JARQ 40 (4), 369 – 378 (2006) http://www.jircas.affrc.go.jp

A Takagi-Sugeno Fuzzy System for the Prediction of River Stage Dynamics

Muhammad AQIL[1], Ichiro KITA[2]*, Akira YANO[2] and Soichi NISHIYAMA[3]

[1] The United Graduate School of Agricultural Sciences, Tottori University (Tottori, Tottori 680–8573, Japan)
[2] Faculty of Life and Environmental Science, Shimane University (Matsue, Shimane 690–8504, Japan)
[3] Faculty of Agriculture, Yamaguchi University (Yamaguchi, Yamaguchi 753–8515, Japan)

Abstract

An algorithm for real-time prediction of river stage dynamics using a Takagi-Sugeno fuzzy system is presented in this paper. The system is trained incrementally each time step and is used to predict one-step and multi-step ahead of river stages. The number of input variables that were considered in the analysis was determined using two statistical methods, i.e. autocorrelation and partial autocorrelation between the variables. Effectiveness of the identification technique was demonstrated by a simulation study on the river stage of the Cilalawi River in Indonesia. The numerical results of the Takagi-Sugeno fuzzy modeling method were compared with the results of a conventional linear regression model. Through inspection of the results it was found that the Takagi-Sugeno fuzzy approach was more accurate in predicting one-step and multi-step ahead of river stage dynamics than the conventional multiple linear regression approach. The Takagi-Sugeno fuzzy system was able to make a proper fuzzy rule from the training data set, which might be considered as one of the main drawbacks of the Takagi-Sugeno fuzzy system. Yet, more substantial improvement certainly should be pursued through further research to improve the forecast results at greater lead times.

Discipline: Agricultural engineering
Additional key words: linear regression, multi-step ahead, time series

Introduction

The design, planning, and operation of river systems depend largely on relevant information derived from the forecasting and estimation of extreme events. Reliable flood forecasts are particularly important for improving public safety and mitigating economic damages caused by inundations. During the past few decades, a great deal of research has been devoted to the modeling and forecasting of river flow dynamics. Such efforts have led to the formulation of a wide variety of approaches and the development of a large number of models. The existing models for river flow forecasting may broadly be grouped under two main categories namely, physically based models and black-box models. Due to the realistic representation of watershed topography and ability to capture the surface and ground water interaction, the more reasonable method to predict a flood is the distributed and physically based model. However, extensive topographic, meteorological, and hydrologic data are required to describe the runoff process and time is also required to calibrate conceptual models (especially distributed models), which are important factors to be considered in their practical applications. Thus, the implementation and calibration of conceptual models can typically present various difficulties[4,6]. In this context data-driven models, which can discover relationships from input-output data without having the complete physical understanding of the system, may be preferable. While such models do not provide any information on the physics of the hydrologic processes, they are in particular, very useful for river flood forecasting where the main concern is accurate predictions of a flood at specific watershed locations[7].

The analysis and design of the Takagi-Sugeno fuzzy system have been studied for decades since this system was introduced[11]. The Takagi-Sugeno model can be implemented based on a neural network-driven fuzzy reasoning system. It is also well known that the Takagi-Sugeno fuzzy system is very suitable for applications in hydrological studies. There are two reasons for using the

*Corresponding author: fax +81–852–32–6030; e-mail kita@life.shimane-u.ac.jp
Received 10 November 2005; accepted 27 March 2006.

Takagi-Sugeno fuzzy system for hydrological studies: (1) the simplicity of the inference procedure, and (2) the possibility to incorporate a general condition on the physical structure of the system into the fuzzy system. Therefore, during recent decades, a number of simulation studies can be noticed[1,3,4,8,12,13], which are dedicated to the Takagi-Sugeno fuzzy system. However, detailed studies that compare the Takagi-Sugeno fuzzy model with the traditional model for river stage estimation under limited data, especially in developing and under-developed countries, are less available. This study addresses this need with a detailed comparison of two models in the context of river stage estimation. A Takagi-Sugeno fuzzy model and linear regression model were trained incrementally using the same data sets and were used to predict one-step and multi-step ahead of river stage dynamics.

Takagi-Sugeno fuzzy system

The fuzzy inference system proposed by Takagi and Sugeno, known as the TS model in fuzzy system literature provides a powerful tool for modeling complex nonlinear systems. The basic idea of the Takagi-Sugeno model is the fact that an arbitrary complex system is a combination of mutually inter-linked subsystems[5]. Schematic representation of a Takagi-Sugeno fuzzy system is shown in Fig. 1.

Given properly defined input variables and membership functions, the Takagi-Sugeno fuzzy rules for a system considered herein are in the form of

$$\begin{aligned}&R_i : \text{IF } x_1 \text{ is } A_{i1} \text{ and } \ldots \text{ and } x_m \text{ is } A_{im} \text{ THEN}\\ &y_i = a_{i1}x_1 + \ldots + a_{im}x_m + a_{i0}\end{aligned} \quad (1)$$

where R_i $(I = 1, 2, \ldots, c)$ denotes the ith fuzzy rule, x_j $(j = 1, 2, \ldots, m)$ are the input (antecedent) variables, y_i are the rule output variables, $A_{i1}, \ldots, A_{im}$ are fuzzy sets defined in the antecedent space, and $a_{i1}, \ldots, a_{im}, a_{i0}$ are the model consequent parameters that have to be identified in a given data set. For a given input crisp vector $x = (x_1, \ldots, x_m)^T$, the inferred global output of the Takagi-Sugeno model is computed by taking the weighted average of the individual rules' contributions

$$\hat{y} = \frac{\sum_{i=1}^{c} \tau_i(x) \cdot y_i}{\sum_{i=1}^{c} \tau_i(x)} \quad (2)$$

where $\tau_i(x)$ is the degree of fulfillment of the ith fuzzy rule, defined by

$$\begin{aligned}&\tau_i(x) = \mathrm{Min}\{\mu_{A_{i1}}(x_1) \ldots \mu_{A_{im}}(x_m)\} \text{ or}\\ &\tau_i(x) = \mu_{A_{i1}}(x_1) \cdot \mu_{A_{i2}}(x_2) \ldots \mu_{A_{im}}(x_m) \quad i = 1, 2, \ldots, c\end{aligned} \quad (3)$$

for the minimum and product conjunction operators, respectively. $\mu_{A_{ij}} : R \rightarrow [0,1]$ is the membership function of the antecedent fuzzy set A_{ij}.

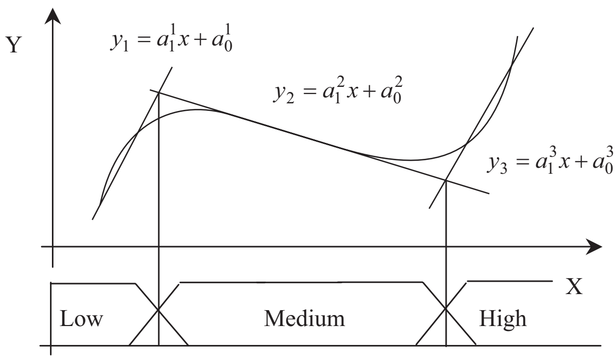

Fig. 1. Schematic representation of a Takagi-Sugeno model

Study watershed and modeling

1. The available data

To illustrate the practical application of the Takagi-Sugeno fuzzy system, the Cilalawi River Basin, located in the West Java Province of Indonesia was used as a research area. The climate of the catchment is generally dry, except during the monsoon months from December to April. The annual precipitation is 3,000 mm in the mountainous area and 2,500 mm in the lowland. Normally, around 70% of the precipitation falls during the rainy season whereas 30% falls during the dry season. The location of the study area is shown in Fig. 2. Water resources in the study area are operated and managed by Perum Jasa Tirta II, a public corporation formed in 1967. There are three large multipurpose reservoirs from upstream to downstream (Saguling, Cirata, and Jatiluhur) that regulate the water flow and are the main source of water supply in Jakarta City and West Java Province through the Tarum Canal. The upper part of the Citarum River has many tributaries, and one of the main tributaries is the Cilalawi River. The total drainage area of the Cilalawi River Basin is approximately 60.17 km^2. The river stages data of the Cilalawi River are available for the hydrological year of 2002 with a sampling interval of 6-min.

In this study, the performance of the Takagi-Sugeno fuzzy system was examined on hourly intervals. To achieve this, the 6-min data series was converted into average hourly data before proceeding onto the Takagi-Sugeno fuzzy inference network. The data were divided into three independent subsets: 4,000 data sets for a train-

Fig. 2. Map of the river basin under consideration

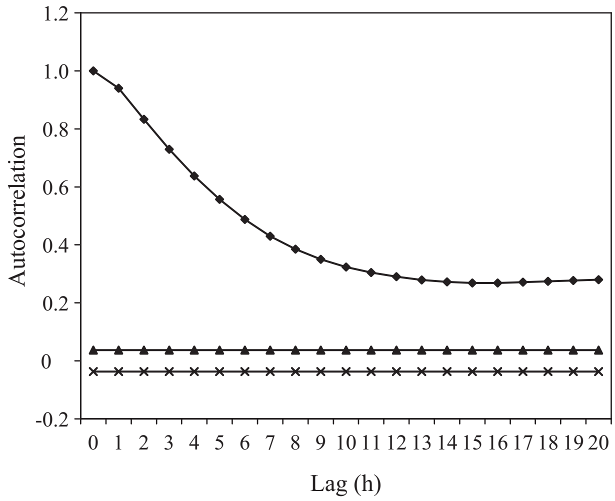

Fig. 3. Autocorrelation plot of the river stage series

◆ : Autocorrelation coefficient,
▲ : 95% Confidence band,
✕ : 95% Confidence band.

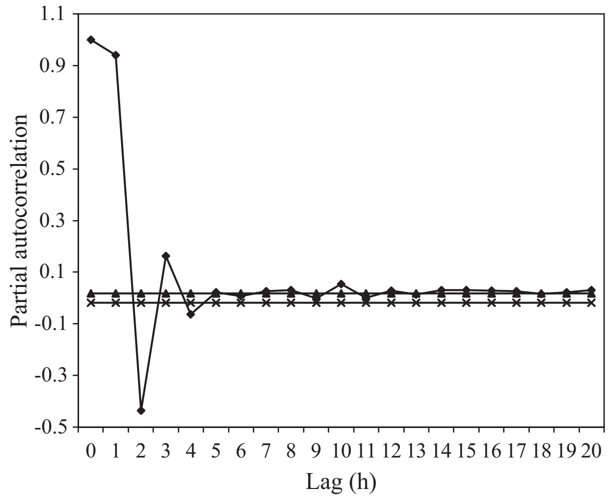

Fig. 4. Partial autocorrelation plot of the river stage series

◆ : Partial autocorrelation coefficient,
▲ : 95% Confidence band,
✕ : 95% Confidence band.

ing subset; 2,500 data sets for the verification; and 2,000 data sets for testing the model.

2. Selection of antecedent river stage inputs to the model

One of the most important steps in the model development process is the determination of significant input variables. The parameters that need to be selected in the input variable are the number of river stage values at different intervals (of time) that have a significant influence on the predicted river stage. In this study, the number of parameters corresponding to different antecedents was determined by two statistical methods, i.e. autocorrelation function (ACF) and partial autocorrelation function (PACF) between the variables. The ACF and PACF are generally used to gather information about the autoregressive process of the data series[10]. The number of antecedent river stages that should be included in the input variables is usually determined by placing a 95% confidence interval on the autocorrelation and partial autocorrelation plots.

The ACF and the corresponding 95% confidence intervals of the river stage series for lag 0 to lag 20 are presented in Fig. 3. Similarly, the PACF and the corresponding 95% confidence intervals of the river stage series are presented in Fig. 4. The ACF of Fig. 3 showed a significant correlation at 95% confidence level interval up to 14-h of river stage lag. In addition, the PACF showed significant correlation up to lag of 3 (3-h). Results of correlogram plots of the data series shown in Figs. 3 and 4 imply that incorporating the river stage values up to lag 3-h can best represent the process in the catchment area under examination. Therefore, in this study, three antecedent values of river stage denoted by RS(t-2), RS(t-1) and RS(t), respectively, were selected as inputs for modeling the river stage.

3. Identification of a Takagi-Sugeno fuzzy system

A Takagi-Sugeno fuzzy model can be identified through a two step procedure: (1) identification of the antecedent membership functions of the fuzzy sets in the premise of each rule; and (2) identification of the parameters in the consequence of each rule. The antecedent membership functions of a fuzzy rule define local region, while the consequent describes the behavior within the region via various constituents.

Computation of the antecedent membership functions of the fuzzy sets was done using a grid partitioning method, where each antecedent variable is independently partitioned. The derivation of membership functions depends on the expert's a priori knowledge about the model under consideration. However, no specific knowledge is available for many systems and in such cases the domain of the antecedent variables can simply be partitioned into a number of equally spaced and equally shaped membership functions[2]. If the measured input/output data of a process are available, the shape and location of the membership functions can be created.

In addition to the consequent parameters estimation, a global least squares method is used. Consider $\{(x_i, y_i) | i = 1, 2, \ldots, n\}$ is a set of n input-output data pairs of a system. Let X denote the matrix whose ith row is the input vectors x_i and let Y denote the vector column having y_i as its ith component. Let W_i denote the $n \times n$ real matrix having the normalized degree of fulfillment β_{ij} as its jth diagonal element, where

$$\beta_{ij} = \frac{\tau_i(x_i)}{\sum_{k=1}^{c} \tau_k(x_j)} \tag{4}$$

Let $\theta_i = [a_{i1}, a_{i2} \ldots a_{im}, a_{i0}]$ denote the vector of consequent parameters of the ith rule. In order to estimate the off-set term a_{i0}, a unitary column I is appended to the matrix X, to produce the extended matrix $X_c = [X, I]$. Then, the parameter vector θ_i is calculated as the least-square solution

$$\theta_i = [X_c^T \cdot W_i \cdot X_c]^{-1} X_c^T \cdot W_i \cdot Y \tag{5}$$

For further details on alteration from the equation parameter optimization into the least squares estimation on these techniques, we refer to Babuška[2].

4. Examples of river stage estimation

In order to demonstrate the effectiveness of the fuzzy method, a data set from the Cilalawi River was used. Based on ACF and PACF analysis, it was found that incorporating the river stage values up to lag 3-h can best represent the process in the catchment area under examination. Therefore, the 3 antecedent values of river stage (RS) denoted by RS(t-2), RS(t-1), and RS(t), respectively, were selected as inputs for modeling river stage. We then

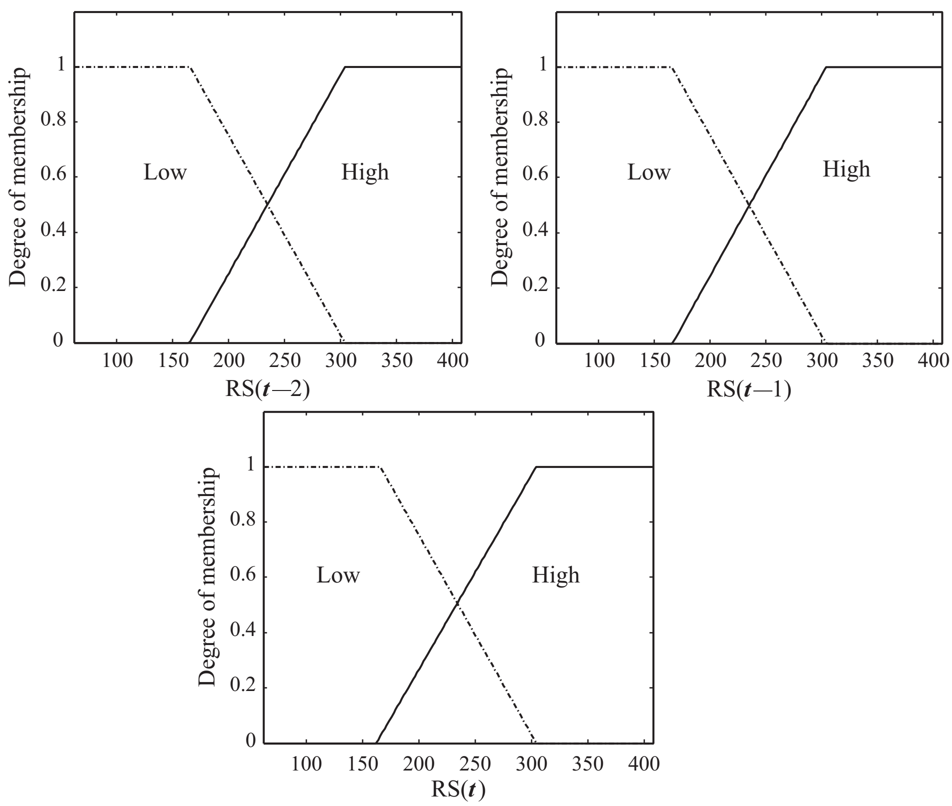

Fig. 5. Membership functions for input parameters

partition all inputs into two domains (i.e. low level and high level) and a trapezoidal type membership function was assigned for all fuzzy sets. The membership functions of river stage are shown in Fig. 5. The number of rules defined in this model is a product of the number of membership functions in each input. Therefore, the model contains 8 (2 × 2 × 2) rules and the descriptions of each rule are explicitly depicted in Table 1.

To demonstrate how to compute the output ($\hat{y}$) for a given data, a sample inference is shown in Table 2 for three given antecedent RS values of 234.9, 232.6, and 235.1 cm. The inputs intersect the membership functions at some membership levels, and fuzzy subsets of low and high for RS(*t*-2), RS(*t*-1), and RS(*t*) are triggered. In the antecedent part of Table 2, membership values for each fuzzy subset can be seen. Then the membership values from each rule are propagated to the consequents. For the first rule, each input yielded the fuzzy membership values of 0.51, 0.53, and 0.50, respectively. The fuzzy prod operator then simply calculates the value of the degree of fulfillment as 0.135 and fuzzy operation for this rule is complete. This operation is conducted for each rule and the inferred global output of the Takagi-Sugeno fuzzy system is calculated by the fuzzy mean weighted

$$\hat{y} = \frac{\begin{array}{c}0.135 \times 228.32 + 0.132 \times 350.23 + 0.117 \times 1{,}111.5 - 0.115 \times 1{,}512.7 \\ + 0.127 \times 245.49 + 0.125 \times 500.47 - 0.110 \times 282.18 + 0.108 \times 1{,}239.1\end{array}}{0.135 + 0.132 + 0.117 + 0.115 + 0.127 + 0.125 + 0.110 + 0.108} = 237.29 \text{ cm} \tag{6}$$

Table 1. Fuzzy scaling rules for the model

	RS(*t*–2)	RS(*t*–1)	RS(*t*)	y_i
Rule 1	Low	Low	Low	y_1= 0.13 × RS(*t*–2) – 0.66 × RS(*t*–1) + 1.47 × RS(*t*) + 5.7
Rule 2	Low	Low	High	y_2= 0.98 × RS(*t*–2) – 0.24 × RS(*t*–1) + 2.97 × RS(*t*) – 522.4
Rule 3	Low	High	Low	y_3= 0.72 × RS(*t*–2) + 4.26 × RS(*t*–1) + 8.93 × RS(*t*) – 2,148
Rule 4	Low	High	High	y_4= –7.47 × RS(*t*–2) + 5.45 × RS(*t*–1) + 4.14 × RS(*t*) – 1,999
Rule 5	High	Low	Low	y_5= –0.53 × RS(*t*–2) + 0.64 × RS(*t*–1) + 1.18 × RS(*t*) – 56.3
Rule 6	High	Low	High	y_6= –1.09 × RS(*t*–2) – 3.27 × RS(*t*–1) + 14.59 × RS(*t*) – 1,913
Rule 7	High	High	Low	y_7= 1.23 × RS(*t*–2) – 5.41 × RS(*t*–1) – 4.38 × RS(*t*) + 1,717
Rule 8	High	High	High	y_8= –1.37 × RS(*t*–2) – 4.47 × RS(*t*–1) + 0.36 × RS(*t*) + 2,516

Table 2. A sample of the Takagi-Sugeno inference system

	$\mu_{A_{i1}}(x_1)$	$\mu_{A_{i2}}(x_2)$	$\mu_{A_{i3}}(x_3)$		$\tau_i(x)$
Rule 1	0.51	0.53	0.50	228.3	0.51 × 0.53 × 0.50 = 0.135
Rule 2	0.51	0.53	0.49	350.2	0.51 × 0.53 × 0.49 = 0.132
Rule 3	0.51	0.46	0.50	1,111.5	0.51 × 0.46 × 0.50 = 0.117
Rule 4	0.51	0.46	0.49	–1,512.7	0.51 × 0.46 × 0.49 = 0.115
Rule 5	0.48	0.53	0.50	245.4	0.48 × 0.53 × 0.50 = 0.127
Rule 6	0.48	0.53	0.49	500.4	0.48 × 0.53 × 0.49 = 0.125
Rule 7	0.48	0.46	0.50	–282.18	0.48 × 0.46 × 0.50 = 0.110
Rule 8	0.48	0.46	0.49	1,239.	0.48 × 0.46 × 0.49 = 0.108
	RS(*t*–2)=234.9	RS(*t*–1)=232.6	RS(*t*)=235.1		$\hat{y}$ = 237.29

5. Multiple linear regression model development

Multiple linear regression analysis is a method used to model the linear relationship between a dependent variable and one or more independent variables. The objective of the multiple linear regression analysis is to determine the values of the parameters of the regression equation and then to quantify the goodness of fit in respect of the dependent variable. For further details on this technique, we refer to Snedecor and Cochran[9]. The derived linear regression model for river stage estimation is as follows:

$$\hat{y} = 1.360 + 7.900x_1 + 0.149x_2 - 0.598x_3 \quad (7)$$

where $\hat{y}$ is the predicted variable, 1.360 is the intercept, (7.900, 0.149, and –0.598) are the regression coefficients, and $x_1 - x_3$ are independent variables referring to the river stage values. Multiple linear regression analysis was developed and tested with the same data sets used for the Takagi-Sugeno fuzzy model, and the developed regression was referred to as a trained model. Then the predictive ability of the model was validated and tested with the same data sets used to test the Takagi-Sugeno fuzzy model, thus making the model results comparable, and referred to as validated and tested models. The validated and tested models are indicative of the model capability to simulate the river stage dynamics since the data are independent of the data used for model development.

6. Model performance

The performances of the models developed in this study were assessed using various standard statistical performance evaluation criteria. The statistical measures considered were coefficient of correlation (CORR), mean absolute percentage error (MAPE), and root mean square error (RMSE). Statistical performance measures are listed in Table 3.

Results and discussion

The available data set was divided into three sets called training, verification, and testing data sets. A grid partition method was used to create the initial membership function matrix using trapezoidal functions for each of the input variables. We selected three membership functions for the river stage at *t-2*, *t-1*, and *t*, respectively, with the number of membership functions for each input fixed at 2. As the parameters in the premise membership functions are adjusted, the grid evolves. After computing the gradient vector of the parameters of the membership functions, the model employed an optimization technique to adjust the parameters to reduce some error measures (usually defined by the sum of the squared difference between actual and desired outputs). Since this model consists of two membership functions for each input, the river stage estimation can be performed using 8 rules.

To ensure the accuracy of the developed river stage dynamics based on Takagi-Sugeno fuzzy and multiple regression models and to survey the spread of the values, the observed and predicted results using the developed models are compared in Fig. 6. The line of best fit using the plotted points was calculated using the regression. In each of the scatter diagrams, the closer the points fall on a straight line, the more accurate the tested model was. In

Table 3. List of the performance measures

Statistical parameter	Expression
Coefficient of correlation (CORR)	$\text{CORR} = \left[\dfrac{\sum_{i=1}^{n} (y_i^o - \overline{y^o})(\hat{y}_i^p - \overline{\hat{y}^p})}{\sqrt{\sum_{i=1}^{n} (y_i^o - \overline{y^o})^2} \sqrt{\sum_{i=1}^{n} (\hat{y}_i^p - \overline{\hat{y}^p})^2}} \right]$
Mean absolute percentage error (MAPE)	$\text{MAPE} = \dfrac{1}{n} \sum_{i=1}^{n} \left\| \dfrac{\hat{y}_i^p - y_i^o}{y_i^o} \right\| \times 100$
Root mean square error (RMSE)	$\text{RMSE} = \sqrt{\dfrac{\sum_{i=1}^{n} (y_i^o - \hat{y}_i^p)^2}{n}}$

y_i^o and $\hat{y}_i^p$ are the observed and predicted stages at time *t* respectively, $\overline{y^o}$ and $\overline{\hat{y}^p}$ are the mean of the observed and predicted river stages, and *n* is the number of data points.

addition, the values of the performance indices of the Takagi-Sugeno and multiple linear regression models during the training, verification, and testing stages are shown in Table 4. As displayed in Fig. 6 and Table 4, it is obvious that although the results obtained by Takagi-Sugeno and multiple linear regression models were relatively successful, the Takagi-Sugeno fuzzy model gives the best fit to the observed results and produced better prediction of river stage than the developed empirical equations.

A more detailed comparison shows that in the training phase, the Takagi-Sugeno fuzzy model improves the multiple linear regression forecast by reductions of about 8.11 and 6.37% in MAPE and RMSE values, respectively. MAPE measures the absolute error as a percentage of the forecast, and RMSE evaluates the residual between observed and predicted river stage. In addition, the improvement of the forecast result regarding the correlation coefficient (CORR) value during the validation phase was approximately 1.06%. CORR evaluates the

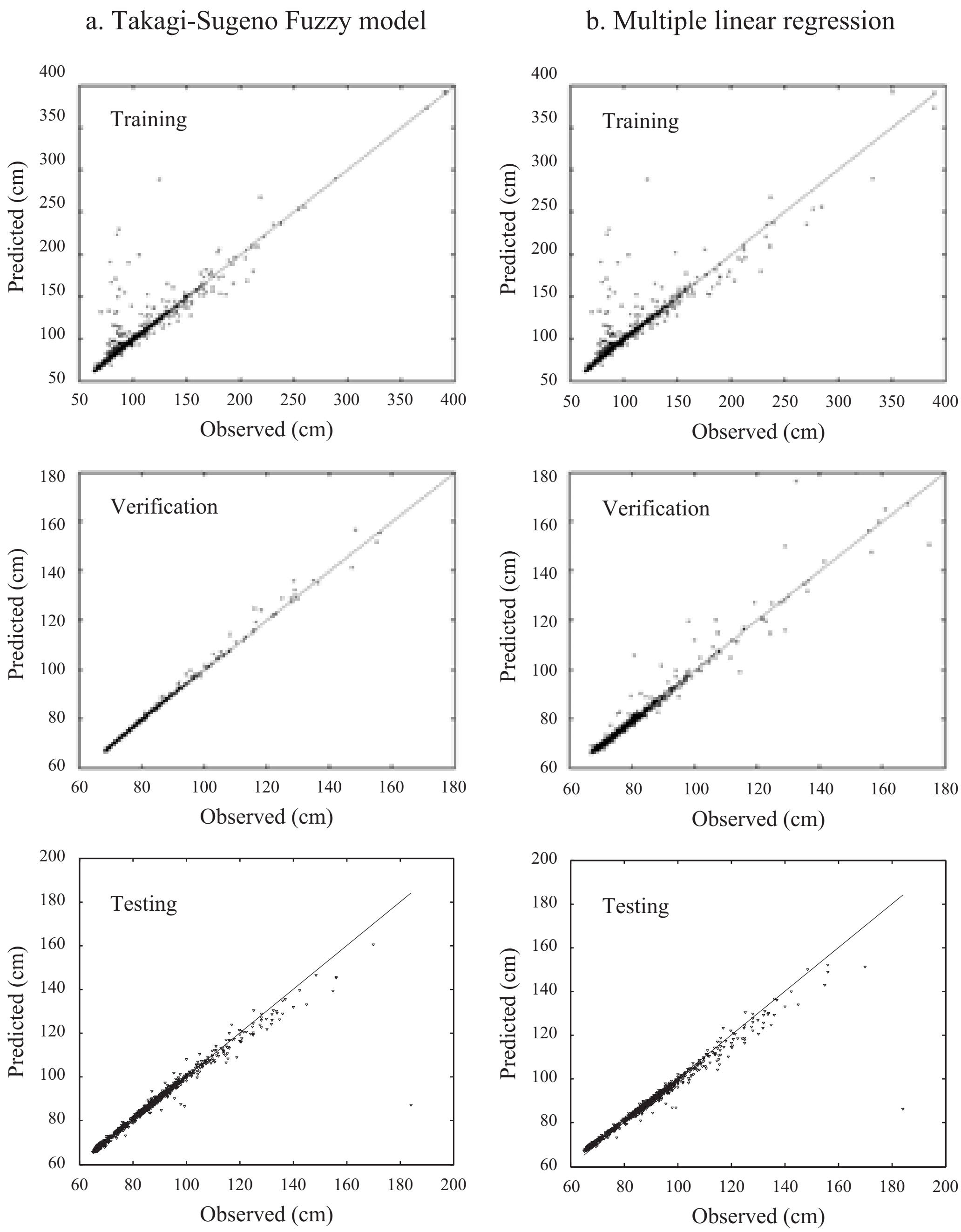

Fig. 6. Scatter plots of observed and predicted river stage during the training, verification, and testing periods

Table 4. Performance of the Takagi-Sugeno model at different phases

Period		Performance					
		Takagi-Sugeno fuzzy			Multiple linear regression		
		CORR	MAPE	RMSE	CORR	MAPE	RMSE
Training		0.95	2.15	7.49	0.94	2.34	8.00
Verification		0.97	1.34	3.12	0.96	2.15	3.55
Testing	t+1	0.98	1.15	2.65	0.96	1.73	5.02
	t+2	0.96	2.56	4.11	0.95	3.78	5.51
	t+3	0.95	4.71	4.50	0.94	5.70	6.64

linear correlation between the observed and predicted river stage. In the verification phase, the Takagi-Sugeno fuzzy model improves the multiple linear regression forecast by reductions of about 37.67 and 12.11% in MAPE and RMSE values, respectively. In addition, the improvement of the forecast result regarding the CORR value during the verification phase was approximately 1.04%.

As the next task, a multi-step ahead testing is performed. Although all models have been trained and verified, it is desirable to investigate their performance in multi-step ahead prediction. In this particular case, the predicted outputs at time $t+i$ were recursively used to predict the river stage at $t+i+1$, for any value of step i. At 1-h (t+1) lead-time, it was obvious that Takagi-Sugeno fuzzy and multiple linear regression models do not provide estimations deviating significantly from the actual values (Table 4). The Takagi-Sugeno fuzzy model improves the multiple linear regression forecast result by reductions of about 33.53 and 47.21% in MAPE and RMSE values, respectively. In addition, the improvement of the forecast result regarding the CORR value was 2.08%. In the forecast of 2-h (t+2) lead-time, the Takagi-Sugeno fuzzy model was shown to be more efficient where the model enforces the predictions to follow more precisely the observed reality. This is supported by the reasonably lower values of performance indices compared to the traditional multiple linear regression. The Takagi-Sugeno fuzzy model improves the multiple linear regression forecast result by reductions of about 32.25 and 25.41% in MAPE and RMSE values, respectively, at 2-h lead-time. In addition, the improvement of the forecast result regarding the CORR value was 1.05%. Finally, at 3-h (t+3) lead-time, the forecast results of the traditional linear regression approach start to deviate from the actual data, and it was clearly less effective than the Takagi-Sugeno fuzzy model. At 3-h lead-time, the Takagi-Sugeno fuzzy system still shows good performance where the CORR, MAPE, and RMSE results were 0.95, 4.71, and 4.50, respectively, which were better than those obtained by the multiple linear regression approach (0.94, 5.70, and 6.64, respectively). The Takagi-Sugeno fuzzy model improves the multiple linear regression forecast results by reductions of about 17.37 and 32.23% in MAPE and RMSE values, respectively. In addition, the improvement of the forecast result regarding the CORR value was 1.06%. The main advantage of Takagi-Sugeno fuzzy models over linear regression models is that they can inherently detect and incorporate non-linear relationships between input variables. Another purported advantage of Takagi-Sugeno fuzzy models is the ability to recognize the relationship between input and output data sets without specifying an a priori relationship.

To get a brief picture of the general performance of the constructed models, partly produced hydrograph plots of the observed river stage and the one-step ahead, two-step ahead and three-step ahead river stage predictions in advance are shown in Fig. 7. Overall results of the forecasts indicate that as the time step increases further on, both models tended to provide shifted (late) predictions. Fig. 7 clearly shows that although the river stage hydrographs of the Takagi-Sugeno fuzzy and multiple linear regression models do not deviate significantly at 1-h lead time, at 2-h and 3-h lead times, the multiple linear regression predictions start to move away from the actual data. The model appeared to generally over predict the lower and under predict the higher river hydrographs for this time period. This trend also occurred in the Takagi-Sugeno fuzzy model; although the magnitude of the error forecast was relatively lower than those of multiple linear regression. There are some reasons why both models failed to obtain a good result. Firstly, the worst forecast particularly at larger lead-times was due to the error accumulations. The error on predicted river stage RS(t+1) for hour (t+1) will definitely affect the forecasted river stage RS(t+2) on the hour (t+2). Similarly, the error on predicted river stage RS(t+2) for hour (t+2) will also affect the forecasted river stage RS(t+3) on the hour (t+3). In this way, the error is accumulated as the lead-time increases and this error accumulation is the obvious reason for an increasing trend in error with increase in lead-

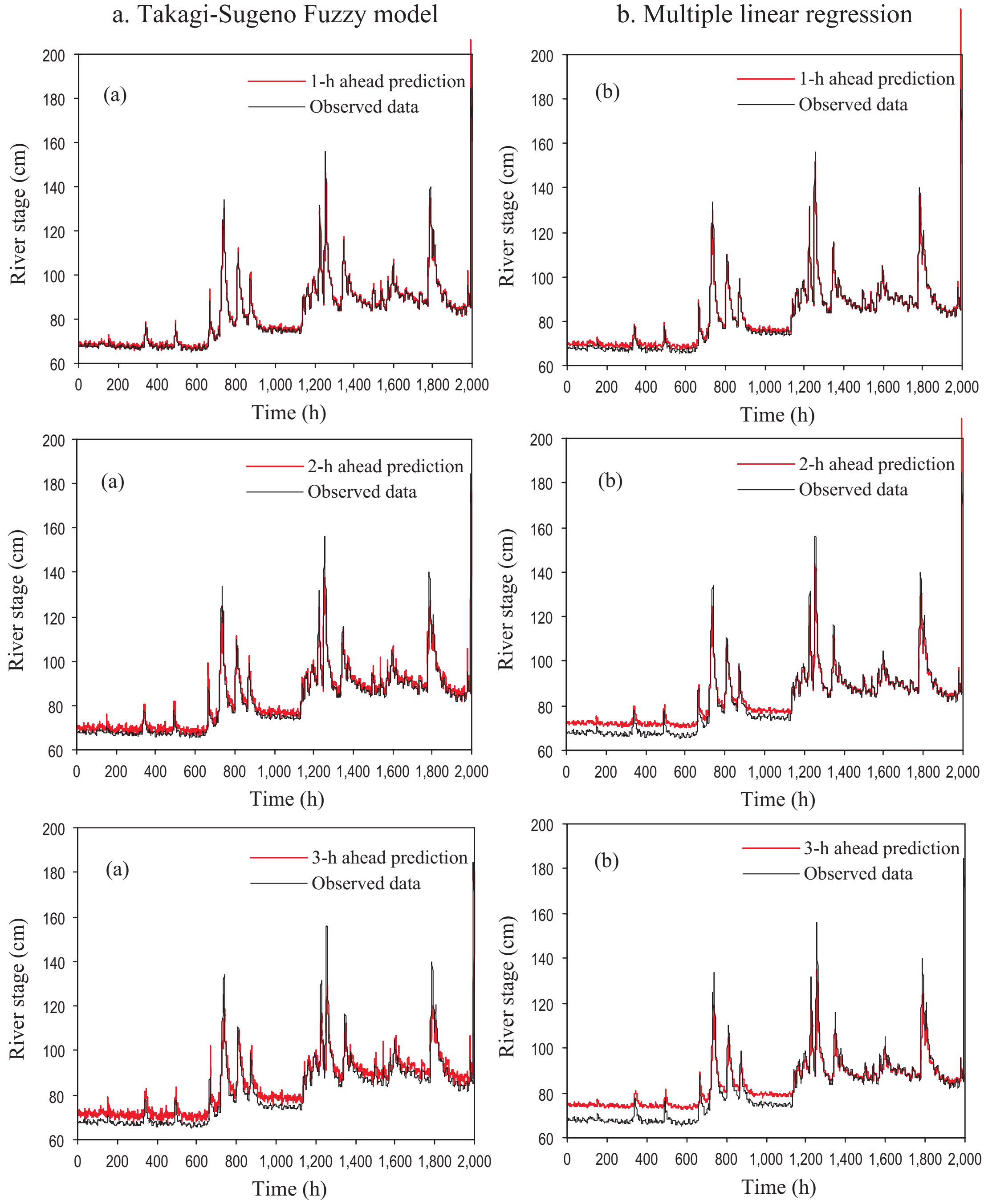

Fig. 7. Partly produced hydrograph plots of the testing results of 1-h, 2-h, and 3-h predictions in advance by Takagi-Sugeno fuzzy and multiple linear regression models

time. Secondly, there are some missing values in the data series particularly the data used to train the network which made the models very sensitive. This caused the models not to be able to predict the extreme high and low values. Therefore, our next task should be to give rather more importance to the model's ability in multi-step-ahead predictions. It would also be advisable to replace the missing values in the data sets in future research.

Conclusions

This paper presents an algorithm for real-time prediction of river stage dynamics using a rule based fuzzy system. A Takagi-Sugeno fuzzy system is trained incrementally each time step and is used to predict one-step and multi-step ahead of river stages. The number of input variables that were considered in the analysis was determined using two statistical methods, i.e. autocorrelation and partial autocorrelation between the variables. The research was illustrated through a case study of developing a Takagi-Sugeno fuzzy model for river stage prediction in the Cilalawi River of Indonesia. Further, a conventional linear regression model was developed with the same data set and compared to the results of the Takagi-Sugeno fuzzy modeling method. It was shown that the Takagi-Sugeno fuzzy approach was more accurate in predicting river stage dynamics than the conventional multiple linear regression approach. In predictions up to 3-h ahead, the Takagi-Sugeno fuzzy system still shows good performance where the CORR, MAPE, and RMSE results were 0.95, 4.71, and 4.50, respectively, which were better than those obtained by the multiple linear regression approach (0.94, 5.70, and 6.64, respectively). Therefore we conclude that the constructed Takagi-Sugeno fuzzy system can efficiently deal with vast and complex input-output patterns, and has a great ability to learn and build up an adaptive network-based fuzzy system for prediction, and the prediction results provide a useful guidance or reference for flood control operations. Yet, more substantial improvement certainly should be pursued through further research to improve the forecast results at greater lead times.

Acknowledgments

The authors would like to extend grateful thanks to Perusahaan Umum Jasa Tirta II (PJT II), West Java Indonesia for the cooperation and assistance during the data collection exercise. In a special way the authors wish to acknowledge the assistance of Mr. Andri Sewoko in providing the data on which this research was based.

References

1. Abebe, A. J., Solomatine, D. P. & Venneker, R. G. W. (2000) Application of adaptive fuzzy rule-based models for reconstruction of missing precipitation events. *Hydrol. Sci. J.*, **45**, 425–436.
2. Babuška, R. (1998) Fuzzy modeling for control. Kluwer, Boston, MA, pp.260.
3. Chang, F. -J. & Chen, Y. C. (2001) A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *J. Hydrol.*, **245**, 153–164.
4. Dou, C., Woldt, W. & Bogardi, I. (1999) Fuzzy rule-based approach to describe solute transport in the unsaturated zone. *J. Hydrol.*, **220**, 74–85.
5. Finol, J., Guo, Y. K. & Jing, X. D. (2001) A rule based fuzzy model for the prediction of petrophysical rock parameters. *J. Pet. Sci. Eng.*, **29**, 97–113.
6. Hu, T. S., Lam, K. C. & Ng, S. T. (2001) River flow time series prediction with a range-dependent neural network. *Hydrol. Sci. J.*, **46**, 729–745.
7. Nayak, P. C. et al. (2005) Short-term flood forecasting with a neurofuzzy model. *Water Resour. Res.*, **41**, W04004(1–16).
8. Schulz, K. & Huwe, B. (1997) Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach. *J. Hydrol.*, **201**, 211–229.
9. Snedecor, G. W. & Cochran, W. G. (1981) Statistical methods. Iowa State University Press, Iowa.
10. Sudheer, K. P., Gosain, A. K. & Ramasastri, K. S. (2002) A data-driven algorithm for constructing artificial neural network rainfall–runoff models. *Hydrol. Process.*, **16**, 1325–1330.
11. Takagi, T. & Sugeno, M. (1985) Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Syst. Man Cyber.*, **15**, 116–132.
12. Vernieuwe, H. et al. (2005) Comparison of data-driven Takagi-Sugeno models of rainfall-discharge dynamics. *J. Hydrol.*, **302**, 173–186.
13. Xiong, L., Shamseldin, A. Y. & 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**, 196–217.