Far East Journal of Mathematical Sciences (FJMS)

Published Online: March 2015
http://dx.doi.org/10.17654/FJMSMar2015_671_684
Volume 96, Number 6, 2015, Pages 671-684 **ISSN: 0972-0871**

PARAMETER ESTIMATION BASED ON PARTIAL L-MOMENTS METHOD FOR CENSORED SAMPLES

Zahrahtul Amani Zakaria, Ani Shabri* and Mustafa Mamat

Faculty of Informatics and Computing
Universiti Sultan Zainal Abidin K. Terengganu
Malaysia
e-mail: zahrahtulamani@unisza.edu.my

*Department of Mathematics
Faculty of Science
Universiti Teknologi Malaysia
Skudai, Johor, Malaysia

Abstract

Estimation of flood magnitude is a crucial component in planning, designing, and managing of water resources projects. The main focus in hydrologic design is the estimation of high flow quantile. L-moments, popular among hydrologist in flood estimation is known to be oversensitive towards the lower part of the distribution and gives insufficient weight to large sample values. As an alternative, the method of partial L-moments (PL-moments) is proposed to give weightage to the upper part of distribution and large values in censored sample. In this paper, three widely used distributions are selected namely; generalized extreme value (GEV), generalized logistic (GLO) and generalized Pareto (GPA) distribution, for the analysis of censored flood samples. Monte Carlo simulations are

Received: April 12, 2014; Accepted: May 24, 2014

2010 Mathematics Subject Classification: 62N01, 62F10.

Keywords and phrases: censored samples, floods, PL-moments, probability distribution, simulation study.

Communicated by K. K. Azad

conducted to illustrate the performance of PL-moments compared to simple L-moments in fitting each distribution to its samples. Finally, both simple L-moments and PL-moments are used to fit the GLO distribution to two data sets of annual maximum flow series of River Ketil in Kedah and River Gemencheh in Negeri Sembilan, Malaysia.

Introduction

The purpose of analyzing hydrological extreme events such as annual maximum series of floods is, in most cases, to predict magnitude of flood of relatively large return period such as 100 years and above [1]. Hence, it is actually advantageous to intentionally censor (or eliminate) low-value observations because using only the larger value flood ensures that the extrapolation to large return periods flood is carried out by exploring the trend of these larger flows only. Cunnane [2] suggested that in such cases a censored sample should be used and the analysis will be based on only those floods whose magnitudes have exceeded a certain threshold.

Since L-moments were first introduced by [3] as a parameter estimation method, it has been widely applied in many fields of hydrology. Although L-moments result in quite efficient estimate in parameter estimation, this may not be so for predicting large return period events. The question arose whether L-moments are oversensitive to the lower part of distributions and give insufficient weight to large data values that actually contain useful information on the upper distribution tail [4, 5].

Partial L-moment (PL-moments) which are variant of L-moments and analogous to partial probability weighted moments (PPWMs), were first introduced by [1] to deal with censored samples. PL-moments are introduced for characterizing the upper part of distributions and larger events in data. Using PL-moments reduce undesirable influences that small sample events may have on the estimation of large return period events. A number of studies have used PL-moments in application of censored data [1, 6, 7, 8, 9, 11] and showed favorable results such that PL-moments could constitute a valid tool.

A long record of past data is able to well extrapolate the future events and thus produce high accuracy of flood estimation. However, in relatively young country like Malaysia, such hydrological data and information are limited. The available streamflow records are all too often inadequate or unavailable to allow for reliable flood estimation at a location of interest.

The simulation technique generates a series of data which imitate the natural properties of the real world data structure using particular mathematical modeling. In hydrology, Monte Carlo simulation techniques were widely used to establish the properties of a frequency analysis procedure, or to compare two or more procedures [11]. The Monte Carlo simulation generates synthetic flows from various background distributions. These samples were in turn fit with various assumed distributions.

In this paper, simulation study is carried out to investigate the sampling properties of the proposed parameter estimation methods of L-moments and PL-moments. PL-moments and L-moments are compared using both simulated and real data to investigate their sampling properties and ability in fitting GEV, GLO and GPA distributions. Then, two data sets of annual maximum flow series of River Ketil in Kedah and River Gemencheh in Negeri Sembilan, Malaysia are used as a case study.

Parameter Estimation using PL-Moments

L-moments are expressed by [3] as a linear combination of probability weighted moments as follows

$$\beta_r = \int_0^1 [x(F)] F^r dF, \tag{1}$$

where $F = F(x)$, $x(F)$ is an inverse distribution function or so-called quantile function of random variables X and r is real number. [1, 6] extended the concept of L-moments to PL-moments as follows:

$$\delta'_r = \int_{F_0}^1 x(F) F^r dF. \tag{2}$$

However, [7] made a new modification on the former definition introduced by [1]. The formally definition of PL-moments given by [7] as follows:

$$\beta'_r = \frac{1}{1-(F_0)^{r+1}} \int_{F_0}^{1} x(F) F^r dF, \tag{3}$$

where $r = 0, 1, 2, \ldots$ denotes the order of PL-moments. When $F_0 = 0$, the β'_r becomes the ordinary β_r. In general, the rth PL-moments can be written in the term of β'_r as

$$\lambda'_{r+1} = \sum_{k=0}^{r} (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \beta'_k. \tag{4}$$

The PL-moment ratios are defined as $\tau'_r = \lambda'_r / \lambda'_2$ for $r \geq 3$.

The parameter estimation of GEV distribution is derived as follows:

GEV distribution

$$\beta'_r = \xi + \alpha H(r, F_0, k), \tag{5}$$

where $\gamma[1+k, -\ln(r+1)(F_0)]$ is the Incomplete Gamma function and

$$H(r, F_0, k) = \frac{1}{k}\left\{1 - \frac{\gamma[1+k, -(r+1)\ln(F_0)]}{(r+1)^k [1-(F_0)^{r+1}]}\right\}. \tag{6}$$

The first four PL-moments can be obtained by substituting equation (5) into equation (4) yields

$$\lambda'_1 = \beta'_0 = \xi + \alpha H(0, F_0, k),$$

$$\lambda'_2 = \alpha[H(1, F_0, k) - H(0, F_0, k)],$$

$$\lambda'_3 = \alpha[2H(2, F_0, k) - 3H(1, F_0, k) + H(0, F_0, k)],$$

$$\lambda'_4 = \alpha[5H(3, F_0, k) - 10H(2, F_0, k) + 6H(1, F_0, k) - H(0, F_0, k)]. \tag{7}$$

The details of the estimation of GEV distribution and also other distributions of GLO and GPA, can be found in [13].

Simulation Study

A good parameter estimation technique for censored sample should yield results in agreement with those obtained from the complete sample [12]. For this purpose, Monte Carlo simulation is carried out to investigate the sampling properties on different quantile estimators, $x(F)$ obtained from the method of PL-moments compared with those from L-moments. A total number of 5,000 samples of size $n = 15,$ 30 and 50 are simulated to represent small, medium and large samples of maximum stream flows data in Peninsular Malaysia. The levels of censoring, F_0 being tested are in the range of $F_0 = 0.0,$ 0.1, 0.2, 0.3, 0.4 and 0.5. PL-moments at zero censoring $(F_0 = 0.0)$ are equivalent to the simple L-moments.

The performance of the estimation methods is assessed by evaluating the relative bias (RBIAS), relative root mean square error (RRMSE) and mean absolute error (MAE), as follows:

$$RBIAS = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} \left(\frac{x(F)_m^S - x(F)^C}{x(F)^C} \right),$$

$$RRMSE = \sqrt{\frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} \left(\frac{x(F)_m^S - x(F)^C}{x(F)^C} \right)^2},$$

$$MAE = \frac{1}{N_{sim}} \sum_{m=1}^{N_{sim}} | \, x(F)_m^S - x(F)^C \, |,$$

where N_{sim} is the number of generated samples 5,000, $x(F)_m^S$ and $x(F)^C$ are simulated and calculated quantiles of design floods, respectively; estimated at different recurrence intervals of 20, 50, 100 and 200 years.

In practice, the true underlying distribution function that represents a data series is never known. However, it is still valuable to investigate the sampling properties on how quantile estimation is affected by different levels of censoring, F_0 when the population distribution function is known. Therefore in this study, three selected probability distribution functions namely GEV, GLO and GPA distributions were chosen to represent the known parent distribution. The parameters for each distribution are known as ξ, α and k which represents location, scale and shape parameters of the distribution respectively. The values of location and scale parameters are set as $\xi = 0$ and $\alpha = 1$, respectively. In flood frequency analysis, the shape parameter, k is usually in the range of -0.4 to 0.4 [1].

Results and Discussion

In estimating the bias of the estimators, the level of censoring plays an important role since the results obtained at different censoring levels produce different RBIAS value for the considered probability distribution.

Based on the analysis in Figures 1, the results indicate that using small sample size lead to a greater RBIAS for all k values regardless whether L-moments or PL-moments are used. Figures 1 (a)-(b) show that RBIAS on the GEV quantile estimates obtained from PL-moments with level of censoring of $F_0 \leq 0.3$, for sample sizes of 15-50 are almost similar to that for simple L-moments. In certain cases, PL-moments for the GEV distribution with $F_0 \leq 0.3$ produce smaller RBIAS compared to L-moments for negative k value $(k = -0.2)$. However, for the case of positive k value $(k = +0.2)$, only PL-moments with $F_0 = 0.3$ lead to a smaller RBIAS than L-moments. When the censoring level $F_0 > 0.3$, the quantile estimator becomes more negatively biased as F_0 increases.

Results for GLO quantile estimator show that the RBIAS produced by the method of PL-moments with the level of censoring up to 0.4 is very similar to that of simple L-moments. However, using PL-moments at $F_0 = 0.4$ result in smaller RBIAS compared to L-moments regardless of the k values. For the censoring level of $F_0 > 0.4$, the negative values of RBIAS from PL-moments increase appreciably with the increase in F_0.

Based on Figures 1 (e)-(f), the lower RBIAS values can be produced by using PL-moments with $F_0 \leq 0.2$ compared to simple L-moments for all k values are used. However for PL-moments with $F_0 > 0.2$, as noted above, the negative value of bias increases sharply.

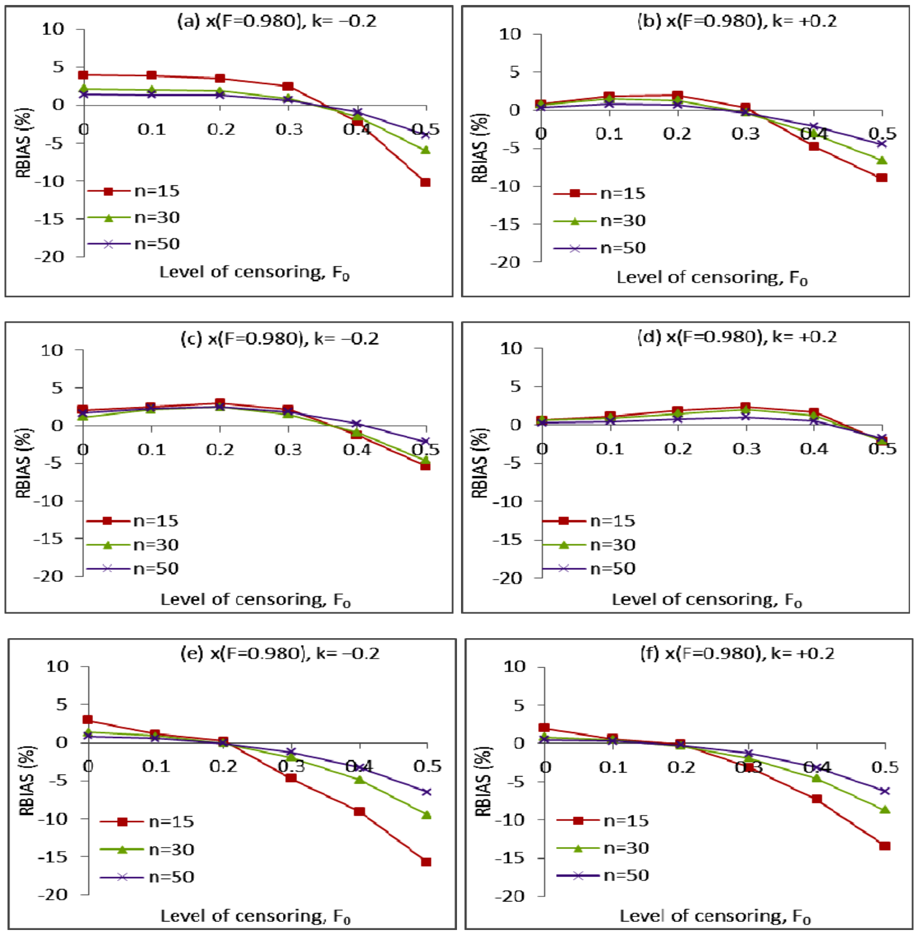

Figure 1. RBIAS for (a)-(b) GEV, (c)-(d) GLO and (e)-(f) GPA quantile estimators of $x(F = 0.980)$ plotted against censoring level (F_0) for shape parameter, $k = -0.2$ and $k = +0.2$.

The overall results show that at certain level of censoring, F_0, PL-moments estimate are almost unbiased to that simple L-moments. In summary, the RBIAS on quantile estimates obtained from PL-moments with level of censoring of $F_0 \leq 0.3$ for GEV distribution, $F_0 \leq 0.4$ for GLO distribution and $F_0 \leq 0.2$ for GPA distribution, exhibit almost similar pattern to that for simple L-moments, for sample sizes: $n = 15$, 30 and 50 and quantile estimates: $x(F = 0.950)$, $x(F = 0.980)$, $x(F = 0.990)$ and $x(F = 0.995)$, considered in this study. These observations reveal that the method of PL-moments with censoring from left up to a certain level will not practically add any bias over that by the method of simple L-moments and even lead to a smaller bias at certain cases.

In order to obtain more comprehensive results for RRMSE and MAE, the analyses focus up to certain censoring level of PL-moments (based on results of RBIAS) estimated for different sample size *n* and quantiles for positive and negative values of shape parameter *k*. The analysis for GEV distribution from $F_0 = 0.1$ to 0.3 yielded 48 combinations of 4 shape parameter values (either positive $k = 0.1$, 0.2, 0.3, 0.4 or negative $k = -0.1$, -0.2, -0.3, -0.4) and 4 quantiles function $(F = 0.950, 0.980, 0.990, 0.995)$. For the GLO distribution from $F_0 = 0.1$ to 0.4 and GPA distribution from $F_0 = 0.1$ to 0.2, analysis of 4 shape parameter values and 4 quantiles function yielded 64 and 32 combinations for the respective distributions.

The results are simplified in Table 1 in order to get a clear picture of performances of PL-moments compared to L-moments based on RRMSE and MAE values. In order to better illustrate the results, the cases of PL-moments superior to L-moments are marked in bold.

From Table 1, it is shown that when the parent distribution is known, PL-moments is outperformed L-moments for positive values of shape parameter *k*, for most of sample sizes (15-50) based on RRMSE and MAE values. However, as *k* goes negative, PL-moments results in greater RRMSE and MAE than using L-moments method.

Table 1. Performances of PL-moments compared to L-moments for simulation of known parent distribution based on RRMSE and MAE

		RRMSE		MAE	
		k value		k value	
Dist.	**n**	**k** ≤ -0.2	**k** ≥ 0.2	**k** ≤ -0.2	**k** ≥ 0.2
GEV ($F_0 \le 0.3$)	15	19/48	16/48	14/48	17/48
	30	7//48	**25/48**	9//48	**28/48**
	50	7/48	**26/48**	6/48	**28/48**
GLO ($F_0 \le 0.4$)	15	19/64	**34/64**	18/64	18/64
	30	16/64	**34/64**	12/64	**37/64**
	50	13/64	**36/64**	14/64	**38/64**
GPA ($F_0 \le 0.2$)	15	3/32	**29/32**	1/32	**24/32**
	30	3/32	**30/32**	2/32	**25/32**
	50	1/32	**31/32**	2/32	**29/32**

Case Study

To illustrate the application of fitting distribution to censored flood samples using PL-moments approach, two sets of annual maximum flow series are presented here. The first data comes from 5608418 River Ketil in Kedah, Malaysia, with 30 annual maximum flows covering 1980-2009. The station has a catchment area of 1000 km^2. The second data comes from 2525415 River Gemencheh in Negeri Sembilan, Malaysia, with 32 annual maximum flows covering 1961-1992. The station has a catchment area of 453 km^2. The flood data were obtained from Department of Irrigation and Drainage, Ministry of Natural Resources and Environment, Malaysia.

Among three distributions under study, GLO distribution appears to be the best fitted distribution to both data series. Figure 2 shows the GLO distribution curves fitted to the data series of 5608418 River Ketil by L-moments and PL-moments at different censoring level, F_0 ranging from

0.1 to 0.6. The observed data values are plotted against the corresponding quantile for the 70% of highest quantile values.

From these fitted plots, it is generally observed that the frequency curves obtained by L-moments are significantly influenced by small annual maximum flows, leading to poor prediction of large return period events. In contrast, the curves fitted by PL-moments method $(F_0 = 0.1 \text{ to } 0.6)$ better capture the trends shown by the high quantiles. However, with increasing F_0 values from 0.5 to 0.6, the PL-moments results are getting poor in low quantiles. Therefore, the PL-moments method with censoring threshold of $F_0 = 0.1$ to 0.4 are good enough to yield a satisfactory result that fitted the observed data better than L-moments in high quantile estimation. This is in agreement with the earlier results that PL-moments method of GLO distribution with level censoring up to 0.4 lead to almost unbiased over that the method of simple L-moments.

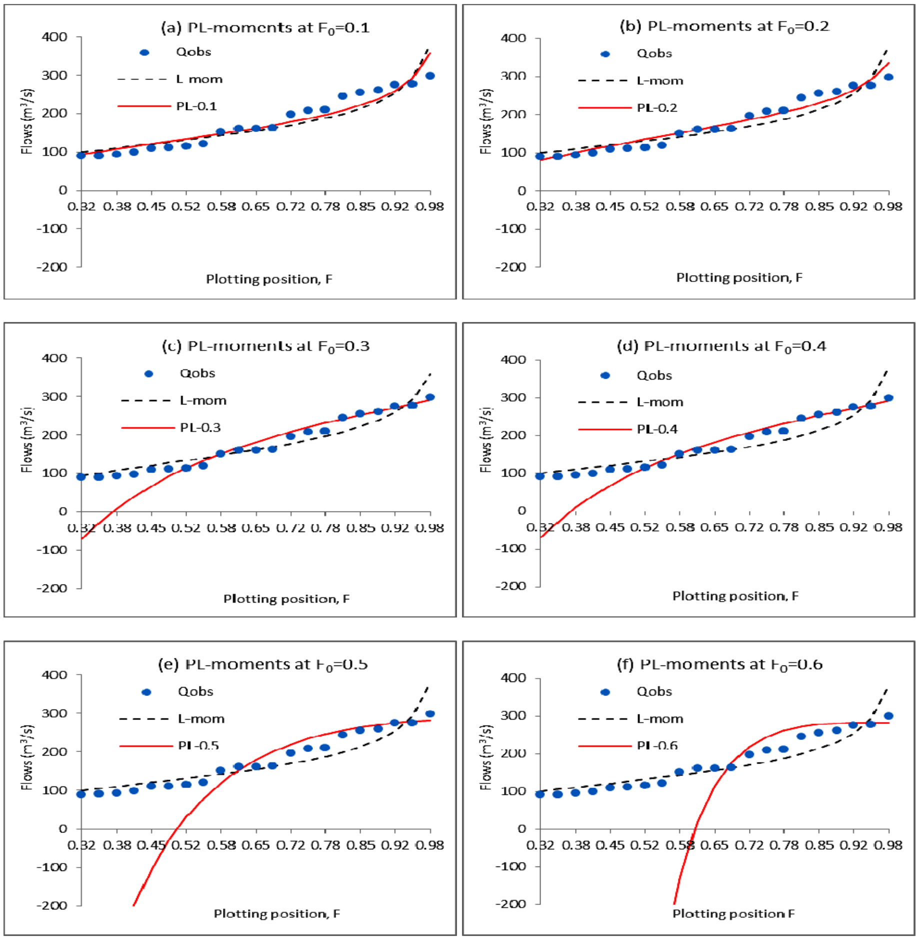

Figure 2. Fitting the GLO distribution to annual maximum flows of 5608418 River Ketil (catchment area 1000 km^2) for 70% of highest quantile values for different censoring levels, F_0.

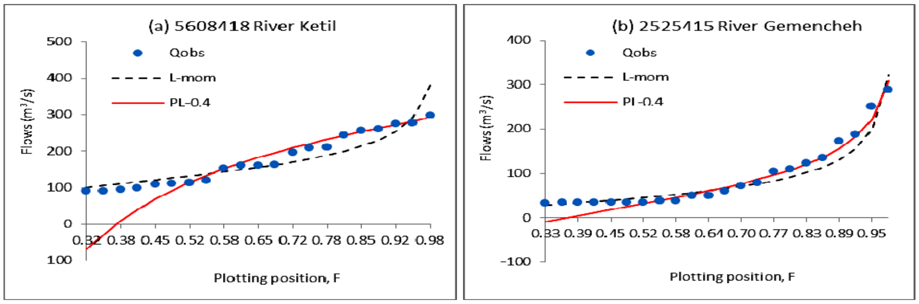

Figure 3. Fitting the GLO distribution to annual maximum flows of (a) River Ketil in Kedah (catchment area 1000 km^2) for sample size, $n = 30$ and (b) River Gemenceheh in Negeri Sembilan (catchment area 453 km^2) for sample size, $n = 32$.

Figure 3 shows the GLO distribution curves fitted to the both data series of 5608418 River Ketil and 2525415 River Gemencheh by L-moments and PL-moments at $F_0 = 0.4$. From these fitted plots, the PL-moments $(F_0 = 0.4)$ estimate of large return period events are less influenced by the small annual maximum flows. This seems to suggest that the PL-moments method would improve the estimation of floods of larger return periods.

Conclusion

For all three-parameter (GEV, GLO and GPA) distributions considered in this study, the PL-moments with censoring from left up to a certain level produce smaller bias compared to L-moments for any range of k values. By censoring the sample from left, PL-moments approach eliminates some small sample and gives more weight to the larger sample compared to L-moments which encompass the entire sample. Thus, PL-moments may reduce the influence of the small sample events might have when estimating the large return period events.

Analysis of annual maximum streamflow series data shows that the PL-moments approach is quite effective in fitting GLO distribution (identified as best fit distribution in this case) to large flood data, and even produces better performance than the simple L-moments method. Using

PL-moments reduces the undesirable influences that the small events may have on the estimation of large return-period events compared to using L-moments. Our evaluations support the finding of previous studies (e.g. [2, 8]) that analysis of censored samples would improve the estimation of floods of large returns periods.

Acknowledgements

The authors thankfully acknowledged the financial support provided by Ministry of Higher Education, Malaysia and Universiti Sultan Zainal Abidin, Malaysia. The authors also would like to thank the Department of Irrigation and Drainage, Ministry of Natural Resources and Environment, Malaysia for providing the floods data and Universiti Teknologi Malaysia.

References

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

[2] C. Cunnane, Review of Statistical Methods for Flood Frequency Estimation in Hydrologic Frequency Modeling, D. Reidel, Dordrecht, 1987, pp. 49-95.

[3] J. R. M. Hosking, L-moments: analysis and estimation of distributions using linear combinations of order statistics, J. Roy. Statist. Soc. 52 (1990), 105-124.

[4] Q. J. Wang, LH-moments for statistical analysis of extreme events, Water Resources Research 33(12) (1997), 2841-2848.

[5] B. Bobée and P. F. Rasmussen, Recent advances in flood frequency analysis, U.S. Nalt. Rep. Int. Union Geol. Geophys. Rev. Geophys. 33 (1995), 1111-1116.

[6] Q. J. Wang, 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, J. Hydrology 120 (1990b), 115-124.

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

[8] K. P. Bhattarai, Partial L-moments for the analysis of censored flood samples, Hydrological Sciences J. 49(5) (2004), 855-868.

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

[10] J. Deng and M. D. Pandey, Using partial probability weighted moments and partial maximum entropy to estimate quantiles from censored samples, Probabilistic Engineering Mechanics 24 (2009), 407-417.

[11] J. R. M. Hosking and J. R. Wallis, Regional Frequency Analysis: An Approach Based on L-moments, Cambridge University Press, 1997.

[12] R. E. Thompson, E. O. Voit and G. I. Scott, Statistical modelling of sediment and oyster PAH contamination data collected at a South Carolina Estuary (Complete and Left-censored Samples), Environmetrics 11(1) (2000), 99-119.

[13] Z. A. Zakaria and A. Shabri, Regional frequency analysis of extreme rainfalls using partial L-moments method, Theoretical and Applied Climatology 113(1-2) (2013), 83-94.