Journal of
Applied Sciences

science
alert

ANSI*net*
an open access publisher
http://ansinet.com

Application of Multi Criteria Method to Identify the
Best-fit Statistical Distribution

[1]Ani Shabri and [2]Abdul Aziz Jemain
[1]Department of Mathematic, Technology University of Malaysia, Malaysia
[2]Sciences Mathematic Studies Center, Universiti Kebangsaan Malaysia

Abstract: Generally, researchers are faced to identify the true statistical distributions for the analysis of a various hydrologic data sets. Using traditional statistical analysis methods one choose a hypothesized distribution to describe the observed data, estimate the distribution parameters and then apply the goodness of fit test such as the Chi Square test (CS) or Kolmogorov Smirnov (KS) test. For more accurate, several factors or criteria should be considered in selection of the best distribution. However when more than two criteria are used to identify the best distribution, it is more difficult and more subjective. In this paper, we propose a new Multi Criteria Decision Making method (MCDM) based on nonlinear programming for selection of the best distribution to fit a set of data. The Generalized Extreme Value (GEV), Generalized Pareto (GP), Pearson 3 (P3) and Lognormal 3 (LN3) are used and their goodness of fit has been examined by various test statistics. A numerical example is used to illustrate the applicability of the proposed approach.

INTRODUCTION

In order to do effective planning, design and management of water resources engineering such as water supplies, hydropower, irrigation systems, etc., the statistics of annual streamflow series are required, which can be understood by conducting frequency analysis on annual streamflow. The purpose of Flood Frequency Analysis (FFA) is to select an appropriate distribution type for representing a hydrological variable of interest at a site or in a region. There is a wide range of flood frequency models developed in hydrology. In the open literature, there are a few studies to identify the probability type of annual streamflow in the world (Yue and Wang, 2004). The search for the proper distribution function has been subjective of several studies.

The commonly used procedure is to first choose a hypothesized distribution, then estimate the parameters of the hypothesized distribution using the moment method, the maximum likelihood method or L-moment method and finally apply the goodness of fit such as Kolmogorov-Smirnov test (KS) or Chi-square (CS) test to see if the hypothesized distribution can be rejected or not (Wang *et al.*, 2004).

Cunnane (1985) discussed factors affecting the choice of a distribution for FFA, including the method of parameter estimation, treatment of outliers, inclusion of large historical flood values, data transformations and causal compositions of flood population. He concluded that distribution choice could not be based on theoretical arguments alone or one criterion.

Many researchers proposed several criteria good test statistics in selection of the best statistical distribution to fit the data. For example, Turkman (1985) used the Akaike's Information Criteria (AIC) for the choice of extremal models and analyzed its effectiveness in choosing the most likely among the Gumbel, Frechet and Weibull distributions. Onoz and Bayazit (1995) considered the CS, KS, AD (Anderson-Darling) and PPCC tests for evaluating the suitability of seven distributions for the flood data from 19 stations all over the world. Kim and Heo (2002) used the statistic CS, KS, CvM (Cramer von Mises) and PPCC tests to see if the hypothesized of the Gamma, GEV, Gumbel, Log-Gumbel, Lognormal, Log- Pearson or Weibull distribution can be rejected or not to fit the annual maximum flood of Goan station in Korea. Zalina *et al.* (2002) applied the PPCC test, the Root Mean Square Error (RMSE), the relative root-mean square error (RRMSE) and the Min Absolute Error (MAE) to assess the capability of eight distributions to describe the annual extreme rainfall data of Peninsular of Malaysia.

In these procedures, the choice of the hypothesized distribution is often based on the rule of thumb. The selected statistical distributions for the same data may be different for different analysis. This will occur when two

---

Corresponding Author: Ani Shabri, Department of Mathematic, Technology Universiti of Malaysia, Malaysia

or more statistical distributions were tested by only one or two test statistics. However when more than two test statistics are used to identify the best distribution, it is more difficult and more subjective.

The selection of the "best" distribution from a set of test statistics is a multi-criteria decision making problem. Many methods have been proposed to solve a multi-criteria decision making problem such as the least deviation method (Xu and Da, 2004), linear goal programming method (Fan *et al.*, 2004, 2005) and the fuzzy majority approach (Chiclana *et al.*, 1998; Herrera *et al.*, 2001).

In this study, a new method based on a non linear programming is proposed to solve multi criteria decision making problem. In the proposed approach, a non linear programming model is constructed to integrate the fuzzy preference relation and to compute the collective ranking values of the alternatives. Once the collective ranking values are known, the selection of the best distribution can be obtained.

Five selection test statistics, namely the MSEC (mean square relative error in cumulative distribution function, CDF), MAE, RMSE, KS and NIFL (normalized likelihood function index) test and four distributions, namely the Generalized extreme value (GEV), Generalized Pareto (GP), Pearson 3 (P3) and Lognormal 3 (LN3) distributions are considered in the illustration of the proposed approach. The performance of the new approach is compared to the fuzzy majority approach that proposed by Chiclana *et al.* (1998).

PROBABILITY DISTRIBUTION FUNCTIONS OF FLOOD FREQUENCY ANALYSIS

Four of the commonly distributions used in hydrology are considered in this study (e.g., Hosking, 1990; Hosking and Wallis, 1997; Sveinsson *et al.*, 2002). The forms of the probability density function of these distributions are follows:

Lognormal (LN3):

$$f(x) = \frac{1}{k(x-\xi)\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}\{\ln(x-\xi)-\alpha\}^2\right] \quad (1)$$

Generalized Extreme Value (GEV):

$$f(x) = \frac{1}{\alpha}\left[1-k\left(\frac{x-\xi}{\alpha}\right)\right]^{1/k-1} \exp\left\{-\left[1-k\left(\frac{x-\xi}{\alpha}\right)\right]^{1/k}\right\} \quad (2)$$

Pearson Type Three (P3):

$$f(x) = \frac{1}{\alpha\Gamma(k)}\left(\frac{x-\xi}{k}\right)^{k-1} \exp\left(-\frac{x-\xi}{\alpha}\right) \quad (3)$$

Generalized Pareto (GP):

$$f(x) = \frac{1}{\alpha}\left[1-\frac{k}{\alpha}(x-\xi)\right]^{1/k-1} \quad (4)$$

where, α, ξ and k are scale, location and shape parameters respectively. Parameters of these distributions are estimated by the L-moments method. We do not give here any equations for the parameter estimation of the considered distributions because they are well known and commonly appear in many publications (Rao *et al.*, 1997; Sveinsson *et al.*, 2002; Hosking, 1990; Hosking and Wallis, 1997).

GOODNESS OF FIT CRITERIA

Five criteria, which measure the relative goodness of fit test statistics, were employed for comparison of the probability distributions for fitting on a data set. The set of criteria as follows:

a. The Mean Square Relative Error (MSEC) In CDF statistic test

The MSEC test is based on mean square distance between the hypothesized CDF and the empirical CDF (Whalen *et al.*, 2002). The MAE is given by:

$$MSEC = \sum_{i=1}^{n}\left(\frac{i-1}{n-1}-F(x_i)\right)^2 \quad (5)$$

b. The Mean Absolute Error (MAE) statistic test

The MAE test is based on the mean absolute distance between the observed data values, x_i and the estimated data values, $\hat{x}_i$, respectively (Zalina *et al.*, 2002). The MAE is given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i-\hat{x}_i}{x_i}\right| \quad (6)$$

c. The Root Mean Square Error (RMSE) statistic test

The RMSE test is based on the mean square distance between the observed data values, x_i and the estimated data values, $\hat{x}_i$, respectively (Guo *et al.*, 1996). The RMSE is given by:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \hat{x}_i}{x_i}\right)^2} \qquad (7)$$

d. The Kolmogorov-Smirnov (KS) test statistics

The KS test statistic is use to measure the maximum deviation between the hypothesized CDF and the empirical CDF at the same observation point or among a pair of consecutive observation points (Wang, 2004). The KS is given by:

$$\text{KS} = \max_{1\le i\le n-1}\{|\,F(x_i) - i/n\,|, |\,F(x_i) - (i-1)/n\,|\} \qquad (8)$$

e. The Normalized Likelihood Function Index (NLFI) test statistics

The NLFI test statistic is the mean of the logarithm of the probability distribution function, f(x) (Jain and Singh, 1987). The NLFI test statistic is given by:

$$\text{NLFI} = \left|\frac{1}{n}\sum_{i=1}^{n}\log f(x)\right| \qquad (9)$$

The smaller value of the test statistics in case of criteria (a)-(e) is the better the sample data fits the chosen distribution.

MAKING THE MEASUREMENT VALUE UNIFORM

In MCDM approach, the rating or the performance measurement value given by each test statistics to the distributions is used in the field. The higher the rating, the better the distribution satisfies the test statistics. The ratings assumed are given in the range between 0 and 1.

However, in our study the five criteria test statistics may take values out of the range [0, 1]. In addition, based on each of the five criteria, smaller the value, the better the distribution fit to the sample. There is need to standardize these measures such that they are all in the range of [0, 1] and the higher the measurement value, the better the fit. The transform function based on the Cauchy distribution function is used to standardize the measures (Wang *et al.*, 2004). The Cauchy distribution function is given by:

$$r(v) = \frac{1}{1 + t\,v^2}(t > 0, v > 0) \qquad (10)$$

where, v, is the test statistic value of a hypothesized distribution and t is a constant determined empirically to make sure that the performance measures will spread in the range [0, 1].

MULTIPLE CRITERIA

Multi Criteria Decision Making (MCDM) is an important part of modern decision science. It has been extensively applied to various areas such as economic analysis, urban or regional planning and forecasting etc. In a MCDM problem, we have a set of alternatives to be analyzed according to different purposes in order to select the best one (Chiclana *et al.*, 2001). A decision maker is often faced with the problem of selecting or ranking alternatives associated with non-commensurate and conflicting attributes (Fan *et al.*, 2004).

Presentation of the problem: Let $X = \{x_1, x_2, K, x_n\}$, $(n \ge 2)$ be a finite set of the distributions and $E = \{e_1, e_2, K, e_m\}$, $(m \ge 2)$ be a finite set of test statistics. Let $C = (c_1, c_2, K, c_m)$ be the weight vector of the test statistics, where $\sum_{h=1}^{m} c_h = 1$, , c_h, ≥ 0, $h = 1, K, m$ and c_h denotes the important degree of test statistic c_h and is usually determined by the decision makers.

The problem concerned in this paper is to rank distributions or to select desirable distributions among a finite set X based on fuzzy preference relations.

The fuzzy majority approach: Chiclana *et al.* (1998) proposed the fuzzy majority approach to solve the Multi Criteria Decision Making (MCDM) problem for the aggregation and to select desirable of the information in decision making. The approach consist the three steps: (i) uniform the preference information through a transform function, (ii) aggregate the uniformed preference information into a collective one and (iii) rank alternatives or select the most desirable alternative of distributions by means of the weighted average operator.

In generally, the information provided by a set of criteria is supposed to be of a diverse nature. To make the information uniform, a transformation function is used to transform the preference utilities into the format of the preference relations. The transform function is called the preference degree of the alternative x_i over x_j is given by:

$$p_{ij}^k = \frac{(u_i^k)^2}{(u_i^k)^2 + (u_j^k)^2} \qquad (11)$$

where, u_i^k represents the measurement value given by the test statistics e_k.

Since the important degree of each test statistics is not equal, hence the fuzzy majority criteria with the fuzzy quantifier such as most, at least half or as many as possible, is used to aggregate the uniformed preference information into a collective one. The collective fuzzy preference relation is obtained as follows:

$$p_{ij}^c = \Phi\left(p_{ij}^1, \dots, p_{ij}^m\right) = \sum_{k=1}^{m} c_k \cdot q_{ij}^k \qquad (12)$$

where, q_{ij}^k is the k th largest value in the set $\{p_{ij}^1, \dots, p_{ij}^m\}$ and Q(r) is given by:

$$Q(r) = \begin{cases} 0, & 0 \le r < a, \\ \dfrac{r-a}{b-a}, & a \le r \le b, \\ 1, & b < r \le 1. \end{cases} \qquad (13)$$

$$a, b \in [0,1] \text{ and } c_k = Q(k/m) - Q((k-1)/m) \qquad (14)$$

Thus by Eq. 13, the fuzzy preference relations, P^1, K, P^m, are aggregated into a collective fuzzy preference relation as follows:

$$P^* = (p_{ij}^k)_{n \times n} \qquad (15)$$

In order to the rank the distributions or select the most desirable distribution (s), the fuzzy quantifier also is used to compute the collective ranking values of distribution, i.e.,

$$w_i^* = \Phi\left(p_{ij}^k, j=1,\dots,n, j \ne i\right)^{`} \qquad (16)$$

where, w_i^* is the collective ranking value of the I th distribution. From Eq. 16, we can obtain that the collective ranking value vector of the distributions is

$$w^* = (w_1^*, \dots, w_n^*)^T \qquad (17)$$

According to the obtained ranking values, the ranking of the distributions or selection of the most desirable distribution(s) is done. The greater the ranking value w_i^* is the better the corresponding alternative x_i will be.

The proposed approach: Suppose the collective ranking value of alternative x_i is w_i (i = 1, K, n) and w_i is unknown variable where wi ≥ 0, $\sum_{i=1}^{n} w_i = 1$. The problem in this research is how to determine the collective ranking values of alternatives based on the preference relation provided by the decision makers.

In order to do that, we proposed the fuzzy preference relation $P^k = [p_{ij}^k]_{n \times n}$ (Chilana *et al.*, 1998) where

$$p_{ij}^k \approx \frac{w_i^2}{w_i^2 + w_j^2} \qquad i,j=1,n, k=1,K \; m; i \ne j \qquad (18)$$

From Eq. 18, the deviation degree between p_{ij}^k and $w_i^2/(w_i^2 + w_j^2)$ is given by:

$$g_{ij}^k(w) = w_i^2 - p_{ij}^k(w_i^2 + w_j^2)$$

The collective deviation degree g_{ij}^k as follows

$$z_{ij}(w) = \sum_{k=1}^{m} g_{ij}^k(w) = \sum_{k=1}^{m} c_k \left| w_i^2 - p_{ij}^k(w_i^2 + w_j^2) \right| \qquad (19)$$

To make the group consensus better, we can minimize z_{ij} (w) by assessing the collective ranking values w_i(i = 1, K, n). Thus, the following multiple objective constrained optimization model:

$$\text{Minimize } z_{ij}(w) = \sum_{k=1}^{m} c_k \left| w_i^2 - p_{ij}^k(w_i^2 + w_j^2) \right| \qquad (20a)$$
$$ij = 1, K, n, i \ne j$$

$$\text{Subject to } \sum_{i=1}^{n} w_i = 1, \qquad (20b)$$

$$wi \ge 0, \; i = 1, K, n \qquad (20c)$$

Solution to the above minimization problem is found by solving the following non linear programming model:

$$\text{Minimize } z = \sum_{i=1}^{n} \sum_{j=1, i \ne j}^{n} \left(s_{ij} u_{ij}^+ + t_{ij} u_{ij}^- \right) \qquad (21a)$$

$$\text{Subject to } \sum_{k=1}^{m} c_k [w_i^2 - p_{ij}^k(w_i^2 + w_j^2)] - u_{ij}^+ + u_{ij}^- = 0 \qquad (21b)$$

$$\sum_{i=1}^{n} w_i = 1, \qquad (21c)$$

$$wi \ge 0, \; i = 1, K, n \qquad (21d)$$

$$u_{ij}^+ \ge 0, u_{ij}^- \ge 0, \text{ and } u_{ij}^+ u_{ij}^- = 0, j=1,\dots,n, i \ne j \qquad (21e)$$

Where, s_{ij} is the weighting factor corresponding to positive deviation u_{ij}^+ and t_{ij} is the weighting factor corresponding to positive deviation. By solving model (21), the collective ranking value vector, $w^* = (w_1^*, w_2^*, K, w_n^*)$, can be obtained. The greater the ranking value w_i^* is the better the corresponding alternative x_i will be.

Based on Eq. 19, the total deviation degree of group consensus is given by:

$$D(w) = \sum_{i=1}^{n} \sum_{j=1}^{n} z_{ij}(w) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\sum_{k=1}^{m} c_k \left| w_i^2 - p_{ij}^k(w_i^2 + w_j^2) \right| \right] \qquad (22)$$

Obviously, the smaller the value of total deviation degree D(w) is the better the group consensus are.

ILLUSTRATIVE EXAMPLE

To illustrate the application of MCDM problem, the annual flood peaks series data of the station 3516422-Selangor in Peninsular Malaysia for the past 40 years (1961-2001) was analyzed. The data is given in

Table 1: Recorded annual flood peaks series of the station 3516422-Selangor (values in m³ s⁻¹)

46.07	34.18	61.00	57.50	33.86	63.00	69.68	46.33	51.17	67.22
67.32	13.36	65.96	29.48	23.48	43.60	67.34	83.98	60.88	40.07
62.20	21.63	124.31	25.87	44.41	53.01	61.80	38.11	81.69	20.07
52.92	97.57	67.49	35.95	17.54	84.88	80.25	55.00	96.00	14.48

Table 2: Estimated parameters for candidate distributions

Distribution	GEV	LN3	GP	P3
Shape (k)	0.183	0.119	0.780	31.373
Scale (α)	44.079	52.534	70.047	4.498
Location (ξ)	23.560	24.945	14.661	-87.099

Table 3: Statistics calculated with Eq. 5-9

		Distribution			
No.	Test Statistic	GEV	LN3	GP	P3
1	MSEC	0.02141	0.02107	0.03165	0.02102
2	MAE	0.05927	0.06047	0.06724	0.05998
3	RMSE	0.11534	0.12210	0.08227	0.12038
4	KS	0.08398	0.08113	0.16105	0.08146
5	NLFI	4.6064	4.6072	4.4444	4.6066

Table 4: The rating values obtained with Eq. 10

		Distribution			
No.	Rating	GEV	LN3	GP	P3
1	MSEC	0.9033	0.90469	0.86337	0.9049
2	MAE	0.94405	0.94298	0.93700	0.94341
3	RMSE	0.92253	0.92496	0.92398	0.92468
4	KS	0.89659	0.89119	0.86129	0.89255
5	NLFI	0.81280	0.81277	0.81818	0.81279

Table 1. The parameters of the all distributions were estimated with the L-moments method. The estimated parameters of these four distributions are listed in Table 2.

For the data in Table 3, we then used Eq. 10 to transform them into values in (0, 1) with larger value indicating a better performance against each criteria. The constant values t of Eq. 10 used for the five rows of the data given in the Table 3 are 5, 1, 1, 1 and 0.05, respectively. The transformed performance measures obtained using Eq. 10 are given in the Table 4.

From Table 4, we can see that based on the MSEC test statistic alone, P3 distribution provides the best fit. If only the MAE test statistic is considered, the GEV best fit. If only the RMSE test statistic is considered, the GP distribution provides the best fit. The other results are shown in Table 4.

The fuzzy majority approach: To apply the fuzzy majority approach, a transform function given in Eq. 11 is used to transform the rating values into the formats of fuzzy preference relation. The fuzzy preference relation are given by:

$$P^1 = \begin{bmatrix} 0.500 & 0.499 & 0.523 & 0.499 \\ 0.501 & 0.500 & 0.523 & 0.500 \\ 0.477 & 0.477 & 0.500 & 0.477 \\ 0.501 & 0.500 & 0.523 & 0.500 \end{bmatrix}$$

$$P^2 = \begin{bmatrix} 0.500 & 0.501 & 0.504 & 0.500 \\ 0.499 & 0.500 & 0.503 & 0.500 \\ 0.496 & 0.497 & 0.500 & 0.497 \\ 0.500 & 0.500 & 0.503 & 0.500 \end{bmatrix}$$

$$P^3 = \begin{bmatrix} 0.500 & 0.499 & 0.499 & 0.499 \\ 0.501 & 0.500 & 0.501 & 0.500 \\ 0.501 & 0.499 & 0.500 & 0.500 \\ 0.501 & 0.500 & 0.500 & 0.500 \end{bmatrix}$$

$$P^4 = \begin{bmatrix} 0.500 & 0.503 & 0.520 & 0.502 \\ 0.497 & 0.500 & 0.517 & 0.499 \\ 0.48 & 0.483 & 0.500 & 0.482 \\ 0.498 & 0.501 & 0.518 & 0.500 \end{bmatrix}$$

$$P^7 = \begin{bmatrix} 0.500 & 0.500 & 0.497 & 0.500 \\ 0.500 & 0.500 & 0.497 & 0.500 \\ 0.503 & 0.503 & 0.500 & 0.503 \\ 0.500 & 0.500 & 0.497 & 0.500 \end{bmatrix}$$

Since the important degree of each decision maker is not equal, the fuzzy majority criterion with the fuzzy quantifier "at least half" with the pair of values (0, 0.5) based on the Eq. 14 is used. The weighting vector $c_1 = 0.4$, $c_2 = 0.4$, $c_3 = 0.2$, $c_4 = 0$ and $c_5 = 0$ using the Eq. 14. Then the collective fuzzy preference relation is:

$$P = \begin{bmatrix} 0.500 & 0.503 & 0.521 & 0.502 \\ 0.501 & 0.500 & 0.520 & 0.500 \\ 0.495 & 0.495 & 0.500 & 0.495 \\ 0.501 & 0.501 & 0.520 & 0.500 \end{bmatrix}$$

In order to rank the alternatives or select the most desirable alternatives, the fuzzy quantifier "most" with the pair (0.3, 0.8), i.e., the corresponding with the weighting vector $c_1 = 0$, $c_2 = 0.4$, $c_3 = 0.5$ and $c_4 = 0.1$ is found. From Eq. 17, we obtained that the collective ranking value vector of the alternatives is

$$w^* = (w_1, w_2, w_3, w_4)^T =$$
$$(0.50195, 0.50028, 0.49498, 0.50063)^T$$

Table 5: Comparative results of the two approaches

Approaches based on	Collective ranking values of alternatives				Ranking of alternatives	D(w)
	w_1	w_2	w_3	w_4		
Nonlinear programming	0.25111	0.25095	0.24689	0.25105	$x_1 > x_4 > x_2 > x_3$	0.00778
The fuzzy majority	0.25125	0.25041	0.24776	0.25058	$x_1 > x_4 > x_2 > x_3$	0.00807

According to the obtained ranking value vector w^*, the ranking result of the four distributions is $x_1 > x_4 > x_2 > x_3$. This means that the GEV distribution, provides the best fit to describe the annual flood peaks series data of the station 3516422-Selangor in Peninsular Malaysia.

To obtain the total deviation degree of group consensus, i.e., $D(w^*)$, in Eq. 10, the sum of the entries of ranking value vector w^* must has a sum of 1. However, the sum of the collective ranking value vector w^* had a sum more than 1. The ranking value vector was scaled so that the sum the entries ranking value vector had a sum of 1. The total deviation degree of group consensus after scaled, i.e., $D(w^*) = 0.00807$.

The proposed approach: In this study, the important degree of each test statistic is equal, i.e., $c_1 = K = C_5 = 1/5$. For simplicity, let s_{ij} and t_{ij} be equal to 1. Using Eq. 21a-e, we can set up the following non linear programming model:

$$\text{Minimize } z = \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} \left(u_{ij}^+ + u_{ij}^- \right)$$

Subject to

$$0.4999w_1^2 - 0.5001w_4^2 - u_{41}^+ + u_{41}^- = 0$$

$$0.4915w_1^2 - 0.5085w_3^2 - u_{31}^+ + u_{31}^- = 0$$

$$0.4998w_4^2 - 0.5002w_2^2 - u_{24}^+ + u_{24}^- = 0$$

$$0.4918w_2^2 - 0.5082w_4^2 - u_{14}^+ + u_{14}^- = 0$$

$$0.5002w_2^2 - 0.4998w_4^2 - u_{42}^+ + u_{42}^- = 0$$

$$0.5084w_3^2 - 0.4916w_4^2 - u_{21}^+ + u_{21}^- = 0$$

$$0.5003w_2^2 - 0.4997w_3^2 - u_{12}^+ + u_{12}^- = 0$$

$$0.4916w_4^2 - 0.5084w_3^2 - u_{13}^+ + u_{13}^- = 0$$

$$0.5001w_4^2 - 0.4999w_1^2 - u_{14}^+ + u_{14}^- = 0$$

$$0.4997w_1^2 - 0.5003w_2^2 - u_{21}^+ + u_{21}^- = 0$$

$$0.5085w_3^2 - 0.4915w_1^2 - u_{13}^+ + u_{13}^- = 0$$

$$0.5082w_3^2 - 0.4918w_2^2 - u_{23}^+ + u_{23}^- = 0$$

$$w_1 + w_2 + w_3 + w_4 = 1$$

$$w_i \geq 0, \qquad i = 1, K, n$$

$$u_{ij}^+ \geq 0, u_{ij}^- \geq 0, \text{ and } u_{ij}^+ u_{ij}^- = 0, j = 1, \ldots, n, i \neq j$$

By solving the above non linear programming problem, we have the following results:

$$w_1 = 0.25111, w_2 = 0.25095, w_3 = 0.24689, w_4 = 0.25105$$

Therefore, the collective ranking value vector of alternatives is

$$w^* = (0.25111, 0.25095, 0.24689, 0.25105)$$

According to the obtained ranking value vector w^*, the ranking results of the four distributions is $x_1 > x_4 > x_2 > x_3$ The total deviation degree of group consensus, $D(w^*)$ can be obtained by substituted w^* (i = 1, 2, 3, 4) in Eq. 10, i.e., $D(w^*)=0.00622$.

The computational results of the nonlinear programming approach and the fuzzy majority approach after scaled are compared and shown in Table 5.

It can be noticed from above table that the two collective ranking value vectors obtained by the two approaches are similar. But it can be seen that the total deviation degree of the group consensuses obtained by two approach are different and $D(w^*) < D(w^`)$. This illustrates that the group consensus of the proposed approach is better than one based on the fuzzy majority approach.

CONCLUSIONS

The search for the proper distribution function has been subjective of several studies. The best distribution choice could not be based on one theoretical argument alone or one criterion. However, when more than two criteria are used, it more difficult and more subjective.

In this study, we proposed a new Multi Criterion Decision Making (MCDM) approach for selection of the best distribution to fit a set of data. A numerical example is used to illustrate the use of the proposed method. The approach is based on the non linear programming which can be used to assess attribute weight and then to select

the most desirable distributions. The five commonly distributions considered are: GEV, GL, LN3, GP and P3. For the distributions selection, five criteria test statistics were used.

The performance of the new approach is compared to the fuzzy majority approach. The result show that the new approach performs as well as the fuzzy majority approach.

The proposed approaches provides decision makers to improve quality of decisions by making its more explicit and efficient, especially in order to selection the best distribution in specific field of application.

REFERENCES

Chiclana, F., F. Herrera and E. Herrera-Viedma, 2001. Integrating multiplicative preference relations in a multipurpose decision-making model based on fuzzy preference relations. Fuzzy Sets Sys., 122: 277-291.

Cunnane, C., 1985. Factor affecting choice of distribution for flood series. Hydrol. Sci. J., 30: 25-36.

Chiclana, F., F. Herrera and E. Herrera-Viedma, 1998. Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations. Fuzzy Sets Sys., 97: 33-48.

Fan, Z.P., G.F. Hu and S.H. Xio, 2004. A method for multiple attribute decision-making with the fuzzy preference relation on alternatives. Computers and Industrial Engineering, 46: 321-327.

Fan, Z.P., J. Ma, Y.P. Jiang, Y.H. Sun and L. Ma, 2005. A goal programming approach to group decision making based on multiplicative preference relations and fuzzy preference relations. Eur. J. Operat. Res., (In Press).

Guo, S.L., R.K. Kachroo and R.J. Mngodo, 1996. Nonparametric kernel estimation of flow quantiles. J. Hydrol., 185: 335-348.

Herrera, F., E. Herrere-Viedma F. and Chiclana, 2001. Multiperson decision-making based on multiplicative preference relations. Eur. J. Operat. Res., 129: 372-385.

Hosking, J.R.M., 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. Royal. Stat. Soc., 5: 105-124.

Hosking, J.R.M., J.R. Wallis, 1997. Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press, Cambridge, pp: 223.

Jain, D. and V.P Singh, 1987. Comparison of Some Flood Frequency Distributions Using Empirical Data. Hydrologic Frequency Modeling, Singh V.P., (Ed.), D. Reidel Publishing Company, pp: 467-485.

Kim, K.D. and J.H. Heo, 2002. Comparative study of flood quantiles estimation by nonparametric models. J. Hydrol., 260: 176-193.

Onoz, B. and M. Bayazit, 1995. Best-fit distributions of largest available flood samples. J. Hydrol., 167: 195-208.

Rao, A. R., ASCE. Member and K.H. Hamed, 1997. Regional frequency analysis of wabash river flood data by l-moments. J. Hydrol. Eng., 2: 169-179.

Sveinsson, G.B., J.D. Salas, M. ASCE and D.C. Boes, 2002. Regional frequency analysis of extreme precipitation in northeastern colorado and fort collins flood of 1997. J. Hydrol. Eng., 7: 49-63.

Turkman, K. F., 1985. The choice of extremal models by akaike's information criterion. J. Hydrol., 82: 307-315.

Wang, Y., R.C.M. Yam and M.J. Zuo, 2004. A multi-criterion approach to select of the best statistical distribution. Computers and Industrial Eng., 47: 165-180.

Whalen, T.M., G.T. Savage and G.D. Jeong, 2002. The method of self-determined probability weighted moments revisited. J. Hydrol., 268: 171-191.

Xu, S. and Q. Da, 2005. A least deviation method to obtain a priority vector of a fuzzy preference relation. Eur. J. Operat. Res., 164: 206-216.

Yue, S. and C.Y. Wang, 2004. Possible regional probability type of canadian annual streamflow by l-moments. Water Resources Management, 18: 425-438.

Zalina, D., A.H.M. Kassim, M.N.M. Desa and V.T.V. Nguyen, 2002. Statistical analysis of at-site extreme rainfall processes in peninsular malaysia. Proceedings of the fourth International Conference. Cape Town, South Africa, 274.