DETERMINING PROBABILITY DISTRIBUTION FOR STREAMFLOW REGIONS USING PARTIAL-MOMENTS

[1]ZAHRAHTUL AMANI ZAKARIA, [2]ANI SHABRI&[3]MOHD KHALID AWANG

[1,3]Faculty Informatics and Computing, Univeristi Sultan Zainal Abidin, Terengganu, Malaysia.
[2]Faculty of Science, Univeristi Teknologi Malaysia, Johor, Malaysia.
E-mail: [1]zahrahtulamani@unisza.edu.my, [2]ani@utm.my, [3]khalid@unisza.edu.my

Abstract— An attempt has been made to model the annual maximum streamflow, utilizing the guidelines in the regional flood frequency analysis. The Partial L-moments (PL-moments) at several censoring levels are employed to estimate the regional parameters of three extreme value distributions, namely; generalized extreme value (GEV), generalized logistic (GLO) and generalized Pareto (GPA) distributions. A total number of 18 streamflow stations located throughout the eastern region of Peninsular Malaysia were used as a case study. Firstly, the data is screening out for data verification and quality control. Next, identification of homogeneous regions is made using homogeneity test based on PL-moments. The PL - diagram is then constructed and GEV and GLO distributions appeared to be the acceptable distributions for representing the regional data. However, it is relatively difficult to identify a particular distribution that most fitted the regional data.Thus, goodness-of-fit test (Z-test) is used and the result showed that the most appropriate distribution for modeling maximum streamflow in the East Coast of Peninsular Malaysia, based on PL-moments is the GLO distribution.

Keywords— Partial L-Moments, Frequency Analysis, Parameter Estimation, Censoring Level.

I. INTRODUCTION

Floods are regular natural disasters in Malaysia which happen nearly every year during the monsoon season. The weather in Malaysia is characterized by two monsoon regimes, namely, the Northeast monsoon that brings heavy rainfall, particularly in the East Coast states of Peninsular Malaysia, whereas the Southwest monsoon normally signifies relatively drier weather. Due to the heavy rainfall, the East Coast states like Kelantan, Terengganu and Pahang will be exposed for the possibility of flooding. In this case, analyzing rainfall and streamflow data is important in order to obtain the probability distribution of flood, thus the prediction of flood events and their characteristics can be determined.

One of the most important areas of flood hydrology is flood frequency analysis, which involves the estimation of distributional parameters and the extrapolation of cumulative distribution functions to generate extreme flood values. However, annual flood series are generally insufficient to obtain reliable estimates of the flood quantile or there is no flow record available at the site of interest, especially in developing and undeveloped countries. Therefore, estimation using regional flood frequency analysis (RFFA) is a popular and practical means of providing flood information at sites with little or no flow data available (Jingyi and Hall, 2004). The RFFA allows the estimation of flood quantile at any stream location within a region. Hosking and Wallis (1997) have developed an approach to regional frequency analysis based on L-moments estimation method. Upon the ability of the RFFA to provide higher accuracy of estimated flood quantile than at-site frequency analysis (Kite, 1988), this approach continues to be used worldwide in modeling various floods cases. The study utilized the flood data in Malaysia have been done by Lim and Lye (2003), Zin et al. (2009) and Shabri et al. (2011).

Wang (1990a, b) introduced partial L-moments (PL-moments) which are variants of L-moments and also analogous to the partial probability weighted moments (PPWM). In the case of flood estimation, the interest is focused mainly on the estimation of the right-hand tail of a distribution function. Due to the concern of data on small flood events can sometimes be only of little relevance to the larger ones, the PL-moments method is introduced for characterizing the larger events in data. Using PL-moments may reduce undesirable influences that small sample events may have on the estimation of large return period events.

The PPWM and PL-moments approach related to censored flood data have been investigated by Wang (1990a, b; 1996) and Bhattarai (2004) which utilized PL-moments in fitting generalized extreme value (GEV) distribution to the censored flood samples. Bhattarai (2004) explored different censoring levels of PL-moments and found that sampling properties of PL-moments, with censoring flood samples up to 30% are similar those of simple L-moments. Some other researches can be found in Kroll and Stedinger (1996), Koulouris et al. (1998), Moisello (2007) and recently by Kochanek (2008). These researches contribute to the application of flood data for at-site frequency analysis and thus reveal limited usage of the proposed PL-moments in the regional frequency analysis.

Therefore, this study intends to carry out the RFFA based on PL-moments method for a series of annual maximum streamflow from 18 stations located throughout the East Coast of Peninsular Malaysia. Three extreme value distributions, namely; generalized extreme value (GEV), generalized logistic (GLO) and generalized Pareto (GPA) distributions are used to model the regional data.

Brief introduction on PL-moments method and regional RFFA procedures are described before discussing the results and making concluding remarks.

II. METHODOLOGY

2.1. Methods of PL-Moments

There have been numbers of research discussing on the definition of partial PWMs by Wang (1990a,b; 1996), Hosking (1995) and Koulouris et al. (1998). In the present study, the definition of partial PWMs by Wang (1996) is employed. Wang (1996) defined partial PWMs as extended from the concept of PWMs to be applied to a censored sample

$$\beta_r^{'} = \frac{1}{1-F_0^{r+1}} \int_{F_0}^{1} x(F) F^r dF \quad (1)$$

where $F_0 = F(x_0)$, x_0 being the censoring threshold. Given a complete sample $x_{(1)} \le x_{(2)} \le ... \le x_{(n)}$, the following statistic is defined by Wang (1990a) as an unbiased estimator of $\beta_r^{'}$

$$b_r^{'} = \frac{1}{1-F_0^{r+1}} \sum_{i=1}^{n} \frac{(i-1)(i-2)...(i-r)}{n(n-1)(n-2)...(n-r)} x_{(i)}^{*} \quad (2)$$

where

$x_{(i)}^{*} = 0$ for $x_{(i)} \le x_0$

$x_{(i)}^{*} = x_{(i)}$ for $x_{(i)} > x_0$

The level of censoring, F_0 determine the number of the sample data points to be censored as

$$F_0 = \frac{n_0}{n} \quad (3)$$

where n is the length of the uncensored sample and n_0 is the number of occurrence of values which do not exceed x_0 in the sample (censored data points). In the terms of partial PWMs, the first four PL-moments have the same definition and interpretations as the first four L-moments. When $F_0 = 0$, the PL-moments become the ordinary L-moments as there is no data that being censored.

2.2. Relationships between the PL-moments and the PDFs

In the present study, the generalized extreme value (GEV), Generalized Pareto (GPA) and generalized logistic (GLO) distributions are investigated to find the best fitting probability distribution to the extreme streamflow data.

The parameter estimation for the GEV based on PL-moments can be found in Wang (1996) whereas GLO and GPA distributions can be found in Zakaria et al. (2012). Upon the issues of censoring would improve the estimation of high return period event, this study emphasizes on the censoring at 2.5%, 5% and 10% of the complete data. These levels of censoring would contribute to censoring levels at F_0 = 0.025, 0.05 and 0.1 respectively.

2.3. Regional frequency analysis procedures

Hosking & Wallis (1993; 1997) provided step by step guidelines for performing regional frequency analysis, using the L-moments. The four steps involved in the regional frequency analysis are outlined as follows: (a) screening of the data using discordancy test, (b) identification of homogeneous regions and (c) choice of a regional distribution. The procedures discussed by Hosking & Wallis (1993; 1997) are similarly employed for the PL-moments. For the simplicity of the exposition, the framework of regional frequency analysis will be further discussed in the results and discussion section.

III. RESULTS AND DISCUSSION

3.1. Case Study

The present study is carried out for the eastern region streamflow, located on the East Coast of Peninsular Malaysia. This East Coast region compromises of three states which are Kelantan, Terengganu and Pahang. Records of daily streamflow from 18 sites in the East Coast region were acquired from the Department of Irrigation and Drainage, Malaysia. The data contains measurements of annual daily maximum streamflow in m3/s from 1960 until 2009.

3.2 Regional Homogeneity Analysis

Initially the whole of the East Coast region was assumed as one homogeneous region, and the discordancy test was used for data verification and quality control. Results for PL-moments at F_0 = 0.025, 0.05 and 0.1 shows that the discordance test is passed at all sites with $D_{critical} < 3.0$. Then, the next step is heterogeneity test. For this purpose, simulation with the kappa distribution was performed to conduct *H*-tests for 18 sites of PL-moments. The criteria for assessing homogeneity of a region are similar to those suggested by Hosking and Wallis (1997), as follows

$H_i < 1$ region is acceptably homogeneous

$1 \le H_i < 2$ region is possibly homogeneous

$H_i \ge 2$ region is definitely heterogeneous for $i = 1,2,3$

Table 1: Results of the Homogeneity Test

Method	Homogeneity test		
	H_1	H_2	H_3
PL- 0.025	-1.0368	-0.2410	0.2517
PL-0.05	-0.2151	0.0349	0.3714
PL-0.1	0.2595	-0.4862	-0.1447

The above criteria are applied in the present study to allow a common comparison base for different censoring levels of PL-moments in conjunction with the works of Hosking and Wallis (1993; 1997).

Results of the H_i (i = 1, 2, 3) for all design censoring levels of PL-moments are given in **Table 1**. Based on the results, the study region was identified as an acceptably homogeneous ($H < 1$), therefore a subdivision of the sites into more regions was not necessary.

3.3. Selection of regional distribution function

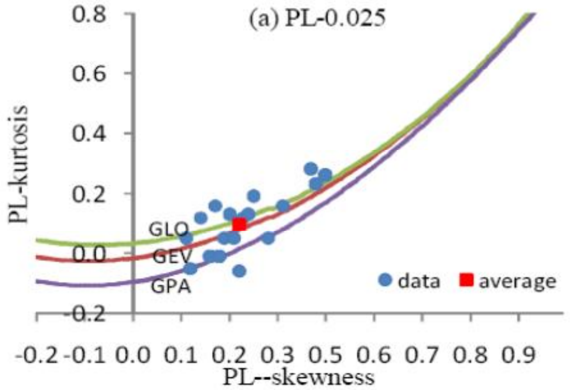
After confirming the homogeneity of the studied region, an appropriate distribution needs to be selected for the regional frequency analysis. Using the previous developed equations for the GEV, GLO and GPA distributions, the PL-diagrams were drawn, which are used for an initial evaluation of the distributions. Fig. 1 illustrates a comparison of the observed and theoretical relations of the distributions. Based on the Fig. 1, GPA distribution appears as a not good candidate for the PL-moments methods in representing maximum streamflow considered in this study. Interestingly, analysis of the PL-diagrams reveal that their sample average values are better described by their theoretical PL-moments of the GEV and GLO distributions for all of the censoring level under investigation.

However, it is relatively difficult to identify a particular distribution that most fitted the regional observed data, which will be evaluated by the goodness-of-fit test (Z-test). Results of the Z-test for the three distributions are given in **Table 2**.

Table 2: Results of the Goodness-of-fit Test.

Method	Z-test		
	GEV	GLO	GPA
PL- 0.025	-0.7175	0.5221	-3.1216
PL-0.05	-1.1516	0.0779	-3.4466
PL-0.1	-1.1766	-0.3283	-2.5782

Based on **Table 2**, for all design methods, the GPA distribution failed the test with a Z-test exceeding the critical value of 1.64. For the L-moments method, it can be seen that the minimum critical Z value for L-moments is achieved by the GEV distribution, with $\left|Z^{DIS}\right|$ of 1.1016 and for the PL-moments of 0.025, 0.05 and 0.1 are achieved by the GLO distribution which have $\left|Z^{DIS}\right|$ of 0.5221, 0.0779 and 0.3283 respectively.



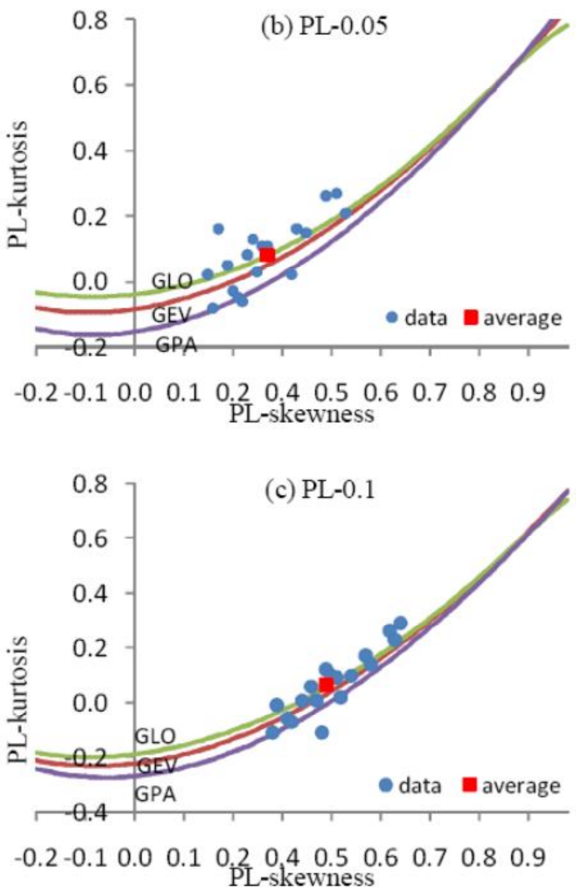


Fig.1. (a - b) PL-diagram of the GEV, GLO and GPA distributions

CONCLUSIONS

The study is intended to model the annual maximum streamflow in the East Coast of Peninsular Malaysia, employing the guidelines in regional frequency analysis. For this purpose, GEV, GLO and GPA distributions are taking into consideration. The developed regional frequency analysis based on L-moments by Hosking and Wallis (1993; 1997) is revisited and the corresponding relationships of regional homogeneity by PL-moments are developed.
The establishment of regional homogeneity was started by first assuming the entire 18 streamflow stations located throughout eastern region catchments as one homogeneous region, based on PL-moments at several design censoring levels. This assumption is statistically accepted when tested using the discordancy test and heterogeneity test. The results of Z-test showed that the most suitable distribution for modeling maximum streamflow in the East Coast of Peninsular Malaysia, based on PL-moments is the GLO distribution.

ACKNOWLEDGMENTS

The authors would like to thank Universiti Sultan Zainal Abidin (UniSZA) for funding this research under University Research Grant Scheme (UniSZA / 14 / GU (022)).

REFERENCES

[1] Bhattarai, K.P, "Partial L-moments for the analysis of censored flood samples" in *Hydrological Sciences Journal,* **49**(5), 855-868 (2004).

[2] Hosking, J.R.M, "The use of L-moments in the analysis of censored data" in *Recent Advances in Life-Testing and Reliability,* edited by N. Balakrishnan (CRC Press, Boca Raton, Fla, 1995), pp. 545-564.

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

[4] Hosking, J.R.M. and Wallis, J.R., *Regional frequency analysis: An approach based on L-Moments.* (Cambridge University Press, UK, 1997).

[5] Jingyi, Z & Hall, M.J., "Regional flood frequency analysis for the Gang-Ming River basin in China" in*Journal of Hydrology,* **296**, 98-117 (2004).

[6] Kite, G.W., "Frequency and risk analysis in hydrology"inWater Resources. (Littleton, Colo, 1988).

[7] Kochanek, K., Strupczewski, W. G., Singh, V. P. and Weglarczyk, S., "The PWM large quantile estimates of heavy tailed distributions from samples deprived of their largest element" in *Hydrological Sciences Journal,* **53**(2), 367-386 (2008).

[8] Koulouris, et al., "L moment diagrams for censored observations" in *Water Resources Research,* **34**(5), 1241-1249 (1998).

[9] Kroll, C.N. and Stedinger, J.R., "Estimation of moments and quantiles using censored samples" in *Water Resources Research,* **32**(4), 1005-1012 (1996).

[10] Lim, Y.H. and Lye, L.M., "Regional flood estimation for ungauged basins in Sarawak, Malaysia." In *Hydrological Sciences Journal,* **48**(1), 79-94 (1998).

[11] Moisello, U., "On the use of partial probability weighted moments in the analysis of hydrological extremes" in *Hydrological Processes,* **21**, 1265-1279 (2007).

[12] Shabri, A., Daud, Z.M. and Ariff, N.M., "Regional analysis of annual maximum rainfall using TL-moments method" in *Theoretical &Applied Climatology*, DOI: 10.1007/s00704-011-0437-5 (2011).

[13] Wang, Q.J., "Estimation of the GEV distribution from censored samples by method of partial probability weighted moments" in *Journal of Hydrology,* **120**, 103-110 (1990a).

[14] Wang, Q.J., "Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution" in *Journal of Hydrology,* **120**,115-124 (1990b).

[15] Wang, Q.J., "Using partial probability weighted moments to fit the extreme value distributions to censored samples" in *Water Resources Research,* **32**(6), 1767-1771 (1996).

[16] Zin, W.W.Z, Jemain, A.A. and Ibrahim, K., "The best fitting distribution of annual maximum rainfalls in Peninsular Malaysia based on methods of L-moment and LQ-moment" in *Theoretical and Applied Climatology,* **96**, 337-344 (2009).

[17] Zakaria, Z.A., Shabri, A. and Ahmad, U.N., "Regional Frequency Analysis of Extreme Rainfalls in the West Coast of Peninsular Malaysia using Partial L-Moments", *Water Resources Management*, **26**, 4417-4433 (2012).

★★★