1, 2, 3, \dots, N/2$

1-level Transform

$$f \xrightarrow{H_1} (a^1/d^1)$$

and so on.

2.5 Adaptive Neuro-Fuzzy Inference System

The adaptive neuro-fuzzy interface system (ANFIS) methods is used worldwide, as an estimator. It is able to approximate real continuous function on a compact set to a degree of accuracy (Jang 1993). The General outlook of fuzzy inference system is seen as a model, which maps input characteristics to input membership functions. After that, these input membership functions relates with rules and then these rules relates to set of output characteristics. At last, it maps output characteristics to output membership functions, and the output membership function to a single output or a decision associated with the output Jang (1993)). There are three main parts of fuzzy systems, fuzzifier, fuzzy database and defuzzifier. Further fuzzy database consists of two main sections, fuzzy rule base, and inference mechanism (Karmakar and Mujumdar 2006; Seyed et al. 2013; Sahay and Srivastava 2014).

2.5.1 ANFIS Architecture

According to the Takagi and Sugeno type, the FIS has two inputs x & y and one output f as shown in (Fig. 3) (Chaturvedi, et al. 2004).

Rule 1 If x is A_1 & y is B_1 then

$$f_1 = p_1x + q_1y + r_1 \tag{2}$$

Rule 2 If x is A_2 & y is B_2 then

$$f_2 = p_2x + q_2y + r_2 \tag{3}$$

Layer 1 Every node i in this layer is a square node with a node function

$$O_i^1 = \mu A_i(x) \tag{4}$$

Where x is the input to node i , and A_i is the linguistic label related to this node function. So, O_i^1 is the membership function of A_i , which expresses the degree to which the given x satisfies the A_i . Mostly, it is supposed $\mu A_i(x)$ to be bell-shaped and maximum value to 1, minimum value to 0, such as the generalization bell function

$$\mu A_i(x) = \frac{1}{1 + \left[\left(\frac{x-c_i}{a_i} \right)^2 \right]^{b_i}} \tag{5}$$

Or Gaussian function

$$\mu A_i(x) = e \left[- \left(\frac{x-c_i}{a_i} \right)^2 \right] \tag{6}$$

where $\{a_i, b_i, c_i\}$ is the parameter set. The bell shaped function behaving according as the values of parameters set change. It exhibiting various forms of membership functions on linguistic label A_i . Parameters in this layer are referred as premise parameter.

Layer 2 Every node in this layer is a circle node. It multiplies the input signals and sends the product out. For instance

$$O_i^2 = w_i = \mu A_i(x) \times \mu B_i(x), \quad i = 1, 2, 3 \dots \tag{7}$$

Each node output represents the firing strength of a rule.

Layer 3 Every node in this layer is a circle node labeled N . The i^{th} node measures the ratio of the i^{th} rule's firing strength to sum of all rule's firing strengths

$$O_i^3 = w_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2, 3 \dots \tag{8}$$

For convenience, outputs of this layer will be called normalized firing strength.

Layer 4 Every node i in this layer is a square node with a node function

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i = \bar{w}_i(p_i x + q_i y + r_i) \tag{9}$$

where \bar{w}_i is the output of the previous layer (i.e. layer 3) and (p_i, q_i, r_i) is the parameter set.

Layer 5 The single node in this layer is a circle node labeled, which computes the whole output as the sum up of all incoming signals i.e.

$$O_i^5 = \sum \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{10}$$

Hence, we have developed an adaptive network that is working equivalent to a type 3 FIS.

3 Results and Discussion

Time series graph of past data of river Yamuna in Delhi is discussed in Fig. 1. This shows the past pattern of the behavior of river quality water.

3.1 Statistical Analysis

3.1.1 Testing of Stationarity of COD Time Series

Many hydrological time series processes may be stationary or non stationary. The presence of non stationarity in a hydrological time series can result from steady natural and man-made changes in the hydrological environment. The suspicion of stationarity is especially essential in studies, which keep focus on forecast of rare event which happen convulsively in the given time series.

Stationarity is the first primary statistic property tested in the time series study. Autocorrelation function (ACF) serves as a root indicator whether non stationarity or stationarity is present in the time series. For stationarity analysis of COD level of river Yamuna at Nizamuddin station autocorrelation function (ACF), partial autocorrelation function (PACF) test have been applied. Wiece (1990) states that if the ACF decays very slowly, then the time series is non stationary and if a fast decay occurs it implies that the samples are stationary. The autocorrelation function graph which is known as correlogram is taken into consideration to visually detect the existence of stationarity. Figure 2 depicts the fast decay of ACF for 24 lags, which reveals the stationarity of monthly COD level.

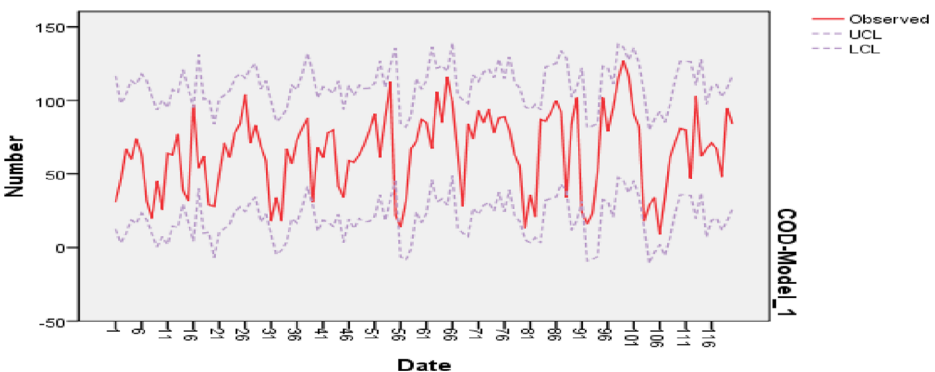


Fig. 1 Time series of monthly level of COD of River Yamuna at Nizamuddin (Delhi)

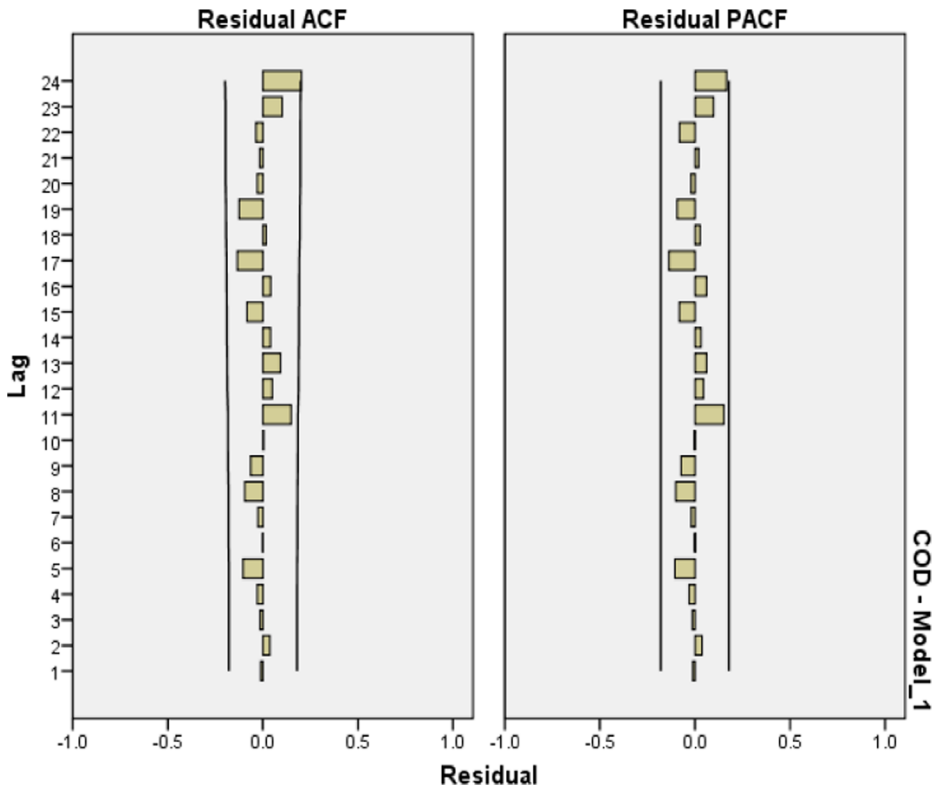


Fig. 2 Correlogram representations of COD level time series for 24 lags

According to the results, null hypotheses is rejected, which has non stationary is not detected and COD level time series is confirmed to be stationary.

3.1.2 Statistical Analysis of COD Level

Statistical properties are presented in Table 1. The minimum value (9 mg/l) and maximum value (127 mg/l) shows that the COD level variation has a long range. It can be seen that the time series show high randomness and scattered distribution about the mean as the standard deviation (27.35 mg/l) and variance (748.29 mg/l) are high. The training data considers the minimum and maximum input of data. This implies the trained neural network is not face any difficulties in extrapolation. The autocorrelation record show the assumption of stationarity is true.

Table 1 Statistical information of monthly COD level of River Yamuna at Nizamuddin (Delhi)

Data	No of Sample	Mean (mg/l)	Min (mg/l)	Max (mg/l)	Standard deviation (mg/l)	Variance (mg/l)	Skewness	Kurtosis	Median
Whole	120	65.06	9	127	27.35	748.29	-0.1989	-0.743	67
Training	108	65	9	127	28.044	786.49	-0.2118	-0.7996	68.5
Testing	9	71.55	47	103	19.31	373.03	0.3784	-0.7515	67

3.2 Neuro -Fuzzy Coupled Model

The neuro fuzzy model is trained with backpropagation for generalization and sugeno inference for specialization. The network is trained for 2,000 epochs with 0.001 tolerance level. The forecasting method uses ANFIS method to predict the next value. Graphical representation of this coupled model of ANN and Fuzzy is shown in Fig. 3.

3.3 Neuro -Fuzzy -Wavelet Coupled Model

Before performing wavelet decomposition, three issues need to be resolved that are selection of mother wavelet, order of mother wavelet and the number of level decomposition. Attributes of the mother wavelet and the characteristic of the signal should be considered carefully, while choosing the appropriate wavelet. Daubechies are the most appropriate for treating random and spike series. For these families of wavelet, the smoothness increases as the order of function increases and suitable wavelet and hence suitable higher order wavelet must be taken into account. In this work wavelet of order 2 to 8 has been considered and it was observed that the decomposed series by wavelet of order 8 are giving the best results.

Figure 4 shows signal approximation, a_1 to a_6 and detailed parts d_1 to d_6 . Approximation curves correspond to low frequency bands and represent the trends of the COD level. Skewness is a measure of the symmetry of the data around the data mean and is zero for an ideal normal curve. Kurtosis refers to the degree of flatness.

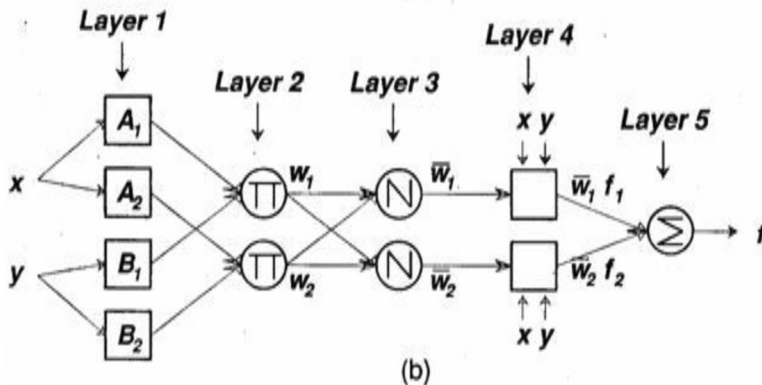
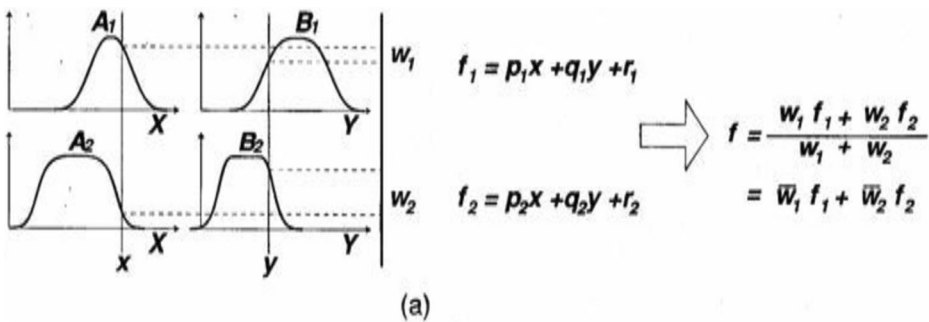


Fig. 3 Graphical representation of Neuro Fuzzy

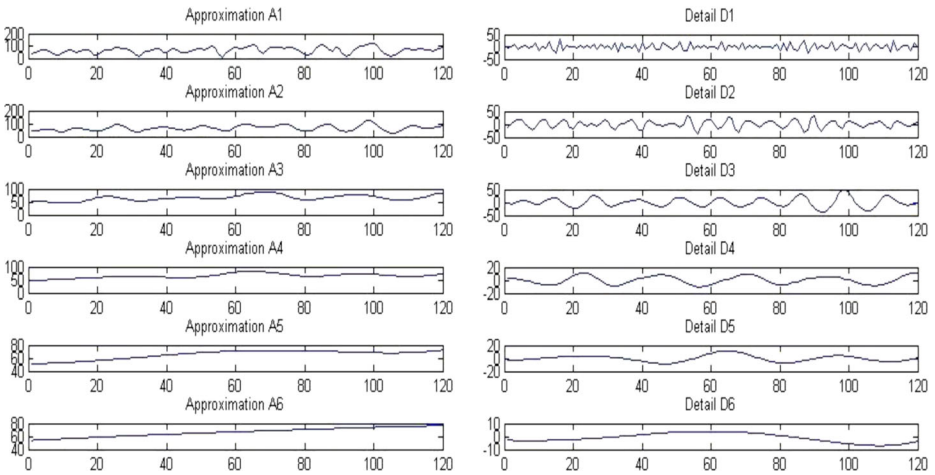


Fig. 4 Six level detailed and approximation coefficient using Daubechies 8 (Db_8) wavelet

Kurtosis of normal distribution curve is three. (Fig. 4) shows that a_1, a_2 and a_3 series are in similar shapes to the original signal than the other approximation level.

Statistical property of a_3 considering skewness shows the best matching and a_4 to a_7 does not provide any Meaningful information shown in Table 2, thus third level describes the regular behavior of COD level series. On examining the detail series d_1 to d_7 from (Fig. 4), it can be seen that range of detail parts is lower in comparison to approximation part. Further range of d_4 to d_7 is lower than d_1 to d_3 , thus detailed part d_4 to d_7 are mainly superficial random noise, as is evident from the Kurtosis characteristic as shown in Table 3.

d_1 to d_3 detects the localized variation and provides good results. On visual and statistical inspection d_1 and d_2 contain useful higher frequency information and shows the irregularities representing random variations. d_3 shows some peaks that allow time localization of peak COD level.

This model presents a different approach, here the idea is to decompose the COD level series using wavelet with daubechies8 wavelet and model it via individual fitting of each level of resolution. That is, each component (a_3, d_3, d_2 and d_1) is modeled separately, and the final result is obtained by adding those four forecasts. (Fig. 5) shows the diagram for this model where input is given to the ANFIS architecture to predict the wavelet coefficient.

Table 2 Skewness of approximation coefficient using Daubechies wavelet

	Db_1	Db_2	Db_3	Db_4	Db_5	Db_6	Db_7	Db_8
a_1	-0.1345	-0.2639	-0.3299	-0.2888	-0.1903	-0.1075	-0.0994	-0.1594
a_2	0.7019	-0.1101	-0.1954	0.4200	0.4305	-0.0641	0.2408	0.4799
a_3	0.2489	0.1206	0.2573	0.2712	0.2935	0.2821	0.3198	0.3178
a_4	-0.2165	0.0268	-0.2191	-0.6398	0.0621	-0.0987	-0.4448	0.1907
a_5	-0.6786	-0.987	-0.5524	-0.5923	-0.1994	-0.4026	-0.2900	-0.9250
a_6	0.1353	-1.1753	-0.4541	0.2866	-0.1876	-0.6772	-0.1025	-0.2282

a_3 , wavelet coefficients chosen according as skewness and kurtosis for futher prediction

Table 3 Kurtosis of detailed coefficient using Daubechies wavelet

	Db ₁	Db ₂	Db ₃	Db ₄	Db ₅	Db ₆	Db ₇	Db ₈
d₁	0.0843	0.1315	0.2469	0.4042	0.9249	0.8475	0.0115	-0.3374
d₂	-0.4001	1.8018	1.2106	-0.0693	-0.4587	-0.1005	-0.5256	-0.1411
d₃	-0.0835	1.6129	0.3871	0.5981	-0.6064	-0.3181	-0.7430	0.1544
d ₄	1.0956	-0.0166	2.8460	-0.4430	0.9574	-0.8021	2.2034	-1.2519
d ₅	-1.8278	0.3101	0.0601	-0.3585	0.4002	-1.0274	-0.8871	-0.2700
d ₆	-1.2199	-0.7587	-0.2034	-1.2052	-1.3946	-0.7003	-0.4499	-1.1854

d1, d2, and d3 are wavelet coefficients choosed according as skewness and kurtosis for futher prediction

The total predicted at any instant (*i*) is given by

$$f(i) = f_1(i) + f_2(i) + f_3(i) + f_4(i) \tag{11}$$

3.4 Experimental Results

The numerical experiments of predicting the future level of COD have been performed. The instrumental record data for training and testing have been collected from 1999 to 2009. 10 years data have been used for learning and the other part of 9 months for testing only. The results of learning and testing have been assessed on the basis of the mean absolute error (MAE) and the average relative error of the actual and the predicted value.

The mean absolute error (Chaturvedi, et al. 2004):

$$MAE = \frac{1}{N} \left(\sum_{i=1}^N |d_i - y_i| \right) \tag{12}$$

where *d* denote the actual measurement, *y* is the predicted value and *N* is the number of days under prediction.

The relative error:

$$\epsilon = \frac{\|d - y\|}{\|d\|} \tag{13}$$

The neuro fuzzy predictor model has been developed and tested. The results of ANFIS prediction is compared with neural network and neuro fuzzy models. The introductory experiment has been performed for the prediction of the level of COD without any previous decomposition. However, the results were not encouraging. The training and testing error is far too large due to lack of generalization ability. This was the main reason to proposed the additional step of decomposing the data into the wavelet coefficients and used neuro fuzzy for predicting the coefficients of each decomposition level.

Neuro fuzzy model is evaluated considering the following combination of input data for COD level forecasting

1. p (t) for prediction of p (t+τ).
2. p (t-τ) and p(t) for prediction of p (t+τ).

3. $p(t-2\tau)$, $p(t-\tau)$ and $p(t)$ for prediction of $p(t+\tau)$.

The performance evaluations measures are mean absolute error (MAE) and relative error (ϵ). The result shows that the best forecast was obtained with past three inputs.

Neuro fuzzy wavelet model is evaluated using wavelet sub series. Past three input, db_8 is used to predict the COD level. The results are confirmed for order 2–8 as shown in Table 4 that past three input, db_8 has shown the best results. The testing data which was not used in learning shows 4.68 % relative error for 9 months ahead prediction. This is very interesting phenomenon of the proposed system of prediction follow the principle of neuro, fuzzy and wavelet coupling. The neuro, fuzzy and wavelet coupled model as described in [Artificial Neural Networks \(ANN\)](#) is implemented to predicting the wavelet coefficients a_3 , d_1 , d_2 and d_3 . Each coefficient is added up to predict the next future instant value. The results of predicted coefficients for training data are shown in the (Figs. 6, 7, 8 and 9).

The trained predicted output is obtained from the decomposed wavelet coefficients by simple summation represented by $S(n)$.

$$S(n) = D_1 + D_2 + D_3 + A_3 \quad (14)$$

The actual trained and predicted trained signal is represented in the (Fig. 10).

The 7 % data which was left for testing is applied to the trained system. The actual testing coefficients obtained from wavelet decomposition are given to the ANFIS system individually. The total predicted and actual testing output is shown in (Fig. 11). Figure 12 shows the validation results of introduced model.

The monthly COD level prediction is also carried out by classical ANN. The ANN was trained Levenberg-Marquardt method. The sigmoid and linear activation functions are used at hidden layer and output neuron respectively. The hidden neuron was changed as (3,3,1), (3,4,1) and (3,6,1). The best results were obtained for (3,4,1) denoting three inputs, four hidden neuron and one output node. The neuro fuzzy wavelet, neuro fuzzy and ANN models are compared to each other in the Table 5. The Neuro fuzzy model shows slightly better results than the ANN, but neuro, fuzzy and wavelet coupled model forecast perfectly than other models.

Table 4 Training and testing results by neuro fuzzy wavelet to predict nine months ahead COD level of river Yamuna using db_2 to db_8 wavelet

Wavelet order	Training results		Testing results	
	MAE (mg/l)	ϵ (%)	MAE (mg/l)	ϵ (%)
db_2	0.2725	0.42	5.8523	8.91
db_3	0.3263	0.5	10.7296	14.99
db_4	0.2005	0.31	2.4061	4.73
db_5	0.2133	0.33	5.0221	7.02
db_6	0.1673	0.26	4.7903	6.69
db_7	0.1782	0.27	4.3801	6.12
db_8	0.1363	0.21	3.3475	4.68

db_8 , wavelet coefficients chosen according as skewness and kurtosis for further prediction

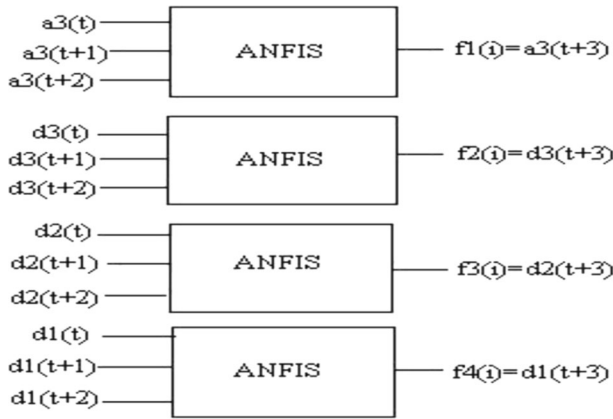


Fig. 5 Forecasting procedure

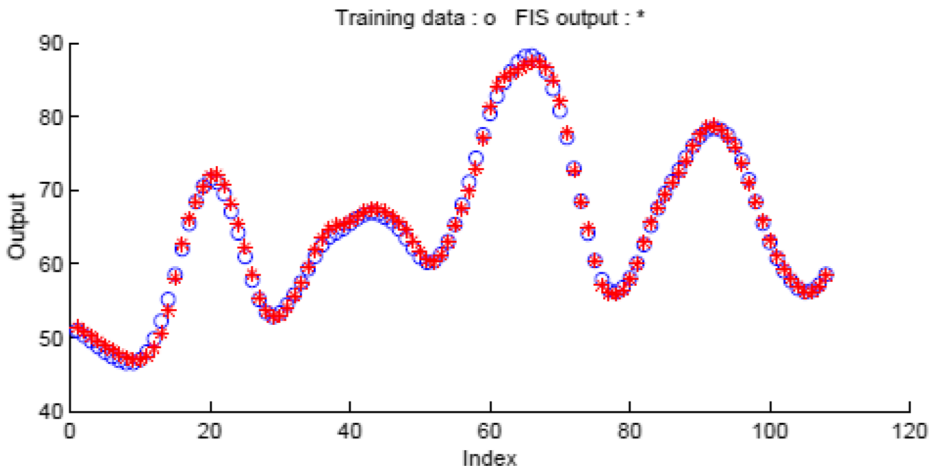


Fig. 6 a_3 actual and trained results

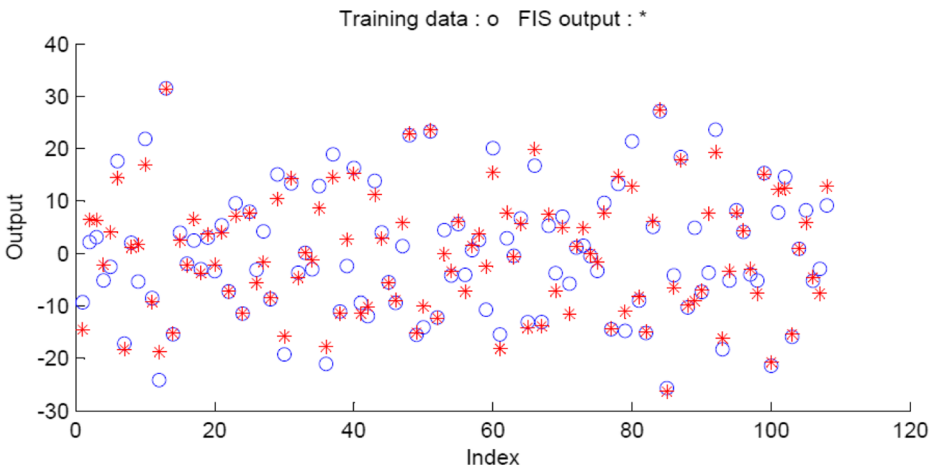


Fig. 7 d_1 actual and trained results

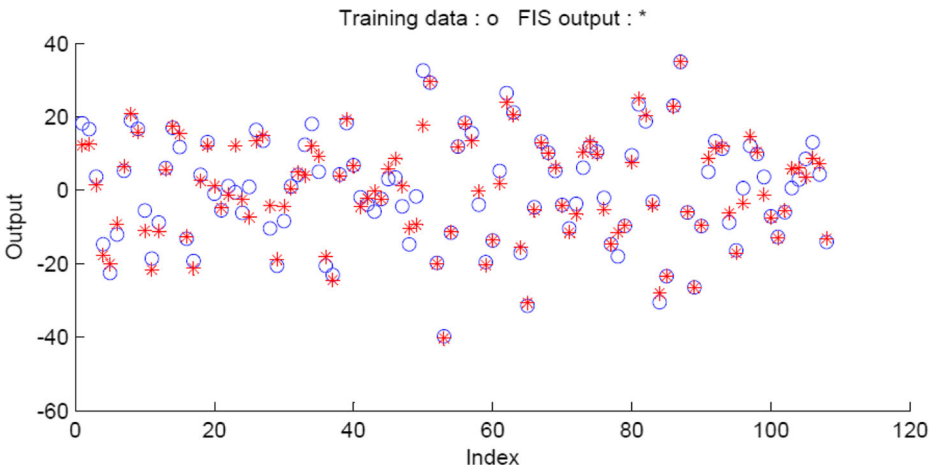


Fig. 8 d_2 actual and trained results

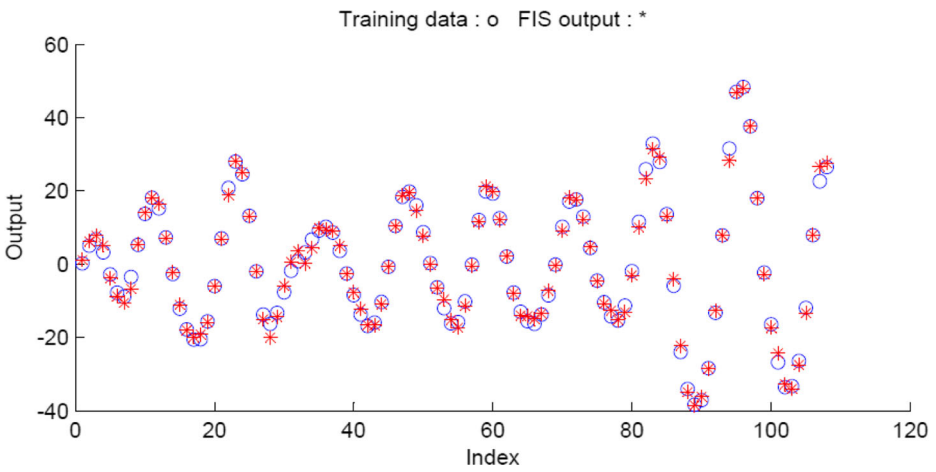


Fig. 9 d_3 actual and trained results

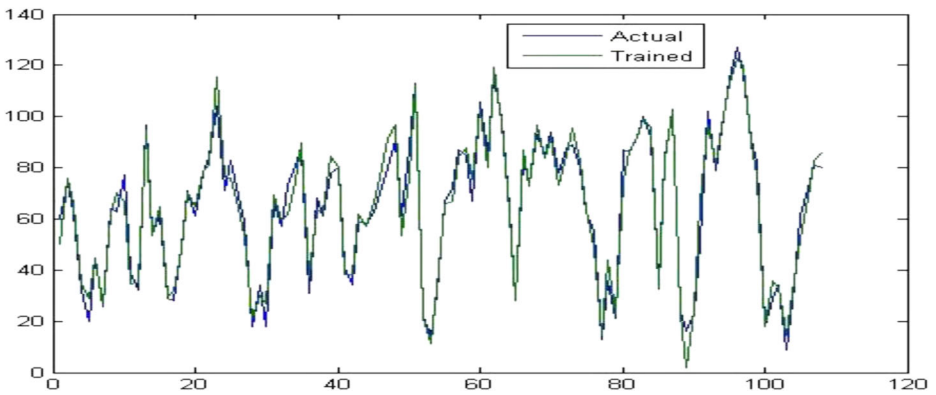


Fig. 10 Actual output and predicted time series of average COD level for training data using neuro fuzzy wavelet

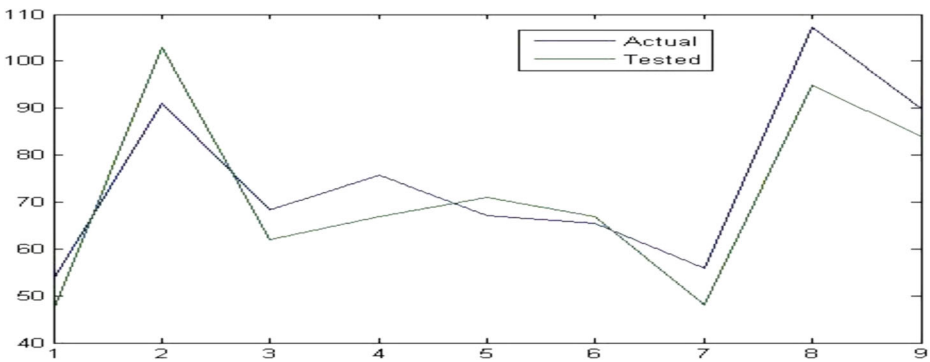


Fig. 11 Actual output and predicted time series of average COD level of testing data using Neuro fuzzy wavelet

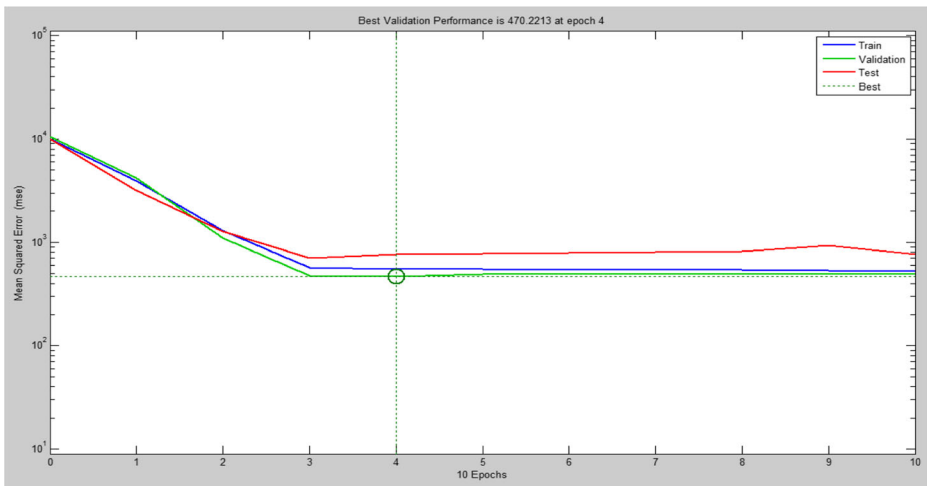


Fig. 12 Validation of model

Table 5 Training and testing results of each model for monthly COD level forecasting

	Training results		Testing results	
	MAE (mm)	ϵ (%)	MAE (mm)	ϵ (%)
Wavelet-neuro-fuzzy	0.1363	0.21	3.3475	4.68
Neuro-fuzzy	0.6725	1.07	5.8442	9.47
Artificial neural network	1.1542	4.49	6.2457	12.28
Multiple regression	3.8465	8.42	9.3594	16.49

The neuro fuzzy wavelet model has given 4.68 % relative error for the testing data and provided a good with the training data (0.21 %). The relative error for the neuro fuzzy, ANN method are 9.47, 12.28 % respectively. Therefore the testing data for these models are not showing good fit in comparison to the neuro fuzzy wavelet model. The mean absolute error for neuro fuzzy wavelet is 3.3475 mg/l.

4 Conclusions

In this work prediction the COD level of river Yamuna for next 9 months which gives satisfactory results on employing wavelet decomposition with neuro fuzzy. The comparison shows that the proposed method gives satisfactory results for prediction. The variability of data was unable to be trained by using only neuro fuzzy. Therefore, the wavelet decomposition and coefficient prediction plays a vital role in the analysis of pollution level in river water.

Acknowledgments Authors are thankful to the University Grant Commission (UGC), Government of India for financial support (F. 41-803/2012 (SR)); Central Pollution Control Board (CPCB), Government of India for providing the research data; Guru Gobind Singh Indraprastha University, Delhi (India) for providing research facilities. The first author is thankful to Sant Baba Bhag Singh Institute of Engineering and Technology for providing study leave to pursue research degree.

References

- Adamowski J, Chan HG (2011) A wavelet neural network conjunction model for groundwater level forecasting. *J Hydrol* 407:28–40
- Aksoy H, Toprak ZF, Aytek A, Ünal NE (2004) Stochastic generation of hourly mean wind speed data. *Renew Energy* 29:2111–2131
- Bhardwaj R, Parmar KS (2013a) Water quality index and fractal dimension analysis of water parameters. *Int J Environ Sci Technol* 10:151–164
- Bhardwaj R, Parmar KS (2013b) Wavelet and statistical analysis of river water quality parameters. *Appl Math Comput* 219:10172–10182
- Bhardwaj R, Parmar KS, Chuha P, Minhas P, Sahota HS (2011) Seasonal variation of physico-chemical parameters and water quality indexing of Harike Lake. *Indian J Environ Protozool* 31:482–486
- Can Z, Aslan Z, Oguz O, Siddiqi AH (2005) Wavelet transform of metrological parameter and gravity waves. *Ann Geophys* 23:659–663
- Chang FJ, Chang YT (2006) Adaptive neuro fuzzy inference system for prediction of water level in reservoir. *Adv Water Res* 29:1–10
- Chaturvedi DK, Singh MM, Kalra PK (2004) Improved generalized neuron model for short term load forecasting. *Int J Soft Comput Fusion Found Methodol Appl* 8:10–18
- Chen HW, Chang NB (2010) Using fuzzy operators to address the complexity in decision making of water resources redistribution in two neighboring river basins. *Adv Water Resour* 33:652–666
- CPCB, Water Quality Status of Yamuna River (1999–2005) (2006) Central pollution control board, ministry of environment & forests, assessment and development of river basin series: ADSORBS/41/2006-07
- Daubechies I (1992) Ten lectures on wavelets. SIAM, Philadelphia
- Diodato N, Guerriero L, Fiorillo F, Esposito L, Revellino P, Grelle G, Guadagno FM (2014) Predicting monthly spring discharges using a simple statistical model. *Water Resour Manag* 28:969–978
- Dökmen F, Aslan Z (2013) Evaluation of the parameters of water quality with wavelet techniques. *Water Resour Manag* 27:4977–4988

- Doyle ME, Barros VR (2011) Attribution of the river flow growth in the Plata basin. *Int J Climatol* 31:2234–2248
- Grapes A (1995) An introduction to wavelets. *IEEE Computational Sci Eng Signal Image Process* 2:50–61
- Hsu K, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall runoff process. *Water Resour Res* 31:2517–2530
- Hung NQ, Babel HS, Weesakul S, Tripathi NK (2009) An artificial neural network model for rainfall forecasting in Bangkok Thailand. *Hydrol Earth Syst Sci* 13:1413–1425
- Jang JSR (1993) ANFIS: adaptive network based fuzzy inference system. *IEEE Trans Syst Manag Cybernet* 23: 665–685
- Jeong C, Shin JY, Kim T, Heo JH (2012) Monthly precipitation forecasting with a neuro-fuzzy model. *Water Resour Manag* 26:4467–4483
- Kahya E, Kalayci S (2004) Trend analysis of streamflow in Turkey. *J Hydrol* 289:128–144
- Karmakar S, Mujumdar PP (2006) Grey fuzzy optimization model for water quality management of a river system. *Adv Water Resour* 29:1088–1105
- Labat D (2008) Wavelet analysis of the annual discharge records of the world's largest rivers. *Adv Water Resour* 31:109–117
- Nayak PC, Sudheer KP, Ranjan DM, Ramasastri KS (2004) A neuro fuzzy computing technique for modeling hydrological time series. *J Hydrol* 291:52–66
- Parmar KS, Bhardwaj R (2014) Water quality management using statistical analysis and time-series prediction model. *Appl Water Sci*. doi:10.1007/s13201-014-0159-9
- Parmar KS, Chugh P, Minhas P, Sahota HS (2009) Alarming pollution levels in rivers of Punjab. *Indian J Env Protozool* 29:953–959
- Partal T, Kisi O (2007) Wavelet and neuro fuzzy conjunction model for precipitation forecasting. *J Hydrol* 342: 199–212
- Rangarajan G, Ding M (2000) Integrated approach to the assessment of long range correlation in time series data. *Phys Rev E* 61:4991–5001
- Sachindra DA, Huang F, Barton A, Perera BJC (2012) Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *Int J Climatol*. doi:10.1002/joc.3493
- Sahay RR, Srivastava A (2014) Predicting monsoon floods in rivers embedding wavelet transform, genetic algorithm and neural network. *Water Resour Manag* 28:301–317
- Seyed AA, Ahmed E, Jaafar O (2013) Improving rainfall forecasting efficiency using modified adaptive neuro-fuzzy inference system (MANFIS). *Water Resour Manag* 27:3507–3523
- Shukla JB, Misra AK, Chandra P (2008) Mathematical modeling and analysis of the depletion of dissolved oxygen in eutrophied water bodies affected by organic pollutant. *Nonlinear Anal: Real World Appl* 9:1851–1865
- Toprak ZF, Sen Z, Savci ME (2004) Comment on longitudinal dispersion coefficients in natural channels. *Water Res* 38:3139–3143
- Toprak ZF, Eris E, Agiralioğlu N, Cigizoglu HK, Yilmaz L, Aksoy H, Coskun G, Andic G, Alganci U (2009) Modeling monthly mean flow in a poorly gauged basin by fuzzy logic. *Clean-Soil Air Water* 37:555–564
- Underwood FM (2012) Describing seasonal variability in the distribution of daily effective temperatures for 1985–2009 compared to 1904–1984 for De Bilt. *Holland Meteorol Appl*. doi:10.1002/met.1297
- Wiece WWS (1990) Time series analysis. Addison Wesley publishing company, New York, p 478
- Yeniguna K, Ecer R (2012) Overlay mapping trend analysis technique and its application in Euphrates Basin. *Turk Meteorol Appl*. doi:10.1002/met.1304
- Yeon IS, Jun KW, Lee HJ (2009) The improvement of total organic carbon forecasting using neural networks discharge model. *Environ Technol* 30:45–51
- Zhang Q, Xu CY, Chen X, Zhang Z (2011) Statistical behaviours of precipitation regimes in China and their links with atmospheric circulation 1960–2005. *Int J Climatol* 31:1665–1678