Desalination and Water Treatment
www.deswater.com

doi: 10.5004/dwt.2019.24549

167 (2019) 385–397
November

A hybrid fuzzy frequency factor based methodology for analyzing the hydrological drought

Christopher Papadopoulos, Mike Spiliotis*, Panagiotis Angelidis, Basil Papadopoulos

Department of Civil Engineering, School of Engineering, Democritus University of Thrace, Kimmeria Campus, 67100 Xanthi, Greece, Tel. +302541079613; emails: mspiliot@civil.duth.gr (M. Spiliotis), cpapadp@civil.duth.gr (C. Papadopoulos), pangelid@civil.duth.gr (P. Angelidis), papadob@civil.duth.gr (B. Papadopoulos)

Received 28 February 2019; Accepted 11 June 2019

ABSTRACT

In this work, a hybrid fuzzy-probabilistic approach is proposed in order to classify the hydrological drought. The analysis focuses on the annual cumulative discharge which is considered to be a random variable. Based on a fuzzified version of the frequency factor method, the fitting between the empirical probabilities and the theoretical probability distribution is investigated with the assumption of Log-Normal and Log-Pearson III. This fitting is achieved by using Tanaka's fuzzy linear regression and hence, all the observed probabilities are included within the produced fuzzy band. Furthermore, a modified fuzzy regression model is also applied. An assumption of the mean value and the standard deviation regarding the log-transformed data can be simultaneously achieved based on the theoretical density probabilities and the sample. Based on the achieved fuzzy frequency factor curve, the fuzzy cumulative annual discharge which corresponds to each threshold of drought can be determined. In order to classify the intensity of hydrological drought, an ascending procedure is proposed by comparing the existing annual cumulative discharge and the fuzzified thresholds of the drought categories. The proposed methodology is applied in the case of the Evros River.

Keywords: Classification of drought; Fuzzy least square regression; Fuzzy linear regression; Log-normal probability distribution; Log-Pearson III probability distribution; Frequency factor; Evros River

1. Introduction

Drought must be considered as a relative condition, rather than an absolute condition. It occurs in both high and low rainfall conditions [1]. In other words, drought occurs when the water availability is below the canonical values which very often are described by the mean value and the standard deviation. Several types of droughts exist, while in this work the hydrological drought is studied.

Mostly, the phenomenon of drought is considered through drought indices aiming at the estimation of drought intensive, however, a few of drought indices have found generalized application [2–4].

The Standardized Precipitation Index, known as SPI, seems to be the most widely-used compared with the existing simple indices to classify the drought events [5]. In brief, the computation of the SPI involves the fitting of gamma probability density function and thereafter the cumulative probability distribution is transformed into the standard normal distribution to yield the SPI [5]. The drought categories which are defined according to SPI are also used to other similar drought indices. Hence, starting initially from a probabilistic approach, many standardized drought indices as SPI [5], RDI [6], SDI [7,8] and others, conclude to an index which in fact, is the standardized normal variable Z.

Severity, temporal and spatial extent of drought are the basic magnitudes of drought phenomenon, which do not

* Corresponding author.

Presented at the 3rd International Conference on Insights on the Water-Energy-Food Nexus (EWaS-3), 27–30 June 2018, Lefkada Island, Greece

have standardized distribution. For this reason, the fitting between the empirical probabilities of a historical sample and the probability density function of population is a crucial point. Several probability distributions have been used to investigate the most proper fitting to the data by using statistical test of suitability. For instance, in meteorological drought analysis, historical sequences of precipitation have been adopted by Yusof [9] on assumptions of exponential, gamma and Weibull distributions. In another research work suggested by Sharma and Panu [10], the river flow data are assumed to follow Pearson Type III distribution in order to predict the return periods of hydrological drought.

Despite the fact that there are several research works on the characterization of drought events in a probabilistic and stochastic manner, the exact derivation of the probabilistic structure of drought characteristics is still an open issue [11]. In this work, the examined hydrological variable is the annual cumulative discharge. First, a hybrid fuzzy probabilistic approach is proposed in order to improve the couple between the observed probabilities and the adopted theoretical probability distribution. Second, based on the widely-used standardized normal thresholds to drought, the corresponding (fuzzy) annual cumulative discharge thresholds are determined. Third, the (crisp) observed cumulative annual discharge is compared with these fuzzy thresholds in order to classify the drought.

Compared with the work by Spiliotis et al. [12], in this work, the log-Pearson III probability distribution is also examined. In addition, a modification of the widely used fuzzy regression model is proposed. Finally, in order to check the proposed approach several additional measures of suitability are proposed.

2. Proposed methodology

2.1. Fundamentals of fuzzy sets and logic

A fuzzy set A on a universe set X is a mapping $A{:}X \to [0,1]$, assigning to each element $x \in X$ a degree of membership $0 \le A(x) \le 1$. The membership function $A(x)$ can be also presented as $\mu_A(x)$.

If A is a fuzzy set, and any number $\alpha \in [0,1]$, by the α-cut, $A[\alpha]$, and the strong α-cut, $A[\alpha]^+$, the crisp sets [13] are defined, respectively:

$$A[\alpha] = \{x \in X : A(x) \ge \alpha\} \tag{1}$$

$$A[\alpha]^+ = \{x \in X : A(x) > \alpha\}_{(\text{strong } \alpha-\text{cut})} \tag{2}$$

The 0-cut can be defined as follows:

$$A[0]^+ = \{x \in X : A(x) > 0\} \tag{3}$$

It is worth noting that by using the α-cut concept we can move from the fuzzy sets to the conventional crisp mathematical methodologies.

A special kind of fuzzy sets is the fuzzy numbers. In this work, fuzzy symmetric triangular numbers are used which are special kinds of fuzzy numbers. The fuzzy symmetric triangular numbers have the following membership function:

$$\mu_A(x) = \begin{cases} 1 - \dfrac{|x-a|}{w}, & \text{if} \quad a - w \le x \le a + w \\ 0, & \text{otherwise} \end{cases} \tag{4}$$
$$w > 0$$

in which a is the centre and w is the spread of the fuzzy number (Fig. 1).

2.2. Observed probabilities and the frequency factor method

Let us study an historical sample. The rank order method involves ordering the data from the largest hydrological value to the smallest hydrological value, assigning a rank of 1 to the largest value and a rank of N to the smallest value. Based on the Weibull [14] empirical distribution is used to compute the plotting position probabilities as follows:

$$P(Q \ge q) = \frac{m}{N+1} \tag{5}$$

Therefore, the cumulative probability of non exceedance probability can be determined as follows [15]:

$$P(Q < q) = 1 - \frac{m}{N+1} \tag{6}$$

Chow [16] suggested a mathematical expression for determining the value of a random hydrological variable based on the adopted probability distribution function.

Hence, a linear relationship between the examined hydrological variable and corresponding values of frequency factor K_T, which is related with the adopted theoretical probability distribution and the return period, can be determined.

$$V_T = \mu + \sigma K_T \tag{7}$$

where μ and σ are the mean value and the standard deviation correspondingly, K_T is the frequency factor.

A used technique in hydrology is to investigate a linear relation between the frequency factor and the hydrological variable instead of the probability plot [17]. Therefore,

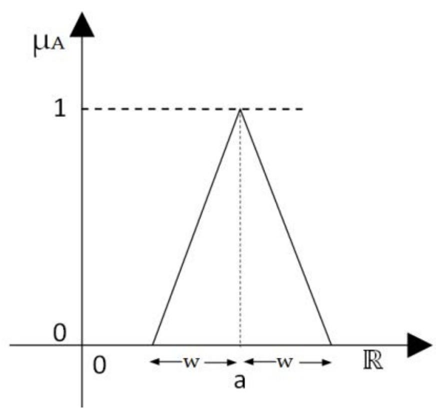

Fig. 1. Fuzzy symmetric triangular number.

a fuzzy linear regression model can be applied in order to identify the relationship between the hydrological variable, which is the annual cumulative streamflow $V_{T,j}$ in this work, and the frequency factor $K_{T,j}$:

$$V_{T,j} = \tilde{a}_1 + \tilde{a}_2 K_{T,j} \tag{8}$$

By comparing Eqs. (7) and (8), the fuzzy coefficients $\tilde{a}_1$, $\tilde{a}_2$ can be seen as a fuzzy estimation of the mean value and the standard deviation, respectively. The index j denotes the observation of each hydrological year.

2.3. Calculation of frequency factor K_T for several theoretical probability distributions

Let the normal theoretical probability distribution. It is obvious that the frequency factor $K_{T,j}$ in the case of a normally distributed variable X, is equivalent to the standardized normal variable $Z_{T,j}$.

$$K_{T,j} = Z_{T,j} \tag{9}$$

Let now the log-normal theoretical probability distribution. To simplify the procedure, it is well-known that in case of the log-normal distribution, the normal distribution with log-transformed data can be used instead of the log-normal distribution; which simply means that log-transformed data are implemented instead of the raw data and hence, the new auxiliary variable $y_{T,j}$, is normally distributed:

$$\ln V_{T,j} = y_{T,j} \tag{10}$$

Subsequently, based on the standardized normal variable Z of the normal distribution it holds:

$$\ln V_{T,j} = y_{T,j} = \bar{y} + s_y \times Z_{T,j} = \lambda + \zeta \times Z_{T,j} \tag{11}$$

In which as λ, ζ state the mean value and the standard deviation of the log transformed variable y correspondingly.

Let the Pearson type III theoretical probability distribution. Pearson type III distribution is derived from Gamma distribution with the scale parameter α and shape parameter β by adding the location parameter ξ. As a result, the Pearson type III distribution is also called as the three parameter gamma [18], whom probability density function and the cumulative distribution function are given as:

$$f_x(x) = \frac{1}{\alpha\Gamma(\beta)}\left(\frac{x-\xi}{\alpha}\right)^{\beta-1}\exp\left(-\frac{x-\xi}{\alpha}\right) \tag{12}$$

$$F_x(x) = \frac{1}{\alpha\Gamma(\beta)}\int_{\xi}^{x}\left(\frac{x-\xi}{\alpha}\right)^{\beta-1}\exp\left(-\frac{x-\xi}{\alpha}\right)dx \tag{13}$$

According to Sakkas [19], the frequency factor K_T in the case of Pearson type III distributed variable X [X ~ P-III (α,β,ξ)] can be estimated as follows:

$$K_T \cong \frac{1}{3\lambda}\left[\left(1+\lambda Z-\lambda^2\right)^3 - 1\right] \tag{14}$$

The parameter λ is given by the following expression:

$$\lambda \simeq \frac{1}{6}C_s \tag{15}$$

where C_s is the skewness coefficient which can be estimated as:

$$C_s \simeq \frac{a}{s^3} \tag{16}$$

Furthermore, a is the asymmetry of a sample which unbiased estimation is:

$$a = \frac{N}{(N-1)(N-2)}\sum_{j=1}^{N}(x_j - \bar{x})^3 \tag{17}$$

where N is the magnitude of the historical sample.

Finally, let us consider the log-Pearson type III theoretical probability distribution. The log-Pearson type III distribution of [X ~ LP-III (α_Y,β_Y,ξ_Y)] corresponds to the Pearson type III distribution if the following transformation is applied $Y = \ln(X)$. In other words, log-transformed data of a historical sample follow the Pearson type III distribution [$\ln(X)$ ~ P-III (α_Y,β_Y,ξ_Y)] [18].

In this work, instead of values of the primary data, the new auxiliary variable $y_{T,j}$ is utilized as in case of the log-normal distribution. The frequency factor $K_{T,j}$ for the Pearson type III which depends on the probability of non-exceedance and the skewness coefficient $C_{s,y}$ are approximated by Eqs. (14)–(17).

2.4. Fuzzy regression methodology

The uncertainty because of the matching between the observed probabilities and the adopted theoretical probability distribution can be treated by using the fuzzy regression suggested by Tanaka [20] and hence, all the observed data will be included in the produced fuzzy band. The main points of the proposed methodology are the following:

- Based on the observed probabilities and the adopted theoretical probability distribution, the frequency factor $K_{T,j}$ is determined for each pair of data.
- Based on Eqs. (7) and (8), a fuzzy linear regression model is implemented in order to determine a fuzzy relationship between the natural log of the cumulative discharge or simply the cumulative discharge and the frequency factor $K_{T,j}$ (which corresponds to a probability). In case of log transformed data according to Eq. (11) it holds [21]:

$$\tilde{y}_{T,j} = \tilde{\bar{y}} + \tilde{s}_y \times K_{T,j} = \tilde{\lambda} + \tilde{\zeta} \times K_{T,j} \tag{18}$$

It should be clarified that since fuzzy symmetric triangular numbers are selected as fuzzy coefficients hence, the mean value and the standard deviation are estimated as fuzzy symmetric triangular numbers.

- A modification of the well-known model of Tanaka [20] is applied in this article. The only difference is focused on the objective function. The problem of fuzzy linear

regression is concluded to a linear programming problem. Here, instead of the objective function of Tanaka [20] (which is the total spread of the produced fuzzy band), the use of a non-linear objective function is proposed. The modified model of fuzzy regression incorporates within the objective function both the total fuzzy spread and the distances between the central values and the observed data.
- The suitability of the proposed model can be estimated based on the magnitude of the fuzziness and furthermore by using mathematical distance norms which incorporate the unbiased estimators and the fuzzy estimation of the mean value and the standard deviation.

Although the frequency factor $K_{T,j}$ (independent variable) and the value of random variable (dependent variable) of the historical sample take only crisp values, the fuzziness arises from the inclusion constraints that is, from the requirement that all the data must be included in the produced fuzzy band. In other words this means that the fuzziness is generated from the (expected) no identical matching between the theoretical probability distribution (in this article the log-normal and the log-Pearson type III distributions will be preferred) and the observed probabilities.

According to the extension principle, in case of fuzzy symmetric triangular numbers as coefficients, the function $\tilde{y}_{T,j}$ will be also a fuzzy symmetric triangular number with the following centre ($Y_j^{h=1}$) and width ($w_{y,j}$) [13,20]:

$$\text{Centre: } Y_j^{h=1} = \lambda_a + \zeta_a \times Z_{T,j} \tag{19}$$

$$\text{Width: } w_{y,j} = w_\lambda + w_\zeta \times \left|Z_{T,j}\right| \tag{20}$$

which as $Y_j^{h=1}$, λ_a, ζ_a state the central values of the corresponding variables and as $w_{y,j}$, w_λ, w_ζ the corresponding spreads are meant.

The concept of inclusion is used to express the inclusion constraints. Thus, the inclusion of a fuzzy set A to a fuzzy set B with the associated degree $0 \le h \le 1$ is defined as follows:

$$A\left[h\right] \subseteq B\left[h\right] \tag{21}$$

A physical interpretation of the level h is that an observation y_j is contained in the support interval of the corresponding fuzzy estimate, which has a degree of membership greater than h_j. The degree of fit of the estimated model to the entire data set is defined as the minimum of all these h_j, which is denoted as h [21].

The produced fuzzy band will contain all the observed data:

$$y_j \in \left[y_{h,j}^L,\ y_{h,j}^R\right] \tag{22}$$

By taking into account the fuzzy arithmetic, for a selected level h, the inclusion constraints, in case that the decision variables (fuzzy coefficients) are selected to be fuzzy symmetric triangular numbers, are equivalent to [20,22,23]:

$$\begin{cases} \text{Inclusion Constraints} \\ \left(\lambda_a + \zeta_a \times Z_{T,j}\right) - \left(1-h\right)\left(w_\lambda + w_\zeta\left|Z_{T,j}\right|\right) = y_{h,j}^L \le \ln V_{T,j} = y_{T,j} \\ \left(\lambda_a + \zeta_a \times Z_{T,j}\right) + \left(1-h\right)\left(w_\lambda + w_\zeta\left|Z_{T,j}\right|\right) = y_{h,j}^R \ge \ln V_{T,j} = y_{T,j} \\ w_\lambda,\ w_\zeta \ge 0 \end{cases} \tag{23}$$

Since the fuzzy regression model of Tanaka [20] is transformed to a constrained optimization problem, the assessment of the suitability of the model is based on the produced fuzzy band. Therefore, a significant small fuzzy band indicates a proper approach. Thus, Tanaka [20] suggested the minimization of the sum of the produced fuzzy semi-spreads for all the data:

$$\min J\left(=\sum_{j=1}^{M} w_{y,j} = \left\{M \times w_\lambda + w_\zeta \sum_{j=1}^{M}\left|Z_{T,j}\right|\right\}\right)\ \left(\text{Objective function}\right) \tag{24}$$

where M is the number of the observed data.

Another interesting point is that Tzimopoulos et al. [24] applies another objective function, based on least squares based model of Diamond [25]:

$$\text{minimize } S = \sum_{j=1}^{M}\left\{\begin{array}{l}\left(y_j - \left(\lambda_a - w_\lambda + \zeta_a Z_{T,j} - w_\zeta\left|Z_{T,j}\right|\right)\right)^2 + \\ \left(y_j - \left(\lambda_a + w_\lambda + \zeta_a Z_{T,j} + w_\zeta\left|Z_{T,j}\right|\right)\right)^2\end{array}\right\} \tag{25}$$

where the first bracket denotes the distance between the observed data and the left bound of the produced fuzzy band whilst the second bracket denotes the distance between the observed data and the right bound of the produced fuzzy band.

However, Tzimopoulos et al. [24] applied their model without the conclusion constraints for crisp data and thus, the model could not lead to fuzzy coefficients. In this work, the objective function of Tzimopoulos et al. [24] is applied together with the inclusion constraints as they hold in case of the Tanaka formulation (Eq. (23)). It is worth mentioning that the approach of Tzimopoulos et al. [24] works well in case of fuzzy data.

It turns out (Appendix) that the objective function given by Eq. (25) includes both the objective function suggested by Tanaka (Eq. (24)) and the distance of the central values with the observed data (Eq. (A6)) and this can be seen as an advantage of the proposed modification. Both the measures S, J can be seen also as measures of suitability. Hence, in general, the measures, J, S express the uncertainty of the produced fuzzy model.

Two additional measures are established, δ_1 and δ_2 which are related to the unbiased estimation of the mean value and the standard deviation compared with the assessment of the same quantities produced by using the fuzzy regression. The third criterion of suitability, δ_1, examines how close is the estimated central values of the mean value and the standard deviation to the unbiased (usual statistical) estimation of the same variables $\hat{\lambda}$, $\hat{\zeta}$:

$$\delta_1 = \sqrt{\left(\lambda_a - \hat{\lambda}\right)^2 + \left(\zeta_a - \hat{\zeta}\right)^2} \tag{26}$$

Regarding the fourth criterion, δ_2, in the nominator, a simple Euclidian distance is used to compare the location of the observed data with the fuzzy output. In the dominator, the observed data are compared with the use of the unbiased estimation of the mean value instead of the fuzzy regression. The proposed measure is similar to the (crisp) measure R^2:

$$\delta_2 = 1 - \frac{\sum_{j=1}^{M}\left(\left(y_{T,j} - Y_{T,j}^{L}\right)^2 + \left(y_{T,j} - Y_{T,j}^{R}\right)^2 + \left(y_{T,j} - Y_{T,j}^{h=1}\right)^2\right)/3}{\sum_{j=1}^{M}\left(y_{T,j} - \bar{y}\right)^2} \quad (27)$$

where $\ln V_{T,j} = y_{T,j}$ is the log-transformed value of annual cumulative discharge of the j hydrological year and, $Y_{T,j}^{h=1}$, $Y_{T,j}^{R}$, $Y_{T,j}^{L}$, are the central value, the right and the left hand boundaries of the estimated values based on the produced fuzzy relationship.

3. Categorization to hydrological drought based on the return period

As aforementioned, the values of Z as they are presented in Table 1 are used to define the thresholds of several drought categories according to the SPI index. These values of Z correspond to some probability degrees. Having finished the fuzzy regression approach, then the values of Z described in Table 1 are used in order to determine the fuzzy thresholds of the categories.

It should be clarified that in case of the log-normal distribution (instead of the Gamma distribution), the normalized variable Z corresponds to a log-transformed cumulative discharge. The (crisp) thresholds of Table 1 lead to a fuzzy log-transformed cumulative discharge (based on Eq. (11)) which can be compared with the current (crisp) real log-transformed value of the cumulative discharge. Therefore, in this article the (fuzzy) thresholds of the annual cumulative discharge, $\ln \tilde{V}_k = \tilde{y}_k$ (considering the k^{th} threshold) are compared with the (crisp) observed annual cumulative discharge.

In case that the log-Pearson type III probability distribution is selected to be examined, a same procedure can be repeated. Based on Eqs. (14)–(17), the thresholds of Table 1 [5] are used to determine the frequency factor and hence, the fuzzy thresholds are calculated accordingly.

Even if, there are many measures to compare fuzzy numbers, there are not all of them suitable to compare a fuzzy number with a crisp number. To address this problem, the reliability measure of Ganoulis [26] is adopted.

Table 1
Classification of hydrological drought based on the random variable Z [5]

Category	Description	Criterion
0	Non-drought	$Z \geq 0.0$
1	Mild drought	$-1.0 \leq Z < 0.0$
2	Moderate drought	$-1.5 \leq Z < -1.0$
3	Severe drought	$-2.0 \leq Z < -1.5$
4	Extreme drought	$Z < -2.0$

Let a system which has a resistance $\tilde{R}$ and a load $\tilde{L}$ as fuzzy numbers. A reliability measure or a safety margin of the system may be defined as being the difference between load and resistance. This is also a fuzzy number given by Ganoulis [26].

$$\tilde{M} = \tilde{R} - \tilde{L} \quad (28)$$

Hence, Ganoulis [26] has proposed a fuzzy measure of risk, r, which is defined as the region of the fuzzy safety margin, where values of $\tilde{M}$ are negative. Mathematically, this may be expressed as follows:

$$r = \frac{\int_{m \leq 0} \mu_{\tilde{M}}(m)dm}{\int_{-\infty}^{+\infty} \mu_{\tilde{M}}(m)dm} \quad (29)$$

A similar approach is proposed by Spiliotis et al. [27] to achieve the comparison between the (crisp) exerted dimensionless shear stress and the (fuzzy) critical dimensionless shear stress.

Let us return to the examined problem. Hence, the authors propose a measure $G_{j,k}$, to indicate the degree according to which of the examined hydrological year, j, has a cumulative annual discharge, $y_j = \ln V_{T,j}$ greater than the examined fuzzy threshold of drought k, $\tilde{y}_k = \ln \tilde{V}_k$. It may be considered (Fig. 2):

$$G_{j,k} = \frac{\int_{y_k \leq y_j} \mu_{\mathrm{thr},k}(y)\, dy}{\int_{-\infty}^{+\infty} \mu_{\mathrm{thr},k}(y)\, dy} = 1 - \frac{\int_{y_k \geq y_j} \mu_{\mathrm{thr},k}(y)\, dy}{\int_{-\infty}^{+\infty} \mu_{\mathrm{thr},k}(y)\, dy}, \qquad S_{j,k} = \frac{\int_{y_k \geq y_j} \mu_{\mathrm{thr},k}(y)\, dy}{\int_{-\infty}^{+\infty} \mu_{\mathrm{thr},k}(y)\, dy} \quad (30)$$

For simplicity, the T– index is omitted in y_j since further, the y_j is used to check the intensity of the hydrological drought and not for another additional statistical analysis.

In the same way, a degree, $S_{j,k}$, according to which of the examined hydrological year, j, has a cumulative annual discharge smaller than the examined fuzzy threshold of drought k can be also considered. In Fig. 2, the grey hatched area (which is marked with (1)) denotes the numerator whilst the dominator is equal to the total area of the membership function.

Therefore for each hydrological year, the comparison is started from the lowest to the upper values, that is, by following an ascending procedure. Therefore, the comparison concludes to a category where based on both the proposed measures.

4. Implementation of the proposed methodology: annual cumulative streamflow time sequence

The case under investigation is the northern region of Prefecture Evros (Fig. 3). The annual cumulative streamflow, which is derived from the monthly discharges of the Evros

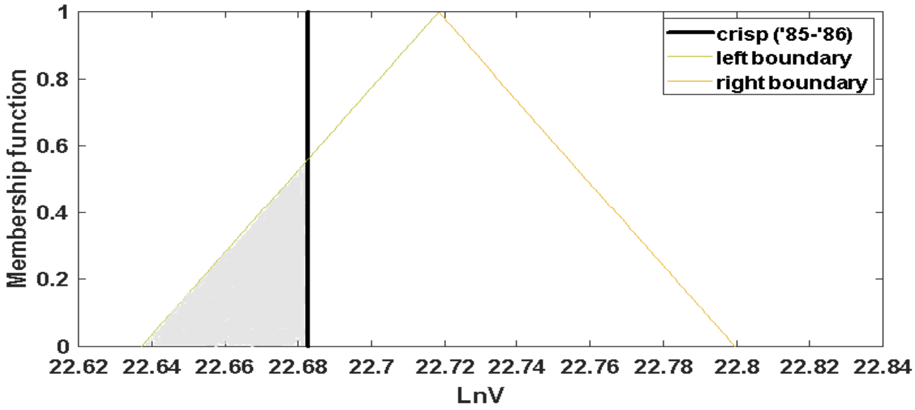

Fig. 2. Measure value resulting from the ratio of the hatched area to the total area in the case that it is less than 0.50 for the year 1985–1986.

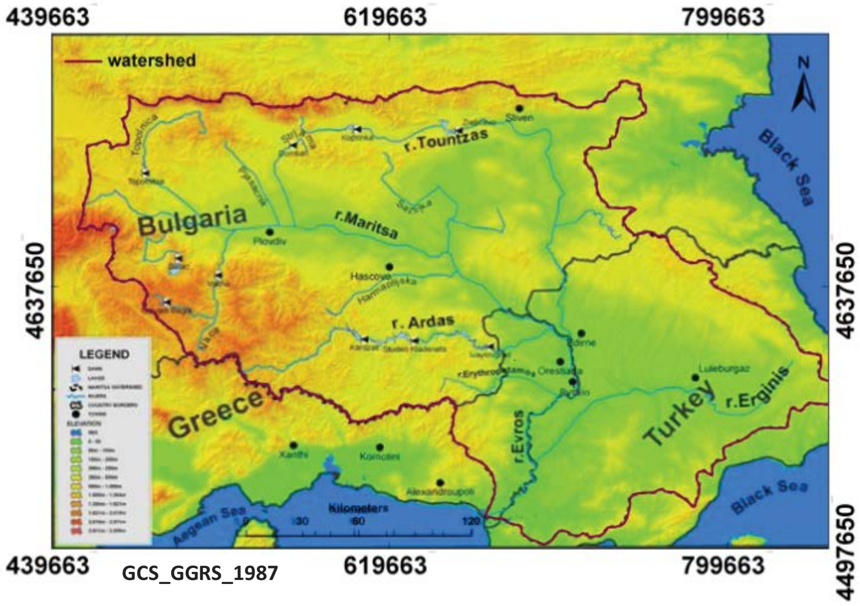

Fig. 3. Case of the trans-boundary Evros River and its tributaries. The examined data are derived from Pythio's bridge (41°21′43.51″ N 26°37′51.67″ E) (from Angelidis et al. [28]).

River at Pythio's bridge, is studied. The Evros River (Maritsa or Meric) is one of the largest river of Balkan Peninsula in terms of length, since it crosses the Bulgarian, Greek and Turkish borders. It rises in the Rila Mountains in Western Bulgaria and has its outlet in the Aegean Sea. It serves as a natural borderline between Greece and Turkey [28]. Its total watershed area is equal to 53,000 km^2, while the 6% of this is in Greek territory.

4.1. Methodology steps

The proposed methodology is implemented using the following steps:

- Based on the monthly discharges, the annual cumulative volumes of streamflow are calculated and then they are transformed to logarithmic values.
- A choice must be made regarding the theoretical probability distribution. As aforementioned, this hypothesis is checked later based on the suitability measures. In our case, the log-transformed theoretical distribution probabilities fit better to the historical sample.
- The fuzzy linear regression model of Tanaka [20] and the proposed modification of the objective function are applied (Figs. 4 and 5) between the frequency factor $K_{T,j}$ and the log-transformed annual cumulative discharge $y_{T,j}$, under the assumption that the observed probabilities follow either the log-normal distribution or the log-Pearson type III distribution. The produced fuzzy coefficients can be seen as a fuzzy assessment of the mean value and the standard deviation of the log-transformed data.
- The suitability of the proposed models is checked according to the value of the objective function, J and the proposed objective function, S. These measures are related to the

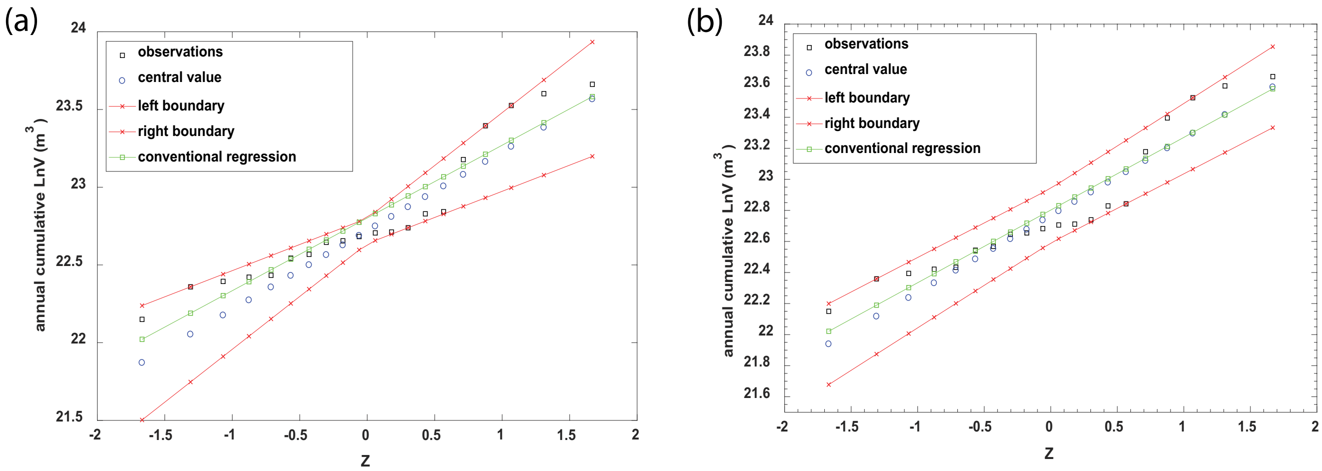

Fig. 4. Observations, fuzzy and conventional regression between the log transformed values of annual cumulative streamflow and the frequency factor $K_{T,j} = Z_{T,j}$ on the assumption of log-normal distribution (a) based on Tanaka's model and (b) based on the modified fuzzy regression model.

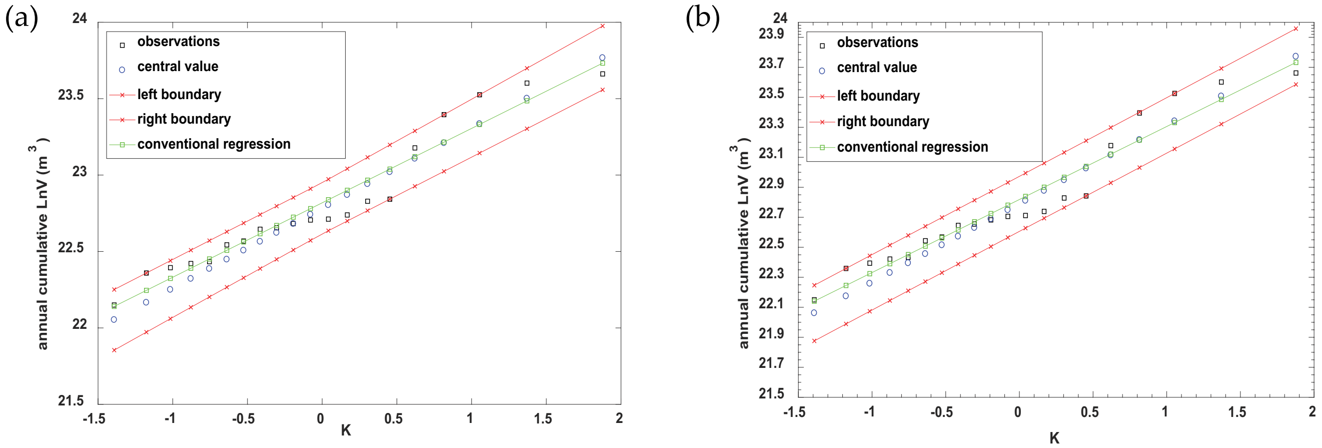

Fig. 5. Observations, fuzzy and conventional regression between the log transformed values of annual cumulative streamflow and the frequency factor $K_{T,j}$ on the assumption of log Pearson type III distribution (a) based on Tanaka's model and (b) based on the modified fuzzy regression model.

produced uncertainty. Furthermore, the measures δ_1 and δ_2 are implemented in order to compare the fuzzy solution with respect to the unbiased estimation of the mean value and the standard deviation.

- Based on drought classification (Table 1), the ln of the annual cumulative discharge which correspond to the normalized variable Z equal to −2, −1.5, −1 and 0 (Table 1) are calculated based on the produced fuzzy relationship. Therefore, according to the corresponding values of Z, the thresholds of drought are fuzzified.
- The observed cumulative annual discharge is compared with the aforementioned fuzzified thresholds of drought. From a mathematical point of view, the observed cumulative annual discharges are crisp numbers whilst the thresholds are fuzzy numbers. As aforementioned, the comparison starts from the lowest to the upper values, that is, by following an ascending procedure.

The results of the applied fuzzy models on assumption of log-normal (LN) and log-Pearson type III (LP III) distributions are presented in Table 2. Hence, two types of probability distributions are examined and simultaneously two fuzzy regression models which differ only in the objective function.

It is worth mentioning that the fuzziness $J_{\ln}$ decreases where the Log-Pearson type III distribution is adopted for both of the two fuzzy models, that is, the Tanaka's model and the modified fuzzy regression model. Hence, the Log-Pearson type III distribution is preferred based on the produced uncertainty.

Another interesting point is that based on the suitability measures, δ_1 and δ_2, the modified fuzzy least square regression model seems more appropriate than the Tanaka's model. In addition, based either on the measure J or on measure S, the Log-Pearson type III distribution is promised. Therefore the fuzzy solution, which is based on the

Table 2
Unbiased estimation of the mean value and standard deviation, fuzzy coefficients and suitability measures for both of two fuzzy regression models on the assumption of log-normal and log-Pearson type III distributions

Fuzzy regression coefficients and suitability measures		Tanaka model (min J)		Modified fuzzy regression model (minS)	
		LN	LP III	LN	LP III
Mean value (unbiased)	λ	22.80	22.80	22.80	22.80
Standard deviation (unbiased)	ζ	0.44	0.44	0.44	0.44
Central value of λ	λ_a	**22.72**	22.78	22.77	22.79
Spread of λ	w_λ	0.171	0.167	0.175	0.183
Central value of ζ	ζ_a	0.51	0.52	0.50	0.52
Spread of ζ	w_ζ	0.081	0.022	0.051	**0.002**
Total fuzziness	J	4.09	**3.66**	4.24	3.69
Least square objective function	S	3.01	1.88	2.63	**1.87**
Suitability measure	δ_1	0.11	0.09	**0.07**	0.08
Suitability measure	δ_2	0.678	0.804	0.724	**0.806**

Log-Pearson type III distribution and the least square based objective function, seems to be preferable. However, based only on measure δ_2, the solution which is based on the log-normal probability distribution and the modified fuzzy regression is preferred.

In Figs. 4 and 5, the results of the fuzzy regression applied by the two fuzzy models are illustrated, in case of the log-normal and the log-Pearson type III distribution. In Fig. 6, a fuzzy linear regression based on Tanaka's model is applied on assumption of normality in order to be obvious that the normal distribution does not provide good fit to the observed data due to the un-functional significant uncertainty of the produced fuzzy curve.

Next, the observed (crisp) cumulative annual discharge is compared with the aforementioned drought thresholds. It should be justified that in this step the fuzziness is utilized in order to categorize the drought. Even if crisp values of

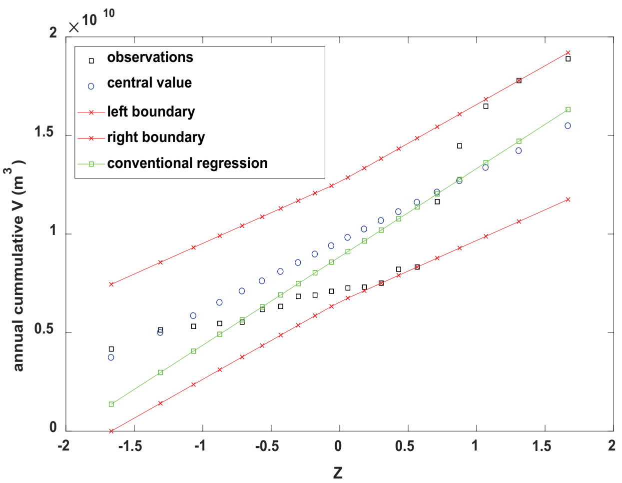

Fig. 6. Observations, fuzzy and conventional regression between the log transformed values of annual cumulative streamflow and the frequency factor $K_{T,j}$ on the assumption of normal distribution based on Tanaka's model.

Z are adopted, the fuzziness appears at the corresponding thresholds of volumes.

From a mathematical point of view, the observed cumulative annual discharges are crisp numbers whilst the thresholds are fuzzy numbers. As aforementioned, the comparison is started from the lowest to the upper values, that is, by following an ascending procedure.

There are also some cases where the comparison between the fuzzy threshold and the crisp annual cumulative discharge is not precise. Hence, the following criterion is adopted in this analysis:

$$\ln V_j = y_j \text{ overcomes } \tilde{y}_k \text{ if it holds: } G_{j,k} = \frac{\int_{y_j \geq \ln V_k} \mu_{\text{thr},k}(y)\,dy}{\int_{-\infty}^{+\infty} \mu_{\text{thr},k}(y)\,dy} > 0.5 \quad (31)$$

In Figs. 7 and 8, the fuzzified thresholds obtained based on the values of Z (Table 1) and the produced fuzzy curves of fuzzy linear regression models are presented, in the case both of log-normal and log-Pearson type III distribution. It is observable that because of the overlapping of the fuzzy $\ln \tilde{V}_k$ values (Fig. 7), the frontiers between the categories are overlapped to some degree. In any case, it seems more reasonable to adopt fuzzy thresholds among the categories of drought, compared with crisp thresholds, as the conventional methodology does. It is worth noting that although there are some cases where the comparison between the fuzzy threshold and the crisp annual cumulative discharge is not precise, in most of them, the $G_{j,k}$ index has values which are discernibly different from 0.5.

As it is aforementioned, on the assumption of log-Pearson type III distribution, the thresholds of each drought category are fuzzified based on the produced fuzzy relationship between the annual cumulative discharge $\ln V_j$ and the frequency factor $K_{T,j}$ which is a function of the variable $Z_{T,j}$. Thus, the frequency factor $K_{T,j}$ is calculated for the Z_T values of –2, –1.5, –1, 0, 1, 1.5 and 2 (crisp thresholds of drought categories).

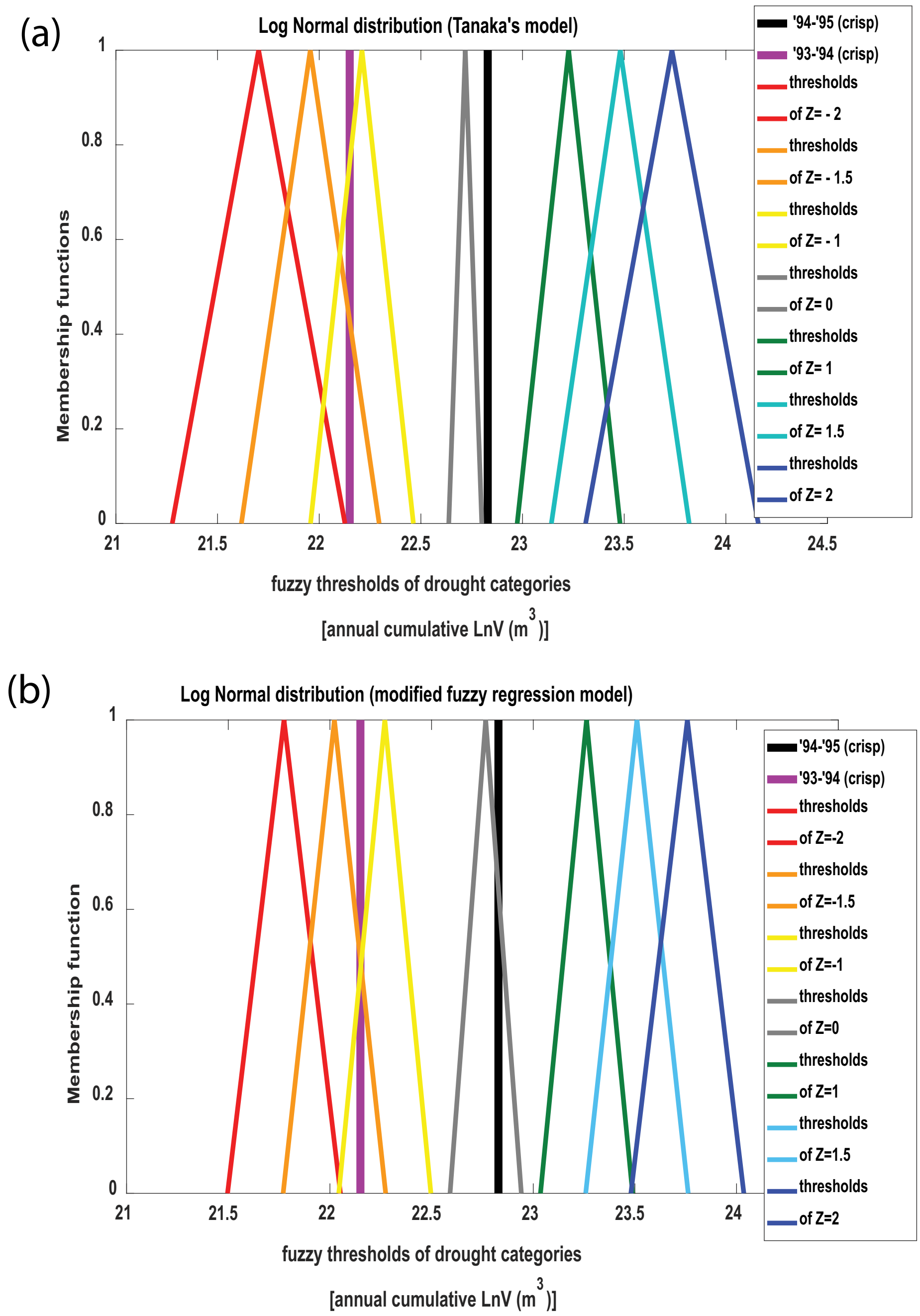

Fig. 7. Fuzzy thresholds of drought categories and the (crisp) annual cumulative discharge for the hydrological years 1994–1995 and 1993–1994, in the case of log-normal distribution (a) based on Tanaka's model and (b) based on the modified fuzzy regression model.

In Table 3, the results of the classification of hydrological drought in all cases are presented. For reasons of spatial organization, only the values of the $G_{j,k}$ and the $S_{j,k-1}$ measures in the case of the log-Pearson type III distribution based on the modified fuzzy regression are presented.

From Table 3, it is concluded that the hydrological years are classified in the same drought categories in almost all cases. It is also pointed out that on assumption of log-normal distribution, based on the modified fuzzy regression model, there are two hydrological years (1987–1988 and 1994–1995) classified as mild drought years, in contrast to the assumption of the Log-Pearson type III where these years are classified as non-drought years.

5. Concluding remarks

The uncertainty of the coupling between the observed probabilities and the adopted theoretical probability distribution can be treated by using the fuzzy regression model of Tanaka [20], where all the observed data are included in the produced fuzzy band. By using either the log-normal probability distribution or the log-Pearson III together with the

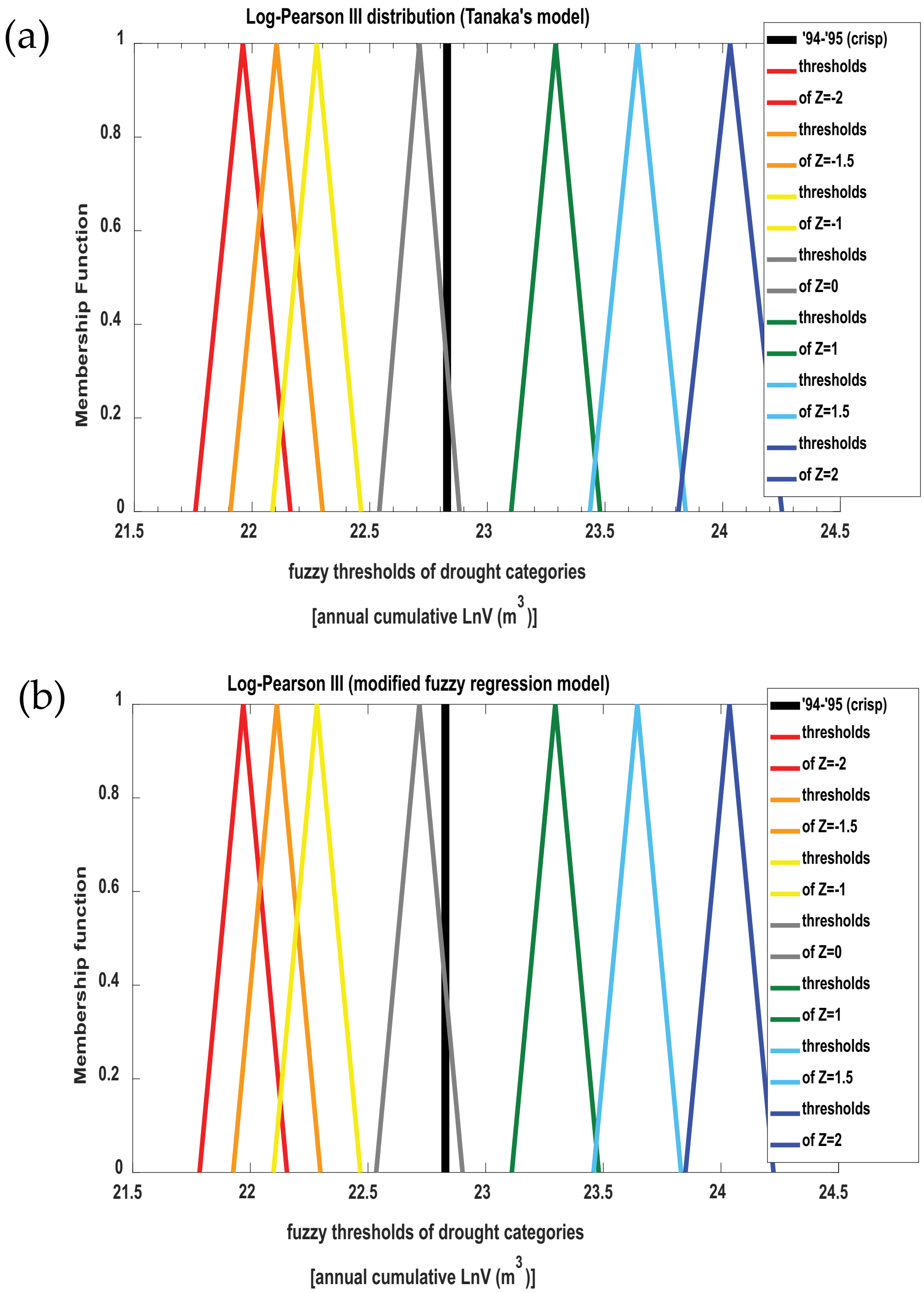

Fig. 8. Fuzzy thresholds of drought categories and the (crisp) annual cumulative discharge for the hydrological year 1994–1995, in the case of log-Pearson type III distribution (a) based on Tanaka's model and (b) based on the modified fuzzy regression model.

fuzzy regression, an estimation of the mean value and the standard deviation can be achieved simultaneously.

In addition, a modification of the Tanaka's model is proposed. Even if the inclusion constraints are kept, the objective function is changed and hence, it includes both the total semi-widths and the distance of the observed data with the produced central values as it is proved in the Appendix.

Four criteria of suitability are established in order to check the suitability of the adopted theoretical probability density with fuzzy numbers as parameters. The first criterion of suitability is based on the width of the produced fuzzy band and the third is based on the distance between the unbiased estimation of the mean value and the standard deviation with the central values of the estimated fuzzy quantities. Two other new measures are proposed. The first one is the least squares measure which incorporates both the distance of the real data with the central values and the produced fuzzy width and second the measure δ_2 which is explained above. Therefore, based on the suitability measures, the log-transformed probability distributions are preferred. The Log-Pearson type III distribution seems to be preferable compared with the log-normal

Table 3
Drought classification based on the fuzzy measures $G_{j,k}$ and $S_{j,k-1}$

Comparison of Crisp and Fuzzy streamflow values							
		Modified fuzzy regression model, LP type III		Modified fuzzy regression model		Tanaka's model	
Hydrological Year	$y_{T,j}$ (m^3)	Greater than the lower thresh. of the category	Smaller than the upper thresh. of the category	LP type III	LN	LP type III	LN
1985–1986	22.6826542	1	0.6800	Mild drought	√	√	√
1986–1987	22.7118034	1	0.5392	Mild drought	√	√	√
1987–1988	22.7391331	0.6039	1	Non drought	Mild drought	√	√
1988–1989	22.5682508	1	0.9838	Mild drought	√	√	√
1989–1990	22.3941092	0.9200	1	Mild drought	√	√	√
1990–1991	22.7059660	1	0.5675	Mild drought	√	√	√
1991–1992	22.5434611	1	0.9999	Mild drought	√	√	√
1992–1993	22.4211663	0.9662	1	Mild drought	√	√	√
1993–1994	22.1499070	0.6800	0.9608	Moderate drought	√	√	√
1994–1995	22.8424773	0.9456	1	Non drought	Mild drought	√	√
1995–1996	23.1777925	1	0.9352	Non drought	√	√	√
1997–1998	23.5255678	1	0.9352	Non drought	√	√	√
1998–1999	23.3953991	0.8896	1	Non drought	√	√	√
1999–2000	22.6456004	1	0.8200	Mild drought	√	√	√
2000–2001	22.3591526	0.8260	1	Mild drought	√	√	√
2001–2002	22.4334485	0.9820	1	Mild drought	√	√	√
2003–2004	22.8287708	0.9200	1	Non drought	√	√	√
2004–2005	23.6016702	1	0.7112	Non drought	√	√	√
2005–2006	23.6618808	0.5860	1	Non drought	√	√	√
2006–2007	22.6549724	1	0.7887	Mild drought	√	√	√

distribution regarding the proposed hybrid fuzzy–probabilistic approach.

Hence, by using the standardized indices, Z_T, to categorize the drought, the corresponding thresholds of drought can be determined as fuzzy numbers (based on the achieved fuzzy regression) which represent the cumulative annual discharges. The proposed methodology is successfully applied in the case of the Evros River where the (crisp) cumulative discharge is compared with the aforementioned fuzzy thresholds of drought. This comparison takes place by using the proposed measure in this article of comparison between fuzzy number and crisp number, which exploits all the information of the membership function. Finally, an efficient classification of drought is achieved following the proposed methodology although the fuzziness is taken into account.

Acknowledgments

This research is co-financed by Greece and the European Union (European Social Fund - ESF) through the Operational Programme "Human Resources Development, Education and Lifelong Learning" in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY).

References

[1] D. Wilhite, M.V.K. Sivakumar, R. Pulwarty, Managing drought risk in a changing climate: the role of national drought policy, Weather Clim. Extremes, 3 (2014) 4–13.
[2] G. Rossi, A. Cancelliere, At-Site and Regional Drought Identification by REDIM model, G. Rossi, A. Cancelliere, L.S. Pereira, T. Oweis, M. Shatanawi, A. Zairi, Eds., Tools for Drought Mitigation in Mediterranean Regions, Water Science and Technology Library, Springer, Netherlands, 2003, pp. 37–54.
[3] J.D. Salas, C. Fu, A. Cancelliere, D. Dustin, D. Bode, A. Pineda, E. Vincent, Characterizing the severity and risk of drought in the Poudre River, Colorado, J. Water Resour. Plann. Manage., 131 (2005) 383–393.
[4] Z. Tarawneh, J.D. Salas, The Occurrence Probability and Return Period of Extreme Hydrological Droughts, 13th International Water Technology Conference, IWTC 13, Hurghada, Egypt, 2009.
[5] P. Angelidis, F. Maris, N. Kotsovinos, V. Hrissanthou, Computation of drought index SPI with alternative distribution functions, Water Resour. Manage., 26 (2012) 2453–2473.
[6] G. Tsakiris, D. Pangalou, H. Vangelis, Regional drought assessment based on the reconnaissance drought index (RDI), Water Resour. Manage., 21 (2007) 821–833.
[7] I. Nalbantis, G. Tsakiris, Assessment of hydrological drought revisited, Water Resour. Manage., 23 (2009) 881–897.
[8] T.B. McKee, N.J. Doesken, J. Kliest, The Relationship of Drought Frequency and Duration to Time Scales, Proceedings of the 8th Conference on Applied Climatology, American Meteorological Society, Boston, 1993, pp. 179–184.
[9] F. Yusof, Use of statistical distribution for drought analysis, Appl. Math. Sci., 6 (2012) 1031–1051.

[10] T.C. Sharma, U.S. Panu, Predicting return periods of hydrological droughts using the Pearson 3 distribution: a case from rivers in the Canadian prairies, Hydrol. Sci. J., 60 (2015) 1783–1796.
[11] A. Cancelliere, J.D. Salas, Drought probabilities and return period for annual streamflows series, J. Hydrol., 391 (2010) 77–89.
[12] M. Spiliotis, Ch. Papadopoulos, P. Angelidis, B. Papadopoulos, Hybrid Fuzzy-Probabilistic Analysis and Classification of the Hydrological Drought, V. Kanakoudis, S. Tsitsifli, Eds., 3rd EWaS International Conference, Lefkada, Greece, 2018.
[13] M. Spiliotis, V. Hrissanthou, Fuzzy and Crisp Regression Analysis Between Sediment Transport Rates and Stream Discharge in the Case of Two Basins in Northeastern Greece, Regression Analysis: Introduction, Applications and Theory, Nova Science Publishers, New York, NY, USA, 2018.
[14] W. Weibull, A Statistical Theory of the Strength of Materials, Generalstabens Litografiska Anstalts Förlag, Stockholm, 1939.
[15] V. Chow, D. Maidment, L. Mays, Applied Hydrology, Hill International Ed., McGraw, Singapore, New York, NY, USA, 1988, pp. 394–398.
[16] V.T. Chow, A general formula for hydrologic frequency analysis, Trans. Am. Geophys. Union, 32 (1951) 231–237.
[17] M. Spiliotis, P. Angelidis, B. Papadopoulos, A Hybrid Fuzzy Regression-Based Methodology for Normal Distribution (Case Study: Cumulative Annual Precipitation), L. Iliadis, I. Maglogiannis, V. Plagianakos, Eds., 14th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Springer International Publishing, IFIP AICT-519, Rhodes, Greece, 2018, pp. 568–579.
[18] M. Naghettini, Fundamentals of Statistical Hydrology, Springer International Publishing, Switzerland, 2017.
[19] J. Sakkas, Applied Hydrology, Aivazis Ed., Thessaloniki, 2007.
[20] H. Tanaka, Fuzzy data analysis by possibilistic linear models, Fuzzy Sets Syst., 24 (1987) 363–375.
[21] M. Spiliotis, B. Papadopoulos, A Hybrid Fuzzy Probabilistic Assessment of the Extreme Hydrological Events, Proceedings of the 15th International Conference of Numerical Analysis and Applied Mathematics (ICNAAM 2017 Thessaloniki, Greece, 25–30 September 2017), AIP Conference Proceedings 1978, 290011, 2018.
[22] H. Moskowitz, K. Kim, On assessing the *H* value in fuzzy linear regression, Fuzzy Sets Syst., 58 (1993) 303–327.
[23] G. Tsakiris, D. Tigkas, M. Spiliotis, Assessment of inter connection between two adjacent watersheds using deterministic and fuzzy approaches, Eur. Water, 15 (2006) 15–22.
[24] C. Tzimopoulos, K. Papadopoulos, B.K. Papadopoulos, Models of Fuzzy Linear Regression, M. Rassias, V. Gupta, An Application in Engineering Mathematical Analysis, Approximation Theory and Their Applications, Springer International Publishing, Switzerland, 2016, pp. 693–714.
[25] P. Diamond, Fuzzy least squares, Inf. Sci., 46 (1988) 141–157.
[26] J. Ganoulis, Integrated Risk Analysis for Sustainable Water Resources Management, L. Linkov, A.B. Ramadan, Eds., Comparative Risk Assessment and Environmental Decision Making, Vol. 38, Kluwer Academic, Norwell, MA, USA, 2004, pp. 275–286.
[27] M. Spiliotis, V. Kitsikoudis, V.S. Ozgur Kirca, V. Hrissanthou, Fuzzy threshold for the initiation of sediment motion, Appl. Soft Comput., 72 (2018) 312–320.
[28] P. Angelidis, M. Kotsikas, N. Kotsovinos, Management of upstream dams and flood protection of the transboundary River Evros/Maritza, Water Resour. Manage., 24 (2010) 2467–2484.
[29] A.B. Ubale, S.L. Sananse, A comparative study of fuzzy multiple regression model and least square method, Int. J. Appl. Res., 2 (2016) 11–15.

Appendix

We start with the total produced fuzziness which is analyzed either around the central values or the observed data (Fig. A1) (similar considerations can be found in [29] but with different purposes):

$$\sum_{j=1}^{M}\left[Y_j^R - Y_j^L\right]^2 = \sum_{j=1}^{M}\left[Y_j^R - Y_j^L\right]^2 \Leftrightarrow \sum_{j=1}^{M}\left[\left(Y_j^R - Y_j\right)+\left(Y_j - Y_j^L\right)\right]^2 = \sum_{j=1}^{M}\left[\left(Y_j^R - Y_j^{h=1}\right)+\left(Y_j^{h=1} - Y_j^L\right)\right]^2 \Leftrightarrow \tag{A1}$$

$$\Leftrightarrow \sum_{j=1}^{M}\left(Y_j^R - Y_j\right)^2 + \sum_{j=1}^{M}\left(Y_j - Y_j^L\right)^2 + 2\sum_{j=1}^{M}\left[\left(Y_j - Y_j^L\right)\left(Y_j^R - Y_j\right)\right] = \sum_{j=1}^{M}\left(Y_j^R - Y_j^{h=1}\right)^2 + \sum_{j=1}^{M}\left(Y_j^{h=1} - Y_j^L\right)^2 + 2\sum_{j=1}^{M}\left[\left(Y_j^{h=1} - Y_j^L\right)\left(Y_j^R - Y_j^{h=1}\right)\right] \tag{A2}$$

where Y_j is the observed value of dependent hydrological variable and, $Y_j^{h=1}$, Y_j^R, Y_j^L are, the central value of the estimated hydrological variable, the upper bound of the produced fuzzy band and the bound limit of the produced fuzzy band, respectively. Solving for $\sum_{j=1}^{M}\left(Y_j - Y_j^L\right)^2 + \sum_{j=1}^{M}\left(Y_j^R - Y_j\right)^2$ Eqs. (A1) and (A2) can be written equivalently as:

$$\sum_{j=1}^{M}\left(Y_j - Y_j^L\right)^2 + \sum_{j=1}^{M}\left(Y_j^R - Y_j\right)^2 = \sum_{j=1}^{M}\left(Y_j^{h=1} - Y_j^L\right)^2 + \sum_{j=1}^{M}\left(Y_j^R - Y_j^{h=1}\right)^2 + 2\sum_{j=1}^{M}\left[\left(Y_j^{h=1} - Y_j^L\right)\left(Y_j^R - Y_j^{h=1}\right)\right] - 2\sum_{j=1}^{M}\left[\left(Y_j - Y_j^L\right)\left(Y_j^R - Y_j\right)\right] \tag{A3}$$

by developing the last two terms it results

$$2\sum_{j=1}^{M}\left[\left(Y_j^{h=1} - Y_j^L\right)\left(Y_j^R - Y_j^{h=1}\right)\right] - 2\sum_{j=1}^{M}\left[\left(Y_j - Y_j^L\right)\left(Y_j^R - Y_j\right)\right] = 2\sum_{j=1}^{M}\left(Y_j^{h=1}Y_j^R - \left(Y_j^{h=1}\right)^2 - Y_j^L Y_j^R + Y_j^L Y_j^{h=1}\right) - 2\sum_{j=1}^{M}\left(Y_j Y_j^R - Y_j^2 - Y_j^L Y_j^R + Y_j^L Y_j\right) \tag{A4}$$

Consequently, by eliminating the terms of Y_j^R, Y_j^L and by making common factors the terms of $Y_j^{h=1}$ and Y_j, Eq. (A4) can be written as:

$$2\sum_{j=1}^{M}\left[Y_j^{h=1}\left(Y_j^R - Y_j^{h=1} + Y_j^L\right)\right] - 2\sum_{j=1}^{M}\left[Y_j\left(Y_j^R - Y_j + Y_j^L\right)\right] = 2\sum_{j=1}^{M}\left[Y_j^{h=1}\left(Y_j^R - Y_j^{h=1} + Y_j^L\right) - Y_j\left(Y_j^R - Y_j + Y_j^L\right)\right] \tag{A5}$$

multiplying and dividing the parentheses with the number two (2) and placing where $Y_j^{h=1}=\dfrac{Y_j^R+Y_j^L}{2}$, which holds in the case that fuzzy symmetric triangular numbers are selected as fuzzy coefficients (which is done in this article), it is concluded to:

$$2\sum_{j=1}^{M}\left[Y_j^{h=1}\left(2\frac{Y_j^R+Y_j^L-Y_j^{h=1}}{2}\right)-Y_j\left(2\frac{Y_j^R+Y_j^L-Y_j}{2}\right)\right]=$$
$$2\sum_{j=1}^{M}\left[Y_j^{h=1}\left(2Y_j^{h=1}-Y_j^{h=1}\right)-Y_j\left(2Y_j^{h=1}-Y_j\right)\right]= \tag{A6}$$

$$2\sum_{j=1}^{M}\left[\left(Y_j^{h=1}\right)_j^2-2Y_j^{h=1}Y_j+Y_j^2\right]=2\sum_{j=1}^{M}\left(Y_j^{h=1}-Y_j\right)^2$$

Therefore the following relation is modulated:

$$\sum_{j=1}^{M}\left(Y_j-Y_j^L\right)^2+\sum_{j=1}^{M}\left(Y_j^R-Y_j\right)^2=\sum_{j=1}^{M}\left(Y_j^{h=1}-Y_j^L\right)^2+$$
$$\sum_{j=1}^{M}\left(Y_j^R-Y_j^{h=1}\right)^2+2\sum_{j=1}^{M}\left(Y_j^{h=1}-Y_j\right)^2 \tag{A7}$$

In case of triangular symmetric fuzzy numbers are used as fuzzy coefficients (as in this article) then:

$$\sum_{j=1}^{M}\left(Y_j^{h=1}-Y_j^L\right)^2=\sum_{j=1}^{M}\left(Y_j^R-Y_j^{h=1}\right)^2=\sum_{j=1}^{M}w_{y,j}^2 \tag{A8}$$

The total sum of the difference between the observed value and the upper bound raised to the second power plus the difference between the observed value and lower bound raised to the second power is equal to the total sum of the difference between the central value of the estimated hydrological variable and the upper bound raised to the second power plus the difference between the central value of estimated hydrological variable value and the lower bound raised to the second power plus the (double) difference between the central value of estimated hydrological variable and the observed data raised to the second power. Hence, this is the reason why Eq. (A7) is premised in this article instead of the standard objective function according to the study by Tanaka [20] formulation, since it includes not only the total fuzziness but also the distance of the central values from the observed data.

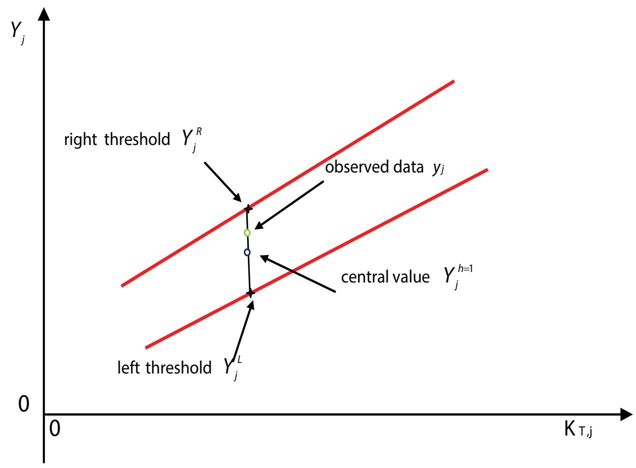

Fig. A1. Graphical representation of Eq. (A1).