Flood Flow Frequency Model Selection Using L-moment Method in Arid and Semi Arid Regions of Iran

A.R. Keshtkar*, A. Salajegheh, M. Najafi Hajivar

* Assistant Professor, International Desert Research Center (IDRC), University of Tehran, Tehran, Iran
Associate Professor, Faculty of Natural Resources, University of Tehran, Karaj, Iran
MSc. Graduate, Natural Resources and Watershed Management Office, Koohrang, Iran

Received: 3 November 2009; Received in revised form: 2 March 2012; Accepted: 4 April 2012

Abstract
Statistical frequency analysis is the most common procedure for the analysis of flood data at a gauged location that in first step it is needed to select a model to represent the population. Among them, the central moment has been the most common and widely used, and with the using of computers, the application of the maximum likelihood has increased. This research was carried out in order to recognition of suitable probability distributions with pervious common methods. In order to investigate of suitable probability distribution for flood flow, using L-moment method through the existing hydrometric stations in the region, 17 hydrometric stations were selected for peak discharges data studies. According to results of this research for peak discharge, LP3 distribution and ordinary moment method, P3 distribution and L-moment method, LN2 and LN3 distributions and ordinary method have been suitable distinguished for 53%, 35%, 6% and 6% of stations, respectively. We concluded that L-moment method is suitable to determine peak series probability distributions in the Iran central plateau and P3 is the best probability distribution for modeling peak series in this region.

Keywords: Flood flow; Frequency distribution function; Linear moment; Maximum likelihood; Ordinary moment

1. Introduction
Hydrological frequency analysis is one of the essential tasks in hydrological engineering design. It is the work of determining the magnitude of hydrological variables corresponding to given frequencies or recurrence intervals (McCuen, 2003). Procedures involved in frequency analysis include: collection a random sample of the interested hydrological variables; finding the best-fit-distribution for the samples by a good-of-fit (GOF) test or other appropriate methods; and determining the magnitude of the hydrological variable corresponding to a given probability of exceedance using the best-fit-distribution. Chow (1951) proposed the following general equation for hydrologic frequency analysis of a random variable X with mean μ and standard deviation σ:

\[ X_T = \mu + K_T \sigma \]  

(1)

Where \( K_T \), the frequency factor is a function of the return period T and is dependent on the distribution type of X. Frequency factors of distributions commonly used in hydrologic frequency analysis are themselves variables and their distributions have been identified (Kite, 1988). Upon collection of a random sample, one must decide the type of distribution which best characterizes the random sample and determine the corresponding \( K_T \) value.

Two GOF tests, namely the chi-square test and the Kolmogorov-Smirnov test, are often used for the selection of probability distributions for hydrological variables (Haan, 2002). Another method of goodness-of-fit test is the method based on ordinary moment ratio diagrams (D’Agostino and Stephen, 1986). Moment ratios are unique properties of probability distributions and sample moment ratios of ordinary skewness and kurtosis have
been used for selection of probability distribution (Kottegoda, 1980 and D’Agostino and Stephen, 1986). In recent years there have been many applications of L-moments diagrams for selecting various distributional alternatives in a region and the skewness and kurtosis L-moment-ratio diagram (LMRD) was suggested as a useful tool for discrimination between candidate distributions (Hosking and Wallis, 1987; Hosking, 1990; Hosking and Wallis, 1993; Vogel and Fennessey, 1993; Hosking and Wallis, 1997 and Peel et al., 2001). The L-moments uniquely define the distribution if the mean of distribution exists, and the L-skewness and L-kurtosis are much less biased than the ordinary skewness and kurtosis (Hosking and Wallis, 1997).

Estimation of flood flow is often required for watersheds with insufficient or nonexistent hydrometric information particularly in arid and semi-arid regions. Because parametric methods require a number of assumptions, nonparametric methods have been investigated as alternative methods. L-moment diagrams and associated goodness-of-fit procedure (Wallis, 1988; Hosking, 1990; Chowdhury et al., 1991; Pearson, 1992; Vogel et al., 1993a; Daviau et al., 2000; Peel et al., 2004; Yurekli et al., 2005; Chen et al., 2006; Eslamian and Feizi, 2007 and Salajegheh et al., 2008) have been advocated for evaluating the suitability of selecting various distributional alternatives for modeling flows in a region. For example Wallis (1988) found an L-moment diagram useful for rejecting Jain and Singh’s (1987) conclusion that annual maximum flood flows at 44 sites were well approximated by a Gumbel distribution and for suggesting a general extreme value (GEV) distribution instead. Vogel et al., (1993b) used L-moment diagrams to show that two and three-parameter log-normal models (LN2 and LN3), the LP3 and the GEV distributions were all acceptable models of flood flows in the southwestern United States. Gholami et al., (2001) used L-moment diagrams to show that the Gumbel distribution was acceptable for annual maximum series (AMS) in the north of Iran.

Base on the climate conditions and annual average precipitation, Iran is one of the arid and semi-arid countries of the world, which is mostly encountered with lack of water. Most of the rivers in arid regions of Iran are seasonal and their flood flows may become unavailable during a short time of rainfall seasons and because of some special geological problems of this region most of the permanent rivers contain saline water and are useless. The objective of the present study is to introduce and evaluate the suitable probability distributions for modeling peak flows in arid and semi-arid regions of Iran.

2. Materials and Methods

2.1. Study area

The Iran central plateau watershed where is located in center of Iran, has considered as a arid and semi-arid region of Iran. The Iran central plateau watershed is 854000 Km² in area and has seven sub watersheds and several catchments in different size. For the purpose of this research 17 catchments were chosen where permanent river and hydrometric station had. The selection was based on Ministry of Agriculture information that no significant abstraction exist in the contributing catchments upstream of these stations. The location of the hydrometric stations utilized in this research is given in table 1 and the spatial distribution shown in fig.1.

<table>
<thead>
<tr>
<th>Station</th>
<th>River</th>
<th>Longitude (E)</th>
<th>Latitude (N)</th>
<th>Altitude (m)</th>
<th>Catchment area (Km²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarab hendeh</td>
<td>Golpaiegan</td>
<td>50</td>
<td>21</td>
<td>2000</td>
<td>817</td>
</tr>
<tr>
<td>Abgarm</td>
<td>Kharood</td>
<td>48</td>
<td>18</td>
<td>1560</td>
<td>2450</td>
</tr>
<tr>
<td>Dehsomee</td>
<td>Kordan</td>
<td>50</td>
<td>57</td>
<td>1410</td>
<td>360</td>
</tr>
<tr>
<td>Sim</td>
<td>Karaj</td>
<td>51</td>
<td>09</td>
<td>1790</td>
<td>725</td>
</tr>
<tr>
<td>Dodahak</td>
<td>Ghomrood</td>
<td>50</td>
<td>34</td>
<td>1470</td>
<td>8851</td>
</tr>
<tr>
<td>Chamriz</td>
<td>Kor</td>
<td>52</td>
<td>07</td>
<td>1800</td>
<td>3390</td>
</tr>
<tr>
<td>Balf</td>
<td>Solami</td>
<td>56</td>
<td>32</td>
<td>2000</td>
<td>935</td>
</tr>
<tr>
<td>Safarzadeh</td>
<td>Halirood</td>
<td>57</td>
<td>33</td>
<td>920</td>
<td>8420</td>
</tr>
<tr>
<td>Dehrood</td>
<td>Shoor</td>
<td>57</td>
<td>44</td>
<td>1000</td>
<td>1321</td>
</tr>
<tr>
<td>Adoori</td>
<td>Talango</td>
<td>58</td>
<td>07</td>
<td>1690</td>
<td>276</td>
</tr>
<tr>
<td>Jirofto</td>
<td>Hafkoosk</td>
<td>57</td>
<td>11</td>
<td>2600</td>
<td>225</td>
</tr>
<tr>
<td>Gharatiolarab</td>
<td>Chary</td>
<td>56</td>
<td>02</td>
<td>1853</td>
<td>173</td>
</tr>
<tr>
<td>Godarzarch</td>
<td>Abbakshia</td>
<td>56</td>
<td>34</td>
<td>2200</td>
<td>1144</td>
</tr>
<tr>
<td>Namrood</td>
<td>Namrood</td>
<td>52</td>
<td>39</td>
<td>1810</td>
<td>587</td>
</tr>
<tr>
<td>Simindasht</td>
<td>Delichai</td>
<td>52</td>
<td>31</td>
<td>1435</td>
<td>2254</td>
</tr>
<tr>
<td>Benkooh</td>
<td>Habterood</td>
<td>52</td>
<td>52</td>
<td>1000</td>
<td>3209</td>
</tr>
<tr>
<td>Senobar</td>
<td>Shastdare</td>
<td>59</td>
<td>06</td>
<td>1760</td>
<td>152</td>
</tr>
</tbody>
</table>
2.2. Methods

In this research, several methods were carried out in order to propose the most suitable probability distributions for peak series. From the hydrometric stations existing in the watershed, 17 stations were selected for analysis. The peak flows data were available for 17 gauging stations, with stream flow records of 20 years or more. Most of the data were recorded between 1972 and 2000.

Then, data were analyzed to choose an identical data period. Therefore, the peak series data from 1971 to 1998 were taken as the 25-year data period of each station.

The missing data were generated and completed using regression relationships between the stations. For each distribution, the values of residual sum of squares (RSS) were calculated using the ordinary moment and L-moment methods. The RSS values of the two methods were compared, suitable distributions for each station were chosen according to the lowest RSS. The best of probability distribution was applied to estimate $T$-year peak series.

2.3. The ordinary moment ratio diagram

Parameters of a probability distribution can be expressed in terms of its moments ($\mu_r'$) or central moments ($\mu_r$) defined as

$\mu_r' = E(X^r)$

$\mu_r = E(X - \mu_1)^r$  \hspace{1cm} (2b)

For $r = 1, 2$ therefore, the shape of a probability distribution can be characterized by the moments of that distribution. This is achieved using the relationships between the standardized coefficients of skewness ($\sqrt{\beta_1}$) and kurtosis ($\beta_2$) which are defined as

$\sqrt{\beta_1} = \mu_3/(\mu_2)^{3/2}$ \hspace{1cm} (3a)

$\beta_2 = \mu_4/\mu_2^2$ \hspace{1cm} (3b)

Theoretical $\sqrt{\beta_1}$ - $\beta_2$ relationships of various distributions (fig. 2) are known as the moment-ratio diagram. Hereafter, the standardized
moments ($\sqrt{\beta_1}$) and ($\beta_2$) will be referred to as the skewness and kurtosis, respectively.

Given a random sample $\{x_1, x_2, \ldots, x_n\}$, the sample estimates of $\sqrt{\beta_1}$ and $\beta_2$ are defined as

$$
\sqrt{\beta_1} = \frac{m_1}{m_2}
$$

(4a)

$$
b_2 = \frac{m_4}{m_2^2}
$$

(4b)

Where

$$
m_r = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^r
$$

(4c)

$$
S = \sqrt{\frac{m_2}{m_2}}
$$

(4d)

Sample estimates of ($\sqrt{\beta_1}$) and ($\beta_2$) do not always coincide with the theoretical points or curves of their parent distributions, and method of GOF test using ($\sqrt{\beta_1}$, $\beta_2$) have thus been developed for normal distribution (D’Agostino and Pearson, 1973 and Bowman and Shenton, 1975). Bowman and Shenton (1986) developed acceptance regions for ($\sqrt{\beta_1}$, $\beta_2$) based GOF test using stochastic simulation. The sample moments ($\sqrt{\beta_1}$, $\beta_2$) tend to have high variances, especially when the sample size is small. Therefore, it is often difficult to distinguish among candidate distributions. Even though the joint distribution of the ordinary sample skewness and sample kurtosis is asymptotically normal, such asymptotic property is a poor approximation in small and moderately samples, particularly when the underlying distribution is even moderately skew (Hosking and Wallis, 1997).

Using stochastic simulation, Wu (2005) studied the distribution of ($\sqrt{\beta_1}$, $\beta_2$) for normal distribution with respect to sample sizes varying from 20 to 1000. As can be seen in fig. 2, for smaller sample sizes (for example $n=20$) the distribution of $\sqrt{\beta_1}$ given $\beta_2$ changes from unimodality to bimodality as $\beta_2$ increases. Thus, closed curves encompassing certain percentages of ($\sqrt{\beta_1}$, $\beta_2$) samples of the normal distribution are difficult to be expressed by mathematical equations when the sample size is small.

$L$-moment analysis

$L$-moment is alternative system of describing the shapes of probability distributions. $L$-moments are linear combinations of order statistics which are robust to outliers and virtually unbiased for small samples, making them suitable for flood frequency analysis, including identification of distribution and parameter estimation (Hosking, 1990 and Hosking and Wallis, 1993). $L$-moments are defined as linear combinations of probability weighted moments (PWM):

$$
\beta_r = E\{X[F(x)]^r\}
$$

(5)

Where $\beta_r$ is the $r$th order PWM and $F(x)$ is the cdf of $X$. When $r=0$, $\beta_0$ is the mean stream flow. Hence a sample estimate of the first PWM, which we term $b_0$ is simply the sample mean. Nevertheless, unbiased estimators are often preferred in goodness-of-fit evaluations such as $L$-moment diagrams. Unbiased sample estimates of the PWM for any distribution can be computed from

$$
b = \frac{1}{n} \sum_{i=1}^{n} x_i
$$

(6a)

$$
b_1 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})
$$

(6b)

$$
b_2 = \frac{1}{n-2} \sum_{i=1}^{n} (x_i - \bar{x})^2
$$

(6c)

$$
b_3 = \frac{1}{n-3} \sum_{i=1}^{n} (x_i - \bar{x})^3
$$

(6d)

Where $x_i$ represents the ordered stream flows with $x_1$ being the largest observation and $x_n$ the smallest. The PWM estimators in eqn. (6) can be more generally described using

$$
b_r = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sqrt{r}}\right)^r
$$

(7)

For a random variable $X$ with quantile function $x(u)$, Hosking and Wallis (1997) defined the $L$-moments ($\lambda_r, r = 1, 2, \ldots$) as

$$
\lambda_r = \int x(u) d\mu_u
$$

(8a)

Where

$$
p_r = \sum_{u=0}^{\infty} d\mu_u x_u
$$

(8b)

$$
p_{r+k} = (-1)^{r-k} \binom{r+k}{r} (-\frac{\theta}{\alpha})^{r-k} \gamma (r+k)
$$

(8c)

The $L$-moments can also be expressed in terms of the probability weighted moments defined by
Greenwood et al., (1979), and for any distribution, the first four L-moments are easily computed from PWMs using

\[ \lambda_1 = \beta_0 \]  
\[ \lambda_2 = 2\beta_1 - \beta_0 \]  
\[ \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \]  
\[ \lambda_4 = 30\beta_3 - 12\beta_2 + \beta_0 \]  

Where \( \beta_r, r = 0,1,2, \ldots, \) are probability weighted moments defined by

\[ \beta_r = \int_{-\infty}^{\infty} x^r f(x) dx \]  

In terms of linear combination of order statistics, the L-moments can also be expressed by

\[ \lambda_i = \frac{\lambda_i}{\lambda_2} \]  

Theoretical relationships between L-skewness \( (\tau_3) \) and L-kurtosis \( (\tau_4) \), i.e. the L-moment ratio diagram of several probability distributions have been given by Hosking (1990 and 1991) and can be used to distinguish different probability distributions.

### L-moment ratio diagrams

Analogous to the product moment ratio; coefficient of variation \( C_v = \frac{\sigma}{\mu} \), skewness \( \gamma \) and kurtosis \( \kappa \), Hosking (1990) define the L-moment ratios

\[ \tau_r = \frac{\lambda_r}{\lambda_2} \]  

Where \( \tau_r, r = 1, \ldots, 4 \) are the first four L-moments and \( \tau_2, \tau_3 \) and \( \tau_4 \) are the L-coefficient of variation \( (L-C_v) \), L-skewness and L-kurtosis, respectively. The first L-moment \( \lambda_1 \) is equal to the mean stream flow \( \mu \); hence it is a measure of location. Hosking (1990) shows that \( \lambda_2; \tau_3 \) and \( \tau_4 \) can be thought of as measures of a distribution’s scale, skewness and kurtosis, respectively, analogous to the moment’s \( \sigma, \gamma \) and \( \kappa \) respectively.

![Fig. 2. L-moment ratio diagram of various distributions](image-url)
3. Results

According to results of this research for peak discharge and based on computed residual sum of squares (RSS), LP3 distributions and ordinary moment method, P3 distribution and L-moment method, LN2 and LN3 distributions and ordinary moment method, have been suitable distinguished for 53%, 35%, 6% and 6% of stations, respectively (Table 2). Estimated $T$-year peak series using the best of probability distribution have shown in Table 3.

Table 2. Suitable probability distributions and methods for peak series

<table>
<thead>
<tr>
<th>Suitable Method</th>
<th>Suitable Probability Distribution</th>
<th>RSS</th>
<th>Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>L-moment</td>
<td>P3</td>
<td>40</td>
<td>Sarab hendeh</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>15</td>
<td>Abgarm</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LN3</td>
<td>5.9</td>
<td>Dehsomme</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>8.2</td>
<td>Sira</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>18</td>
<td>DFalhak</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>8.9</td>
<td>Chamriz</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>16.9</td>
<td>Baft</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>10.5</td>
<td>Safarzadeh</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>9.4</td>
<td>Dehrood</td>
</tr>
<tr>
<td>L-moment</td>
<td>P3</td>
<td>34.4</td>
<td>Adoori</td>
</tr>
<tr>
<td>L-moment</td>
<td>P3</td>
<td>11.2</td>
<td>Jirofto</td>
</tr>
<tr>
<td>L-moment</td>
<td>P3</td>
<td>10.6</td>
<td>Gharatolarab</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>14.1</td>
<td>Godarzarch</td>
</tr>
<tr>
<td>L-moment</td>
<td>P3</td>
<td>24.6</td>
<td>Namrood</td>
</tr>
<tr>
<td>L-moment</td>
<td>P3</td>
<td>22.8</td>
<td>Simindasht</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LP3</td>
<td>12.1</td>
<td>Bonkooh</td>
</tr>
<tr>
<td>Ordinary moment</td>
<td>LN2</td>
<td>15.3</td>
<td>Senobar</td>
</tr>
</tbody>
</table>

Table 3. Estimated $T$-Year peak series

<table>
<thead>
<tr>
<th>Station</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarab hendeh</td>
<td>80</td>
<td>120</td>
<td>144</td>
<td>162</td>
<td>188</td>
<td>205</td>
</tr>
<tr>
<td>Abgarm</td>
<td>73</td>
<td>90</td>
<td>99</td>
<td>107</td>
<td>121</td>
<td>130</td>
</tr>
<tr>
<td>Dehsomme</td>
<td>57</td>
<td>77</td>
<td>89</td>
<td>101</td>
<td>116</td>
<td>128</td>
</tr>
<tr>
<td>Sira</td>
<td>87</td>
<td>120</td>
<td>142</td>
<td>159</td>
<td>192</td>
<td>214</td>
</tr>
<tr>
<td>DFalhak</td>
<td>341</td>
<td>398</td>
<td>433</td>
<td>457</td>
<td>500</td>
<td>527</td>
</tr>
<tr>
<td>Chamriz</td>
<td>47</td>
<td>82</td>
<td>108</td>
<td>144</td>
<td>171</td>
<td>238</td>
</tr>
<tr>
<td>Baft</td>
<td>273</td>
<td>403</td>
<td>496</td>
<td>571</td>
<td>601</td>
<td>821</td>
</tr>
<tr>
<td>Safarzadeh</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>495</td>
</tr>
<tr>
<td>Dehrood</td>
<td>41</td>
<td>59</td>
<td>71</td>
<td>80</td>
<td>95</td>
<td>181</td>
</tr>
<tr>
<td>Adoori</td>
<td>41</td>
<td>59</td>
<td>71</td>
<td>80</td>
<td>95</td>
<td>495</td>
</tr>
<tr>
<td>Jirofto</td>
<td>30</td>
<td>51</td>
<td>60</td>
<td>68</td>
<td>77</td>
<td>181</td>
</tr>
<tr>
<td>Ghariatolarab</td>
<td>39</td>
<td>52</td>
<td>60</td>
<td>66</td>
<td>74</td>
<td>20</td>
</tr>
<tr>
<td>Godarzarch</td>
<td>30</td>
<td>51</td>
<td>60</td>
<td>68</td>
<td>77</td>
<td>105</td>
</tr>
<tr>
<td>Namrood</td>
<td>39</td>
<td>52</td>
<td>60</td>
<td>66</td>
<td>74</td>
<td>20</td>
</tr>
<tr>
<td>Simindasht</td>
<td>30</td>
<td>51</td>
<td>60</td>
<td>68</td>
<td>77</td>
<td>105</td>
</tr>
<tr>
<td>Bonkooh</td>
<td>97</td>
<td>140</td>
<td>169</td>
<td>198</td>
<td>237</td>
<td>266</td>
</tr>
</tbody>
</table>

For two methods sum of scores obtained in order to better comparison between distributions. First rank was gave to each distribution that RSS estimated was lowest and fifth rank for each distribution that RSS estimated was highest and in equal scores, ranking was similar. Finally sum of scores were computed for any distribution that according this method distribution was best that was accepted lowest score (Fig. 3).

4. Conclusion

The goal of this study was to select a set of suitable probability observed distributions for modeling peak series in arid and semi-arid regions of Iran. L-moment diagrams revealed that the Pearson type 3 (P3), log Pearson type 3 (LP3), Gumbel (G) distributions all provide acceptable approximations to the distribution of peak series in the study area (Figure 1), which means that other three and two-parameter alternatives LN3 and LN2 are not acceptable for the most parts of the study region. These results are similar to the other studies such as Pearson et al., 1991; Vogel et al., 1993b; Vogel and Wilson, 1996; Adamowski, 2000 and Kjeldson et al, 2002 that were used L-moment diagrams for choosing various distributional alternatives for annual maximum data in a region. Of all the models evaluated, the P3 distributions probably provide the best description of the distribution.
of peak series across the entire Iran central plateau, however, separation of central plateau into broad homogeneous regions can improve our ability to discriminate among potential flood flow frequency models such that the P3 distribution provides the best approximation to the distribution of peak series in the most of this region. Of the models tested, the LP3, P3 and LN2 distributions provides the best approximation to the distribution of peak series using ordinary moment method. Finally we concluded that L-moment method is suitable to determine peak series probability distributions in the Iran central plateau and P3 is the best probability distribution for modeling peak series in arid and semi-arid regions of Iran.

![Fig. 3. Sum of scores obtained from peak series for different statistical distribution in L-moment and ordinary moment](image)

**References**


