FREQUENCY ANALYSIS OF ANNUAL RUNOFF IN PENINSULAR MALAYSIA USING A REGIONAL APPROACH BASED ON L-MOMENTS

(Date received: 10.05.12/Date accepted: 22.12.12)

Aminuddin Mohd. Baki[1], Ismail Atan[2] and Nora Farina Mohd. Halim[3]

Faculty of Civil Engineering, Universiti Teknologi MARA (UiTM)
40450 Shah Alam, Selangor, Malaysia.
E-mail: [1]amin_ba_ki@yahoo.co.uk, [2]ismailatan585@gmail.com, [3]norafarina@ymail.com

ABSTRACT

Estimation of runoff is very important for flood estimation. Thus, runoff is very important in designing and managing water resources projects involving natural hazards. Region refers to a set of sites of which is expected to have data drawn in a synonymous frequency distribution. In this study, a 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 [1] is a very reliable method for assessing exceedance probabilities of extreme environmental events when data is available from more than one site. The General Logistics (GLO) and Generalized Extreme Value (GEV) were found to be suitable for the region in peninsular Malaysia [2]. The best fit distributions of annual runoff in Malaysia are determined based on L-moment ratio diagram and statistical tests. Numerical analysis was performed on some stations in every state in Peninsular Malaysia. Thus, the analysis performed are using data from homogeneous regions established using L-moments procedure. From the analysis, probability plots and comparison between at site and regional estimation shows that most of the station are almost the same and lies on the 45^o lines meaning that the site are fitted distribution.

Keywords: *Homogeneous, L-Moment, Regionalisation, Regional Frequency Analysis*

1.0 INTRODUCTION

Flooding occurs when a watercourse is unable to convey the quantity of runoff flowing downstream. The frequency with which this occurs is described by a return period. Regionalization is the procedure to find natural groups of watersheds with homogeneous hydrologic response, and finds applications in hydrologic design, planning and management of water resources systems [3]. The regional frequency analysis is widely used in flood analysis. The approach based on the theory of L-moments was developed by Hosking and Wallis [1] is a very reliable method for assessing exceedance probabilities of extreme environmental events when data is available from more than one site. The method of L-moments is analogous to the method of ordinary moments. The main advantage of L-moments is that, being a linear combination of data, they are less influenced by outliers.

In recent years, researchers either hydrologist or engineer used different procedures because of the different parameters used and resulting to the different results. The reliability of the estimate are largely depends on the quality and the length of the hydrological records used in the analysis. The accuracy is the parameter that most design engineer had problem in analyses that can cause errors in estimation of hydrological parameters. Frequency analysis method using data from single site will limit the prediction value and lead to a large sampling errors [4].

This paper contains the application of the methodology proposed by Hosking and Wallis [1] for the estimation of the annual runoff distribution in any Malaysian rivers by using hydrometric and climatic data of Malaysian gauged stations. The objectives of this study are to identify the homogeneous region and to compare between regional estimation and at-site estimation analysis.

2.0 METHODOLOGY IN REGIONAL FREQUENCY ANALYSIS BY USING L-MOMENTS

2.1 L-Moment Definition and Terminology

L-moments are statistics used to summarize the shape of a probability distribution and also define as linear combination of Probability Weighted Moments (PWMs) [5]. 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 [6].

For a random variable X, the first four population L-moments are [7]:

$$\lambda 1 = EX$$
$$\lambda 2 = (EX2{:}2 - EX1{:}2) / 2$$
$$\lambda 3 = (EX3{:}3 - 2EX2{:}3 + EX1{:}3) / 3$$
$$\lambda 4 = (EX4{:}4 - 3EX3{:}4 + 3EX2{:}4 - EX1{:}4) / 4, \qquad (1)$$

where Xk: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:
$\lambda 1$ = mean, L-mean or L-location,
$\lambda 2$ = L-scale.

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

$$\perp = \frac{y^{I}}{y^{3}}, \tag{2}$$

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
$\tau 3$ = L-skewness,
$\tau 4$ = L-kurtosis.

2.2 L-Moments Ratio Diagram (LMRD)

L-Moments ratio diagram is a graphical analysis of distribution fitting where several distributions can be compared in a single diagram. Theoretical relationship between $\tau 3$ and $\tau 4$ have been derived for various distribution [8]. 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 $\tau 3$ and $\tau 4$ for the region to a theoretical line or point as well as on the variability of the mean values.

2.3 Index Flood Method

The *T*-year event *XT* defined as the event of exceeded on average once every *T* years [9] 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

$$XT,i \;= ui\, ZT \tag{5}$$

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

2.4 Regional Frequency Analysis

There are five steps of the regional frequency analysis methodology which :

i. Screening of the data
ii. Identification of the homogeneous region
iii. Choose frequency distribution
iv. Estimation of the chosen parameters
v. Compute quintile estimation

Since these procedures are presented in detail by Hosking and Wallis [3], only a summary is reported below.

i. Screening of data

Test the gross outliers, inconsistencies, shifts, and trends [10]. Then, discordance measure is used to identify sites from the group of a given sites that are grossly discordance with the group as whole. The discordance measure is single statistics based on difference between L-Moment ratios of the site and the average L-Moment ratio of the group of same sites. These statistics can also be used to identify erroneous data.

ii. Identifying homogeneous regions

Assign the sites to region. Region defined as a set of sites in which the frequency distributions are considered to be appropriately similar. Region is the fundamental unit of the regional frequency analysis. The sites can be divided into region within the homogeneity criteria. It is exactly certified approximate homogeneity is sufficient enough to ensure that the regional frequency analysis is much more accurate compare to at site analysis.

iii. Choose the Frequency distribution

Single frequency distribution is fitted to the data from several sites of homogeneous region in regional frequency analysis. A distribution must be chosen which provides a good fit to the data because the true distribution of the runoff is not known. The parameter distributions that can be used to consider the quintile estimation in this study are Generalized Extreme Value (GEV), the Lognormal, Logistic, and Pearson Type 3. To help in selection of the appropriate distribution, Z-statistic was used. The Z-statistic is a goodness-of-fit measure for those parameter distributions that measures how good the theoretical L-Kurtosis of fitted distribution matched the regional average of the observed data.

iv. Parameter estimation

L-Moment algorithm is used to estimate the chosen distribution parameters and obtain the quintile estimate. It involves fitting the chosen distribution using L-Moment method of the distribution to the sample L-Moment derived from the observed data. Furthermore, the sample of L-Moment ratios from each site in the homogeneous region is weighted according to the length and combined to give regional average L-Moment ratios. Then, the regional mean is set and regional quintile estimates are derived.

v. Compute quintile estimates

To obtain the runoff estimation at ungauged sites, the regional quintile estimates derived from data at gauged sites within the region are interpolated to all ungauged sites in that particular region.

2.5 Discordancy Measures

This measure is used to identify site that are outliers with the group as a whole. If a single site does not appear to belong to the cloud of $\tau 3$ and $\tau 4$ 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.

$$Di = \frac{1}{3} N (ui - \bar{u})^T K^{-1} (ui - \bar{u}) \quad (6)$$

where ui = vector of L-moment ratios for station i,

K = covariance matrix of ui,
$\bar{u}$ = mean of vector ui.

The station i is declared to be discordant, if Di is greater than the critical value of the discordancy statistic given in a tabular form by [6].

2.6 Identification of Homogeneous Region using 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 [11] 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 [1]. 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:

$$H = \frac{(V - \mu v)}{\sigma v} \quad (7)$$

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,

$1 \leq H < 2$, region is possibly heterogeneous,

$H \geq 2$, region is definitely heterogeneous.

2.7 Computer Programs for Method of L-Moments

All of the processes have been programmed by [12] (lmoments: version 3.04) based on FORTRAN-77. The FORTRAN-77 routines and subroutines are combined under a program which has been compiled and adapted into an executable form in this study [13].

3.0 STUDY AREA, AVAILABILITY AND ACQUISITION OF DATA

This study involves data acquisition and numerical modeling. The data acquisition involves acquiring the annual runoff data of selected location all over Malaysia from Department of Irrigation And Drainage, Ampang. Numerical modeling involved analyzing the annual runoff data to find the L-Skewness and L-Kurtosis from different sites of annual runoff data which has historical data for 20 years and above. The study area was divided into three homogeneous region which is southern, east coast and west coast region of peninsular Malaysia as shown in Figure 1. Figure 2 shows the flow of work done in order to construct the Regional Frequency Analysis of flow. The amounts of stations with homogenous data were 38 stations. Six (6) stations employed in REGION I (SOUTH), twenty three (23) stations employed in REGION II (WEST) and nine (9) stations employed in REGION III (EAST). Table 1 show the total sites and data involved in each state while Table 2 to Table 4 shows the Summary of selected Automatic Runoff Stations of R1, R2 and R3 of Peninsular Malaysia.

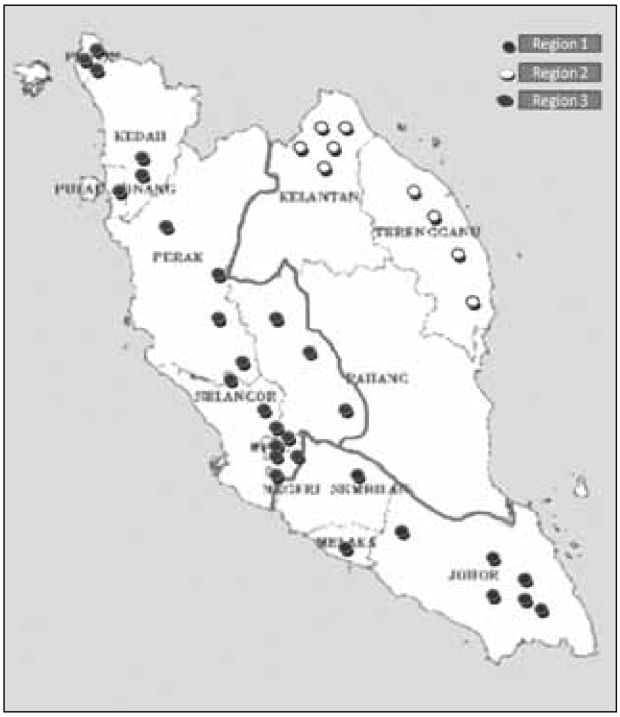


Figure 1: Selected hydrological gauge stations

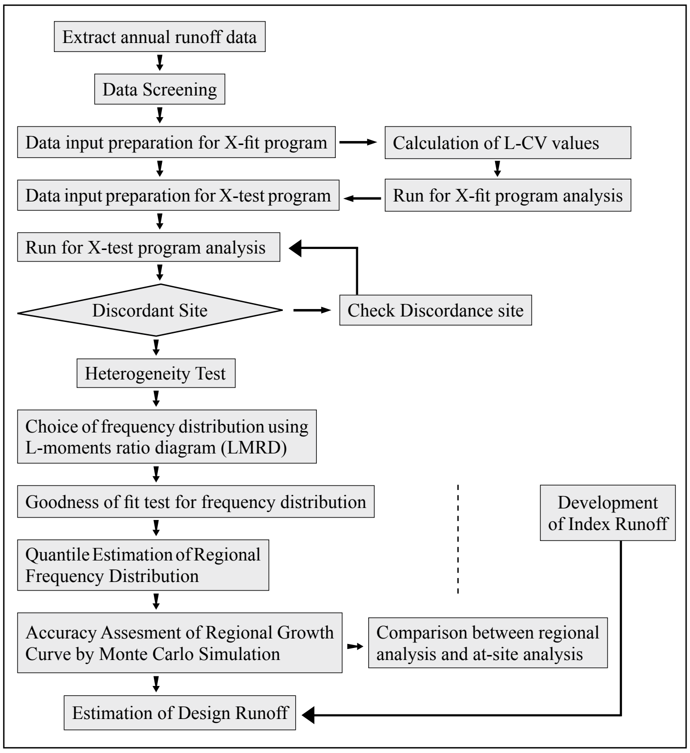


Figure 2: Constructing the Regional Flood Frequency Analysis for runoff

Table 2: Summary of selected Automatic Runoff Stations of R1 of Peninsular Malaysia

No.	Station Id.	Station Name	No. of years in Analysis
1	2527411	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,Yong 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, Negeri Sembilan	45

Region 1 (R1) – Johor, Melaka, and Negeri Sembilan (Southern of peninsular)

Table 3: Summary of selected Automatic Runoff Stations of R2 of Peninsular Malaysia

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 Uda, Pulau Pinang	47
6	5505412	Sg. Muda at Ldg.Victoria, Pulau 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
12	2816441	Sg. Langat at Dengkil, Selangor	50
13	2918401	Sg. Semenyih at Kg. Rinching, Selangor	36
14	2918443	Sg. Semenyih at Semenyih, Selangor	14
15	3615412	Sg. Bernam at Tanjung Malim, Selangor	50
16	3813411	Sg. Bernam at JAM.SKC, Selangor	49
17	3118445	Sg. Lui at Kg. Lui, Selangor	45
18	3116430	Sg. Klang at JAM. Sulaiman	37
19	3116432	Sg. Klang at Leboh Pasar, K. Lumpur	14
20	3116433	Gombak at Jln. Tun Razak	49
21	3519426	Sg. Bentong at JAM. Kuala Marong	41
22	4019462	Sg. Lipis at Benta, Pahang	45
23	4218416	Sg. Jelai at Kuala Medang, Pahang	29

Region 2 (R2) – Perlis, Kedah, Pulau Pinang, Perak, Selangor, Wilayah Persekutuan Kuala Lumpur and Pahang (West coast)

Table 4: Summary of selected Automatic Runoff Stations of R3 of Peninsular Malaysia

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, Kelantan	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, Terengganu	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

Region 3 (R3) – Kelantan, Terengganu and Pahang (East coast)

4.0 RESULT AND DISCUSSION

The weighted mean of L-Skewness and L-Kurtosis were plotted into the L- Moment Ratios Diagram. This result was obtained by running the X-Test and X-Fit Program using FORTRAN Software.

Based on Figures 2 and 3, it can be summarized that Generalized Logistic Distribution (GLO) is the best distribution for regional frequency analysis for Regions 1 and 2. Figure 4 show that Generalized Extreme Value (GEV) is the best distribution for Region 3. The Goodness of fit Test (Z^{DIST}) should be less than or equal to 1.64 but if the value exceeded 1.64 the distribution with value Z^{DIST} close to it will be selected as a parent distribution. From Table 1, it is shown that for Region 1 and 2 represented by GLO distribution and for Region 3 represented by GEV distribution.

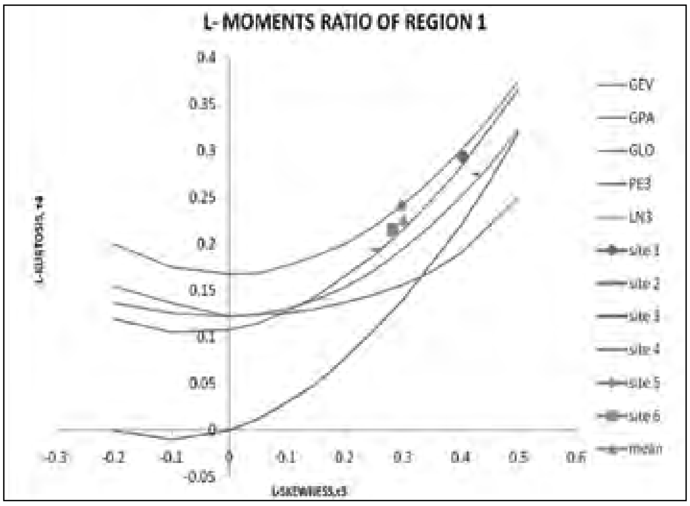


Figure 2: L-Moment Ratio Diagram of Region 1

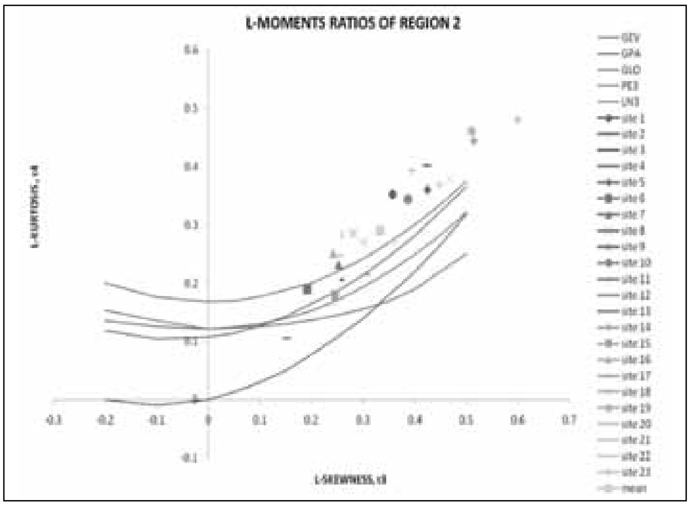


Figure 3: L-Moment Ratio Diagram of Region 2

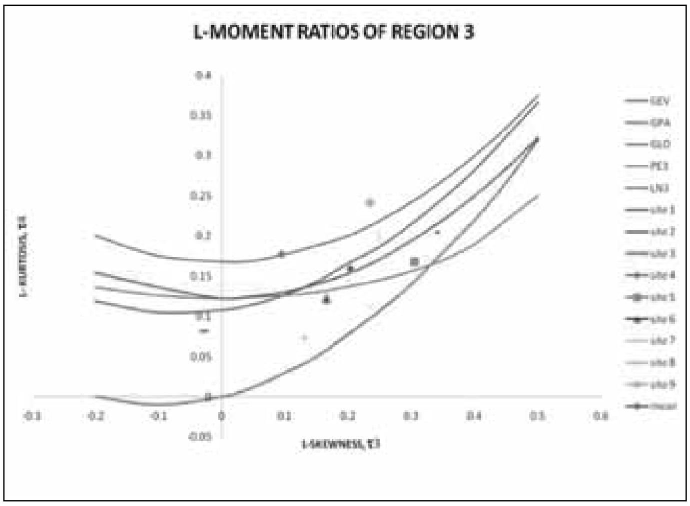
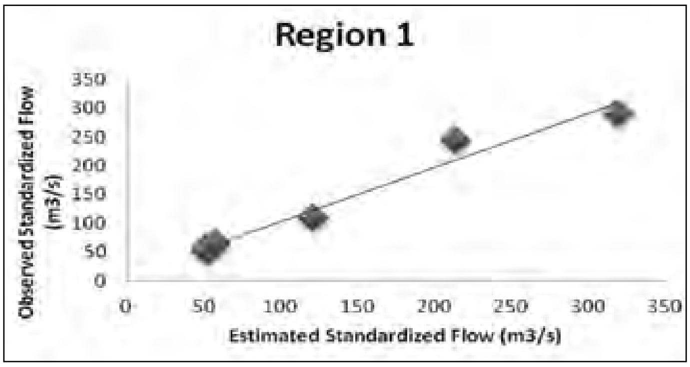
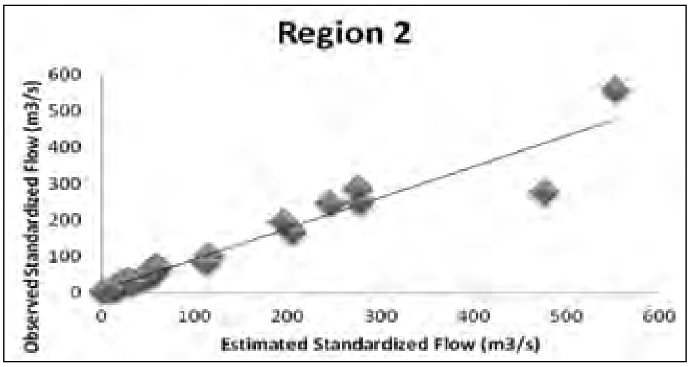


Figure 4: L-Moment Ratio Diagram of Region 3

Table 5: Goodness of fit Test (Z^{DIST})

Z^{DIST}	R1	R2	R3
Generalized Logistic	0.11[a]	-3.63[a]	1.66
Generalized Extreme Value	-0.59[b]	-4.66[b]	0.08[a]
Lognormal	-1.15[c]	-5.77[c]	-0.30[b]
Pearson Type III	-.2.11[d]	-7.65[d]	-1.09[c]
Generalized Pareto	-2.48	-7.66	-3.54[d]





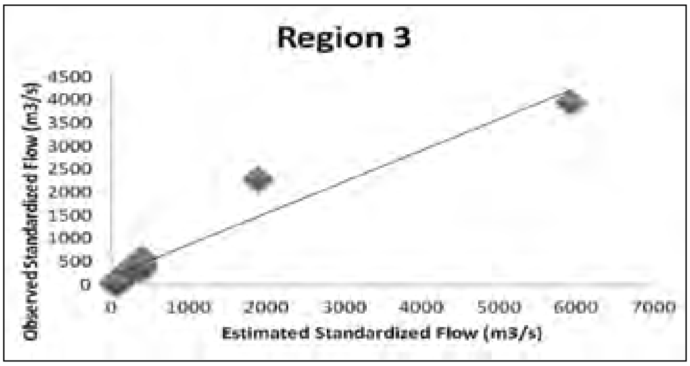


Figure 5: Comparison with regional and at-site analysis

5.0 CONCLUSION

In this study, the regionalisation frequency analysis is based on the Index flood approach. The index flood was replaced by the index runoff consist of two major parts which is the establishment of the regional growth curves or frequency curves and the estimation of the value of index runoff. This study focused on identifying the homogeneous region. The result presented in this study based on Goodness of Fit Test and the Probability plots indicates that the Generalized Logistic (GLO) and Generalized Extreme Value (GEV) distribution were found to be the most suitable distribution in Peninsular Malaysia.

The advantages of using regional frequency analysis approach are evident in many studies that have been done before [14];[15] and [16]. The regional L-Moment algorithm has been successfully implemented in most of the studies on frequency analysis. It is because by using L-Moment, it generally reduces sampling error. This approach reduces at site estimates errors. It can also be used to generate runoff information at an ungauged site.

6.0 ACKNOWLEDGEMENT

This study was carried out as a requirement for the degree of Master of Science in Civil Engineering Universiti Teknologi Mara (UiTM). 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 to 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

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

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

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

[4] M.Z.M. Amin, "Regional flood frequency analysis in the East Coast of Peninsular Malaysia by L-moments approach", M.S thesis (unpublished), National University of Ireland, Galway, 2000.

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

[6] J.R.M. Hosking and J.R. Wallis, "Some statistics useful in regional frequency analysis", *Water Resource Research*, 29(2), pp 271-281, 1993.

[7] J.R.M. Hosking, "The theory of probability weighted moments," *IBM Research*, Yorktown Heights, New York, IBM Research Report. RC 12210, 1986.

[8] J.R.M. Hosking, "L-moments: analysis and estimation of distribution using linear combinations of order statistics", *Journal of the Royal Statistical Society*, B,52(1), pp 105-124, 1990.

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

[10] J.R. Wallis, D.P. Lettenmaier and E.F. Wood, "A daily hydroclimatological data set for continental United States," *Water Resources Research*, 27, 1657-63,1991.

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

[12] J.R.M. Hosking, "Fortran routines for use with the method of L-moments, Version 3.04". *Research Report* RC 20525, IBM Research Division, T.C. Watson Research Center, Yorktown Heights, N.Y, 2005.

[13] Basic Fortran Concepts. What is Fortran?, http://programming.ccp14.ac.uk/fortranresources/~cgp/prof77/node2. html, 2001

[14] R.M. Vogel, W.O. Thomas, and T.A. McMahon, "Flood-flow frequency model selection in southwestern United States". *Journal of Water Resources Planning and Management* 119 (3), pp 353-366, 1993.

[15] H. Madsen, C.P. Pearson, and D. Rosbjerg, *Comparison of annual maximum series and partial duration series method for modelling extreme hydrologic events* 2. Regional modelling Water Resources Research 33 (4), pp 759-769, 1997.

[16] S. Mkhandi and S. Kachroo, "Regional flood frequency analysis for southern Africa", *Southern African FRIEND, Technical Documents in Hydrology No. 15 UNESCO*, Paris, France, 1997.

PROFILES

ASSOC. PROF. IR. DR AMINUDDIN MOHD. BAKI currently is the Head Centre of Study Water Resource and Environmental System (WRES) Faculty of Civil Engineering, Universiti Teknologi MARA Shah Alam. His areas of expertise are Wastewater Engineering and Hydrology. He has involved in many Profesional Organization for example Member of MWA, WEF, IWA, ISRM, ISSMGE, MyWP, PHEAM, ASI, and MIM.

ASSOC. PROF. DR ISMAIL ATAN received his Msc (Engineering math) in 1990 and his Phd also in Engineering Math in 1998 both in the United Kingdom. He is now a senior lecturer in Faculty of Civil Engineering, Universiti Teknologi MARA Shah Alam. His areas of expertise is on Stochastic Hydrology.

NORA FARINA MOHD. HALIM currently is a Postgraduate Student in Faculty of Civil Engineering, Universiti Teknologi MARA Shah Alam. Area of interest is in Water Resource Engineering.