Theor. Appl. Climatol. 90, 227–233 (2007)
DOI 10.1007/s00704-006-0267-z
Printed in The Netherlands

Theoretical and Applied Climatology

Hydraulics Division, Faculty of Civil Engineering, Istanbul Technical University, Maslak, Istanbul, Turkey

Fuzzy logic model of lake water level fluctuations in Lake Van, Turkey

A. Altunkaynak and **Z. Şen**

With 7 Figures

Received April 13, 2005; revised August 9, 2005; accepted July 12, 2006
Published online June 27, 2007

Summary

Lake Van is one of the largest terminal lakes in the world. In recent years, significant lake level fluctuations have occurred and can be related to global climatic change. This fluctuation sometimes exhibits abrupt shifts. Floods originating from the lake can cause considerable damage and loss in agriculture and urban areas. Therefore, water level forecasting plays a significant role in planning and design. This study is aimed at predicting future lake levels from past rainfall amounts and water level records. A dynamical change of the lake level is evaluated by the fuzzy approach. The fuzzy inference system has the ability to use fuzzy membership functions that include the uncertainties of the concerned event. This method is applied for Lake Van, in east Turkey. Furthermore, model capabilities are compared with ARMAX model. It is shown that lower absolute errors are obtained with the Takagi-Sugeno fuzzy approach than with the ARMAX model.

1. Introduction

It is possible to re-establish all hydrological models using fuzzy logic principles and systems. For this purpose, the relationship between hydrological variables and logical structures has to first be defined. If there is persistence in the time series it is expected that high values follow high values, and small values follow small values. In the classical approach such a relationship can be defined by linear equations. Markov and autoregressive integrated moving average (ARIMA) models are used frequently for defining stochastic processes. These models are restricted by assumptions such as linearity, normality, homoscedasticity etc. It is however, possible to eliminate such assumptions through fuzzy logic principles. Time series modelling in different orders, previously modelled by classical methods, can be achieved by applying the fuzzy logic approach.

Several methods have been introduced to model lake level fluctuations. Artificial neural network modelling was performed by Altunkaynak (2006) for Lake Van. Altunkaynak et al. (2003) used the triple diagram model which is based on geostatistical principles and can also be used as an alternative for the second order Markov process. Şen et al. (2000) identified suitable models and estimates for lake level fluctuations and their parameters for trend, periodic and stochastic parts. A second order Markov model was found to be suitable for the stochastic part. Slivitzky and Mathier (1993), stated that most of the modelling of levels and flow series on the Great Lakes have assumed time series stationarity using either Markov or ARIMA models presented by Box and Jenkins (1976). Multivariate models using monthly lake levels fail to adequately reproduce the statistical properties and persistence of basin supplies (Loucks, 1989; Iruine and Eberthardt, 1992). In contrast, spectral analysis of water levels indicated the possibility of significant trends in lake level hydrological variables (Privalsky, 1990; Kite, 1990).

The theory behind the fuzzy set is based on set membership which is the key to decision making when faced with uncertainty. For crisp sets, an element x in the universe X is either a member of some crisp set A or it is not. In the fuzzy sets, an element belongs to some fuzzy set with degrees of membership which take values from continuous interval [0–1]. Ross (1995) discussed models with essentially two different kinds of information: fuzzy membership functions, which represent similarities of objects to ambiguous properties, and probabilities, which provide knowledge about relative frequencies. He also stated that fuzzy models are not replacements for probability models. Sometimes they work better and sometimes they do not. However, a growing body of evidence suggests that fuzzy approaches to real problems are an effective alternative to traditional methods.

The main purpose of this study is to develop an estimation procedure independent of the autocorrelation concept. A fuzzy logic model is suggested and then used to predict monthly lake level fluctuations. The method is applied to water level fluctuations in Lake Van, which is located in eastern Turkey.

2. Study area features

Turkey's largest lake, Lake Van, has a surface area of approximately 3500 km^2 and a depth of 443 m. It is located in eastern Turkey (38.5° N and 43° E). Lake Van is the largest soda lake in the world, has a drainage area of 12500 km^2 and is governed by a continental climate. Most precipitation falls during winter in the form of snow and rain and summer is characterized by warm and dry conditions with an average temperature of 20 °C. Detailed study area features can be found in previous studies such as those of Şen et al. (1999, 2000), and Altunkaynak et al. (2003).

Water level has risen by about 2 m in Lake Van during the last decade which has led to flood damage along the shore. Figure 1 shows monthly lake level fluctuations during the 30 years from 1965 to 1994.

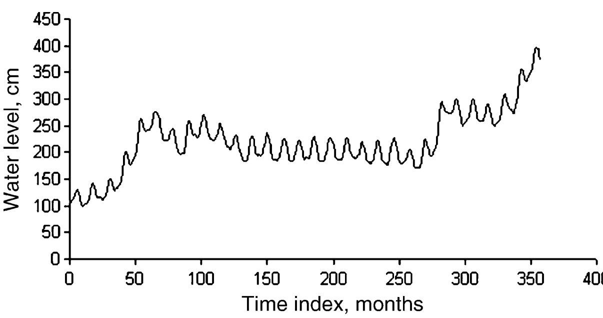

Fig. 1. Water level fluctuation of Van lake

3. Fuzzy logic methodology

A fuzzy inference system is defined as non-linear mapping from a given input space to an output space. Fuzzy Set theory allows the user to capture uncertainties in data. The following steps are necessary for the successful application of fuzzy inference. These are:

(i) Fuzzification is the process of changing crisp input data to fuzzy membership functions such as high, medium, low, heavy, light, hot, warm, big, small, etc. There are three typical membership functions that are commonly used, these are triangular-shaped, Gaussian-shaped, bell-shaped and trapezoid-shaped.
(ii) Construction of fuzzy IF-THEN rules requires expert knowledge and/or measured data. The rules relate the combined linguistic subsets of input variables to the output fuzzy sets.
(iii) The implication part of a fuzzy system is the determination of output weighting factors from the antecedent part.
(iv) Finally, the result is a fuzzy set, and therefore, requires defuzzification to obtain a crisp value, which is required by the administrator or engineer.

A detailed account of fuzzy logic was originally presented by Zadeh (1968). The fuzzy logic approach is applied to various engineering problems (Mamdani, 1974; Pappis and Mamdani, 1977; Ross, 1995; Xiong et al., 2001). Şen (1998) estimated solar radiation from sunshine duration by fuzzy inference system. Şen and Altunkaynak (2004) used fuzzy logic for rainfall-runoff modelling. Altunkaynak et al. (2005a, b) proposed Takagi Sugeno (TS) fuzzy models to predict future monthly water consumption totals for

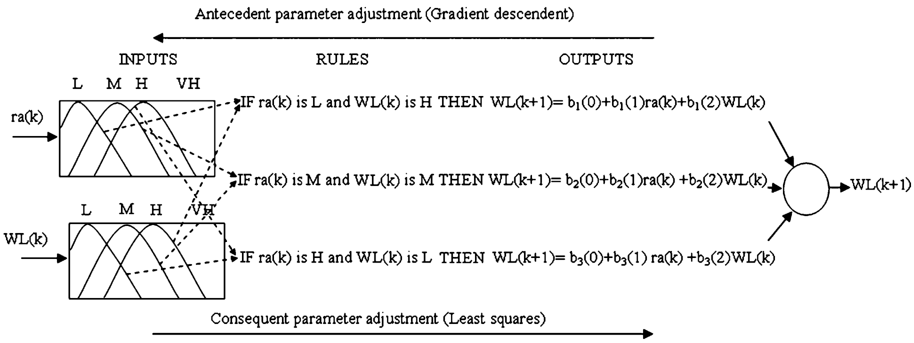

Fig. 2. ANFIS architecture of the input–output system

Istanbul, and dissolved oxygen concentrations in the Golden Horn.

The basis of fuzzy logic is to consider the system states in the form of subsets, each of which is labelled with convenient fuzzy words such as "low", "medium", "big", as previously described. In this way, the variable is considered not as a global quantity but in partial groups, which provide better justification of sub-relationships between two or more variables on the basis of fuzzy words. A small number of fuzzy subgroups leads to unrepresentative predictions whereas a large number leads to many calculations. In practical studies, in most cases, the number of subgroups initially selected is 3 or 4.

There are two main approaches, the Mamdani and the Takagi-Sugeno (TS) approaches, which are frequently used in the application of the fuzzy logic control and forecasting. Those two approaches differ from each other in handling the consequent part. For the Mamdani approach, the outcome of each IF-THEN rule will be a fuzzy set for the output variable, and so at the end of the inference a set of results is obtained. Therefore, the step of defuzzification is indispensable in obtaining crisp values of the output variable. However, in the TS method, the output of each IF-THEN inference rule is a linear function of the input variables rather than a fuzzy set for the output variable. In this paper the TS method is used. Determination of the TS fuzzy model parameters has remained still as a main problem. Jang (1993) proposed a method called as Adaptive Neural Network based Fuzzy Inference System (ANFIS) to estimate the parameters of the membership and the consequent functions. The general scheme of the ANFIS is shown in Fig. 2. Figure 3 displays the TS fuzzy model input–output diagram of the fuzzy inference system. Inputs and their membership functions appear to the left of the fuzzy inference system structural characteristics, while outputs and their membership functions appear on the right.

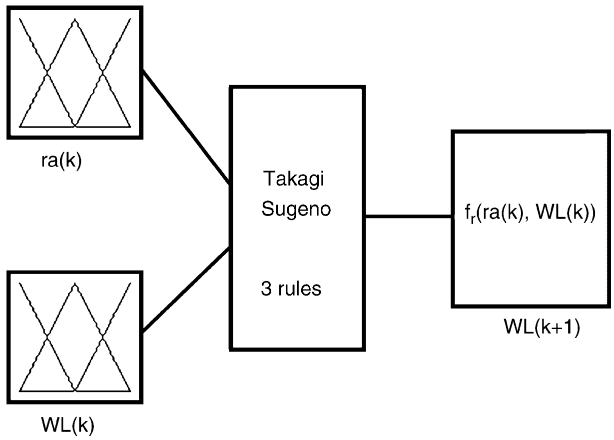

Fig. 3. TS fuzzy model input–output diagram

For the TS fuzzy system (Takagi and Sugeno, 1985), the IF-THEN control rules are given in the form of

$$R_r : \text{IF } (x_1 \text{ is } A_r^{(1)}, x_2 \text{ is } A_r^{(2)}, \ldots, x_p \text{ is } A_r^{(p)}) \\ \text{THEN } y_r = f_r(x_1, x_2, \ldots, x_p)$$

where $A_r^{(i)}$ is a fuzzy set corresponding to a partitioned domain of the input variable x_j in the r*th* IF-THEN rule, p the number of input variables, $f_r(.)$ a function of the p input variables, and y_r is the output of the r*th* IF-THEN inference rule R_r.

If there are $R_r(r = 1, 2, \ldots, n)$ rules in the above mentioned form, the general algorithm of a TS fuzzy inference system would be given as follows.

1. y_i is calculated by the function f_i for each rule R_i.

$$y_r = f_r(x_1, x_1, \ldots, x_p) = b_r(0) + b_r(1)x_1 + \cdots + b_r(p)x_p \quad (1)$$

2. The weights are calculated as,

$$r_r = (m_1{}^r \Lambda\ m_2{}^r \Lambda \cdots m_k{}^r)^* R^r \quad (2)$$

where $m_1{}^r, m_2{}^r, \ldots, m_k{}^r$ denote the α cut-off membership functions relative to input values for the r*th* rule. Λ symbol shows minimum operation and R^r is occurrence probability which taken as 1 in this paper for simplicity.

3. The weights r_r and output of each rule y_r inferred from n implications are used to calculate final output y as,

$$y = \frac{\sum_{r=1}^{n} r_r{}^* y_r}{\sum_{ri=1}^{n} r_r} \quad (3)$$

4. Fuzzy model development

In this work, the TS fuzzy model has been implemented to develop a model that was used to predict changes in water level due to rainfall. In order to apply the TS fuzzy logic approach, it is necessary to divide the data into training and testing parts. Herein, 24 months (1993–1994) are set aside for testing (prediction) whereas all other data values are employed for training. The training set was used to construct a fuzzy rule-based model. Basically, the crucial process of developing the predictive model is to identify the selection of input variables among the available variables. It is well known that effective variable selection can substantially improve model performance and generalization. In general, the model for a multiple input and single output (MISO) system can be written in the following form

$$y(k+1) = f[y(k), y(k-1), \ldots, y(k-n_y+1), \ldots, u_m(k-n_{k_m}), \ldots, u_m(k-n_{k_m}-n_{u_m}+1)] \quad (4)$$

where y(k) is the output vector; u(k) is the input vector with m inputs; n and k denote the number of data samples and the discrete time samples, respectively; n_y and n_{u_m} = maximum lags considered for the output, and input terms, respectively; n_{k_m} represents the pure time delay between change in the inputs and the observed change in the output. Determining the best structure of the model for complex systems is a very difficult task. Here, the identification of variables is not difficult because antecedent rainfall and water levels are taken as inputs, and the future water level is an output. Thus, the next step is to determine the time delay between the change in rainfall and the observed change in water level determined using the cross-correlation technique. It is found that the cross-correlation value was high when $n_{k_m} = 1$. Once the delay is obtained, a selection of the model orders (n_y and n_{u_m}) is needed. The selected regressors were WL(k), which represent the water level at time k, and ra(k) is the rainfall at time k. Mathematically, the MISO fuzzy model of the water level is described by

$$WL(k+1) = f[WL(k), ra(k)] \quad (5)$$

where WL($k+1$) = predicted water level at time k; f (.) = TS fuzzy model; ra(k) refers to the past inputs between a change in rainfall and the observed change in the water level; and WL(k) means that the previous water level has a direct effect on the future water level. The ANFIS method is then used to adjust the parameters of the membership function and the output. After the optimization process is complete, the optimized four membership functions for each input variable are defined as shown in Fig. 4. Three optimized fuzzy rules

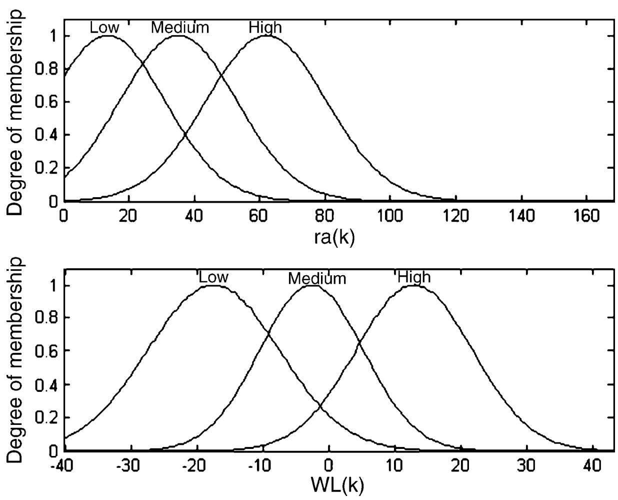

Fig. 4. Membership functions of inputs

Table 1. Fuzzy rule base

Rule	Description
1	IF ra(k) is Low and WL(k) is High THEN WL($k+1$) = 0.233*ra(k) + 0.7613*WL(k) −3.098
2	IF ra(k) is Medium and WL(k) is Medium THEN WL($k+1$) = 0.3858*ra(k) + 0.4635*WL(k) −11.48
3	IF ra(k) is High and WL(k) is Low THEN WL($k+1$) = 0.1274*ra(k) + 0.6043*WL(k) −12.45

with three fuzzy subsets for each variable are extracted and are shown in Table 1.

5. Fuzzy model prediction and interpretation

Lake Van water level records are used for the implementation of the TS fuzzy logic method so as to identify common behaviour of the three variables taken from historical rainfall and water level time series data. The first two variables represent the past rainfall and lake levels and third indicates future lake levels. Hence, the model has three parts, namely, observations (recorded time series) as input, fuzzy model as response, and the output as prediction.

The lake level data show non-stationary behaviour that can clearly be seen in Fig. 1. In order to treat non-stationarity the data are divided into four different sections and then the trend is removed from each. This procedure is explained in more detail is Altunkaynak (2006). Figure 5 exhibits detrended and deseasonalised lake water level fluctuations. After implementing the aforementioned procedure, monthly lake water level fluctuations are ready for the application of the TS fuzzy model.

There are three fuzzy rules for the verbal expression of the relationship between rainfall and lake levels. An example for the inference algorithm is given in Table 2. Given $ra(k) = 71.1$ and $WL(k) = 291$, which is transformed to $WL(k) = -15.549$ when trend and periodicity are removed. The second and third columns of the table indicate the input values of the fuzzy set. Two rules are triggered and the output value is computed from the weighted average of these two output functions. The fifth and sixth columns show the degree of fulfillment (DOF) values, and the seventh column is the result of the multiplication of these values, which correspond to the *prod* operator in fuzzy inference system. Each rule contributes to the result with different weights. Rule 3 has a stronger effect on output than Rule 1. The overall result is computed as $y = (0.0555 + -11.1274)/(0.0340 + 0.8701)$. By using the total values in Table 2, the result is $y = -11.0719/0.9041 = -12.25$. After adding the trend and periodicity, the real value is 297.50 cm.

The parameters shown in Table 1 are small indicating agreement with the model. If the specific rule among the fuzzy rules provides a significant contribution to the water level, the DOF of that rule is closer to 1, if not, the DOF is closer to 0. The DOF of each rule for the input data given is computed in the seventh column of Table 2. The DOF of Rule 3 (0.870) has a higher value, which means a significant contribution to the overall output of a fuzzy model, than Rule 1. It is evident that the output from Rule 3 provides more accurate results.

The prediction results are shown in Table 3 with corresponding absolute errors defined as follows

$$MABE = \frac{1}{n_p}\sum_{i=1}^{n_p} |H_{pi} - H_{mi}| \tag{6}$$

where H_{pi} and H_{mi} are the predicted and measured water level values at month i. Furthermore, fuzzy model results are compared with Auto Regressive Moving Average with Exogenous input (ARMAX) model. The ARMAX (1,1,0,1) type is selected. The fuzzy approach gives 6.60 for MABE whereas for the corresponding ARMAX

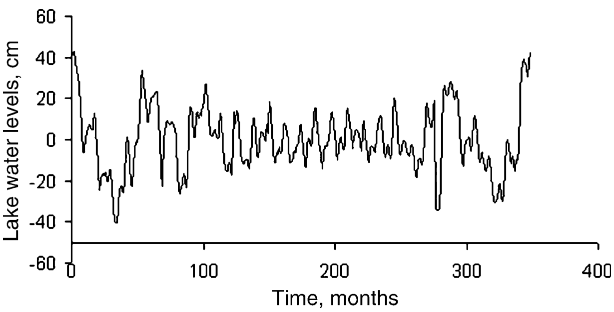

Fig. 5. Detrended and deseasonalised lake water levels

Table 2. A sample calculation for given inputs

R.N	$ra(k) = 71.1$	$WL(k) = -15.549$	$WL(k+1)$	Degree of fulfillment		r_r	$y_r{}^*r_r$
				$ra(k)$	$WL(k+1)$		
R1	Low	High	1.631	0.133	0.256	0.0340	0.0555
R3	High	Low	−12.788	0.887	0.981	0.8701	−11.1274
Total							−12.25

Table 3. Comparison of fuzzy logic and ARMAX model lake water level predictions

Observation (cm)			Prediction (cm)		Absolute error (%)	
ra(k)	WL(k)	WL($k+1$)	Fuzzy	ARMAX (1,1,0,1)	Fuzzy	ARMAX (1,1,0,1)
46.2	257	258	265.7	260.9	7.72	2.92
23.9	258	262	267.9	274.4	5.90	12.36
33.7	262	273	274.0	283.3	1.04	10.33
64.7	273	291	286.0	284.5	5.02	6.54
71.1	291	304	297.5	284.8	6.23	19.23
38.6	304	309	307.7	288.3	1.32	20.73
5.7	309	303	299.9	294.8	3.12	8.17
1.8	303	291	291.1	295.8	0.09	4.78
34.2	291	281	288.5	292.8	7.54	11.79
7.8	281	280	274.3	286.3	5.70	6.27
103.3	280	277	282.8	281.3	5.80	4.29
30.8	277	274	282.9	279.2	8.95	5.18
21.7	274	285	282.4	292.9	2.64	7.90
22.5	285	293	294.5	295.8	1.50	2.76
41.2	293	305	305.8	298.5	0.76	6.47
113.4	305	332	315.7	309.4	16.28	22.61
75.2	332	354	339.2	313.4	14.84	40.56
20.5	354	355	346.8	319.7	8.25	35.30
1.9	355	354	339.6	332.0	14.38	22.00
0	354	343	335.4	339.9	7.59	3.09
2.5	343	333	326.0	334.3	6.96	1.31
52.1	333	335	330.2	332.6	4.76	2.44
137.7	335	340	324.6	327.2	15.43	12.77
Average					2.09	3.67

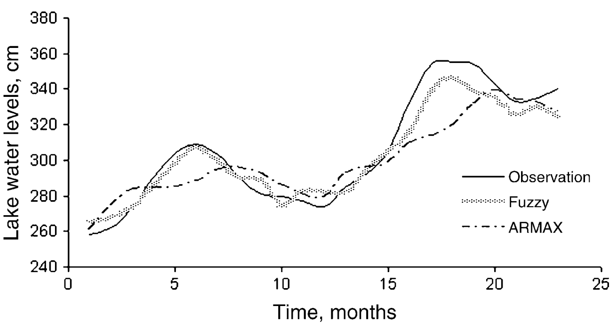

Fig. 6. Observed and predicted lake levels time series for the test data

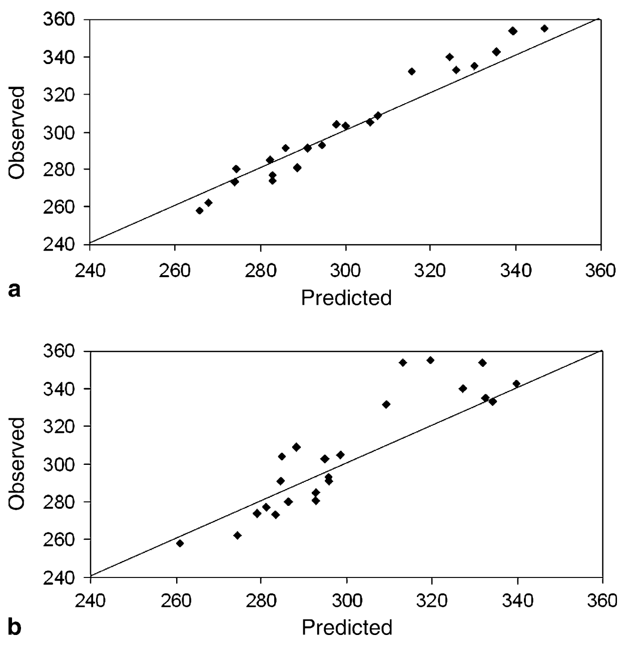

Fig. 7. Model verification for **a**) Fuzzy and **b**) ARMAX

model the value is 11.73. Hence the fuzzy method yields a smaller prediction error and is therefore preferable. Figure 6 indicates the observed and predicted water level (WL($k+1$)) values using the fuzzy model approach. It is obvious that they follow each other very closely and on average observed and predicted lake level series have almost the same statistical parameters.

In order to further show the verification of the TS fuzzy logic approach for lake level predictions, the test data are plotted against the predictions in Fig. 7. It is clear that almost all the points are around the 45° line, and hence the model is not biased. Predictions are successful at low or high values and have a high correlation with observations.

6. Conclusions

Fuzzy logic modelling can easily be used for problems with incomplete, ill-defined and inconsistent information. The TS fuzzy model based on lake water level and rainfall values is presented for the purpose of generating predictions from past records. The model relates historical water levels and rainfall values to future water levels for Lake Van, eastern Turkey. The TS fuzzy model also provides logical rules that can be used for further interpretation. Predictions are obtained for two years (24 months) and the average absolute error obtained is 6.60 cm. The procedure presented in this paper can be used for the prediction of any hydrological variable. In addition, the TS Fuzzy model results were compared with the ARMAX model.

References

Altunkaynak A (2007) Forecasting surface water level fluctuations of Lake Van by artificial neural networks. Water Resources Management 21(2): 399–408

Altunkaynak A, Özger M, Şen Z (2003) Triple diagram model of level fluctuations in Lake Van, Turkey. Hydrol Earth Sys Sci 7(2): 235–244

Altunkaynak A, Özger M, Çakmakcı M (2005a) Water consumption prediction of Istanbul city by using fuzzy logic approach. Water Resources Management 19: 641–654

Altunkaynak A, Özger M, Çakmakcı M (2005b) Fuzzy logic modeling of the dissolved oxygen fluctuations in Golden Horn. Ecol Model 189: 436–446

Box GFP, Jenkins GM (1976) Time series analysis forecasting and control. San Francisco: Holden Day, 560 pp

Degens ET, Kurtman F (1978) The geology of Lake Van. The Mineral Research and Exploration Institute of Turkey, Rep 169, 158 pp

Frederick KD, Gleick PH (2001) Potentional impacts on US Water Resours. In: Claussen E (ed) Climate Change Science Strategres and solutions. Virginal: Pew Center on Global Climate Chage Arlington

Hubert P, Carbomnel JD, Chaouche A (1989) Segmentation des series hydrometeoraloques: application a des series de precipitation et de debits de L'Afrique de L'ouest. J Hydrol 110: 349–367

Iruine KN, Eberthardt AK (1992) Multiplicative seasonal ARIMA models for lake Erie and Lake Ontario water levels. Water Reso Bull 28(3): 385–396

Kadıoğlu M, Sen Z, Batur F (1997) The greatest soda-water lake in the world and how it is influenced by climatic change. Ann Geophysicae 15: 1489–1497

Kempe S, Khoo F, Gürleyik V (1978) Hydrography of lake Van and its drainage area. In: Degen ET, Kurtman F (eds) Geology of Lake Van. The Mineral Research and Exploration Institute of Turkey, Rep 169: 30–45

Kite V (1990) Time series analysis of Lake Erie levels. In: Hurtmann HC, Donalhue MJ (eds) Proc Great Lakes Water Level Forecasting and Statistics Symposium. Ann Arbor, Michigan: Great Lake Commission, pp 265–277

Loucks FD (1989) Modeling the Great Lakes hydrologic-hydraulic system. Ph D Thesis, University of Wisconsin, Madison

Mamdani EH (1974) Application of fuzzy algorithms for simple dynamic plant. Proc IEE 121: pp 1585–1588

Mathier L, Fagherazzi L, Rasson JC, Bobee B (1992) Great Lakes net basin supply simulation by a stochastic approach. INRS-Eau Rapp Scientifique 362, INRS-Fau, Sainte-Foy, 95 pp

Pappis CP, Mamdani EH (1977) A fuzzy controller for a traffic junction. IEEE Trans Syst Man Cybern, pp 7, 10, pp 707–717

Privalsky V (1990) Statistical analysis and predictability of Lake Erie water level variations. In: Hurtmann HC, Donalhue MJ (eds) Proc Great Lakes Water Level Forecasting and Statistics Symposium. Ann Arbor, Michigan: Great Lake Commission, pp 255–264

Ross TJ (1995) Fuzzy logic with engineering applications. USA: McGraw-Hill, 600 pp

Şen Z (1998) Fuzzy algorithm for estimation of solar irradiation from sunshine duration. Solar Energy 63(1): 39–49

Şen Z, Kadıoğlu M, Batur E (1999) Cluster regression model and level fluctuation features of Van Lake, Turkey. Ann Geophysicae 17: 273–279

Şen Z, Kadıoğlu M, Batur E (2000) Stochastic modeling of the Van Lake monthly fluctuations in Turkey. Theor Appl Climatol 65: 99–110

Şen, Z, Altunkaynak A (2004) Fuzzy awakening in rainfall-runoff modeling. Nord Hydrol 35(1): 31–43

Slivitzky M, Mathier L (1993) Climatic changes during the 20th century on the Laurentian Great Lakes and their impacts on hydrologic regime. Deauaille, France: NATO Adv Study Inst

Sneyers R (1992) On the use of statistical analysis for the objective determination of climate change. Meteorol Z 1: 247–256

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

Vannitseni S, Demaree G (1991) Detection et modelisation des secheresses an Sahel-proposition d'une nouvelle methodologie. Hydra Continent 6(2): 155–171

Xiong L, Shamseldin AY, O'Connor MK (2001) A non-linear combination of the forecasts of rainfall-runoff models by the first order Takagi-Sugeno fuzzy system. J Hydrol 245: 196–217

Zadeh LA (1968) Fuzzy algorithms. Inform Control 12: 94–102

Authors' address: Abdüsselam Altunkaynak (e-mail: altunkay@itu.edu.tr), Zekai Şen, Hydraulics Division, Faculty of Civil Engineering, Istanbul Technical University, Maslak 34469, Istanbul, Turkey.