Scientific Research and Essays Vol. 7(24), pp. 2112-2121, 28 June, 2012
Available online at http://www.academicjournals.org/SRE
DOI: 10.5897/SRE11. 2164
ISSN 1992-2248 ©2012 Academic Journals

Full Length Research Paper

Neuro-fuzzy modelling and forecasting in water resources

Hadi Galavi[1]* and Lee Teang Shui[2]

[1]Department of Civil, Faculty of Engineering, University Putra Malaysia (UPM), Malaysia.
[2]Department of Biological and Agricultural Engineering, Faculty of Engineering, University Putra Malaysia (UPM), Malaysia.

Accepted 28 May, 2012

Recently, with rapid development in computer science and technology, artificial intelligent (AI) models that emulate human thinking ability and brain structure are increasingly used in hydrological forecasting context. Neuro-fuzzy (N-F) models or specifically adaptive neuro-fuzzy inference systems (ANFIS) are rapidly becoming conventional in either academic or industrial applications. Although, there is a common network structure among ANFIS models, there is no one-fits-all ANFIS architecture for every case. Moreover, it is discussed that in many application, theory does not guide in model building process by either suggesting the relevant model input variables or correct functional form and model configuration. This paper is focused on the application of ANFIS in water resources context and reviews the common architecture of ANFIS models been used in this area of research. The aim is to familiarize the new researchers with ANFIS application process in water resources studies.

Key words: Neuro-fuzzy modelling, adaptive neuro-fuzzy inference systems (ANFIS), hybrid learning algorithm, subtractive clustering.

INTRODUCTION

In the past decades, vast literature had drawn attention to the Box-Jenkins method for modelling of hydrologic events. This method in the class of empirical models opposes physical models to avoid considering complicated physical processes among the hydrological variables for modelling a system. Empirical or black box models deal with the undertaken system as a black box which the input to and output from the box are solely taken into account. Acceptable forecasting results have been reported for the application of Box-Jenkins or commonly said autoregressive integrated moving average (ARIMA) models in modelling of water resources systems (Ahmed and Sarma, 2007). However, a major problem with such models is that they are linear and stationary assumption based so that the modelling of a non-linear hydrological phenomenon will probably produce less accurate and reliable forecasts.

Recently, artificial intelligent (AI) models that emulate human thinking ability and brain structure are increasingly used in hydrological forecasting context. Construction of these models depends on the data and there is no need of prior knowledge of the system under consideration of the so called data driven models. However, there has been little discussion on computer paradigms, which are often highly data dependant and their performance relies on model specification and ability to cope with dynamic changes of events. Model specification is based on the existing knowledge of hydrological system and data dependency is referred to the availability of data. One of the most significant findings in dealing with the aforementioned debate is soft computing technique. This technique offers a more flexible, less assumption-dependent, and potentially self adaptive approach for modelling of water-level and other dynamic and nonlinear hydrological processes (See and Openshaw, 1999). The artificial neural networks (ANN) and neuro-fuzzy (N-F) model that is a hybrid method by preserving the learning

*Corresponding author. E-mail: hadigalavi@gmail.com. Tel: +60172652966.

ability of ANNs and fuzzy reasoning are of the two important AI models in which the latter one is considered under soft computing class of models.

ANN in the class of black box models became popular among the hydrologists in the recent past. A large number of processing elements with their inter-connections constitutes the artificial neural networks. ANNs, or shortly neural network (NN), follow the cognition process of the human brain. Neural networks are helpful and efficient to cope with the systems that the characteristics of the processes are difficult to be described by deterministic or stochastic equations. It has been demonstrated that neural networks do not require either a priori detailed understanding of physical characteristics of the catchments or extensive data pre-processing. Moreover, NNs can handle incomplete, noisy and ambiguous data (Singh and Deo, 2007; Thirumalaiah and Deo, 1998). In recent years, there has been an increasing amount of literature on the application of ANN models in water resources engineering and managements and many other aspects of hydrology. Moreover, existing literature on rainfall-runoff modelling (Tokar and Johnson, 1999), flood forecasting (Toth et al., 2000), water quality assessment (Maier et al., 2004), evapotranspiration study (Kumar et al., 2002), groundwater prediction (Daliakopoulos et al., 2005), soil water evaporation (Han and Felker, 1997), and prediction of sediment volume (Kisi, 2007) strongly support the potential of neural networks in water resources modelling. To have a comprehensive review of this technique in water resources applications, readers are referred to the article of Maier and Dandy (2000).

On the other hand, N-F models are rapidly becoming conventional in either academic or industrial application when compared to other nonlinear identification techniques. The forte of N-F systems, which has made it unique, is that the semantic transparency of rule-based fuzzy systems is combined with the learning ability of neural networks (Zounemat-Kermani and Teshnehlab, 2008). Thanks to preserving ANNs ability and fuzzy logic, N-F models can be regarded as a gray box technique (Babuška and Verbruggen, 2003). The method of applying the learning ability of artificial neural networks to the fuzzy models or fuzzy inference systems (FIS) is called N-F modelling (Jang et al., 1997). Moreover, N-F models describe the systems using fuzzy *if–then* rules, represented in an adaptive network structure that is trained by a NN learning algorithm. The gray box models, N-F models, are more comprehendible to users than completely black-box models, such as ANNs. The N-F models or generally called adaptive neuro-fuzzy inference systems (ANFIS) (Jang, 1993), benefits from less training time because of their smaller dimensions and the network initialization with parameters relating to the problem domain. Such results put emphasis on the advantages of combination of fuzzy logic and neural network technology as it provides an accurate initialization of the network in terms of the parameters of the fuzzy reasoning system (Aqil et al., 2007; Yurdusev and Firat, 2009).

The N-F approach and particularly ANFIS, as a multilayer feed forward network with the ability to combine the verbal power of a fuzzy system with numeric power of a neural system is becoming a powerful alternative in modelling numerous processes (Chang and Chang, 2006). More recently, literature has found the application of ANFIS in many fields, such as, regional electricity loads (Ying and Pan, 2008), ophthalmology (Güler and Übeyli, 2005), reservoir operation (Dubrovin et al., 2002), wind speed (Sfetsos, 2000), evaporation (Kisi, 2006), river flow (Firat, 2008) prediction, etc.

Soft computing by combining several different computing paradigms, such as ANN, fuzzy-logic (FL) and genetic algorithm (GA) tries to find a new well-suited model to the system (Zadeh, 1994). Several attempts have been made to compare the ANFIS capabilities as a soft computing method with model driven methods, such as ARIMA models that have been conventionally used in water resources forecasting context. Although, literature has demonstrated the superiority of ANFIS models against ARIMA models (Firat, 2008), its superiority is influenced by the case study conditions and in some cases it does not outperform the ARIMA models (Altunkaynak et al., 2005). Moreover, there is no standard model building to cope with all possible case study conditions. However, extensive literature has suggested that N-F approach can be an effective alternative to the real problems when compared with conventionally used models (Altunkaynak and Sen, 2007).

ANFIS models performance in river flow forecasting context has shown significant improvement in terms of computational speed, forecast errors, efficiency and peak flow estimation against ARIMA and ANN models (Shu and Ouarda, 2007). ANFIS models beside preserving reasonable forecasts accuracy in stream flow prediction have shown the capability to estimate peak flows more effectively than the low flows (Swain and Umamahesh, 2004). This investigation was done using two different membership functions in fuzzification step of model building for peak flows and low flows. In contrary to the aforementioned findings, consistent underestimation of peak flows is also reported, while low and medium flows forecasts have been more accurate (Aqil et al., 2007). This contradiction in prediction result implies that ANFIS models are case-specific and result may be different at various case studies.

The effectiveness of human knowledge interference in ANFIS modelling was investigated by Chang and Chang (2006) in a study on prediction of water level at Shihmen reservoir, Taiwan. Two different N-F models were constructed based on the knowledge of case study. One was developed to forecast the water level solely based on upstream flow pattern of the reservoir, while human knowledge interfered model additionally entered current

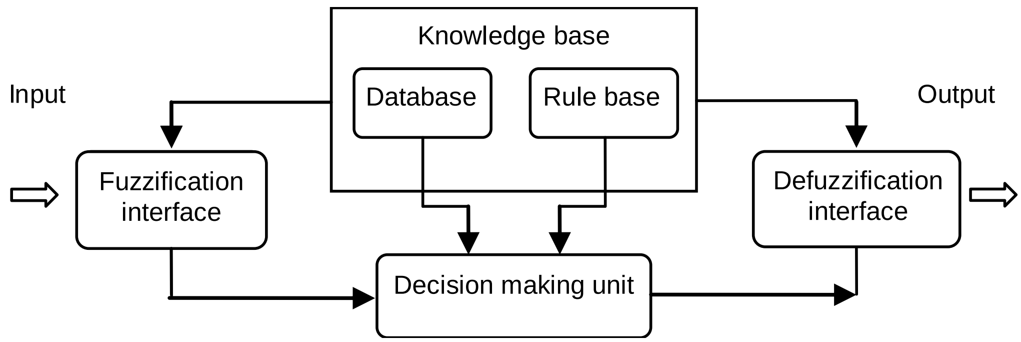

Figure 1. Fuzzy inference system (Jang, 1993).

outflow of the reservoir into the model. Using the same ANFIS architecture for both models, results demonstrated that expert's knowledge based model produces consistently superior precision than the other one. The different steps in ANFIS modeling is described subsequently.

ANFIS

N-F modelling technique benefits from neural networks advantages as well as fuzzy inference system (FIS) capabilities. Characteristics such as learning capabilities, optimization potential and connectionist structure of neural networks in conjunction with FIS abilities to follow human rule thinking way and ease of utilizing experts' knowledge construct the N-F approach. FIS is a framework to establish a simulation of the undertaken system behavior as *if-then* rules using the experts' knowledge or past available data. In addition, it is a process of nonlinear mapping from the input variables to the output variable using fuzzy logic (Jang, 1993). This mapping routine is accomplished using fuzzy *if-then* rules and every rule explains the local behavior of the mapping routine. The premise part of *if-then* rule defines the fuzzy region of the input space and the consequent part specifies the corresponding output of the system. Therefore, the number of fuzzy *if-then* rules defines the FIS efficacy. Practically, a FIS consists of five functional components as shown in Figure 1.

The rule base block contains the fuzzy *if-then* rules, and membership functions (MF) of the fuzzy sets used in the fuzzy rules is defined in the database block. Decision making unit performs the inference operation upon the rules. The fuzzification unit transforms the crisp inputs into fuzzy sets. However, FIS is able to take either fuzzy or crisp values as model inputs, whereas the output is always a fuzzy set; hence, the defuzzification interface turns the fuzzy set into a crisp output. Usually, knowledge base unit is referred to both rule base and database units jointly (Nayak et al., 2005).

Several types of FIS have been proposed based on the specification of consequent part of *if-then* rules and the defuzzification method. Takagi-Sugeno (TS) model and Mamdani model are two of the commonly used fuzzy inference engines (Swain and Umamahesh, 2004; Zounemat-Kermani and Teshnehlab, 2008. TS-FIS using a first-order polynomial of the input variables constitutes the output function. A TS-FIS with such output function is called the *first order* TS-FIS, whereas if the model output is constant then it is named the *zero-order* TS-FIS (Takagi and Sugeno, 1985).

An ANFIS first introduced by Jang (1993), basically, is a multilayer adaptive network that the membership functions optimization of antecedent part of TS-FIS is done using a feed forward neural network. TS-FIS is the inference engine of the ANFIS model. The neuro fuzzy approach overcomes the main drawback of fuzzy logic modelling in terms of the lack of systematic procedure for designing a fuzzy controller, by means of the neural networks learning ability (input-output pairs), self organizing the structure and adaption in an interactive routine (Chang and Chang, 2006). Moreover, both prior knowledge and data processing can be used in construction of the ANFIS method. Expert knowledge of the nature is involved in the form of fuzzy *if-then* rules and their associated parameters (the membership functions and consequent parameters) are fine tuned using data processing (numeric power of NNs). It provides the possibility to interpret the extracted results from N-F models, which is not possible with pure black box models such as NNs. In this perspective, an expert can modify the rules or even add some rules based on his knowledge to expand the validity of the model (Babuška and Verbruggen, 2003).

The relationships between variables of ANFIS models are described using *if–then* rules with ambiguous predicates, such as: if today's water-level is high, then it is highly likely that tomorrow's water-level will be high. This rule is rather a qualitative way to define the relationship between the today's water level and tomorrow's water level. To build an operational model, the meaning of the terms 'high' should be defined more precisely. The data that are considered 'high' constitute a fuzzy set. Membership functions (MF) define a value between [0, 1] to the crisp data of the considered fuzzy sets (Jang et al., 1997). Therefore, a membership value of 0 denotes the non-membership, and value 1 denotes the complete membership in the fuzzy set; a degree between 0 and 1 means partial membership in the fuzzy set. There are two commonly used methods to define the MFs parameters value, which are the back propagation algorithm and hybrid-learning algorithm that provide the network optimization (Firat and Güngör, 2007).

Fuzzy *if-then* rules or fuzzy conditional statements are expressions of the form *IF A THEN B*, where A (high) and B (high) are labels of the fuzzy sets. Fuzzy *if-then* rules are relevant to the imprecise modes of reasoning. In an environment of uncertainty and imprecision, fuzzy *if-then* rules facilitate the human expertise to make decision. In the aforementioned fuzzy conditional statement, both premise part (*A*) and consequent part (*B*) are fuzzy sets. On the other hand, there are fuzzy *if-then* rules proposed by Takagi and Sugeno (1985) that fuzzy sets are involved only in the premise part and the consequent part is described by a non-fuzzy equation. Fuzzy *if-then* rules are the core of the fuzzy inference system. In a research (Mehta and Jain, 2009), merits of different types of fuzzy inference systems were studied. It was demonstrated that the ANFIS model that operates using TS-FIS gave better results with minimum error. On the other hand, Fuzzy Mamdani (FM) model that operates using Mamdani FIS provided the capabilities to change the number of categories and type of the MF of the developed model if there is any drastic change in the present conditions of the case study.

Number and type of membership functions (MF) beside the number of rules are two important determinants in ANFIS architecture, for which compatibility of model to the system under study depends on them. There is no standard method for selecting the type and number of MFs required at network to minimize the output error measure or maximize performance index (Babuška and Verbruggen, 2003). These determinant factors may vary in each different case study. However, selection of appropriate MF at fuzzification part of model building is another important debate in model construction. There exist literature that introduces the triangular membership functions (MF) as a well suited MF to the practical applications (Kisi, 2006). Contrarily, there are some literature that do not only recommend Gaussian MF but have also reported slightly better performance when compared with triangular MF in practice (Aqil et al., 2007; Shu and Ouarda, 2007; Zounemat-Kermani and Teshnehlab, 2008). Overall, fuzzy membership functions can take many forms and the one that gives the minimum mean square error can be the best choice.

Model input selection and data division

The first step to be taken in N-F modelling is the model input selection. To date, in many of the artificial intelligence applications there is no confirmed theory for model building process to suggest the relevant input variables or the correct functional form, thus the input selection remained a controversial issue in modelling and forecasting events (Zounemat-Kermani and Teshnehlab, 2008). Some researchers have used previous knowledge of their case study to select relevant inputs (Shu and Ouarda, 2007) and some by testing the network performance at various input structures select the best model inputs combination (Firat and Güngör, 2007; Nayak et al., 2005). Accurate model input selection can simply influence the model architecture, processing speed, number of *if-then* rules and the parameters to be estimated. Although, ANFIS as well as neural networks has the ability to determine the critical model inputs, presenting a large number of inputs to the network usually increases the network size with increase in the number of rules, so that it decreases the processing speed (Nayak et al., 2005). However, this method can be useful in the absence of prior knowledge about the system. Beside the importance of input selection, the problem is more highlighted in time series application where appropriate number of lags has to be selected for the input.

In N-F modeling, no rigorous criterion has been suggested for input selection. A statistical approach which is based on autocorrelation function (ACF) and partial autocorrelation function (PACF) between the variables has been repeatedly suggested and employed for input selection in N-F modelling (Nayak et al., 2004; Sudheer et al., 2002). In modelling of river flow, the number of antecedent flow values that have a significant influence on the predicted flow value, can be determined by placing a 95% confidence interval on each of ACF and PACF plots. The graphical interpretation of these plots is the same as the traditional statistical methods, whereas the model performance is possible to improve in soft computing methods.

The model training and testing are the two inevitable steps in model building process. The training subset is utilized to optimize the model, and the testing subset is to check the performance and consequently the generalize-tion ability of the built model (Mehta and Jain, 2009). It is worth mentioning that testing subset should be independent from the training set to check the generalization ability of the model beyond the training data set. It is stressed that the available training dataset should cover all the characteristics of the system under study for effective estimation of the event (Mehta and Jain, 2009). One of the methods to tackle with this issue is to use the k-fold cross-validation method upon the data set (Altunkaynak and Sen, 2007). This method randomly splits the dataset into several desired subsets. Thereafter, an iterative method by holding out one of the subsets for the test set, selects the rest of them for the training set. Accordingly, several combination of training/testing subsets will be created so that the best training set would be selected based on the model

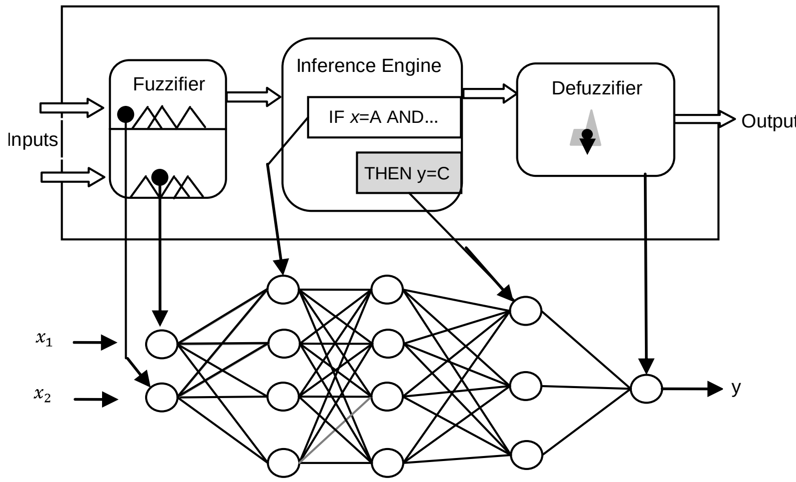

Figure 2. Structure of Hybrid N-F System (Ponnambalam et al., 2003).

performance (Firat and Güngör, 2007).

Training, in detail, is a process to calibrate the value of modifiable parameters in the network such as connection weights and bias via presentation of input-output pairs to the network. It is aforesaid that AI models are data dependent or data driven, thus in these models, no functional relationship is considered between the independent and dependent variables in advance. The functional relationship is settled by the data in training (or calibration) process (Coulibaly et al., 2000). It is argued by Wang et al. (2006) that training is basically a matter of nonlinear optimization that minimizes the error between the constructed network output and the observed output.

Data preparation has also been a controversial issue in the application of artificial intelligence techniques. However, N-F models in comparison with traditional statistical models have the ability to cope with non-stationarity and non-linear patterns in the data series. In addition, one of the preferences of the ANFIS models is that the probability distribution of the data set does not have to be identified in advance (Babuška and Verbruggen, 2003). In this regard, the successfulness of ANFIS models in dealing with non-transformed data, is supported in many attempts in water resources modelling (Firat and Güngör, 2007; Yurdusev and Firat, 2009).

ANFIS architecture and algorithm

The philosophy behind ANFIS is that the adaptive network is used to search for suitable fuzzy decision rules that function well on the tasks in question. An adaptive network is a feed forward structure that by means of selection of modifiable parameters determines the overall output behavior. Thus, ANFIS methodology is designed to create a FIS, using given input-output data set and adjust the membership function parameters by either a back-propagation algorithm solely or a hybrid algorithm of back-propagation and least squares estimator (Swain and Umamahesh, 2004). Generally, in hybrid algorithm, a NN back-propagation learning algorithm determines the membership function parameters and the consequent parameters are determined using least square method (Shu and Ouarda, 2007). A simple visual structure of ANFIS soft computing technique is shown to illustrate the hybrid structure of FIS and an adaptive network (Figure 2). In the illustration of the structure of hybrid N-F system for simplicity, only two input variables and one output are considered representing forecasting of one-day water level using two days before data. It is presumed that based on time series analysis, today's water level has relationship with hydrological events of two days before, representing in water level. The x_1 and x_2 are the input variables that represent the water level of one and two days before the output *y*, that is, the present water level. The connections between the adaptive network and fuzzy inference system are demonstrated as shown in Figure 2. It shows how the fuzzification and defuzzification components as well as fuzzy inference engine work in an adaptive network.

In order to demonstrate the architecture of ANFIS, which is a mapping of FIS to neural network structure, a first order TS fuzzy model is considered (Figure 3, part a) that produces a set of parameters in the consequent part with an equation representing the role of each entry on the output. For simplicity, two input variables (x_1 as the first lag of water level and x_2 as the second one) and one output (*f*) that represent the present water level are considered in model architecture explanation. Assuming each input variable has two membership functions. Figure 3 (part a) shows the TS fuzzy *if-then* rules construction routine that can be expressed as:

Rule 1: If x_1 is A_1 and x_2 is B_1 then $f_1 = p_1x_1 + q_1x_2 + r_1$

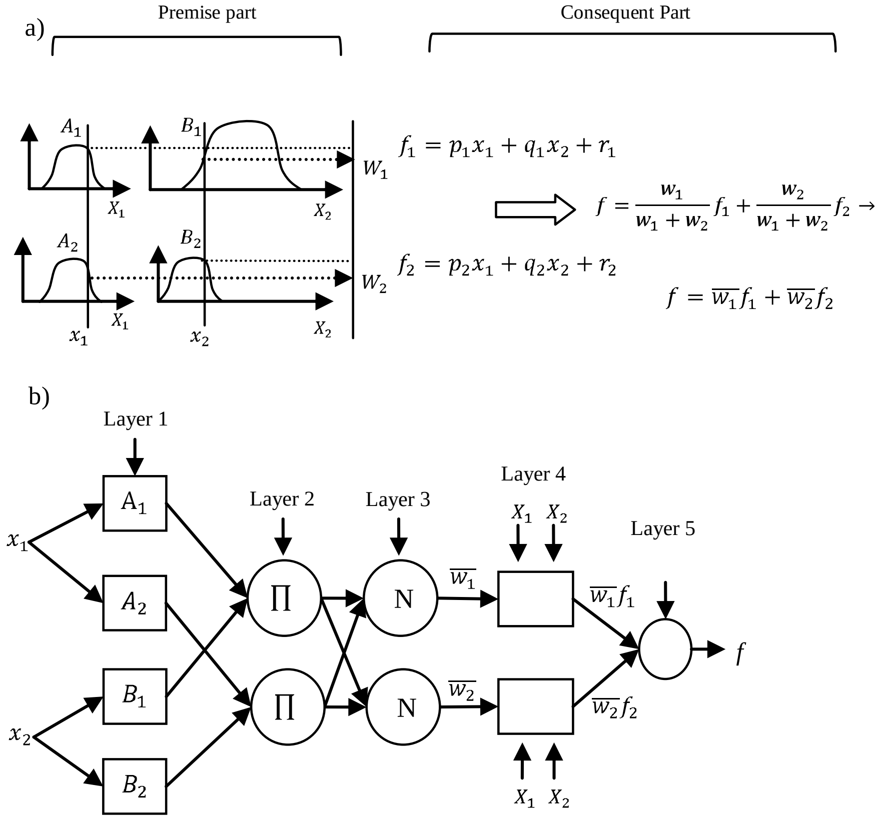

Figure 3. (a) Takagi-Sugeno fuzzy if-then rule and fuzzy reasoning mechanism; (b) equivalent ANFIS architecture (Aqil et al., 2007).

Rule 2: If x_1 is A_2 and x_2 is B_2 then $f_2 = p_2x_1 + q_2x_2 + r_2$

where (p, q, r) are constructing the linear output of fuzzy *if-then* rules. The linguistic labels, A and B, are the fuzzy sets defined by MFs of each input variable that may represent the cluster of low water levels (A_1, B_1) and high water levels (A_2, B_2). Therefore, rule number one says that, if water level at the previous day is low and water level of the second lag is low as well, then the present water level is a function of these variables with their corresponding parameters. The overall output is weighted average of all rules outputs. Since there is no systematic way to decide on the type and shape of the MFs that have the best performance in the defined FIS, the effective method is to use an adaptive neural network model trained by given input-output data to optimize the MFs. Such method is called ANFIS and its architecture is as shown in Figure 3, part b. The ANFIS building with all the relationships between the input and output of each five layers are described as follows (Figure 3, part b): the first layer involves input nodes and every node is an adaptive node. The output node is defined based on the shape of membership function such as,

$$O_{1,i} = \mu_{A_i}(x_1) \qquad for\ i = 1, 2\ or \tag{1}$$

$$O_{1,i} = \mu_{B_{i-2}}(x_2) \qquad for\ i = 3, 4 \tag{2}$$

where the crisp water level data (inputs) are x_1 and x_2 to the node, A_i and B_{i-2} are the fuzzy sets defined by μ_{A_i} and μ_{B_i} membership functions, respectively. Any appropriate membership function regarding the problem can be employed in this node. Among the continuous and piecewise differentiable MFs, bell-shaped and Gaussian MFs due to their smoothness and concise notation are more commonly used (Chang and Chang, 2006). Assuming a Gaussian MF, the output can be computed as (Jang et al., 1997):

$$O_{1,i} = \mu_{A_i}(x_1) = exp\left[\frac{-0.5(x_1 - c_i)^2}{\sigma_i^{\,2}}\right] \tag{3}$$

and $O_{1,i}$ with generalized bell-shaped MF can be computed as (Jang et al., 1997):

$$O_{1,i} = \mu_{A_i}(x_1) = \frac{1}{1+\left|\frac{x_1-\delta_i}{a_i}\right|^{2b_i}} \tag{4}$$

The same function can be applied to define the output $O_{1,i}$ with fuzzy set of B_j, $(j = i - 2)$ at any of these

MFs with only substituting x_1 by x_2 in each equation. Parameters sets of (c_i, σ_i) and (δ_i, a_i, b_i) are of the so called premise (antecedent) parameters of fuzzy *if-then* rules which define the shape of MFs.

In the second layer, the output node is a fixed node which the degree of fulfillment of each input pair to the i_{th} rule is defined using a fuzzy T-norm operator ($\wedge$) that describes the fuzzy intersection (AND) operator. Thus, each output node $O_{2,i}$ represents firing strength of a rule by multiplying incoming signals:

$$O_{2,i} = w_i = \mu_{A_i}(x_1) \wedge \mu_{B_i}(x_2), \qquad i = 1,2. \tag{5}$$

The third layer is to calculate the normalized firing strength of each rule as:

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,2. \tag{6}$$

The contribution of each rule $(\overline{w}_i)$ in the model output is determined in forth layer. Thus, the output of this layer is:

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x_1 + q_i x_2 + r_i), \tag{7}$$

where (p_i, q_i, r_i) are the parameters set of the so called consequent parameters of the FIS.

The single node at the fifth layer computes the overall output of the ANFIS by summing all the incoming signals. The selected operator transforms the fuzzy output into a crisp value. The output of this layer can be defined as:

$$O_{5,i} = \sum \overline{w}_i f_i = \frac{\sum w_i f_i}{\sum w_i} \tag{8}$$

In other words, the network output is the weighted average of all rules outputs (Kazeminezhad et al., 2005). Now, the adaptive network is constructed and the next step is to approximate the parameters of the FIS and train the network based on the supervised learning algorithm to be able to find the precise value of the aforementioned parameters.

Fuzzy clustering

To approximate the FIS parameters in which the number of rules is modifiable, it is required to optimize the premise parameters that define the shape of MFs and consequent parameters that define the final output of ANFIS model. In ANFIS model, as input variables can be clustered into several class values to construct fuzzy *if-then* rules in which the fuzzy rules are built through various MF parameters, the number of parameters to be determined increase as the number of rules increase (Aqil et al., 2007). To keep the ANFIS model as fast and efficient as possible, construction methods based on the fuzzy clustering originated from data analysis and pattern recognition should be taken into consideration. Basically, fuzzy membership shows the degree to which a given input is parallel and similar to some prototypical objects. An appropriate distance measure efficiently helps to acquire the degree of similarity precisely. Therefore, the degree of similarity within each cluster improves and consequently it may reduce the number of clusters and rules.

The subtractive fuzzy clustering method is one of the most efficient and commonly used methods in this context. Moreover, it was noted that ANFIS with subtractive clustering method operated with better results than grid partitioning method in fuzzy inference engine of the model (Mehta and Jain, 2009). Subtractive clustering method is used to establish the fuzzy rule base, because of its capability to determine the number of clusters automatically. Each data point is assumed as potential cluster centre in this method and its likelihood is measured based on the density of its surrounding data points. The potential is calculated for the data instead of the grid points known in data space, thus, according to each cluster potential, the clusters are selected from the system training data. Each data point $((x_1, x_2, \ldots, x_n)$ in an n-dimensional space) is considered as candidate for the cluster centre. The subtractive clustering algorithm approximates the potential of the data point x_i on the basis of its location among the other data points as (Nayak and Sudheer, 2008):

$$p_i = \sum_{j=1}^{n} e^{-\alpha \left\| x_i - x_j \right\|^2} \tag{9}$$

where $\alpha = (\frac{2}{r_a})^2$, and r_a is the cluster radius to define the neighborhood, $\left\| x_i - x_j \right\|$ is the Euclidean distance. Thus, the measure of distance between the data point and its surrounding data points is its potential to be a cluster centre and thus its value is higher as the more neighbor's data points exist. The data point with highest potential among the other data points is considered as a first cluster centre. Denoting x_1^* and p_1^* as the location of the first cluster centre and its potential value respectively, so as to avoid the first cluster centre neighborhoods to be chosen as the second centre, the density measure (potential) of every data point is recalculated by

$$p_i^{new} = p_i - p_1^* e^{-\beta \| x_i - x_1^* \|^2} \tag{10}$$

where $\beta = (\frac{2}{r_b})^2$ and $r_b = \eta \times r_a$ so that r_b is the cluster radius to define the neighborhood by means of η which is a positive constant and is called the squash factor. After this reduction, the data point with highest potential value is selected as the next cluster centre again, x_2^*, p_2^*. The process continues until kth cluster centre is obtained and each data point's potential is revised using:

$$p_i^{new} = p_i - p_k^* e^{-\beta\|x_i - x_k^*\|^2} \quad (11)$$

where x_k^* is the location of kth cluster centre with its potential p_k^*. The process of obtaining the new cluster centre follows the given algorithm (Ren et al., 2006):

if $\frac{p_k^*}{p_1^*} > \bar{\varepsilon}$,
x_k^* is accepted as a cluster centre and continue
else if $\frac{p_k^*}{p_1^*} < \underline{\varepsilon}$
x_k^* is rejected and stop the clustering process.
else let d_{min}= minimal distances between x_k^* and all previously found cluster centers.
if $\frac{d_{min}}{r_a} + \frac{p_k^*}{p_1^*} \geq 1$
x_k^* is accepted as a cluster centre and continue.
else x_k^* is rejected and set p_k^* to 0. Select the data point with the next highest potential as the new x_k^* and reset.
end if
end if

where $\bar{\varepsilon}$ and $\underline{\varepsilon}$ are the accept ratio and reject ratio, respectively. Founder of this method (Chiu, 1994) has suggested the indicative values of the afore-used parameters, ($\bar{\varepsilon}$, $\underline{\varepsilon}$, r_a, η), which are shown in Table 1. Finally, the process is stopped after generating sufficient number of cluster centers and each of them can be reasonably used in a fuzzy rule, premise part, to describe the system behavior. Given an input vector x, the degree to which rule i is fulfilled is defined as:

$$w_i = e^{-\alpha\|x - x_i^*\|^2} \quad (12)$$

where α is a constant defined earlier.

Estimation of parameters (Network training)

An overall view of the fuzzy clustering reveal that two types of parameters need to be estimated for the model construction, nonlinear premise parameters of membership functions (T_1) and linear consequent parameters in TS fuzzy model (T_2). Usually, the estimation of these parameters is done using the ANFIS hybrid learning algorithm (Jang, 1993) and is called network training that is analogous to model optimization in a general view. The theory behind hybrid learning is that the model parameters can be divided into two subsets (T_1, T_2), and elements in the second set are the only linear parameters. Therefore, for given fixed values in T_1, in forward pass least squares method with the objective function of minimizing the sum of squared error can be employed to estimate the consequent parameters in T_2. Thereafter, when T_2 parameters are fixed in backward pass, the gradient descent method is used to update the premise parameters in T_1; Table 2 summarizes the activities in each pass. The objective is to train the built adaptive network in order to acquire convenient unknown functions from the training dataset. For this purpose, it has been shown that the hybrid learning algorithm is much faster and reliable to obtain the proper values of the parameters rather than either strict gradient descent or least square methods (Güler and Übeyli, 2005; Swain and Umamahesh, 2004). The adaptive network has merely one output which is a function of the set of input variables ($\vec{I}$) and the set of parameters (T) (Aqil et al., 2007):

$$output = f(\vec{I}, T) \quad (13)$$

Assuming there exists a K function so that the combined function of $K \circ f$ is linear in some of the elements of T, then the least-square method can identify these elements. Splitting the parameter set T into two subset T_1 and T_2 where operator $\oplus$ represent the direct sum so that:

$$T = T_1 \oplus T_2 \quad (14)$$

So, the composite function $K \circ f$ in a linear function with the elements T_2 and by applying K to the output equation is:

$$K(output) = K \circ f(\vec{I}, T) \quad (15)$$

where $K(output)$ is linear in T_2 elements. Now, given values of T_1element, assigning P training data into the Equation 15, the matrix equation can be obtained as:

$$AZ = B \quad (16)$$

where Z is the unknown vector whose elements are parameters in T_2, the set of linear (consequent) parameters. Let $|T_2| = N$, number of linear parameters, and P number of training data pairs, then the dimension of A, Z, and B are respectively $P \times N$, $N \times 1$ and $P \times 1$. Usually, $P > N$, and therefore, this is an over-determined problem and generally there is no exact

Table 1. Definition and recommended values for parameters in SC (Ren et al., 2006).

Symbol	Definition	Recommend value
$\bar{\varepsilon}$	Accept ratio, which specifies a threshold to certainly reject the data point below it.	0.5
$\underline{\varepsilon}$	Reject ratio which specifies a threshold to accept the data point above it as a definite cluster centre.	0.15
r_a	Hypersphere cluster radius in data space which specifies a neighborhood so that the data points off radius has little effect upon the potential.	[0.25, 0.5]
η	Squash factor $\eta = \frac{r_b}{r_a}$ where r_b defines a neighborhood supposed to have the measurable reductions in potential.	1.25

Table 2. Hybrid learning procedure for ANFIS.

Parameter	Forward pass	Backward pass
Premise	Fixed	Gradient descent
Consequent	Least squares estimator	Fixed
Signals	Node outputs	Error signals

solution to Equation 16. Instead, a least-squares estimate of Z and Z^* can be employed to minimize the squared error of $\|AZ - B\|$. In fact this is a standard problem which is taken into account for the preparation of linear regression, adaptive filtering and signal processing. Z^* can be described by pseudo-inverse of Z:

$$Z^* = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}B \quad (17)$$

where A^{T} is the transpose of A and $= (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$ is the pseudo-inverse of A if $A^{\mathsf{T}}A$ is nonsingular. The recursive least-square estimator (RLS) can also be used to calculate Z^* (Jang, 1993).

Now, a neural network back-propagation learning algorithm (gradient descent) and least square estimator method can be combined to tune up the parameters of the constructed adaptive network. As aforementioned, for the applied hybrid learning into the epochs, each epoch consists of a forward pass and backward pass as summarized in Table 2. More specifically, in the forward pass, output nodes go forward until layer four and the linear (consequent) parameters are determined using the least-squares method. In the backward pass, the error signals propagate backward and the non-linear (premise) parameters are updated using gradient descent (Zounemat-Kermani and Teshnehlab, 2008).

CONCLUSION

A review of ANFIS models application process in water resources studies was organized to explain the fundamental actions taken in this technique in a straight forward and practical manner. An insight to the ANFIS model building process from data preparation and model input selection to fuzzy clustering and network training was provided for water resources engineers and practitioners. Authors aimed to highlight some significant conclusions drawn from previous significant researches on ANFIS models application in water resources studies. It is recommended that care should be taken in FIS construction to choose a model with proper number of MFs and rules in order to save computation time and boost operation speed with an inclusive embrace of case study characteristics.

To this end, although there is optimum values recommended to be used in subtractive clustering method for FIS construction, it is advisable to test several values of the parameters to find the best compatible FIS to the case study conditions. Moreover, different MFs have different effects on fuzzification of the crisp data and the degree to be attributed to the crisp values may differ at different MFs. Type of functional MFs in different applications and different case studies may differ and there should be a comparative study to select a suitable MF for the available conditions. In general, ANFIS models are case-specific and their performance is influenced by case study conditions. To build an operational model, case study conditions should be fully covered by precise model input selection and proper network training.

Randomization of the paired input-output data in selection of best training/testing subsets is generally recommended. Randomization may help in coverage of the all system under study characteristics. Therefore, training (optimization) of the model and its generalization ability testing are under relatively similar conditions as case study.

REFERENCES

Ahmed JA, Sarma AK (2007). Artificial neural network model for synthetic streamflow generation. Water Resourc. Manage., 21(6): 1015-1029.

Altunkaynak A, Özger M, Cakmakci M (2005). Water consumption prediction of Istanbul city by using fuzzy logic approach. Water Resourc. Manage., 19(5): 641-654.

Altunkaynak A, Sen Z (2007). Fuzzy logic model of lake water level fluctuations in Lake Van, Turkey. Theoretical and Appl. Climatol., 90(3): 227-233.

Aqil M, Kita I, Yano A, Nishiyama S (2007). Analysis and prediction of flow from local source in a river basin using a neuro-fuzzy modeling tool. J. Environ. Manage., 85(1): 215-223.

Babuška R, Verbruggen H (2003). Neuro-fuzzy methods for nonlinear system identification. Annu. Rev. Control, 27(1): 73-85.

Chang F, Chang Y (2006). Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Adv. Water Res. 29(1): 1-10.

Chiu S (1994). Fuzzy model identification based on cluster estimation. J. Intelligent Fuzzy Syst., 2(3): 267-278.

Coulibaly P, Anctil F, Bobee B (2000). Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. J. Hydrol., 230(3-4): 244-257.

Daliakopoulos I, Coulibaly P, Tsanis I (2005). Groundwater level forecasting using artificial neural networks. J. Hydrol., 309(1-4): 229-240.

Dubrovin T, Jolma A, Turunen E (2002). Fuzzy model for real-time reservoir operation. J. Water Res. Plan. Manage., 128(1): 66-73.

Firat M (2008). Comparison of artificial intelligence techniques for river flow forecasting. Hydrol. Earth Syst. Sci., 12(1): 123-139.

Firat M, Güngör M (2007). River flow estimation using adaptive neuro fuzzy inference system. Math. Comput. Simul., 75(3-4): 87-96.

Güler I, Übeyli E (2005). Automatic detection of ophthalmic artery stenosis using the adaptive neuro-fuzzy inference system. Eng. Appl. Artif. Intell., 18(4): 413-422.

Han H, Felker P (1997). Estimation of daily soil water evaporation using an artificial neural network. J. Arid Environ., 37(2): 251-260.

Jang J (1993). ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans. Systems, Man, Cybernetics, 23(3): 665-685.

Jang J, Sun C, Mizutani E (1997). Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. IEEE Trans. Automatic Control, 42(10): 1482-1484.

Kazeminezhad M, Etemad-Shahidi A, Mousavi S (2005). Application of fuzzy inference system in the prediction of wave parameters. Ocean Eng., 32(14-15): 1709-1725.

Kisi Ö (2006). Daily pan evaporation modelling using a neuro-fuzzy computing technique. J. Hydrol., 329(3-4): 636-646.

Kisi Ö (2007). Development of streamflow-suspended sediment rating curve using a range dependent neural network. Int. J. Sci. Technol., 2(1): 49-61.

Kumar M, Raghuwanshi N, Singh R, Wallender W, Pruitt W (2002). Estimating evapotranspiration using artificial neural network. J. Irrigation Drainage Eng., 128(4): 224-233.

Maier H, Dandy G (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Model. Software, 15(1): 101-124.

Maier H, Morgan N, Chow C (2004). Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. Environ. Model. Software, 19(5): 485-494.

Mehta R, Jain S (2009). Optimal operation of a multi-purpose reservoir using neuro-fuzzy technique. Water Res. Manage., 23(3): 509-529.

Nayak P, Sudheer K (2008). Fuzzy model identification based on cluster estimation for reservoir inflow forecasting. Hydrol. Process., 22(6): 827-841.

Nayak PC, Sudheer KP, Ramasastri KS (2005). Fuzzy computing based rainfall-runoff model for real time flood forecasting. Hydrol. Process., 19(4): 955-968.

Nayak PC, Sudheer KP, Rangan DM, Ramasastri KS (2004). A neuro-fuzzy computing technique for modeling hydrological time series. J. Hydrol., 291(1-2): 52-66.

Ponnambalam K, Karray F, Mousavi S (2003). Minimizing variance of reservoir systems operations benefits using soft computing tools. Fuzzy Sets Syst., 139(2): 451-461.

Ren Q, Baron L, Balazinski M (2006,). Type-2 Takagi-Sugeno-Kang fuzzy logic modeling using subtractive clustering. Proceeding of Annual meeting of the North American fuzzy information processing society, June. 3-6 Montreal, Canada, pp. 120-125.

See L, Openshaw S (1999). Applying soft computing approaches to river level forecasting. Hydrol. Sci. J. 44(5): 763-778.

Sfetsos A (2000). A comparison of various forecasting techniques applied to mean hourly wind speed time series. Renewable Energy, 21(1): 23-35.

Shu C, Ouarda T (2007). Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. J. Hydrol., 349: 31-43.

Singh P, Deo MC (2007). Suitability of different neural networks in daily flow forecasting. Appl. Soft Comput. J., 7(3): 968-978.

Sudheer K, Gosain A, Ramasastri K (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrol. Process., 16(6): 1325-1330.

Swain P, Umamahesh N (2004). Stream flow forecasting using neuro-fuzzy inference system. Indian J. Power River Valley Dev., 53(5/6): 81-85.

Takagi T, Sugeno M (1985). Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, Cybernetics, 15: 116-132.

Thirumalaiah K, Deo MC (1998). River stage forecasting using artificial neural networks. J. Hydrol. Eng., 3(1): 26-32.

Tokar A, Johnson P (1999). Rainfall-runoff modeling using artificial neural networks. J. Hydrol. Eng., 4(3): 232-239.

Toth E, Brath A, Montanari A (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. J. Hydrol., 239: 132-147.

Wang W, Gelder PHV, Vrijling JK, Ma J (2006). Forecasting daily streamflow using hybrid ANN models. J. Hydrol., 324(1-4): 383-399.

Ying L, Pan M (2008). Using adaptive network based fuzzy inference system to forecast regional electricity loads. Energy Conver. Manage., 49(2): 205-211.

Yurdusev M, Firat M (2009). Adaptive neuro fuzzy inference system approach for municipal water consumption modeling: An application to Izmir, Turkey. J. Hydrol., 365(3-4): 225-234.

Zadeh LA (1994). Soft computing and fuzzy logic. IEEE Software. 11(6): 48-56.

Zounemat-Kermani M, Teshnehlab M (2008). Using adaptive neuro-fuzzy inference system for hydrological time series prediction. Appl. Soft Comput. J. 8(2): 928-936.