Flood Frequency Analysis of Annual Maximum Stream Flows using L-Moments and TL-Moments Approach

Ummi Nadiah Ahmad
Faculty of Science, Universiti Teknologi Malaysia
umminadiahahmad@gmail.com

Ani Shabri
Faculty of Science, Universiti Teknologi Malaysia
ani@mel.fs.utm.my

Zahrahtul Amani Zakaria
Faculty of Informatics, Universiti Sultan Zainal Abidin
amanizakaria@gmail.com

Abstract
Flood frequency analysis (FFA) is the estimation of how often a specified event will occur. Before the estimation can be done, analyzing the stream flows data are important in order to obtain the probability distribution of flood. By knowing the probability distribution, prediction of flood events and their characteristics can be determined. The aim of this study is to perform the FFA of annual maximum stream flows over station in Negeri Sembilan, Malaysia by using the L-Moments (LMOM) and TL-Moments with trimming only one smallest value (TLMOM1) approach. The most suitable distribution was determined by the use of MADI and moments ratio diagram. The result shows that GLO distribution to be the best distribution fitted the data of annual maximum stream flows for stations in Negeri Sembilan, Malaysia.

Keywords: L-Moments, TL-Moments, Generalized Logistic Distribution

1 Introduction

Floods are one of the natural disasters that occur not only in Malaysia, but also in other part of the world. It is also the most costly natural hazard since its ability to destroy human properties and lives. That is why a prevention act has to be taken
in order to help our country from unnecessary cost and economic loss as well as
preventing danger due to overflow of water in the country. Building a beverage,
dig a river and build a water reservoirs are a few examples of prevention act that
can be taken in order to prevent a flood from happening again.

However, these methods have its own disadvantages since this project requires a
lot of money and at the same time it can destroy the ecosystem of the river itself.
Therefore, in order to reduce the weakness, the ability to estimate the magnitude
of the flood in certain return period can help the hydrologist to design the
hydrological projects. Flood frequency analysis is the most suitable method in this
case.

The frequency analysis is the estimation of how often a specified event will occur
(Hosking and Wallis, 1997). The probability for future events can be predicted by
fitting the past observations to selected probability distributions. Chow et. al
(1988), relate the magnitude of these extreme events to their frequency occurrence
through the use of probability distributions.

Before the estimation can be done by using the FFA, analyzing the rainfalls and
stream flows data are important in order to obtain the probability distribution of
flood and other phenomenon related to them. By knowing the probability
distribution, prediction of flood events and their characteristics can be determined.
However, their used are often being complicated by certain characteristics of the
data such as the skewness and range of variation.

The data analysis often requires estimation of parameters for a few probability
distributions. In this study, we are focusing on L-Moments (LMOM) and
TL-Moments (TLMOM) methods. LMOM method which was introduced by
Hosking (1990) is a recent development in mathematical statistics which
facilitates the parameter estimation process in frequency analysis (Stedinger and
Lu, 1995). The method of LMOM has becoming a standard procedure in
estimating the parameters of certain statistical distributions and has been used
rapidly in hydrological field (Parida et al., 2008), Betul Saf (2009), (Kumar et al.,
1990).

Elamir and Seheult (2003) introduced an extension of LMOM called TLMOM.
TLMOM assigns zero weight to extreme observations and is claimed to be more
robust against outliers compared to LMOM method. It is also can exist even if the
distribution does not have a mean such as Cauchy distribution (Abdul Moniem,
2007). A few studies have been made by researcher regarding this method
(Asquith, 2007), (Hosking, 2007), (Abdul Moniem, 2007), and (A. Moniem and
M. Selim, 2009). Related issues in TLMOM method is the choice of amount of
trimming. All researcher are using the equal trimming value where one smallest
and one largest value were trimmed from the sample.

However, based on Cunnane (1987) and Wang (1990), they claimed that by
cencoring the data from below might be advantageous since the smaller sample
values have only a nuisance value in the context of upper quantile estimation and also in model form testing and verification. Based on this idea, we are focusing on trimming only one smallest value \((t_1 = 1, t_2 = 0)\) from our conceptual sample which we denote as TLMOM1.

This paper discussed the LMOM and TLMOM1 approaches in flood frequency analysis of annual maximum stream flows over stations in Negeri Sembilan, Malaysia. The concept of this study was to find the best distribution among the selected distribution which are generalized logistic distribution (GLO), generalized extreme value type 1 (GEV) and generalized pareto distribution (GPA), whose parameter were estimated using LMOM and TLMOM1 approaches.

2 Methodology

2.1 LMOM Method

LMOM has been defined by Hosking (1990) as a linear combination of probability weighted moments (PWM's). Let \(X_1, X_2, \ldots, X_r\) be a conceptual random sample of size \(r\), and \(X_{1r} \leq X_{2r} \leq \ldots \leq X_{rr}\) denote the corresponding order statistics. The \(r^{th}\) LMOM defined by Hosking (1990) is:

\[
\lambda_r = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k})
\]

The first four LMOM are defined as:

\[
\begin{align*}
\lambda_1 & = E(X_{1,1}) \\
\lambda_2 & = \frac{1}{2} E(X_{2,2} - X_{1,2}) \\
\lambda_3 & = \frac{1}{3} E(X_{3,3} - 2X_{2,3} + X_{1,3}) \\
\lambda_4 & = \frac{1}{4} E(X_{4,4} - 3X_{3,4} + 3X_{2,4} + X_{1,4})
\end{align*}
\]

LMOM ratios \(\tau_3\) and \(\tau_4\) called L-Skewness and L-Kurtosis are calculated as:

\[
\tau_3 = \frac{\lambda_3}{\lambda_2} \quad \tau_4 = \frac{\lambda_4}{\lambda_2}
\]
The sample LMOM can be estimated from a sample order statistics $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ (Asquith, 2007):

$$l_r = \frac{1}{r} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \binom{n-r}{n-j} X_{i,n}$$

(4)

2.2 TLMOM1 Method

TL-Moments method has been introduced by Elamir and Seheult due to awareness towards the outliers. TLMOM is a robust modification of LMOM in which $E(X_{r-k:r})$ is replaced by $E(X_{r+t_1-k, r+t_1+t_2})$ for each $r$ where $t_1$ smallest and $t_2$ largest are trimmed from the conceptual sample. Elamir and Seheult (2003) define $r^{th}$ TLMOM as:

$$\lambda_r^{(t_1,t_2)} = \frac{1}{r} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r+t_1-k, r+t_1+t_2})$$

(5)

In this study, we proposed unequal trimming value where only one smallest value will be trimmed from our conceptual sample $(t_1 = 1, t_2 = 0)$. The first four TLMOM1 can be defined as:

$$\lambda_1^{(1,0)} = E(X_{2,2})$$
$$\lambda_2^{(1,0)} = \frac{1}{2} E(X_{3,3} - X_{2,3})$$
$$\lambda_3^{(1,0)} = \frac{1}{3} E(X_{4,4} - 2X_{3,4} + X_{2,4})$$
$$\lambda_4^{(1,0)} = \frac{1}{4} E(X_{5,5} - 3X_{4,5} + 3X_{3,5} - X_{2,5})$$

(6)

TLMOM1 ratio $\tau_3^{(1,0)}$ and $\tau_4^{(1,0)}$ called TL-skewness and TL-kurtosis are calculated as:

$$\tau_3^{(1,0)} = \frac{\lambda_3^{(1,0)}}{\lambda_2^{(1,0)}} \quad \tau_4^{(1,0)} = \frac{\lambda_4^{(1,0)}}{\lambda_2^{(1,0)}}$$

(7)

The sample TLMOM1 can be estimated from a sample order statistics $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ (Hosking, 2007):
Flood frequency analysis

\[ f_r^{(1,0)} = \frac{1}{n} \sum_{i=2}^{n} \sum_{k=0}^{r-i} (-1)^k \binom{r-i}{k} \binom{r}{k} X_{in} \] \hfill (8)

2.3 Parameter Estimation

Before the analysis can be done, the parameter for each selected distribution needs to be estimated first. In this study, LMOM and TLMOM1 approaches were used to estimate the parameter of three selected distribution which are GLO, GPA and GEV distribution since they are commonly used in flood frequency analysis. GLO and GEV distribution has been proven to be the best distribution to fit the data over stations in Sarawak, Malaysia and United Kingdom (Ashkar and Mahdi, 2006; Lim and Lye, 2003). Table 1 and Table 2 show the parameter estimates for the selected distribution using LMOM and TLMOM1 approaches.

Table 1: LMOM Parameter Estimates for Selected Distribution

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Quantile Functions</th>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLO</td>
<td>[ X(F) = \xi + \frac{\alpha}{K} \left{ 1 - \left( \frac{1-F}{F} \right)^K \right} ]</td>
<td>[ \alpha = \frac{l_2}{\Gamma(1+K)\Gamma(1-K)} ] [ \xi = l_1 + \frac{(l_2 - \alpha)}{K} ] [ K = \frac{-\alpha}{\Gamma(1-K)} ]</td>
</tr>
<tr>
<td>GEV</td>
<td>[ X(F) = \xi + \frac{\alpha}{K} \left{ -\ln F \right}^K ]</td>
<td>[ \alpha = \frac{l_2 K}{\Gamma(1+K)\Gamma(1-2^{-\xi})} ] [ \xi = l_1 + \frac{\alpha(\Gamma(1+K) - 1)}{K} ] [ K = 7.8590C + 2.9554C ] [ C = \frac{2}{3+\xi} ]</td>
</tr>
<tr>
<td>GPA</td>
<td>[ X(F) = \xi + \frac{\alpha}{K} \left{ (1-F)^K \right} ]</td>
<td>[ \alpha = l_2 \left( \frac{(K+1)(K+2)}{4} \right) ] [ \xi = l_1 + l_2 (K+2) ] [ K = \frac{4}{\xi + 1} - 3 ]</td>
</tr>
</tbody>
</table>
Table 2: TLMOM1 Parameter Estimates for Selected Distribution

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Quantile Functions</th>
<th>Parameter Estimates</th>
</tr>
</thead>
</table>
| GLO          | \( X(F) = \xi + \frac{\alpha}{K} \left[ 1 - \left( \frac{1 - F}{F} \right)^K \right] \) | \( \alpha = \frac{4Kl_2}{6\Gamma(2-K)\Gamma(1+K) - 3\Gamma(3-K)\Gamma(K+1)} \)  
\[ \xi = l_1 - \frac{\alpha}{K} + \frac{\alpha\Gamma(2-K)\Gamma(K+1)}{K} \]  
\( K = \frac{4 - 27t_3}{20} \) |
| GEV          | \( X(F) = \xi + \frac{\alpha}{K} \left[ -\ln(1-F)^K \right] \) | \( \alpha = \frac{-2(6)^K l_2}{3\Gamma(1+K)(2^K - 3^K)} \)  
\[ \xi = l_1 + \frac{\alpha}{K} \left( \frac{\alpha\Gamma(K)}{2^K} \right) \]  
\( K = 0.49 - 2.08t_3 + 0.61(t_3)^2 - 0.60(t_3)^3 + 0.48(t_3)^4 \) |
| GPA          | \( X(F) = \xi + \frac{\alpha}{K} \left[ -\ln(F)^K \right] \) | \( \alpha = \frac{l_2(K+1)(K+2)(K+3)}{3} \)  
\[ \xi = l_1 + \alpha \left[ \frac{K+3}{(K+1)(K+2)} \right] \]  
\( K = \frac{4 - 12t_3}{3t_3 + 4} \) |

3 Comparison of Probability Distribution

3.1 Mean Square Deviation Index (MADI)

For the comparison among the probability distribution for fitting the data used in this study, mean square deviation index (MADI), was used. The aim was to test whether a given distribution fits the data acceptably closely and choose from a number of candidate distributions, the one that gives the best fits to the data. The MADI can be calculated by:

\[
MADI = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{x_i - y_i}{x_i} \right| 
\]  
(9)

Where \( x_i \) are the observed value while \( y_i \) are the predicted value with \( N \) number of stations. The smaller the value obtained for MADI of a given distribution, shows that it is more fitted for the actual data. Hence, the distribution with smaller value of MADI shows that it is more fitted to the observed data.
3.2 LMOM and TLMOM1 Ratio Diagram

The easiest way in determining the best distribution to fit the actual data is by using the LMOM and TLMOM1 ratio diagrams. These diagrams give a visual indication of which distributions may be expected to give a good fit to a data samples. The distribution nearest the coordinate of the sample LMOM and TLMOM1 ratios is taken as the best distribution to fit the actual data and the furthest is the distribution least fit to represent the data. Some useful relationships for constructing the LMOM and TLMOM1 ratio diagram are given by Hosking (1990, 2007) and Vogel and Fennessey (1993).

4.0 Case Study

The data of annual maximum stream flows for 11 stations in Negeri Sembilan, Malaysia have been provided for this study by Department of Irrigation and Drainage, Ministry of Natural Resources and Environment, Malaysia which range from 1960 to 2009. The catchment areas of this sites ranging from 13 to 1210km² and their mean annual maximum stream flows vary from 20 to 142.40m³/s. Only data from sites with at least ten years in record length were used in this study, see (Noto and Logia, 2009). The values of MADI for each distribution using different methods were collected and the number of times the distribution obtained a given rank was summed up and the totals of each rank were put in the table.

5.0 Results and Discussion

All the maximum values of annual maximum stream flows for each year for all 11 stations were analyzed using Mathcad and their distributions were compared using MADI and ratio diagram for both LMOM and TLMOM1. Each distribution was ranked according to their MADI from the best fit to the least.

Table 3 and Table 4 show the MADI obtained by using LMOM and TLMOM1 method for all selected distributions which are GLO, GEV and GPA distribution. Table 5 and Table 6 show the ranking of the distribution based on the value of MADI for both LMOM and TLMOM1 methods with 1 being the best distribution. Clearly from the table, GLO distribution shows the best fit for both LMOM and TLMOM1 methods, followed by GEV and GPA distributions.

A convenient way of representing the LMOM of different distribution is the LMOM ratio diagram (Hosking, 1997). The sample of LMOM and TLMOM1 ratio for each distribution is taken for the range $-1 \leq t_3 \leq 1$. For this interval, the
$t_4$ is calculated for all the distribution using their relationship with $t_3$, then the values for the sample LMOM and TLMOM1 ratios were calculated as a point in the diagram (Shabri and Ariff, 2009).

Table 3: MADI for stations in Negeri Sembilan using LMOM

<table>
<thead>
<tr>
<th>Station No</th>
<th>GLO</th>
<th>GEV</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2519421</td>
<td>0.081</td>
<td>0.087</td>
<td>0.113</td>
</tr>
<tr>
<td>2520423</td>
<td>0.148</td>
<td>0.120</td>
<td>7.948</td>
</tr>
<tr>
<td>2524416</td>
<td>0.081</td>
<td>0.084</td>
<td>0.175</td>
</tr>
<tr>
<td>2525415</td>
<td>0.259</td>
<td>0.236</td>
<td>0.140</td>
</tr>
<tr>
<td>2619401</td>
<td>0.046</td>
<td>0.053</td>
<td>0.064</td>
</tr>
<tr>
<td>2619402</td>
<td>0.004</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>2625412</td>
<td>0.143</td>
<td>0.145</td>
<td>0.239</td>
</tr>
<tr>
<td>2719422</td>
<td>0.057</td>
<td>0.058</td>
<td>0.066</td>
</tr>
<tr>
<td>2722413</td>
<td>0.171</td>
<td>0.222</td>
<td>0.280</td>
</tr>
<tr>
<td>2920432</td>
<td>0.166</td>
<td>0.151</td>
<td>0.178</td>
</tr>
<tr>
<td>3022431</td>
<td>0.087</td>
<td>0.099</td>
<td>0.184</td>
</tr>
</tbody>
</table>

Table 4: MADI for stations in Negeri Sembilan using TLMOM1

<table>
<thead>
<tr>
<th>Station No</th>
<th>GLO</th>
<th>GEV</th>
<th>GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2519421</td>
<td>0.092</td>
<td>0.097</td>
<td>0.118</td>
</tr>
<tr>
<td>2520423</td>
<td>0.261</td>
<td>0.121</td>
<td>7.977</td>
</tr>
<tr>
<td>2524416</td>
<td>0.081</td>
<td>0.097</td>
<td>0.274</td>
</tr>
<tr>
<td>2525415</td>
<td>0.548</td>
<td>0.423</td>
<td>0.186</td>
</tr>
<tr>
<td>2619401</td>
<td>0.044</td>
<td>0.05</td>
<td>0.082</td>
</tr>
<tr>
<td>2619402</td>
<td>0.006</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>2625412</td>
<td>0.224</td>
<td>0.276</td>
<td>0.418</td>
</tr>
<tr>
<td>2719422</td>
<td>0.076</td>
<td>0.081</td>
<td>0.091</td>
</tr>
<tr>
<td>2722413</td>
<td>0.320</td>
<td>0.353</td>
<td>0.432</td>
</tr>
<tr>
<td>2920432</td>
<td>0.248</td>
<td>0.181</td>
<td>0.196</td>
</tr>
<tr>
<td>3022431</td>
<td>0.118</td>
<td>0.144</td>
<td>0.214</td>
</tr>
</tbody>
</table>

Table 5: Ranking of the distribution for 11 stations by MADI using LMOM approach (on a scale 1 to 3 with 1 being the best)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Number of times a distribution had the ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>GLO</td>
<td>7</td>
</tr>
<tr>
<td>GEV</td>
<td>2</td>
</tr>
<tr>
<td>GPA</td>
<td>2</td>
</tr>
</tbody>
</table>
Table 6: Ranking of the distribution for 11 stations by MADI using TLMOM1 approach (on a scale 1 to 3 with 1 being the best)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Number of times a distribution had the ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>GLO</td>
<td>7</td>
</tr>
<tr>
<td>GEV</td>
<td>2</td>
</tr>
<tr>
<td>GPA</td>
<td>2</td>
</tr>
</tbody>
</table>

The average values of sample LMOM and TLMOM1 ratios, $t_3$ and $t_4$ were calculated as follows:

$t_3 = 0.338$ and $t_4 = 0.250$

$h_3 = 0.367$ and $h_4 = 0.238$

The sample of L-Skewness and L-Kurtosis values for stations in Negeri Sembilan with at least 10 years of record are shown in Figure 1. Based on the graph, clearly that GLO distribution was appropriate for the data because the average values of sample skewness and kurtosis were consistent with the population skewness and kurtosis of GLO distribution. Figure 2 shows the TLMOM1 ratio diagrams for 11 stations in Negeri Sembilan. The data are on average closer to population TLMOM1 of a GLO distribution. We therefore can conclude that the annual maximum stream flows data for 11 stations over Negeri Sembilan may be well described by a GLO distribution rather than GEV and GPA distributions.
6 Conclusion

By using MADI, LMOM and TLMOM1 ratio diagram, the best distribution to fit the data of annual maximum stream flows over stations in Negeri Sembilan, Malaysia was obtained. Three distributions were involved in these studies which are GLO, GEV and GPA distribution. From the result, GLO distribution consistently shows the best fit followed by GEV distribution for both LMOM and TLMOM1 method since it is always rank the first compared to the other distribution. It is also can be proven by the use of LMOM and TLMOM1 ratio diagram where GLO distribution shows the closest to the point of the sample LMOM and TLMOM1 ratio. However, GPA distribution seems unsuitable since it is ranked the third for MADI which is less frequently among the three distributions.

References


Received: June, 2010