Explaining Internal Behavior in a Fuzzy If-Then Rule-Based Flood-Forecasting Model

P. C. Nayak[1]

Abstract: This paper presents a popular fuzzy rule-based model for river flow forecasting for an Indian basin. To set up the fuzzy rules, a cluster estimation method is adopted to determine the number of rules and the membership functions of variables involved in the premises of the rules. The most appropriate set of input variables was determined by trial and error procedure to test the coherence of the different input variables in forecasting flood. It is observed that the last time steps of measured runoff are dominating the forecast. The developed model is used to forecast up to 12 h in advance. The values of three performance evaluation criteria namely, the coefficient of efficiency, the root-mean-square error and the coefficient of correlation, were found to be very good and consistent for flows forecasted 1 h in advance by the model. The performance is decreasing as the forecast horizon is increasing and a reasonable forecast is obtained up to 9 h ahead. A set of fuzzy rules is extracted and used for understanding of the behavior of the developed model. It is observed that the developed model follows the trend of the input membership grade in antecedent part of the fuzzy model.

DOI: 10.1061/(ASCE)HE.1943-5584.0000146

CE Database subject headings: Floods; Forecasting; Fuzzy sets; Hydrologic models.

Author keywords: Flood forecasting; Fuzzy if-then rules; Membership grade; Knowledge extraction; Fuzzy model.

Introduction

Hydrological modeling becomes one of the important tasks for planning, operation, and control of any water resources projects. The prediction of any hydrologic variable such as precipitation, runoff, river stage, etc., has always been a major problem owing to tremendous spatial and temporal variability of basin characteristics and rainfall patterns, as well as a number of other variables associated with modeling the hydrological processes.

There are basically two approaches for hydrological modeling: the theory-driven (conceptual and physically based) approach and the data-driven (empirical and black box) approach often associated by practitioners with statistical modeling. Conceptual models represent the general internal subprocesses and physical mechanisms of the hydrological cycle, without looking at the spatial variability and stochastic properties of the rainfall-runoff process. The parameters are generally assumed as lumped representation of the basin characteristics. Physically based models are based on the understanding of the underlying physical behavior of the system (hydrological cycle). Although the physics based models provide reasonable accuracy, the implementation and calibration of models can typically present various difficulties (Duan et al. 1992), requiring sophisticated mathematical tools, a significant amount of calibration data, and some degree of expertise and experience with the model. It is also not always possible to represent the complex nonlinear hydrological processes by conventional physically based models. In particular the interaction of surface and ground water and hydrodynamic streamflow routing are important components of an adequate simulation model.

Recently there is growing attention using data-driven fuzzy rule-based modeling approach in hydrologic systems. The main advantages of the fuzzy applications are that the fuzzy theory is more logical and scientific in describing the properties of an object. A fuzzy rule-based model of a system is a qualitative logical description of its behavior using variables that are expressed linguistically by means of labels such as "low," "medium," and "high." An advantage of fuzzy logic for modeling cause and effect relationships is that some of the causes that are not considered in idealized types of models, because of generalization or unavailability of known relationships, may be included, whereas some of the causes that are taken into account in the physically based models may be omitted (Hundecha et al. 2001). They have a potential use in situations where numerical models become computationally intractable. An important contribution of fuzzy systems theory is that it provides a systematic procedure for transforming a knowledge base into a nonlinear mapping. Recently, successful applications of fuzzy rule-based modeling for hydrological purposes have been reported.

Bardossy et al. (1990) reported use of fuzzy regression to hydrology and suggested that there are a number of area in hydrology where fuzzy regression has potential applicability. Subsequently, there have been a lot of successful applications of fuzzy rule-based modeling in hydrology. Fuzzy computing technique has been extensively used for reservoir operation studies (Shrestha et al. 1996; Russell and Campbell 1996; Chuntian 1999; Jairaj and Vedula 2000; Panigrahi and Mujumdar 2000; Chang and Chang 2001; Cheng and Chau 2001; Bagis and Karaboga 2004; Cheng et al. 2005; Chang et al. 2005). Hong et al. (2002) employed a fuzzy model to identify and predict ground water level fluctuations caused by storm water infiltration. Some of the other applications of fuzzy technique to hydrology include rainfall modeling, flow routing, evaporation modeling, sediment

[1]Scientist 'C', Deltaic Regional Centre, National Institute of Hydrology, Siddartha Nagar, Kakinada 533 003, AP, India. E-mail: nayakpc@yahoo.co.in

Note. This manuscript was submitted on November 9, 2006; approved on December 12, 2008; published online on May 13, 2009. Discussion period open until June 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 15, No. 1, January 1, 2010. ©ASCE, ISSN 1084-0699/2010/1-20–28/$25.00.

transport prediction, water quality modeling (Bardossy and Disse 1993; Lee, et al., 1995, Schulz and Huwe, 1997; Cheng et al. 2002, 2005; Deka and Chandramouli 2005; Kisi 2006).

In the context of river flow modeling, the fuzzy computing based studies are limited. See and Openshaw (2000) has combined four individual river flow forecast models in a fuzzy frame work. A fuzzy conceptual rainfall-runoff model was proposed by Ozelkan and Duckstein (2001). Chang et al. (2001) suggested a fusion of neural network and fuzzy arithmetic in a counter-propagation fuzzy neural network for real time flood forecasting. Hundecha et al. (2001) developed fuzzy rule-based models to simulate different physical processes involved in the generation of discharge from rainfall and incorporated them within a conceptual model. Xiong et al. (2001) used Takagi-Sugeno (TS) fuzzy model (Takagi and Sugeno 1985) combined with fuzzy-*c*-means clustering algorithm for flood forecasting. Based on the application of their model to 11 watersheds, they reported that TS fuzzy model's performance is comparable with that of neural network models. Xiong and O'Connor (2002) developed a fuzzy autoregressive-threshold updating model for real-time river flow forecasting. Luchetta and Manetti (2003) reported adaptive clustering based fuzzy logic approach to predict extreme and rare situations of hydrometric levels in Padule di Fucecchio basin. Although, the results of the study are found to be better than that provided by an artificial neural network (ANN) model, the predictions were not really accurate. This fact has been acknowledged by them in the conclusion wherein they suggested for enhancement of forecasting.

Sen and Altunkaynak (2003) illustrated fuzzy modeling approach for rainfall runoff modeling and compared the result with linear regression approach for two basins on European and Asian sides of Istanbul. It is concluded that fuzzy system approach yields comparatively less relative error than a regression approach, which is anticipated since fuzzy systems approach is based on linear regression at local level. Nayak et al. (2004) employed a backpropagation algorithm based fuzzy model for river flow forecasting. Nayak et al. (2005b) developed TS fuzzy rule-based rainfall-runoff model for real time flood forecasting. Vernieuwe et al. (2005) developed TS fuzzy for rainfall-discharge dynamics for Zwalm catchment in Belgium and reported that the fuzzy model is not able to capture the low flow dynamics appropriately. It is also observed that the peak flows are underestimated by the fuzzy models. Chang et al. (2005) proposed fuzzy exemplar-based inference system (FEIS) for flood forecasting. The proposed model is employed to predict streamflow 1 h ahead during flood events in the Lan-Yang River, Taiwan. The results show that the FEIS model performs better than an ANN model. They reported that FEIS provides a great learning ability, robustness, and high predictive accuracy for flood forecasting. Jacquin and Shamseldin (2006) developed TS fuzzy model and applied to six catchment of diverse climatic characteristics and reported that fuzzy model is better than simple liner model, linear perturbation model, and nearest neighbor perturbation model. They reported that fuzzy inference systems are suitable alternative to the traditional methods for modeling nonlinear relationship between rainfall and runoff processes. Some of the rainfall-runoff modeling studies using fuzzy technique are reported by Cheng et al. (2002), Chau et al. (2005), and Chau (2006).

Although a plethora of fuzzy applications have been made in hydrologic system modeling, the hydrologists have not attempted to extract knowledge in terms of fuzzy "if-then" rules. One can interpret the rules of the fuzzy model to infer the system dynamics in flood forecasting. This suggests a new and promising research area in hydrologic system modeling. The major objective of the study was to identify the strength of relationship between individual and/or combination input variables and to interpret fuzzy if-then rules in terms of linguistic variable (e.g., low, medium, and high) in flood forecasting. Subtractive clustering technique is used for model identification in TS fuzzy model for generation of fuzzy if-then rules. The methodology is illustrated through a case study by developing a TS fuzzy model of the Baitarani River basin in India.

Methodology

TS Fuzzy Models

There are various types of fuzzy rule-based models in the literature (e.g., Mamdani and Assilian 1975; Tsukamoto 1979; Takagi and Sugeno 1985), and each of them is characterized by their consequent function only. The TS fuzzy model is resulted from an effort to develop a systematic approach to generate fuzzy rules from a given input-output data set (Takagi and Sugeno 1985), in which the rule consequents are typically taken to be either crisp numbers or linear functions of the inputs. Considering, first order TS model with one input variable X and one output variable Y then the input variable is partitioned into n fuzzy sets $A_1, \ldots, A_n$, the antecedent fuzzy sets. This results in n fuzzy rules of the form

$$R_I\text{:if } X \text{ is } A_i \text{ then } Y = a_i X + b_i \tag{1}$$

where a_i and b_i=parameters of the consequent part of rule R_i. Given a value of x of the input variable X, the resulting value y of the output variable Y is computed as

$$y = \frac{\sum_{i=1}^{n} A_i(x)(a_i x + b_i)}{\sum_{i=1}^{n} A_i(x)} \tag{2}$$

In order to apply a TS model to a p dimensional input space, in particular if $A_i = A_{1,i_1}X, \ldots, XA_{p,i_p}$, with $i_1 \in \{1, \ldots, n_1\}, \ldots, i_p \in \{1, \ldots, n_p\}$ and $n_1, n_2, \ldots, n_p$, the number of fuzzy sets each input variable is partitioned into, the rule reads

$$R_I\text{:if}(X_1, \ldots, X_p) \text{ is } A_I \text{ then } Y = a_{1,i}X_1 + a_{2,i}X_2 + \ldots + a_{pi}X_p + b_i \tag{3}$$

For a p dimensional input vector $x = (x_1, \ldots, x_p)$, $A_i(x)$ is then usually realized as

$$A_i(x) = A_{1,i_1}(x_1) \,.\, A_{2,i_2}(x_2) \ldots \ldots .\, A_{p,i_p}(x_p), \tag{4}$$

when the above type of rules are used, $A_i(x)$ is also the degree of fulfillment $w_i(x)$ of rule i. The resulting output value y is then computed as

$$y = \frac{\sum_{i=1}^{n} w_i(x)(a_{1,i}x_1 + \ldots, a_{p,i}x_p + b_i)}{\sum_{i=1}^{n} w_i(x)} \tag{5}$$

In this way, a weighted average of the individual rule outputs is computed and a nonlinear function can be approximated.

For a given data set, different TS models can be constructed using different identification methods. Grid partitioning and fuzzy clustering are two methods generally used to identify the antecedent membership functions. The consequent parameters are computed using a linear least-squares method. In the present investigation subtractive clustering technique is used for model identification in a fuzzy rule-based model. The procedure for the proposed method is discussed below.

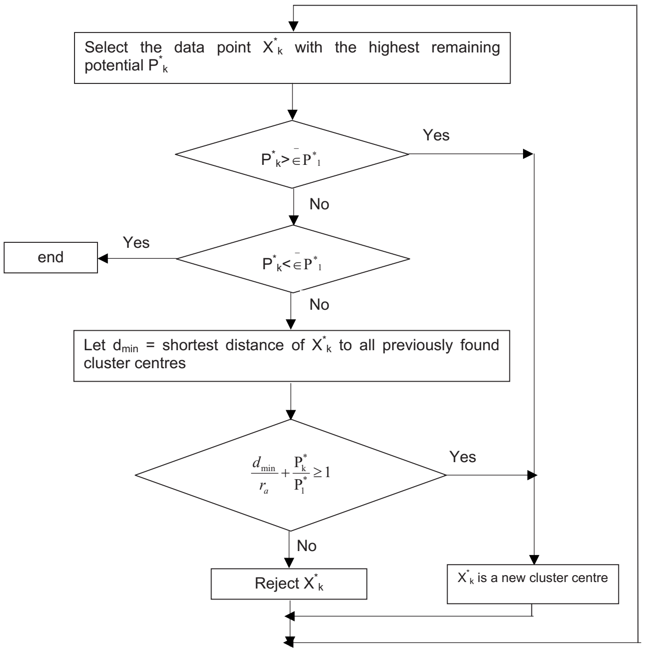

Fig. 1. Cluster estimation algorithm

Cluster Estimation

Subtractive clustering method (Chiu 1994) is an extension of the mountain clustering method (Yager and Filev 1994), where the potential is calculated for the data rather than the grid points defined on the data space. Each historical data point is considered a potential cluster center. A measure of the potential of a data point x_i as a cluster center is defined as

$$P_i = \sum_{j=1}^{n} e^{-\alpha \|x_i - x_j\|^2} \tag{6}$$

where

$$\alpha = \frac{4}{r_a^2} \tag{7}$$

and r_a=positive constant. Thus, the measure of potential for a data point is a function of its distance to all other data points. The data point with highest potential is selected as the first cluster center. Let x_1^* be the location of the first cluster center and p_1^* be its potential value. Before proceeding to find the second cluster center, the amount of potential from the next data point is subtracted, based on its distance from the first cluster center. Then the potential of each data point x_i may be revised by formula

$$p_i = p_i - p_1^* e^{-\beta \|x_i - x_1^*\|^2} \tag{8}$$

where

$$\beta = \frac{4}{r_b^2} \tag{9}$$

and r_b=positive constant. The constant r_b is effectively the radius that defines the neighborhood that will have measurable reductions in the potential of the other data points. To avoid obtaining closely spaced cluster centers, r_b may be set to be somewhat greater than r_a; a good choice is $r_b = 1.5 r_a$

When the potential of all the data points have been revised according to Eq. (8), it then selects the data point with the highest remaining potential as the second cluster center. The process is repeated until a given threshold for the potential is obtained such that $P_k^*/P_1^* < \bar{\varepsilon}$. The choice of $\bar{\varepsilon}$ is an important factor affecting the results; if $\bar{\varepsilon}$ is too large too few data points will be accepted as cluster centers and if $\bar{\varepsilon}$ is too small, too many cluster centers will be generated. Fig. 1 shows the criteria for accepting or rejecting cluster centers. In the present investigation, $\bar{\varepsilon}=0.5$ is considered for analysis as suggested by Chiu (1994).

Model Development

A TS fuzzy model is developed to forecast, the river flow one hour in advance at the Anadapur gauging station of the Baitarani basin in India, covering a catchment area of 8,750 sq. km (Fig. 2). River flow data are available for the monsoon season of 1994 to 1995 at hourly intervals. The model uses observed values of river flow in the past to obtain a forecast of the flow in future. In the TS development the selection of appropriate input variables is important since it provides the basic information about the system being modeled. In addition to the antecedent river flow values, exogenous input variables such as precipitation, evaporation etc., might have an influence on hourly river flows. However, earlier studies reported that TS fuzzy model do not consider the rainfall variability in the data while computing the runoff (Nayak et al. 2005a,b; Nayak and Sudheer 2008) when the antecedent flow values are incorporated. Campolo et al. (1999) suggested that the capacity of a basin to respond to a perturbation is more accurate when recent discharge values are used as input to the model. Further, it significantly reduces the complexity of the model. Therefore, in the current study, the problem has been addressed in a univariate time series approach without exogenous input variables. The parameters that need to be selected in the input vector, hence, are the number of flow values at different time lags that can best represent the time series by a TS model. The current study analyzed individual and different combinations of antecedent flow values and the appropriate input vector has been selected based on the analysis of residuals. The analysis started with one antecedent flow in the input vector and a TS fuzzy model is constructed. The goodness of fit statistics are computed during training and validation for each model, and the best model is selected based on the analysis of residuals. Thus the functional form of the TS fuzzy model is

$$Q(t) = f[Q(t-1), Q(t-2), \ldots, Q(t-n)] \tag{10}$$

where $Q(t-1), Q(t-2), \ldots, Q(t-n)$=dependent variables which represents the state of a certain phenomenon in time.

Different statistical indexes that are employed to estimate the model performance include coefficient of correlation (CORR), efficiency (EFF), root-mean-square error (RMSE). The definitions of different statistical indexes are presented below

$$\text{CORR} = \left[\frac{\sum_{t=1}^{n} (y_t^o - \overline{y^o})(y_t^c - \overline{y^c})}{\sqrt{\sum_{t=1}^{n} (y_t^o - \overline{y^o})^2} \sqrt{\sum_{t=1}^{n} (y_t^c - \overline{y^c})^2}} \right] \tag{11}$$

$$\text{EFF} = 1 - \frac{\sum_{t=1}^{n} (y_t^o - y_t^c)^2}{\sum_{t=1}^{n} (y_t^o - \overline{y^o})^2} \tag{12}$$

Fig. 2. Baitarani River basin map

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{t=1}^{n}(y_t^o - y_t^c)^2}{n}} \tag{13}$$

where y_i^o and y_i^c, respectively, are the observed and computed flow values at time t, and $\overline{y^o}$ and $\overline{y^c}$ are the mean of the observed and computed flow values corresponding to n patterns

Results and Discussions

In subtractive clustering, the procedure automatically determines the number of cluster, which is related to the value of r_a. A larger value of r_a leads to large radius for each cluster and all data can be clustered in a few cases. In contrast, a small value of r_a means that more compact clusters but the number of classes will be larger. The cluster radius may be fixed after several trials. Exponential membership function is employed in the model. It may be noted that the radius specifies the range of influence of the cluster center of each input and output dimension. Assuming that the cluster radius falls within the hyper box of unit dimension, smaller cluster radius will usually yield more clusters in the data, and hence more number of rules. Simultaneously it increases the model complexity and hence violates the principle of parsimony.

To verify the adequacy of information presented to the model and the existence of the prediction limits, different statistical indexes are computed and summarized in the Table 1. In the current investigation, six models are developed using single input variables. The correlation statistics evaluate the linear correlation between the observed and the computed runoff, which is consistent during calibration as well as the validation period and shows more than 0.87 for all models. The fuzzy model performance is very good in terms of the efficiency statistic. It is also observed that as the lag time (in input vector) is increasing (from Models 1 to 6) the efficiency index is decreasing. This observation indicates that last time steps are dominating the runoff predictions. The RMSE statistic is a measure of residual variance and is indicative of the model's ability to predict the peak flow. The RMSE index is varying from 40 to 264 m^3/s compared to peak flow of 4,300 m^3/s during calibration and validation.

In order to check the effect of different individual input variables in forecasting the peak flow, two observed and computed hydrographs are presented in Fig. 3 for Models 1 to 6. From the figures it is evident that there is a shift in computed flow compared to observed flow, this is clearly visible as there is phase lag in the rising and falling limb of observed hydrograph. The phase lag is reducing as the antecedent hour in the input vector is reducing. The falling limb of the hydrograph is minimally

Table 1. Statistical Indexes for Different Models with Single Input Vector

	Input vectors	Correlation		Efficiency (%)		RMSE	
		Calibration	Validation	Calibration	Validation	Calibration	Validation
Model 1	$Q(t-6)$	0.873	0.944	76.185	88.750	264.199	204.380
Model 2	$Q(t-5)$	0.907	0.959	82.240	91.742	228.154	175.101
Model 3	$Q(t-4)$	0.937	0.972	87.829	94.420	188.876	143.943
Model 4	$Q(t-3)$	0.963	0.983	92.648	96.688	146.790	110.900
Model 5	$Q(t-2)$	0.982	0.992	96.497	98.434	101.327	76.251
Model 6	$Q(t-1)$	0.995	0.998	99.020	99.551	53.584	40.815

influenced by any models compared with the rising limb, but $Q(t-1)$, $Q(t-2)$, and $Q(t-3)$ have minimum phase lag compared to other models. Similarly a significant observation is obtained from the hydrographs on peak flow estimation and time to peak. The estimated peak flow is reducing as the lag information is increasing, similarly there is one hour difference in time to peak flow estimation for flow $Q(t-1)$ and it is linearly increasing as difference in time to peak flow is 6 h for flow $Q(t-6)$. The explained variance is estimated for Model 1 to Model 6 during calibration and validation period. It is found that explained variance is nearly 99% for Model 1 during calibration and validation period implying model performance is consistent. As the time lag is increasing model performance is deteriorating, it is found from the Model 6 that explained variance is 74.48 and 78.68% during calibration and validation period, respectively.

Foregoing discussion clearly explains the strength of individual input variables in predicting the peak flow hydrograph at one hour in advance. In order to improve the forecast, combinations of different input variables are proposed, consequently Models 7 to 16 are developed and performance indexes are summarized in the Table 2. It is observed that performance in terms of efficiency statistics is almost static beyond the Model 7. This observation is similar in case of correlation statistics also. It is evident from the table that increasing more input vectors to the model does not increase the model performance. As a result, increasing the input vector will increase the number of fuzzy if-then rules and parameters; simultaneously it increases the model complexity. There may be a number of structure/parameter combinations, all resulting in similar performance. The problem can be addressed by considering minimum structural complexity (in terms of number of rules and parameters) and maximum generalization property for the fitted model. From the table it is also observed that RMSE index is very less compared to earlier developed model. After analyzing the residuals, Model 8, which consists of three antecedent flows as input, showed the highest efficiency and the minimum RMSE, and it is selected as the best-fit model for forecasting the river flow for Baitarani basin. The observed and computed hydrograph for one hour lead forecast from Model 8 during calibration and validation period are presented in Fig. 4. It may be noted that the optimal cluster radius is 0.7 for Model 8 having three numbers of fuzzy if-then rules comprising of three clusters.

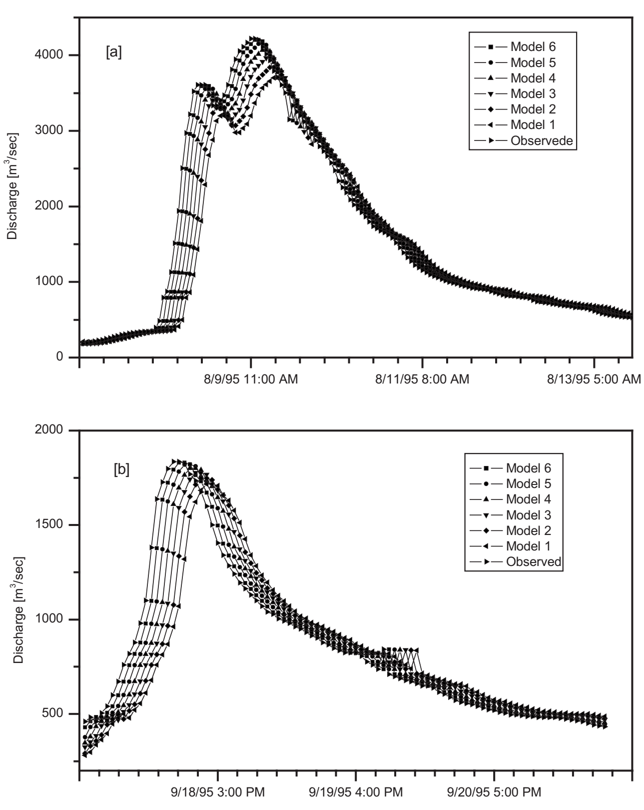

Fig. 3. Observed and computed plots for Models 1 to 6 for two flood events

It may be noted that extreme flood values may be observed beyond the range of the time series in future. To check the prediction limit, the developed model is used for different lead time forecast, keeping other model parameter constant. In order to test the robustness of the model developed, it is important to test the model for higher lead-time forecasting. Consequently, Model 8 has been used to forecast up to 12 h in advance. It may be noted that the input to all these models are the same as that of Model 8. The performance of these models in terms of correlation, RMSE and efficiency statistics against prediction time horizon is presented in Fig. 5. From the Fig. 5 it is observed that as the lead time increasing efficiency is decreasing and RMSE is increasing. The performance of the models in terms of correlation index between the forecasted and observed values of flows is similar to that in terms of the efficiency statistic as presented in the Fig. 5. It is to be noted that even at a 12-h lead time, the forecasted values have good correlation with the observed values. It is to be noted that the efficiency statistic is consistent during training and validation for the fuzzy rule-based model, which indicates a good generalization property for the model.

Understanding the Fuzzy Rules

To analyze the rules developed from the TS fuzzy model, the first task is to find that what has been learnt from the training or what knowledge has gained. The TS fuzzy model usually employs

Table 2. Statistical Indexes for Different Models with Multiple Input Vectors

		Correlation		Efficiency (%)		RMSE	
Mod *M*	Input vectors	Calibration	Validation	Calibration	Validation	Calibration	Validation
Model 7	$Q(t-1)$ and $Q(t-2)$	0.998	0.999	99.632	99.806	32.845	26.808
Model 8	$Q(t-1)$, $Q(t-2)$, and $Q(t-3)$	0.998	0.999	99.680	99.826	30.627	25.428
Model 9	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$, and $Q(t-4)$	0.999	0.999	99.691	99.820	30.087	25.874
Model 10	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$, $Q(t-4)$, and $Q(t-5)$	0.999	0.999	99.691	99.818	30.101	25.965
Model 11	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$, $Q(t-4)$, $Q(t-5)$, and $Q(t-6)$	0.999	0.999	99.706	99.815	29.348	26.179
Model 12	$Q(t-2)$, $Q(t-3)$, $Q(t-4)$, $Q(t-5)$, and $Q(t-6)$	0.995	0.997	98.908	99.389	56.575	47.648
Model 13	$Q(t-1)$, $Q(t-3)$, $Q(t-4)$, $Q(t-5)$, and $Q(t-6)$	0.998	0.999	99.680	99.808	30.608	26.702
Model 14	$Q(t-1)$, $Q(t-2)$, $Q(t-4)$, $Q(t-5)$, and $Q(t-6)$	0.998	0.999	99.674	99.825	30.903	25.491
Model 15	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$, $Q(t-5)$, and $Q(t-6)$	0.999	0.999	99.699	99.814	29.693	26.291
Model 16	$Q(t-1)$, $Q(t-2)$, $Q(t-3)$, $Q(t-4)$, and $Q(t-6)$	0.999	0.999	99.692	99.819	30.061	25.917

if-then rules to reason about data. Fuzzy if-then rules are also known as fuzzy implications or fuzzy conditional statements. They are often employed to capture the imprecise modes of reasoning in an environment of uncertainty. A fuzzy if-then rule associates a condition described using states of variables to a conclusion. It consists of two components: an if part, referred to as the *antecedent*, and a then part, referred to as the *consequent*. All if-then rules follow the following structure:

$$\text{if}\langle\text{antecedent}\rangle\text{then}\langle\text{consequent}\rangle \quad (14)$$

The antecedent and consequent of the fuzzy rule is constructed by a set of fuzzy variables, as well as their values. For example, there are two fuzzy variables, *A* with the value of "high flow," and *B* with the value of "low flow," in the antecedent and one fuzzy variable, *C* with the value of "medium flow," in the consequent. A typical fuzzy rule looks like this

$$\text{if } A \text{ is "high" and } B \text{ is "low" then } C \text{ is "medium"} \quad (15)$$

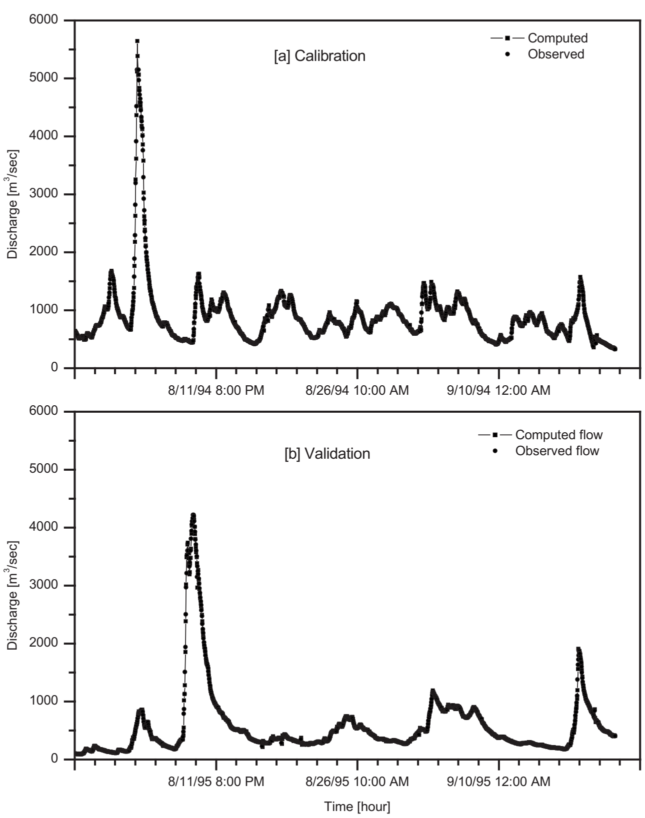

Fig. 4. Observed and computed hydrographs for Model 8 for 1-h lead forecast

Development of Membership Function

Based on the fuzzy modeling approach originally developed by Takagi and Sugeno (1985), the global operation of a nonlinear process is divided into several local operating regions. Within each local region R_i a reduced order linear model is used to represent the processes behavior. Fuzzy sets are used to define the process operating conditions such that the dynamic model of a nonlinear process can be described. Fuzzy regions are represented with different membership grade used in the fuzzy if-then rules, which is the crux of fuzzy modeling approach. Therefore, classification of different local regions in an input/output data set is very important in fuzzy modeling approach and fuzzy clustering technique is widely used for such purposes. In the clustering approach, classification is carried out using different distance measures. The degree of similarity can be calculated by using a suitable distance measure. Based on the similarity, data vectors are clustered such that the data within a cluster are as similar as possible, and data from different clusters are as dissimilar as possible. As described earlier in the current investigation subtractive clustering is used for domain partitioning (low, medium, high, etc.) and least-square estimation (LSE) algorithm is used for consequent parameter estimation. The coordinates of the cluster centers identified by Model 8 in the current study are presented in Table 3. Note that only single dimension for one fuzzy if-then rule corresponding to each variable is presented in the Table 3.

In this experiment, one fuzzy rule obtained after training the Model 8, has been chosen to show the knowledge within the data. The data in each dimension corresponds to one fuzzy variable, denoted by "X_i,"in the antecedent part, which has either three values: high, medium, and low. For the consequent part, the variable has three possible values: low, medium, and high, which indicate the class of the rule represents. The input coordinate for one rule derived from Model 8 is given by

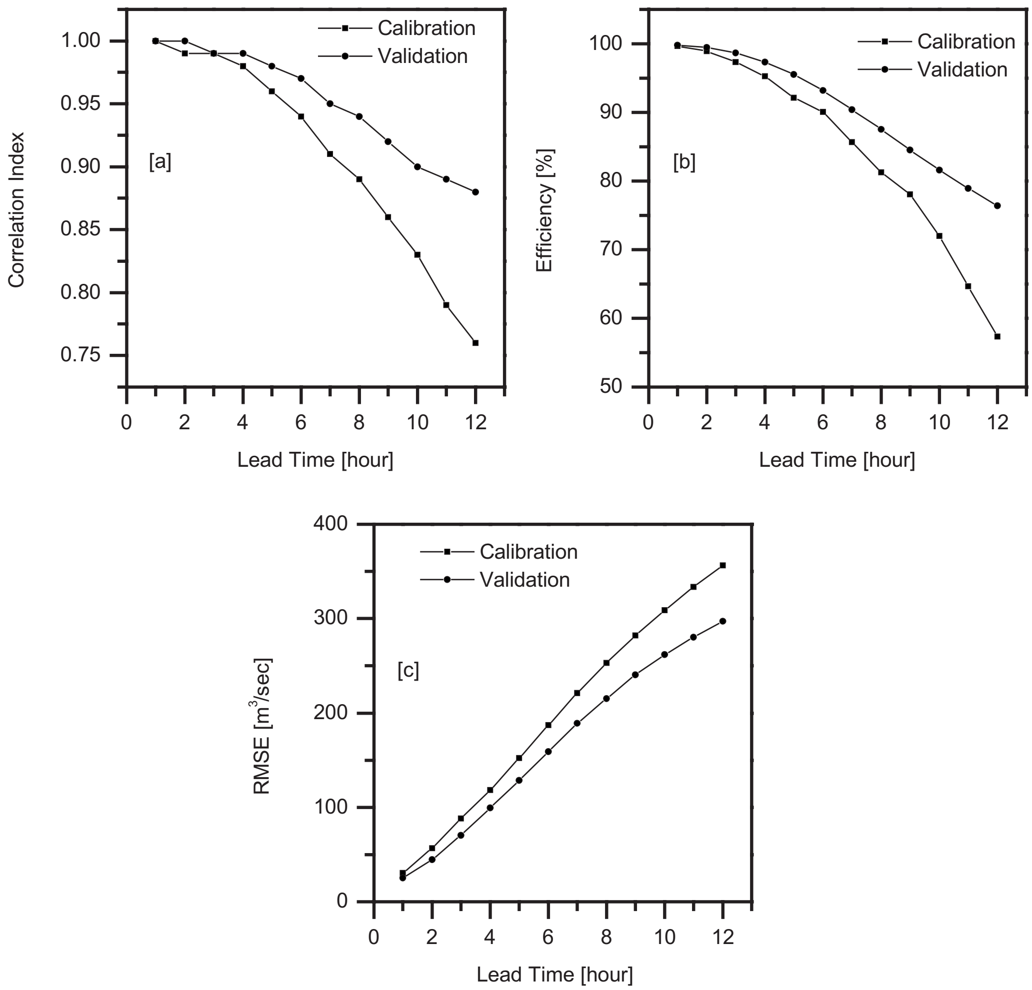

Fig. 5. Plot showing different statistical indexes for Model 8 (a) correlation; (b) efficiency; and (c) RMSE

$$\text{if } [Qt-1) \text{ is } 0.6615 \text{ and } Q\ (t-2) \text{ is } 0.6634$$
$$\text{and } Q(t-3) \text{ is } 0.6653$$

their consequent parameters are derived as

$$\text{then } Q(t) = [1.31Q(t-1) + 0.14Q(t-2) - 0.46Q(t-3)] + 0.005$$

For the first trial, which corresponds to the "low flow" estimated by the model, one of the fuzzy rules is that if $Q(t-1)$ is medium flow; $Q(t-2)$ is high flow; and $Q(t-3)$ is low flow then $Q(t)$ is low flow (obtained after TS fuzzy operation using LSE).

Similarly, in the second trial, which corresponds to the high flow by the model, one of the fuzzy rules is that if $Q(t-1)$ is high flow; $Q(t-2)$ is medium flow; and $Q(t-3)$ is low flow then $Q(t)$ is high flow (obtained after TS fuzzy operation using LSE)

In the above two cases, membership grade (MG) obtained for $Q(t-1)$ and $Q(t-2)$ are kept constant and MG for $Q(t-3)$ is varying from low to high flow for further analysis. It is observed that the variation is very less ($<5\%$) in forecasting runoff. It indicates that $Q(t-3)$ has less influence. It is also observed that while changing the MG at different level of $Q(t-1)$ and $Q(t-2)$, the fuzzy model follows the trend in the antecedent part. In other words, if the MG is in increasing order from $Q(t-2)$ to $Q(t-1)$, then the computed forecast is also increasing, similarly a reverse trend is observed when MG is decreasing from $Q(t-2)$ to $Q(t-1)$. For further analysis, the membership grade ratio between $Q(t-2)$ to $Q(t-1)$ is plotted again the forecast time horizon (see Fig. 6). From the figure it is observed that ratio is almost stable up to 3-h lead forecast and the MG ratio is reducing further. This is because the Model 8 consists of three input vectors with maximum lag up to 3 h. From the discussion it is learnt that the fuzzy if-the rule transfers some knowledge (dividing the data into different flow domain by input space partitioning) that which kind of antecedent flow (low, medium, high, etc.) information may have good performance and which kind may not reasonable in flood forecasting. The variation between different MG obtained after clustering for different input vectors show that last time step are dominating the flow behavior in a fuzzy model. These rules give us a good understanding of the data by varying the membership grade in the fuzzy rule base.

The explained variance is estimated for Model 8 which consists of three input variables $[Q(t-1), Q(t-2), \text{and } Q(t-3)]$ for 1 to 12 h lead forecasting. It is found that explained variance is nearly 100% during calibration and validation period for 1-h lead forecasting, as the lead time is increasing model performance is deteriorating but variance is observed more than 80% during calibration and validation period at 12-h lead forecasting which indicates good performance by the Model 8. This analysis indicates that combination of different input vector generates better flow properties compared to different individual input variable.

Table 3. Coordinates of Cluster Centers

$Q(t-1)$	$Q(t-2)$	$Q(t-3)$
0.6615	0.6634	0.6653
0.6615	0.7364	0.6653
0.6615	0.7364	0.8024

Note: Values are in transformed-standardized-scaled domain.

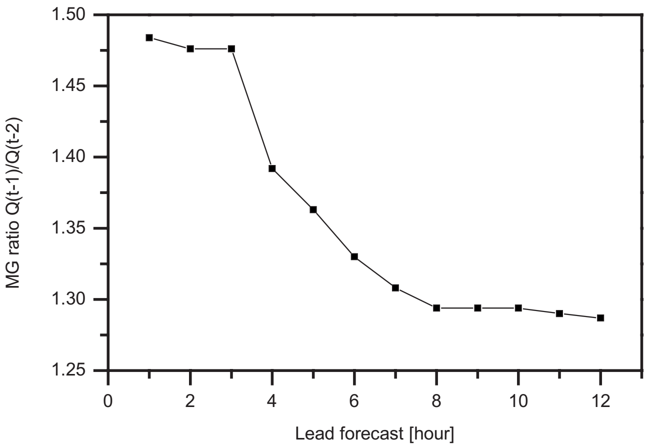

Fig. 6. Plot showing membership grade ratio along the forecast time horizon

Conclusions

This paper discusses an effective technique to distinguish the strength of relationship between input and output variables in a fuzzy rule-based model. The approach is illustrated through a case study by developing a river flow model for Baitarani basin. The influence of each individual input variable on the output variable is assessed. The analysis helps to identify the behavior in estimating the peak flow and time difference to peak flow with respect to individual input variables and their interaction. The explained variance estimated for input of 1-h lag (Model 1) is 99%, as the time lag is increasing model performance is deteriorating. Explained variance estimated by the Model 6 is 74.48 and 78.68% during calibration and validation period respectively. It is also found that as the lag time (in input vector) is increasing, efficiency index is decreasing and similarly, the time to peak flow prediction is also increasing. From the analysis, it is learnt that last time steps of measured runoff values are dominating flow behavior in a fuzzy model. It is also observed that the fuzzy model estimation follows the trend in the antecedent part. While the major objective of this article is to analyze the behavior of fuzzy if-then rule in a flood forecasting model, more rigorous studies are required to draw any concrete evaluation of the physical process.

References

Bagis, A., and Karaboga, D. (2004). "Artificial neural networks and fuzzy logic based control of spillway gates of dams." *Hydrolog. Process.*, 18, 2485–2501.

Bardossy, A., Bogardi, I., and Duckstein, L. (1990). "Fuzzy regression in hydrology." *Water Resour. Res.*, 26(7), 1497–1508.

Bardossy, A., and Disse, M. (1993). "Fuzzy rule-based model for infiltration." *Water Resour. Res.*, 29(2), 373–382.

Campolo, M., Andreussi, P., and Soldati, A. (1999). "River flood forecasting with neural network model." *Water Resour. Res.*, 35(4), 1191–1197.

Chang, F.-J., Hu, H.-F., and Chen, Y.-C. (2001). "Counterpropagation fuzzy-neural network for streamflow reconstruction." *Hydrolog. Process.*, 15, 219–232.

Chang, L.-C., and Chang, F.-J. (2001). "Intelligent control for modelling of real-time reservoir operation." *Hydrolog. Process.*, 15(9), 1621–1634.

Chang, L.-C., Chang, F.-J., and Tsai, Y.-H. (2005). "Fuzzy exemplar-based inference system for flood forecasting." *Water Resour. Res.*, 41, W02005.

Chau, K. W. (2006). "Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River." *J. Hydrol.*, 329(3–4), 363–367.

Chau, K. W., Wu, C. L., and Li, Y. S. (2005). "Comparison of several flood forecasting models in Yangtze River." *J. Hydrol. Eng.*, 10(6), 485–491.

Cheng, C. T., and Chau, K. W. (2001). "Fuzzy iteration methodology for reservoir flood control operation." *J. Am. Water Resour. Assoc.*, 37(5), 1381–1388.

Cheng, C. T., Lin, J. Y., Sun, Y. G., and Chau, K. W. (2005). "Long-term prediction of discharges in Manwan hydropower using adaptive-network-based fuzzy inference systems models." *Lect. Notes Comput. Sci.*, 3612, 1152–1161.

Cheng, C. T., Ou, C. P., and Chau, K. W. (2002). "Combining a fuzzy optimal model with a genetic algorithm to solve multi-objective rainfall-runoff model calibration." *J. Hydrol.*, 268, 72–86.

Chiu, S. (1994). "Fuzzy model identification based on cluster estimation." *J. Intell. Fuzzy Syst.*, 2(3), 267–278.

Chuntian, C. (1999). "Fuzzy optimal model for the flood control system of the upper and middle reaches of the Yangtze River." *Hydrol. Sci. J.*, 44(4), 573–582.

Deka, P., and Chandramouli, V., "Fuzzy neural network model for hydrologic flow routing." (2005). *J. Hydrol. Eng.*, 10(4), 302–314.

Duan, Q., Sorooshian, S., and Gupta, V. K. (1992). "Effective and efficient global optimization for conceptual rainfall runoff models." *Water Resour. Res.*, 28(4), 1015–1031.

Hong, Y.-S., Rosen, M. R., and Reeves, R. R. (2002). "Dynamic fuzzy modeling of storm water infiltration in urban fractured aquifers." *J. Hydrol. Eng.*, 7(5), 380–391.

Hundecha, Y., Bardossy, A., and Theisen, H.-W. (2001). "Development of a fuzzy logic based rainfall-runoff model." *Hydrol. Sci. J.*, 46(3), 363–376.

Jacquin, A., and Shamseldin, Y. (2006). "Development of rainfall-runoff models using Takagi-Sugeno fuzzy inference systems." *J. Hydrol.*, 329(1–2), 154–173.

Jairaj, P. G., and Vedula, S. (2000). "Multireservoir system optimization using fuzzy mathematical programming." *Water Resour. Manage.*, 14(6), 457–472.

Kisi, O. (2006). "Daily pan evaporation modelling using a neuro-fuzzy computing technique." *J. Hydrol.*, 329, 636–646.

Lee, Y. W., Dahab, M. F., and Bogardi, I. (1995). "Nitrate-risk assessment using fuzzy-set approach." *J. Environ. Eng.*, 121(3), 245–256.

Luchetta, A., and Manetti, S. (2003). "A real time hydrological forecasting system using a fuzzy clustering approach." *Comput. Geosci.*, 29(9), 1111–1117.

Mamdani, E. H., and Assilian, S. (1975). "An experiment in linguistic synthesis with a fuzzy logic controller." *Int. J. Man-Mach. Stud.*, 7(1), 1–13.

Nayak, P. C., and Sudheer, K. P. (2008). "Fuzzy model identification based on cluster estimation for reservoir inflow forecasting." *Hydrolog. Process.*, 22, 827–841.

Nayak, P. C., Sudheer, K. P., and Ramasastri, K. S. (2005a). "Fuzzy computing based rainfall-runoff model for real time flood forecasting." *Hydrolog. Process.*, 19, 955–968.

Nayak, P. C., Sudheer, K. P., Rangan, D. M., and Ramasastri, K. S. (2005b). "Short-term flood forecasting with a neurofuzzy model." *Water Resour. Res.*, 41, W04004.

Nayak, P. C., Sudheer, K. P., Rangan, D. M., and Ramasastri, K. S. (2004). "A neuro-fuzzy computing technique for modeling hydrological time series." *J. Hydrol.*, 291(1–2), 52–66.

Ozelkan, E. C., and Duckstein, L. (2001). "Fuzzy conceptual rainfall-runoff models." *J. Hydrol.*, 253, 41–68.

Panigrahi, D. P., and Mujumdar, P. P. (2000). "Reservoir operation modeling with fuzzy logic." *Water Resour. Manage.*, 14(2), 89–109.

Russell, S. O., and Campbell, P. F. (1996). "Reservoir operating rules with fuzzy programming." *J. Water Resour. Plann. Manage.*, 122(3), 165–170.

Schulz, K., and Huwe, B. (1997). "Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach." *J. Hydrol.*, 201, 211–229.

See, L., and Openshaw, S. (2000). "Applying soft computing approaches to river level forecasting." *Hydrol. Sci. J.*, 44(5), 763–778.

Sen, Z., and Altunkaynak, A. (2003). "Fuzzy awakening in rainfall-runoff modeling." *Nord. Hydrol.*, 35(1), 31–43.

Shrestha, B., Duckstein, P. L., and Stakhiv, E. Z. (1996). "Fuzzy rule-based modeling of reservoir operation." *J. Water Resour. Plann. Manage.*, 122(4), 262–269.

Takagi, T., and Sugeno, M. (1985). "Fuzzy identification of systems and its application to modeling and control." *IEEE Trans. Syst. Man Cybern.*, 15(1), 116–132.

Tsukamoto, Y. (1979). "An approach to fuzzy reasoning method." *Advances in fuzzy set theory and application*, M. M. Gupta, R. K. Ragade, and R. R. Yager, eds., North-Holland, Amsterdam, 137–149.

Vernieuwe, H., Georgievab, O., De Baets, B., Pauwels, V. R. N., Verhoest, N. E. C., and De Troch, F. P. (2005). "Comparison of data-driven Takagi–Sugeno models of rainfall-discharge dynamics." *J. Hydrol.*, 302, 173–186.

Xiong, L. H., and O'Connor, K. M. (2002). "Comparision of four updating models for real-time river flow forecasting." *Hydrol. Sci. J.*, 47(4), 621–639.

Xiong, L. H., Shamseldin, A. Y., and O'Connor, K. M. (2001). "A nonlinear combination of the forecasts of rainfall-runoff models by the first order Takagi-Sugeno fuzzy system." *J. Hydrol.*, 245(1–4), 196–217.

Yager, R., and Filev, D. (1994). "Generation of fuzzy rules by mountain clustering." *J. Intell. Fuzzy Sys.*, 2(3), 209–219.