Environmental Sciences Vol.11, No.4 , Winter 2014

21-30

Lag Time Forecasting Using Fuzzy Regression Based Formulae

Moslem Heydari,[1*] Seyed Javad Sadatinejad[2] and Afshin Honarbakhsh[3]

[1] *PHD. Student of watershed management, Faculty of Natural Resources and Earth Sciences, University of Shahrekord*
[2] *Associate Professor, Department of, Faculty of New Sciences and Technologies, University of Tehran*
[3] *Associate Professor, Department of watershed management, Faculty of Natural Resources and Earth Sciences, University of Shahrekord*

Received: 2014/0/0 Accepted: 2014/0/0

Abstract

River basin lag time is an important factor in the linear modelling of river basin response. In this study, the modelling of lag time using fuzzy regression is applied. For this purpose, the data for rainfall-runoff events of Khanmirza basin (nine events) were collected and analysed. Following on, events were divided into two groups: one for formulas based on fuzzy regression and another for the validation of these formulas. The results obtained from this study, based on RE and RMSE statistical measures, showed that the efficiency of newly developed formulas based on fuzzy regression methods is higher than for other formulas used for the calculation of time of concentration.

Key words: Lag time, Fuzzy Regression, Hydrology.

پیش بینی زمان تاخیر با استفاده از روابطی بر مبنای رگرسیون فازی

مسلم حیدری[۱*]،سید جواد ساداتی نژاد[۲] و افشین هنربخش[۳]

[۱] دانشجوی دکتری آبخیزداری، دانشکده منابع طبیعی و علوم زمین، دانشگاه شهرکرد، شهرکرد
[۲] دانشیار گروه، دانشکده علوم و فنون نوین، دانشگاه تهران، تهران
[۳] دانشیار گروه آبخیزداری، دانشکده منابع طبیعی و علوم زمین دانشگاه شهرکرد، شهرکرد

چکیده

زمان تاخیر حوزه آبخیز یک فاکتور مهم در مدلسازی واکنش حوزه آبخیز است. در این مطالعه مدلسازی زمان تاخیر با استفاده از رگرسیون فازی انجام شده است. برای این منظور، داده های بارش - رواناب حوزه آبخیز خانمیرزا جمع آوری و تحلیل شدند. سپس وقایع (۹ واقعه) به دو گروه شامل یک گروه برای ساخت روابطی برمبنای رگرسیون فازی و گروه دیگر برای ارزیابی این روابط تقسیم شدند. نتایج بدست آمده از این تحقیق برمبنای معیارهای آماری خطای نسبی و ریشه میانگین مربع خطا نشان دادند که روابط جدید توسعه یافته برمبنای رگرسیون فازی نسبت به دیگر روابط موجود برای محاسبه زمان تاخیر مناسب تر هستند.

کلمات کلیدی: زمان تاخیر، رگرسیون فازی، هیدرولوژی.

* Corresponding author. *Email Address*: heydari_moslem@yahoo.com

1. Introduction

Lag time is a parameter that often appears in theoretical and conceptual models of river basin behaviour. From the point of view of hydrology, river basin lag time (T_l) is defined as the time distance from the centre of mass excess rainfall to the peak of the hydrograph [15] and is a variable that often used in hydrograph analysis. This variable indicates the hydrological response of river basin to a rainfall event.

The shape of a flood hydrograph is primarily a function of the geometry of the river basin. This is due to the fact that the physical parameters of the river basin all affect outlet runoff of the river basin. Where time distance between rainfall occurrence and runoff generation is small the flood hydrograph displays a small lag and a high, sharp peak discharge; otherwise, the flood hydrograph displays a longer lag and a lower, broader peak discharge. Also, it should be noted that urbanization can affect the shape of the flood hydrograph by increasing the area of impermeable substrate and by reducing the amount of soil infiltration. These changes typically produce a shorter lag and a higher and steeper peak discharge [17, 2, 20, 7, 22].

For estimating lag time, many empirical formulae have been proposed in the literature. Usually, these formulae compute the basin lag as a power function of the basin physical parameters [8, 15, 2, 12, 14].

For small natural drainage basins with simple drainage patterns, the lag time may be very close to the time of concentration. However, it is sometimes difficult to measure the lag time in real world situations [13]. Because many of watersheds are unguaged and there are no data available.

Lag time on the basin is frequently expressed as the mean because it was effective by many factors in the basin (climatic and physical) and changes in any one factor can lead to changes in the basin and eventually result in flooding. This feature matches with fuzzy possibilistic regression. Thus, it is proposed to consider lag time from the viewpoint of fuzzy logic and a method for calculating it using fuzzy sets theory. Today, fuzzy logic has achieved worldwide acceptance and, recently, the fuzzy regression method has been used to reconstruct hydrologic data. First a fuzzy linear regression model with symmetric triangular fuzzy parameters is introduced using linear programming (LP) [24]. Since the membership functions of fuzzy sets are often described as possibility distributions, this approach is usually called ‘possibilistic regression analysis’. The stage-discharge relationship has been determined using fuzzy logic and its results were more accurate than traditional methods and even than artificial neural networks method [10].

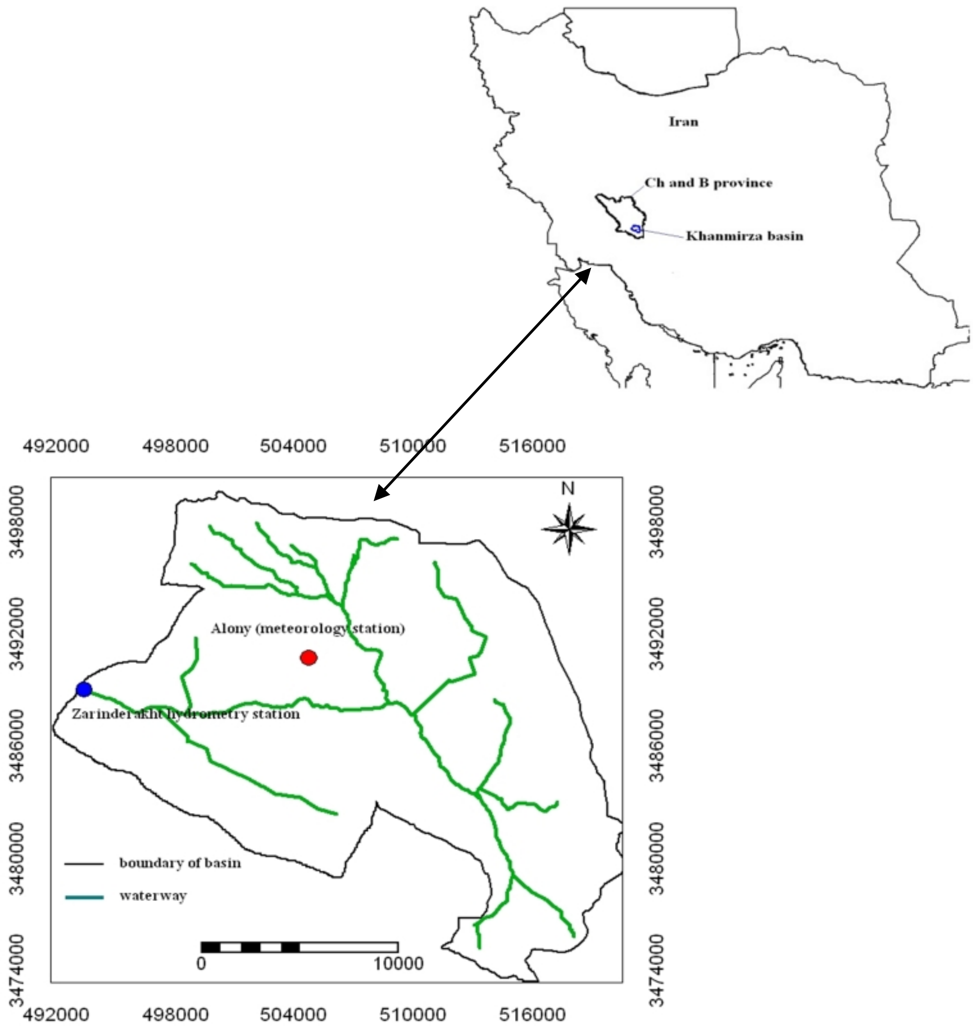
The methods of normal ratio, artificial neural networks and fuzzy logic have been compared for reconstruction of precipitation data in North of Italy. The results showed that fuzzy logic made fewer errors in comparison with the two other methods [1]. An ordered impressionistic fuzzy analysis that has the capability of simulating the unknown relations between a set of meteorological and hydrological parameters has been illustrated [19]. Recently, a new fuzzy method of evaluating hydrological measures in order to deal with uncertainty and impreciseness, based on Buckley’s estimation method, has been developed [21].

Several linear and non-linear models for centralized remote-control systems that can support decision-making by semi-intensive aquaculturists concerning the inflow rates to the ponds have also been evaluated. Results obtained indicated that, in spite of the results being statistically significant, the explained variance levels obtained indicate how difficult it is to capture the aquaculturist’s experience and knowledge concerning the operation of water exchange in their ponds for maintaining the water quality in these production systems [4].

The successful applications of fuzzy logic and its rapid growth suggest that the impact of fuzzy logic will be felt increasingly in coming years. Fuzzy logic is likely to play an especially important role in science and engineering but, eventually, its influence may extend much further. The purpose of this study is the modelling of lag time using fuzzy regression.

2. Materials and Methods

This study was conducted in Khanmirza basin in Iran and was limited to 50°54'50" to 51°12'59" eastern longitude and 31°23'40" to 31°37'53" northern latitude. Using data from a recording float type rain gauge at meteorological stations within the basin, nine events indicating excess and prolonged rainfall, which yielded large floods, were selected (see Figure 1).



Figureure 1. Position of the basin and stations studied.

The hydrograph for each event was obtained using data from the hydrometric station at the basin. Discharge flows were obtained from the stage-discharge relationship produced by the limnograph at this station (see Figures 2 and 3).

In order to determine the time between a mass of excess rainfall, the Φ index (a loss function whose value results in a volume of direct run-off equal to that measured and distributed uniformly across the hydrograph) was calculated and superimposed on the histogram (see Figures 4 and 5). When using this method, the hydrograph should be analysed, excess rainfall determined and an excess rainfall histogram should be produced. Finally, using the hydrograph method, the estimated lag time of events was calculated.

Additionally, using GIS the physical parameters of the basin were computed. These parameters are shown in Table1 below.

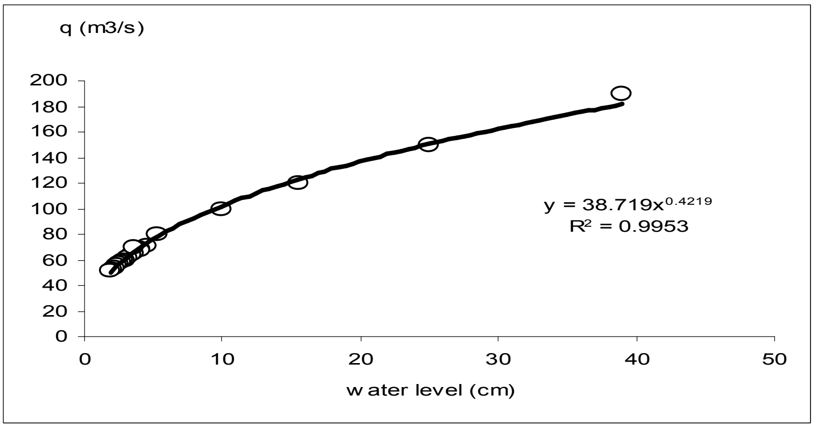


Figure 2- A sample stage - discharge relationship from the basin.

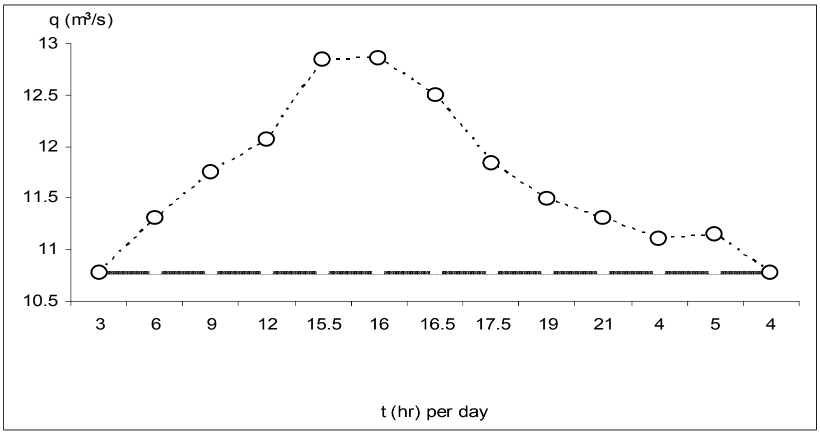


Figure 3- A sample hydrograph from the basin.

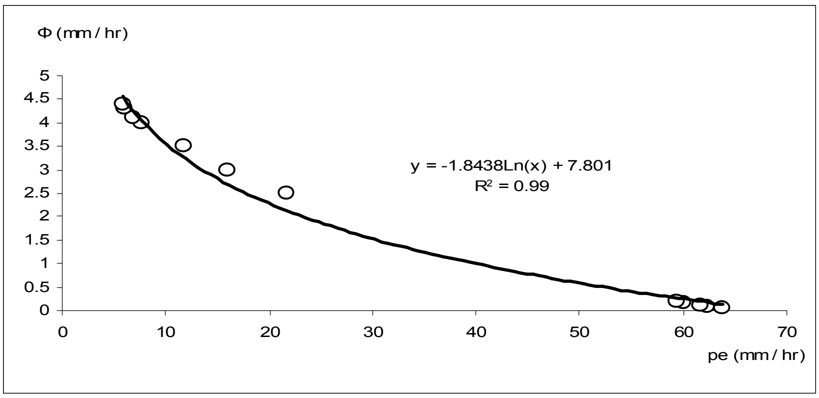


Figure 4. The relationship between the Φ index and excess rainfall (pe) from the basin.

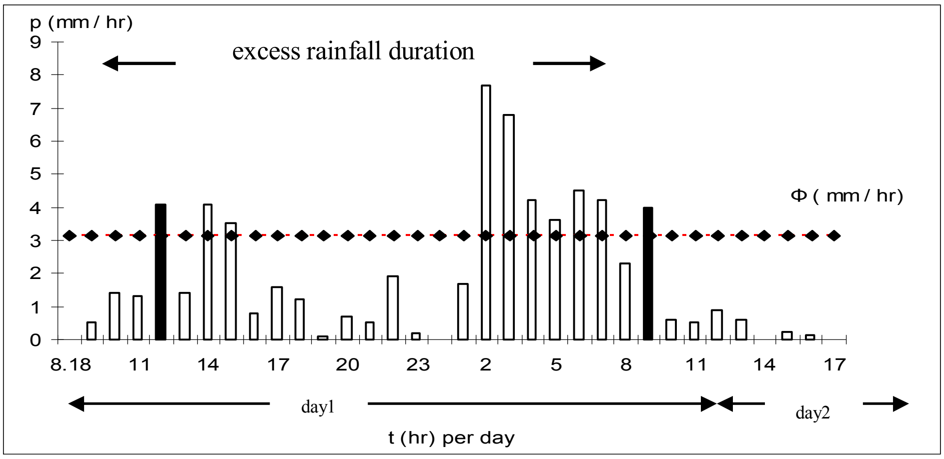


Figure 5. Determination of excess rainfall using the Φ index from the basin.

Table 1- Physical parameters of basin.

Parameters	Value
Area (km2)	390
Perimeter (km)	110
Length of longest flow path (m)	39597
Length of main waterway (m)	39320
Diameter of circle with area equal to area of basin (km)	22
Equivalent diameter (De) or diameter of circle whose perimeter is equal to perimeter of basin (km)	35
Basin length (m)	29618
Basin circularity	0.404
Weighted slope of main water way (%)	2.12
Weighted slope of basin (degree)	15.51
Difference between min and max elevation for basin (m)	1262
Mean basin elevation (m)	1956
Length from outlet to the centroid of basin (km)	12.72
Bifurcation ratio	4.22
Difference between outset and end of main waterway (m)	835

Fuzzy regression

A fuzzy regression (FR) is an input-output relationship in which the input or the output, or both, are fuzzy numbers. Unlike classic linear regression, in which the parameters are assumed to be random variables with probability distribution functions, in a fuzzy regression, the coefficients are subject to the possibility theory. The distribution of these amounts is determined as the membership function.

The fuzzy set theory is a powerful method for analysing statistical data, which includes ambiguity or vagueness as a result of the structure of a process, measurement systems or environmental conditions. The main purpose of fuzzy regression models is to find the best model with the least scope for error. One of the components of fuzzy systems is fuzzy regression. On the whole, it is possible to use fuzzy regression in the following conditions:

1) Insufficiency of the amount of observed data
2) Ambiguity concerning the relationship between dependent and independent variables
3) Inaccuracy of linear theories
4) Failure of errors to comply with normal distribution

Various types of fuzzy regression models have been introduced in the literature and many different methods proposed for estimating fuzzy parameters. This research applied fuzzy possibilistic regression.

The possibilistic approach seeks to minimize the overall fuzziness of the model by minimizing the total spreads of its fuzzy coefficients, subject to including the data points of each sample within a specified feasible data interval. The fuzzy numbers discussed in this paper are assumed to be symmetrically triangular.

In one of the fuzzy possibilistic regression models, the coefficients of regression are fuzzy, while observed input and output are not fuzzy. In this research, the model was applied with the following equation:

$$\tilde{y} = \tilde{A}_0 + \tilde{A}_1 x_1 + \tilde{A}_2 x_2 + \tilde{A}_3 x_3 + \ldots.. + \tilde{A}_n x_n \quad (1)$$

where coefficients of $\tilde{A}_0$, $\tilde{A}_1$,……., $\tilde{A}_n$ are fuzzy numbers , x_1 x_2 x_3 ,……., x_n are independent variables expressed as classic numbers, $\tilde{y}$ is the dependent variable as a fuzzy number and n is the number of variables. It was assumed that m rows were observed with n input variables and one output variable in each row. For a fuzzy number, such as the symmetrical triangle in Figure 6, the membership function is written as the following equation:

$$\mu_{\tilde{A}}(a_i) = \begin{cases} 1 - \dfrac{|P_i - a_i|}{C_i} & P_i - C_i \le a_i \le P_i + C_i \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Where c_i and p_i are the width and centre of the fuzzy number, respectively.

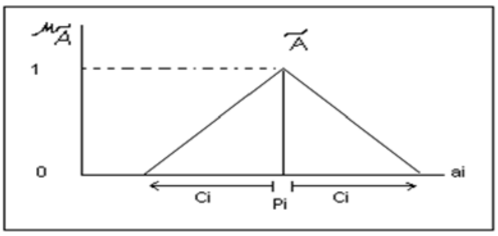


Figure 6. Membership function of a fuzzy number as a symmetrical triangle.

$\tilde{A}$ in Equation 2 is used to indicate the "almost equal to p_i " amount and c_i indicates its degree of being fuzzy, a concept which can be shown as $\tilde{A}_i = (p_i, c_i)$. Thus, the fuzzy regression equation is as follows:

$$\tilde{y} = (p_0, c_0) + (p_1, c_1)x_1 + (p_2, c_2)x_2 + \ldots.. + (p_n, c_n)x_n \quad (3)$$

The membership function of the output fuzzy variable is represented as follows:

$$\mu_{\tilde{y}} = \begin{cases} \max(\min\left[\mu_{\tilde{A}i}(a_i)\right]) & \{a \mid y = f(x,a) \neq \phi\} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Substitution of Eq. 2 in Eq. 4 gives us following equation:

$$\mu_{\tilde{y}}(y) = \begin{cases} 1 - \dfrac{y - p_0 - \sum_1^n p_i x_i}{c_0 + \sum_1^n c_i |x_i|} & x_i \neq 0 \\ 1 & x_i = 0, y = 0 \\ 0 & x_i = 0, y \neq 0 \end{cases} \quad (5)$$

Different algorithms exist for solving the fuzzy linear regression problem, which in this research employed the linear programming problem. With this method, the fuzzy linear regression model changes to a linear programming problem. In this case, the purpose of the regression model is to determine the optimum amounts for $\tilde{A}$ as the membership degree of the fuzzy output variable, which should be greater than a given amount of h as determined by the user. In other words, the following inequality should be correct for m rows of data (j =1, 2, 3, …, m):

$$\mu_{\tilde{y}j}(y_i) \ge h \quad (6)$$

By increasing h , the degree of output fuzziness increases, too. Considering Equation 6, the centre and width of the membership function are equal to $p_0 + \sum_1^n p_i x_i$ and $c_0 + \sum c_i x_i$ respectively.

For determining the regression coefficients, the fuzzy output width is minimized for all data sets. Thus, the subjective function and constrains of linear programming can be represented as below.

Subjective function:

$$\text{Minimize} \quad : \text{mc}_0 + \sum_{j=1}^{m} \sum_{i=1}^{n} c_i \left| x_{ij} \right| \tag{7}$$

Constraints:

$$p_0 + \sum p_i x_{ij} - (1-h)\left[c_0 + \sum c_i x_{ij}\right] \le y_j \tag{8}$$

$$p_0 + \sum p_i x_i + (1-h)\left[c_0 + \sum c_i x_{ij}\right] \ge y_j \tag{9}$$

The above constraints were obtained by substituting Equation 5 into Equation 6. Thus, in order to solve a linear regression problem with fuzzy coefficients and non-fuzzy data, we needed to solve a linear programming model based on equations (7), (8) and (9). Equations 8 and 9 were written separately for each of the observed data pairs. Thus, based on the mentioned equations, the numbers for $2m$ inequalities were established. This was conducted using hydro generator software [17]. The set inequalities were entered into Lingo software and the coefficients of p_i and c_i were obtained. Following on, they were de-fuzzified by the centre of gravity method in order to obtain crisp regression coefficients. The centre of gravity method was based on the following equation:

$$A = \frac{\int \mu\ (x).x.dx}{\int \mu\ (x).dx} \tag{10}$$

Many de-fuzzification approaches have been proposed, among which the centre of gravity (COG) method, otherwise known as the centroid method, shows the most trivial weighted average and has distinct geometrical meaning. In this paper, the COG method was adopted for de-fuzzifying the fuzzy regression coefficients into crisp coefficients.

3. Results and Discussion

Using the hydrograph method, all events were analysed for calculation of the real T_l's, which were then employed to produce and test the new formulae. As stated above, nine events were available for producing and testing the new formulae. Six of these events were randomly selected for creating the formulae and the remaining three were used for testing them. Table 2 presents the observational values of the lag time.

Table 2- Observational values of T_l

	start		T_l (hour)
	date	time	
Events	8/11/1994	3 : 0	11.5
	29/3/1995	22 : 0	9.25
	23/4/1995	24 : 0	8.5
	7/3/1996	24 : 0	10.5
	12/3/1996	17 : 0	10.5
	26/2/1998	22 : 0	9.63
	13/3/1998	13.5 : 0	10.45
	26/3/1998	00 : 45	9.4
	30/3/1998	10 : 0	10

In this article, a new physical parameter is suggested for use in hydrological modelling. This parameter, the equivalent diameter (D_e), is the diameter of the circle whose perimeter equals that of the basin [5]. It is worth noting that another parameter, the elongation ratio, has already been introduced by Schumm (1956) and corresponds to the diameter of a circle whose area is the same as that of the basin. D_e was obtained with the following equation:

$$D_e = 0.3183954 \times P \tag{11}$$

Where:

P = perimeter of basin (km)

In order to develop the formulae for estimating T_l, the relationship between the real T_l and the physical parameters of the basin needed to be investigated. Using the determination coefficient (R^2), the main river length and the equivalent diameter, as defined above, were found to be the most significant investigated physical parameters for this purpose.

According to the fuzzy nature of the time of concentration and using the hydro generator and Lingo software, three formulae were developed for estimating the time of concentration and its minimum, mean and maximum values, as shown below:

$$Tl_{avg} = \sqrt{(0.000241 \times L) \times (0.271428 \times D_e)} \quad (12)$$

$$Tl_{\min} = \sqrt{(0.000241 \times L) \times (0.271428 \times D_e)} - 0.09532 - 1 \quad (13)$$

$$Tl_{\max} = \sqrt{(0.000241 \times L) \times (0.271428 \times D_e)} + 0.09532 + 1 \quad (14)$$

Where:

Tl = lag time (hr)

D_e = obtained from equation 11 (km)

L = length of main water way (m)

The bound width of fuzzy membership function was 0.09532 and 1 was the reduced cost, as the amount that the objective coefficient of the variable would have to improve before it would become profitable to give the variable in question a positive value in the optimal solution (in the maximization and minimization problem). The fuzzy regression-based formulae were capable of estimating the maximum and minimum values of lag time on the basis of real observed values.

It should be noted that according to the SCS, for many cases, the lag time can be related to the concentration time and the relationship-relating lag time and time of concentration can be used to compute lag time using the following formula:

$Tc = 1.67\ T_l$ (15)

Where:

Tc = time of concentration (hr)

Tl = lag time (hr)

Based on the SCS method's applied methodology for deriving the time of concentration and lag time equation, a new formula for estimating lag time was developed. This new formula can be written as follows:

$T_L = 0.8849.(Tc^{0.9683})$ (16)

Where:

T_L = lag time of basin (hour)

Tc = time of concentration (hour) obtained from equation 16 as follows:

$Tc = 0.1244(D_e^{0.3151}).(L^{0.3205})$ [5] (17)

Where:

Tc = time of concentration of basin (hour)

D_e = equivalent diameter□(km^2) that obtained from equation 11 above, and L is main river length in meter.

In order to evaluate and test these formulae, the relationship between the calculated T_l and the real values were compared. Statistical analysis was carried out and the results showed a RMSE of 1.6375, 4.9184 and 0.7561 and RE of 9.182%, 19.2% and 7.3244% for formulae 12, 13 and 14, respectively. Considering the results of the statistical analysis, it can be concluded that the formulae proposed in this work are capable of estimating lag time in a satisfactory manner. In applying these formulae, special attention should be given to regional physical characteristics. In Figure 7, a comparison between the predicted T_l's obtained from fuzzy formulae and the real observed T_l's is shown.

4. Conclusions

According to several applications, such as creating lag structures in order to postpone the peak discharges related to different branches, there is a need for estimating the minimum possible lag time in order to calculate the most dangerous scenarios. Considering the ability of such formulae (the maximum and minimum fuzzy estimate) under these conditions, the low fuzzy formula (Tcmin) can be said to be the best formula for computing lag time.

It should be noted that the application of these models in other basins depends on evaluating of them in each basin. Nonetheless, fuzzy regression analysis has been demonstrated to be a powerful methodology for analysing the vague relationship between a dependent variable and independent variables.

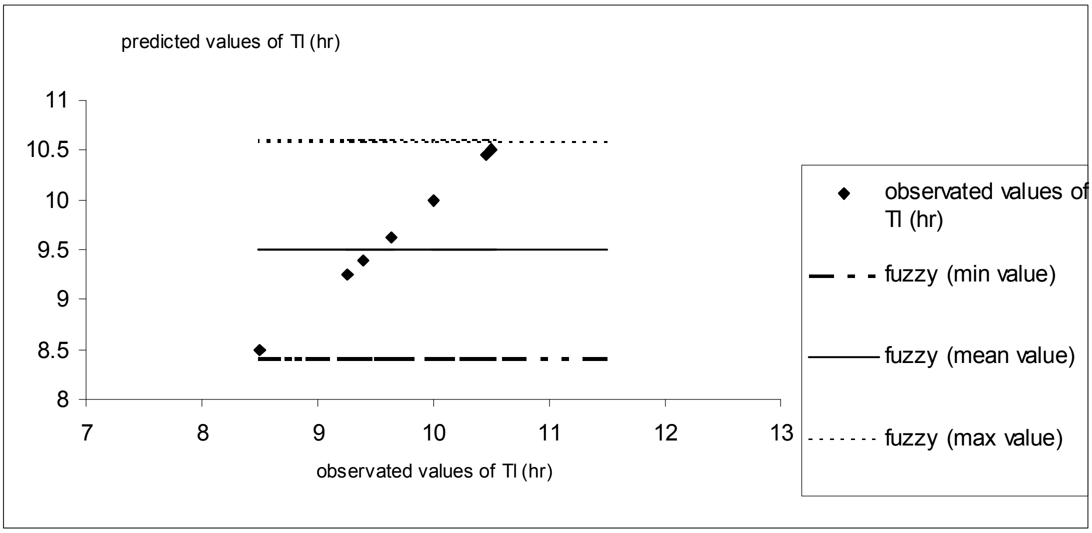


Figure 7. Comparison between predicted T_l's obtained from fuzzy formulae and the real observed T_l's.

References

[1] Abeb, A. J., Solomatine, D. P. and Vennerker, R. G. W. (2000) Application of adaptive fuzzy rule based methods for reconstruction of missing precipitation events. *Hydrological Science Journal*, 45(3):425-436.

[2] Anderson, D. G. (2001) Effects of Urban Development on Flood in Northern Virginia. U.S.G.S Water Supply Paper C.

[3] Gray, D.M. (1961) Synthetic unit hydrographs for small watersheds. Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers, (4):33-54.

[4]Gutierres Estrada, J. C., Pulido Valvo, I., Rosa, I. and Marchini, B. (2011) Modelling inflow rates for the water exchange management in semi-intensive aquaculture ponds. *Aquacultural Engineering*, (48):19-30.

[5] Heydari, M.(2010) Revision of Some Empirical Models for Estimation of Time of Concentration and Survey the Efficiency of Fuzzy Regression Model to Estimate the Time of Concentration in North Karoon River Basin. Msc. Thesis, Shahrekord University (in Persian).

[6]http://www.emsi.com/wmshelp/Hydrologic_Models/Calculators/Computing_Travel_Times/Using_Basin_Data/Equations/Overview_of_Basin_Data_Equations.htm, (assessed: March 31, 2010).

[7] Laenen, A. (1980) Storm Runoff as Related to Urbanization in the Portland, Oregon-Vancouver, Washington Area. U.S.G.S, Water Resource Investigations Open File Report, pp. 80-689.

[8] Laureson, E. M. (1964) A catchment storage model for runoff routing. *Journal of Hydrology*, (2):141-163.

[9] Leopold, L. B. (1968) Hydrology for Urban Land Planning. A Guidebook on the Hydrologic Effect of Urban Land Use. U.S.G.S Circular, at p. 554.

[10] Lohani, A. K., Goel, N. K. and Bhatia, K.S. (2006) Takagi-Sugeno Fuzzy Inference System for Modelling Stage-Discharge Relationship. *Journal of Hydrology*, 333:146-160.

[11] Matreja, K.N. (1990) Applied Hydrology. JaTa McGraw Hill Publishing Company Limited, at p. 204.

[12] Ragan, R. M. and Duru, J O. (1972) Kinematic wave nomograph for time of concentration. *Journal of the Hydraulics Division, Proceedings of the American Society of Civil Engineers*, 98(10):1765-1771.

[13] Ramirez, J. A. (2000) Prediction and Modelling of flood Hydrology and Hydraulics. Chapter 11 of Inland Flood Hazards: Human, Riparian and Aquatic Communities Ed. Ellen Wohl. Cambridge University Press at p. 307.

[14] Rao, A. R. and Delleur, J. W. (1974) Instantaneous unit hydrograph, peak discharges and time lags in urban areas. *Hydrological Sciences Bulletin*, 19(2):185-198.

[15] Rastogi, R.A. and Jones, B.A. (1969) Simulation and hydrologic response of a drainage net of a small agricultural drainage basin. *Transactions of the American Society of Agricultural Engineers*, (12): 899-908.

[16] Roussel, M. C., Thompson, D. B., Fang, X., Cleverland, T. G. and Garcia, C. A. (2005) Time Parameters Estimation for Applicable Texas Watersheds. Department of Civil Engineering, Lamar University, Beaumont, TX: Research Project Summary Report 0-4696-s.

[17] Sadatinejad, S. J. and Shayannejad, M. and Honarbakhsh, A. (2009) Investigation of the Efficiency of the Fuzzy Regression Method in Reconstructig Monthly discharge Data of Hydrometric Stations in Great Karoon River Basin. *J. Agr. Sci. Tech*, (11):111-119.

[18] Sanchez, J.D. and Gomez A.T. (2003) Applications of fuzzy regression in actuarial analysis. Journal of Risk and Insurance, 70(4):797-802.

[19] Saneifard, R. and Ahmadianrad, H. (2012) On the overflow alert system based on fuzzy models. Current research journal of biological sciences, 4(2): 198-201.

[20] Schulz, E.F. and Lopez, O.G. (1974) Determination of urban watershed response time. Hydrology Paper No. 71, Colorado State University, Fort Collins, CO.

[21] Shamskia, N., Rahmati, S. H., Haleh, H. and Rahmati, S.H. (2011) A new approach of fuzzy methods for evaluating of hydrological data. *World Academy of Science, Engineering and Technology*, 5(5): 1051-1059.

[22] Singh, V. P. and Agiralioglu, N. (1982) Lag time for diverging overland flow. *Nordic Hydrology*, (13):39-48.

[23] Sudharsanan, R. M., Krishnaveni, R., and Karunakaran, K. (2010) Derivation of Instantaneous Unit Hydrograph for a sub-basin using Linear Geomorphological Model and Geographic Information Systems 77710. *Journal of Spatial Hydrology*, 10(1): 30-40.

[24] Tanaka, H., Uejima, S. and Asai, K. (1982) Linear regression analysis with fuzzy model. *IEEE Transactions on Systems Man, Cybern*, 12(6):903-907.

[25] Thompson, D.B., Fang, X. and Cleverland, T.G. (2004) Literature Review on Time Parameters for Hydrographs. Department of Civil Engineering, Lamar University, Beaumont, TX.