Regional Flow Frequency Analysis on Peninsular Malaysia using l-moments

*Aminuddin Mohd Baki[1], Doris Asmani Mat Yusof [2], Ismail Atan[3] and *Nora Farina Mohd Halim[4].*

[1]*Envirab Services, P.O Box 7866, 40730 Shah Alam, Selangor*
[2]*Fakulti Kejuruteraan Awam, Universiti Teknologi MARA Pasir Gudang, 81750 Masai,Johor*
[3]*Fakulti Kejuruteraan Awam, Universiti Teknologi MARA Shah Alam, 40000 Shah Alam, Selangor*
[4]*Fakulti Kejuruteraan Awam, Universiti Teknologi MARA Pasir Gudang, 81750 Masai,Johor*

norafarina@ymail.com

Abstract

Through this study, regional estimation has been proposed as an alternative technique to estimate design runoff. Regional frequency analysis of annual runoff in Peninsular Malaysia has been conducted and was based on the theory of L-Moments as developed by Hosking and Wallis (1997) is a very reliable method for assessing exceedance probabilities of extreme environmental events when data is available from more than one site. This recommended technique will definitely enhance the accuracy of the design runoff and at the same time will minimize the uncertainties of observed data. The objectives of this study are to identify the homogeneous region and to establish the best fitted distribution for each region. The study area was separated into five homogeneous regions. The General Logistics (GLO) was found to be the best distribution for the region in peninsular Malaysia followed by Generalized Extreme Value (GEV). The best fit distributions of annual runoff in Malaysia are discovered based on L-moment ratio diagram and statistical tests. Numerical analysis was carried out on some stations in every state in Peninsular Malaysia. Thus, the analysis performed are using data from homogeneous regions determine using L-moments process. The Frequency analysis using a regional approach based on L-moments method was detected to be practical in the estimation of design runoff at ungauged site.

Keywords: L-Moment, Homogeneous, Regional Frequency Analysis, Best Fitted Distributions.

Introduction

Clustering is one of the techniques that are useful to screen catchments in a region into ordinary groups. For each sub-region, growth curves must be constructed and the value of an index flood must be related to catchments characteristics. The regional *L*-moment algorithm may be used as an advantage both to identify an appropriate underlying frequency distribution and to construct sub-regional growth curve (Abdullah et al., 2010). The regional frequency analysis is widely used in flood analysis which is based on the L-moments theory that recently developed by Hosking and Wallis (1997). Being a linear combination of the data and the data which were less influenced by the outliers are the most important advantage of L-moments (Amin, 2001). The advantage of using regional frequency analysis approach over conventional approach is evident and also has been successfully implemented in many studies that have been done before. This paper contains the application of the methodology proposed by Hosking and Wallis (1997) for the estimation of the annual runoff distribution in any Malaysian rivers by using hydrometric and climatic data of Malaysian gauged stations.

This study involves data collection work and numerical modelling. The data collection work involves taking the annual runoff data of selected location all over peninsular Malaysia from Department of Irrigation And Drainage, Ampang. Numerical modelling involved analyzing the annual runoff data to find the L-Skewness and L-Kurtosis from different sites of annual runoff data.

In this study the regional frequency analysis is based on the theory of L-Moment as developed and proposed by Hosking and Wallis (1997). In order to differentiate between the L-Skewness and L-Kurtosis relations of different distributions and data samples, L-Moment ratio diagram can be used (Cannarozzo *et al.*, 2009). This would result in the identification of which distribution may be expected to give a good fit to a data sample.

The amounts of stations with homogenous data were 44 stations. The study area was divided into five homogeneous regions which are R1, R2, R3, R4 and R5 as shown in Figure 1. Six (6) stations employed in REGION I, twelve (12) stations employed in REGION II, eleven (11) stations employed in REGION III, six (6) stations employed in REGION IV and nine (9) stations employed in REGION V. The annual maximum flow data are acquired from the Department of Irrigation and Drainage (DID) with database vary from 20 years to 50 years relying on the establishment of the stations. REGION I, II, III, IV, and V correspond to the South, West and East of Peninsular Malaysia, respectively. These regions were selected because they have similar geology, hydrogeology and climatic aspects. The stations numbers, river names and period of the record are summarized in Table 1 to Table 5 for R1, R2, R3, R4 and R5, respectively.

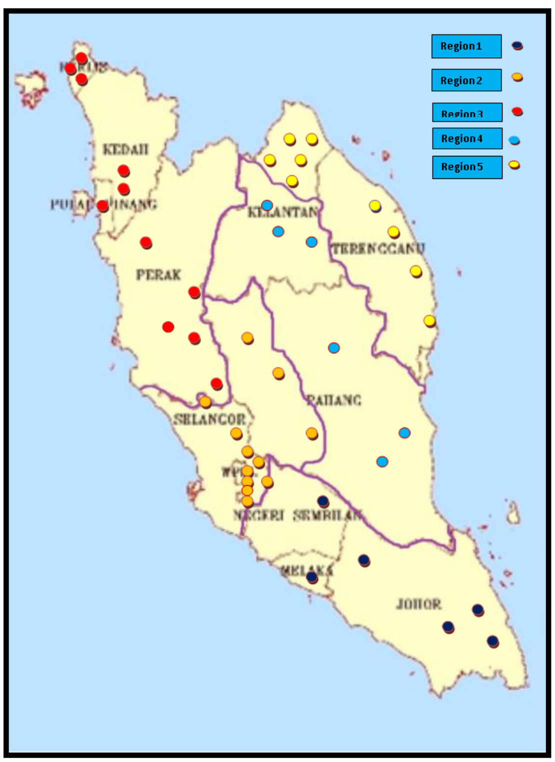

Figure 1. Selected hydrological gauge stations.

Table 1. Summary of selected Automatic Flow Stations for Region 1 (R1)

No	*Station Id.*	*Station Name*	*Noof yearsin Analysis*
1	2527411	Sg. Muar At Buloh Kasap,Johor	44
2	2235401	Sg. Kahang At Bt.26 Jln Kluang	32
3	1931423	Sg. Sembrong At Bt.2 Air Hitam,Y.Peng	30
4	1836402	Sg. Sayong At Jam. Johor Tenggara	33
5	2322413	Sg. Melaka At Pantai Belimbing	50
6	3022431	Sg. Triang At Juntai,N.Sembilan	45

Table 2. Summary of selected Automatic Flow Stations for Region 2 (R2)

No	***Station Id.***	***Station Name***	***No of Years in Analysis***
1	2816441	Sg. Langat At Dengkil,Selangor	50
2	2918401	Sg. Semenyih At Kg.Rinching, Selangor	36
3	2918443	Sg. Semenyih At Semenyih,Selangor	14
4	3615412	Sg. Bernam At Tanjung Malim,Selangor	50
5	3813411	Sg. Bernam At Jam.Skc, Selangor	49
6	3118445	Sg. Lui At Kg. Lui,Selangor	45
7	3116430	Sg.Klang At Jam.Sulaiman	37
8	3116432	Sg. Klang At Leboh Pasar, K.Lumpur	14
9	3116433	Sg. Gombak At Jln. Tun Razak	49
10	3519426	Sg. Bentong At Jam. Kuala Marong	41
11	4019462	Sg. Lipis At Benta, Pahang	45
12	4218416	Sg. Jelai At Kuala Medang, Pahang	29

Table 3. Summary of selected Automatic Flow Stations for Region 3 (R3)

No	***Station Id.***	***Station Name***	***No of Years in Analysis***
1	6502401	Sg. Jerneh At Titi Tampang,Perlis	27
2	6502432	Sg. Tasoh At Titi Baru,Perlis	20
3	6503401	Sg. Arau At Ldg. Tebu Felda,Perlis	27
4	5806414	Sg. Muda At Jeniang, Kedah	47
5	5405421	Sg. Kulim At Ara Kuda,P.Pinang	47
6	5505412	Sg. Muda At Ldg.Victoria, P.Pinang	50
7	3913458	Sg. Sungkai At Sungkai,Perak	49
8	4212467	Sg. Cenderiang At Bt. 32 Jln Tapah	38
9	4310401	Sg. Kinta At Weir G, Tg. Tualang	34
10	4611463	Sg. Kinta At Tg.Rambutan, Perak	49
11	5106433	Sg. Ijok At Titi Ijok,Perak	45

Table 4. Summary of selected Automatic Flow Stations for Region 4 (R4)

No	*Station Id.*	*Station Name*	*No of years in Analysis*
1	5222452	Sg. Lebir At Tualang, Kelantan	34
2	5320443	Sg. Galas At Dabong, Kelantan	36
3	5419401	Sg. Pergau At Bt. Lembu, Kelantan	16
4	3629403	Sg. Lepar At Jam. Gelugor,Pahang	38
5	4223450	Sg. Tembeling At Kg. Merting,Pahang	20
6	3527410	Sg.Pahang At Lubok Paku,Pahang	33

Table 5. Summary of selected Automatic Flow Stations for Region 5 (R5)

No	***Station Id.***	***Station Name***	***No of Years in Analysis***
1	5621401	Sg. Sokor At Kg. Tegawan, Kelantan	15
2	5721442	Sg. Kelantan At Jam.Guillemard, Kel.	50
3	5818401	Sg. Golok At Kg. Jenob, Kelantan	27
4	6019411	Sg.Golok At Rantau, Panjang	45
5	6022421	Sg. Kemasin At Peringat,Kelantan	29
6	4232452	Sg. Kemaman At Rantau, Panjang, Ter	35
7	5130432	Sg. Terengganu At G.Tanggol, Terengganu	49
8	5229436	Sg. Nerus At Kg. Bukit,Terengganu	29
9	5428401	Sg. Chalok At Jam. Chalok,Terengganu	32

Methodology

L-Moment

L-moments are statistics used to summarize the shape of a probability distribution and also define as linear combination of Probability Weighted Moments (PWMs) (Greenwood *et al*., 1979). They are analogous to conventional moments in that they are used to calculate quantities analogous to the mean, standard deviation, skewness and kurtosis of data. In the L-moment field these terms are called L-mean, L-scale, L-skewness and L-kurtosis to distinguish them from the conventional moments and standardized moments. Standardized L-moments are called L-moment ratios (Hosking and Wallis, 1993).

For a random variable X, the first four population L-moments are (Hosking, 1986):

$$\begin{aligned} \lambda_1 &= EX \\ \lambda_2 &= (EX2{:}2 - EX1{:}2) / 2 \\ \lambda_3 &= (EX3{:}3 - 2EX2{:}3 + EX1{:}3) / 3 \\ \lambda_4 &= (EX4{:}4 - 3EX3{:}4 + 3EX2{:}4 - EX1{:}4) / 4, \end{aligned} \quad (1)$$

where X_k:n denotes the kth largest value in an independent sample of size n from the distribution of X.

The first two of these L-moments have conventional names:

λ_1 = mean, L-mean or L-location,
λ_2 = L-scale.

A quantity similar to the coefficient of variation, but based on L-moments is τ defined by

$$\tau = \frac{\lambda_2}{\lambda_1}, \tag{2}$$

which is called the "coefficient of L-variation", or "L-CV".

A set of higher order scaled L-moments, or L-moment ratios, is defined by

$$\tau_r = \frac{\lambda_r}{\lambda_2}, \qquad r = 3, 4, \ldots, \tag{3}$$

and conventional names for the third and fourth order L-moment ratios (r =3, 4) are

τ_3 = L-skewness,
τ_4 = L-kurtosis

Index Flood Method

The *T*-year event *XT* defined as the event of exceeded on average once every *T* years (Stedinger et al., 1993) is given as

$$P[X > XT] = 1/T \tag{4}$$

The parameters of regional frequency distribution can b estimated using L-moment together with regional average sample L-moments ratios. Then, the *T*-year event at site *i* can be estimated as

$$XT,i = u_i\, ZT \tag{5}$$

where u_i is the mean annual flood (MAF) at site *i*, and *ZT* is the regional growth curve.

L-Moments Ratio Diagram

L-Moments ratio diagram is a graphical analysis of distribution fitting where several distributions can be compared in a single diagram. Theoretical relationship between τ_3 and τ_4 have been derived for various distribution (Hosking,1990). The selection of a suitable parametric distribution to describe the data from several sites can be based on the proximity of the mean value of τ_3 and τ_4 for the region to a theoretical line or point as well as on the variability of the mean values.

Discordancy Measures

This measure is used to identify site that are grossly outliers with the group as whole. If a single site does not appear to belong to the cloud of τ_3 and τ_4 points on the L- Moment diagram, a test of discordancy based on L-Moments can be used to determine whether it should be removed from the region. The test of discordancy is applied by calculating the D-statistics, defined terms of L-Moments.

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T K^{-1} (u_i - \bar{u}) \tag{6}$$

where u_i = vector of *L*-moment ratios for station *i*,
K = covariance matrix of u_i,
$\bar{u}$ = mean of vector u_i.

The station *i* is declared to be discordant, if D_i is greater than the critical value of the discordancy statistic given in a tabular form by (Hosking and Wallis, 1993).

Heterogeneity Measure

Heterogeneity measure, *H* compares the inter-site variations in sample L-moments for the group of sites with what would be expected of a homogeneous region. The L-moment tests for heterogeneity fit a four-parameter Kappa distribution (Hosking, 1994) to the regional data set, generate a series of 500 equivalent region's data by numerical simulation and compare the variability of the statistics of the actual region to those of the simulated series (Hosking and Wallis, 1997). The inter-site variation of each generated region is computed and the mean and standard deviation of the computed inter-site variation is determined. Hence, heterogeneity measure is obtained as:

$$H = \frac{V - \mu V}{\sigma V} \tag{7}$$

where *V*= weighted standard deviation of L- coefficient of variation values, μV, σV = the mean and standard deviation of number of simulations of *V*.

The criteria for deciding heterogeneity of a region is as:

H < 1, region is acceptably homogeneous,
1 ≤ H < 2, region is possibly heterogeneous,
H ≥ 2, region is definitely heterogeneous.

Discussions

The result or outcome of the regionalization frequency analysis will be present in tables and graphs. First the weighted mean of L-Skewness and L-Kurtosis must be plotted into the L- Moment Ratios Diagram. This result was obtained by running the X-Test and X-Fit Program using FORTRAN Software**.** The diagram was plotted in order to view the appropriate distributions as guidance through graphical instrument by comparing the fit of several distributions to many samples of data in a group of region. Further analysis to confirm the true frequency distribution was carried out by using Goodness-of Fit Test.

Figure 2 to Figure 6 show the results of weighted mean of L-kurtosis and L-skewness that have been plotted into L-Moment Ratio Diagram for those five regions which are R1, R2, R3, R4 and R5, respectively. Based on Figure 2, it can be summarized that Generalized Extreme Value Distribution (GEV) and Generalized Logistics Distribution (GLO) are the best potential distribution for Region 1. Meanwhile for Figure 3 displays that Generalized Logistics Distribution (GLO) is the best potential distribution for regional frequency analysis for Region 2 same goes to Region 3 and Region 4 that is represented by Figure 4 and Figure 5, respectively. Figure 6 illustrates that although the sites are scattered, from the mean it shows Generalized Extreme Value Distribution (GEV) is the best distribution for Region 5.

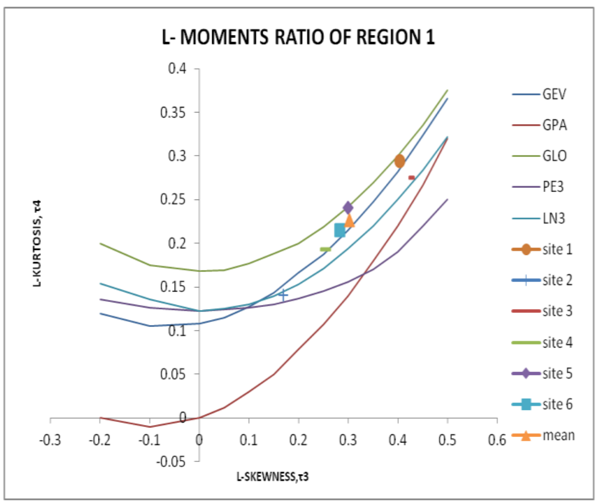

Figure 2. L-Moment Ratio diagram of Region 1

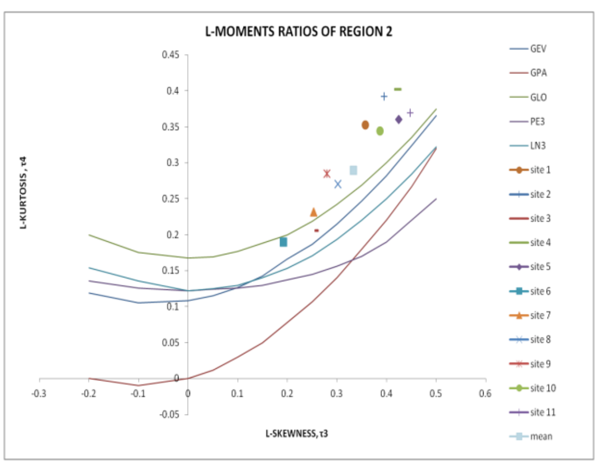

Figure 4. L-Moment Ratio diagram of Region 3

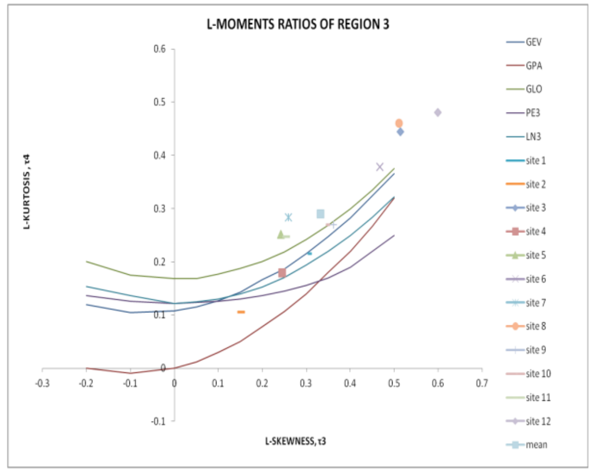

Figure 3. L-Moment Ratio diagram of Region 2

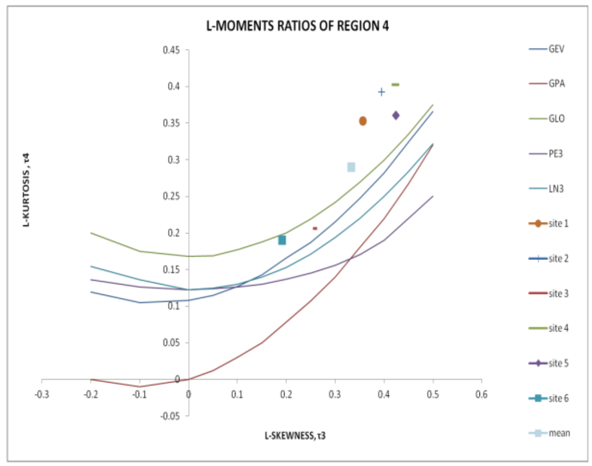

Figure 5. L-Moment Ratio diagram of Region 4

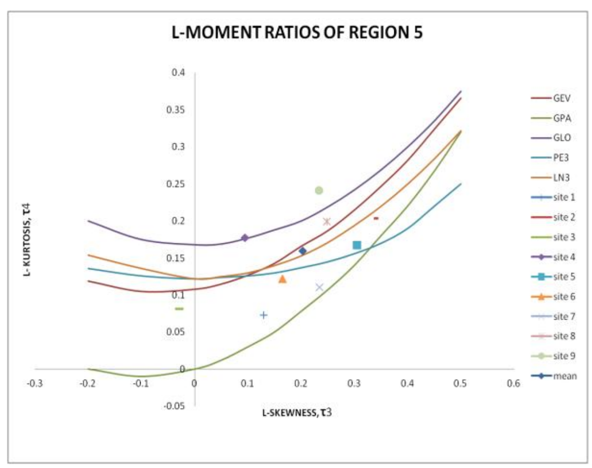

Figure 6. L-Moment Ratio diagram of Region 5

The Goodness of fit Test (ZDIST) should be less than or equal to 1.64 but if the value exceeded 1.64 the distribution with value ZDIST close to it will be selected as a parent distribution. From Table 6 indicates that for R1, generalized logistic (GLO) seems to be the best choice of regional frequency distribution and same goes to R2, R3 and R4 which GLO is the best choice of distribution. Turning to R5, it shows that generalized extreme value (GEV) is the best chosen distribution for that particular region.

Table 6. Goodness-of-Fit Test (Z^{DIST}) for all 5 regions.

(Z^{DIST})	R1	R2	R3	R4	R5
GLO	0.11[a]	0.52[a]	0.98[a]	0.27[a]	1.66
GEV	-0.59[b]	-0.43[b]	-0.08[b]	-0.03[b]	0.08[a]
GNO	-1.15[c]	-1.18[c]	-0.93[c]	-0.12[c]	-0.30[b]
PE3	-.2.11[d]	-2.45[d]	-1.12[d]	-0.96[d]	-1.09[c]
GPA	-2.48	-3.38	-2.32	-2.11	-3.54[d]

Conclusions

In this study, the Regionalization frequency analysis is based on L-Moments method and it has been used for parameter estimation, homogeneity testing and the choice of the regional distribution. The assessment of the value of index runoff was established by applying the index flood approach. Besides focusing on the establishment of the index runoff value, the study was also used in identifying the homogeneous region and the development of the probability plots for the region.
For Region 1, Region 2, Region 3 and Region 4, from the L-moment ratio diagram and also from the goodness-of-fit test revealed that the GLO distribution provides the best approximation of annual maximum flow while for Region 5 the GEV distribution appear to be the best fitted distribution of observed annual maximum flow data.

Acknowledgment

The funding provided by the Research Management Institute Universiti Teknologi Mara [Dana Kecemerlangan, 600-IRDC/ST/DANA 5/3/Dst (20/2010)] and the organisations providing data to the project which is Department of Irrigation and Drainage (DID) Ampang are thanked .The authors wish to thank all family and friends who help to done this research directly or indirectly. Contributions from all relevant parties involved in the study are gratefully acknowledged.

References

Abdullah A. Maimun, Alias Hashim and Jamal I. Daoud (2010). Regionalisation of low flow frequency curves for the Peninsular Malaysia, *Journal of Hydrology,* 381, 174-180.

Amin, M.Z.M. (2001). *Estimation of Design Rainstorm – Regionalization Approach by L-moments Method*. Paper presented at the International Symposium of Achievement of IHP V in Hydrological Research in Southeast Asia and Pacific Hanoi, Vietnam, 19 November 2001.

Cannarozzo M.,Noto L. V., Viola F., La Loggia G. (2009). Annual Runoff regional frequency analysis in Sicily, *Journal of Physics and Chemistry of the Earth,* 34, 679-687.

Greenwood J.A, Landwehr J.M., Matalas N.C, and Wallis J.R, (1979) " Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form," *Water Resources Research*, 15(5), 1049, 81-105.

Hosking J.R.M, (1986) "The theory of probability weighted moments," *IBM Research Division*, Yorktown Heights, New York, IBM Research Report. RC 12210.
Hosking J.R.M, (1990) "L-moments: analysis and estimation of distribution using linear combinations of order statistics," *Journal of the Royal Statistical Society*, B, 52, 105-124.

Hosking J.R.M, (1994) "The four-parameter kappa distribution," *IBM Journal of Research and Development*, 38, 251-8.

Hosking J.R.M and Wallis J.R, (1993) "Some statistics useful in regional frequency analysis," *Water Resource Research*, 29, 271-281.

Hosking, J. R. M., and Wallis, J. R. (1997). Regional frequency analysis : an approach based on L-Moments. *Cambridge University Press,* Cambridge, U.K.

Mohd Halim N.F, Baki A., and Atan I. (2010). *Regionalisation Frequency Analysis of Annual Runoff in Malaysia*. Proceeding of Malaysian Science and Technology Congress (MSTC 2010) Petaling Jaya, Malaysia, 9-11 November 2010.

Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., (1993).Frequency analysis of extreme events. *Handbook of Hydrology*. McGraw-Hill, New York.