Sains Malaysiana 41(11)(2012): 1389–1401

Analysis of *T*-Year Return Level for Partial Duration Rainfall Series

(Analisis Tahap Ulangan *T*-Tahun bagi Siri Hujan Tempoh Separa)

WENDY LING SHINYIE* & NORISZURA ISMAIL

ABSTRACT

This paper aims to estimate the Generalized Pareto Distribution (GPD) parameters and predicts the T-year return levels of extreme rainfall events using the Partial Duration Series (PDS) method based on the hourly rainfall data of five stations in Peninsular Malaysia. In particular, the GPD parameters are estimated using five methods namely the method of Moments (MOM), the probability weighted moments (PWM), the L-moments (LMOM), the Trimmed L-moments (TLMOM) and the Maximum Likelihood (ML) and the performance of the T-year return level of each estimation method is analyzed based on the RMSE measure obtained from Monte Carlo simulation. In addition, we suggest the weighted average model, a model which assigns the inverse variance of several methods as weights, to estimate the T-year return level. This paper contributes to the hydrological literatures in terms of three main elements. Firstly, we suggest the use of hourly rainfall data as an alternative to provide a more detailed and valuable information for the analysis of extreme rainfall events. Secondly, this study applies five methods of parametric approach for estimating the GPD parameters and predicting the T-year return level. Finally, in this study we propose the weighted average model, a model that assigns the inverse variance of several methods as weights, for the estimation of the T-year return level.

Keywords: Generalized Pareto Distribution; parameter estimation; partial duration series; T-year return level

ABSTRAK

Kajian ini bertujuan menganggar parameter Taburan Pareto Teritlak (GPD) dan meramal tahap ulangan T-tahun bagi kejadian hujan melampau menggunakan kaedah siri tempoh separa (PDS) berdasarkan data hujan per jam untuk lima stesen di Semenanjung Malaysia. Secara khususnya, parameter GPD dianggar melalui lima kaedah iaitu momen (MOM), momen kebarangkalian berpemberat (PWM), L-momen (LMOM), TL-Momen (TLMOM) dan kebolehjadian maksimum (ML) dan prestasi tahap ulangan T-tahun untuk setiap kaedah dianalisis berdasarkan ukuran RMSE yang diperoleh melalui simulasi Monte Carlo. Selain itu, kajian ini mencadangkan model purata berpemberat, iaitu suatu model yang mewakilkan pemberat setiap kaedah dengan songsangan varian untuk menganggar tahap ulangan T-tahun. Kajian ini menyumbang kepada literatur hidrologi melalui tiga elemen utama. Pertama, kami mencadangkan penggunaan data hujan per jam sebagai alternatif untuk memberikan maklumat yang lebih bermakna dan menyeluruh bagi analisis kejadian hujan melampau. Kedua, dalam kajian ini kami menggunakan lima kaedah daripada pendekatan berparameter untuk menganggar parameter GPD dan meramal tahap ulangan T-tahun. Akhir sekali, kami mencadangkan model purata berpemberat, iaitu suatu model yang mewakilkan pemberat setiap kaedah dengan songsangan varian untuk penganggaran tahap ulangan T-tahun.

Kata kunci: Penganggaran parameter; siri tempoh separa; Taburan Pareto Teritlak; tahap ulangan T-tahun

INTRODUCTION

Over the past several decades, climate change has been consistently associated with changes in several components of hydrological cycle and system such as precipitation patterns, intensity and extremes (IPCC 2007). As a result, phenomena related to rainfall frequency such as flash floods, landslides, severe erosions and debris flows are also showing changing patterns and affecting all countries over the world, irrespective of their locations on the globe, resulting in potentially huge economic and social implications (Floris et al. 2010). In addition, the changing patterns of rainfall amounts may affect the generation of hydroelectric power, the management and implementation of dams and the management of cooling water (Harasawa & Nishioka 2003). Recently, several areas in Malaysia were also affected by the changes in weather patterns and have been experiencing intense and heavy rainfalls which cause serious flooding and damages to the infrastructure of such areas. As an example, in November 2010, a serious flood occurred in the areas of North Malaysia involving Kedah, Perlis and Kelantan where 50,000 people were evacuated from home. A total of RM26 million have been subsidized by the government to the farmers in the affected areas, as aids and compensations for their severely damaged crops and fields. These events have further proved the importance of observations, analysis and predictions of climate change which can be carried out through hydrological and climatological studies, especially through studies of extreme rainfall events.

Extreme weather event can be defined as 'an event that is rare within its statistical reference distribution at a particular place' (IPCC 2001). The amount of extremely high or low precipitation, leading to flood or drought, is a good example of the many risks of substantial weather. Based on hydrological literatures, Extreme value theory (EVT) is a powerful and robust approach for capturing extreme movements in the tail behaviour of extreme rainfall distributions. One of the important properties in EVT is that the limiting distribution of the extreme data observed over a long period is independent of the underlying distribution itself. Therefore, statistical estimates associated with the fitted distribution, such as the estimation of high quantile or *T*-year return level, can covers conditions beyond the usual phenomenon of rainfall events, which are indicated by the extreme data. In studies of water resource management and financial risk management, EVT is frequently used to obtain probability distributions via two main methods, by fitting the maximum or the minimum data of a random sample or by modeling the distribution of excess data above a certain threshold (Coles 2001; Katz et al. 2002; Kouchak & Nasrollahi 2010; Smith 2001). Previous studies on extreme rainfall in Malaysia was done by Deni et al. (2009) on the trend of wet spells and Zin et al. (2010) on the changes in extreme rainfall events.

The annual maxima series (AMS), a method which fits distributions to the maximum or the minimum, has been considered as one of the most popular approaches of EVT. The most probable reason for using this approach is that the sample data can be easily extracted and the amount of sample data can be largely reduced since only the maximum or the minimum are utilized. However, this approach ignores other important extreme rainfall data, especially those greater than the annual maximum. In addition, since the AMS uses only a single data for each year, the time series of rainfall data has to be long enough so that the sample size is large enough for modeling purposes. Several alternatives have been developed to overcome the disadvantages of the AMS, including the *r*-largest order statistics model (Smith 1986), the method of independent storms (Harris 1999) and the partial duration series (PDS) approach (Hosking & Wallis 1987).

In recent years, the approaches of extreme value studies have changed towards the PDS method, which is also called the peaks over the threshold (POT) method, over the AMS method (Kouchak & Nasrollahi 2010; Pandey et al. 2003; Rasmussen et al. 1994; Todorovic 1978). In particular, the PDS has been recommended for modeling extreme data by several researchers in environmental areas and such studies can be found in Begueria (2005), Lana et al. (2006) and Li et al. (2005). In fact, several studies have demonstrated that the Generalized Pareto distribution (GPD) utilized in the PDS provides a better performance in the fit of the extreme hydrological variable compared to the Generalized Extreme Value (GEV) distribution utilized in the AMS (Cunnane 1973; Madsen et al. 1997). Examples of applications of PDS for the Malaysia's data can be found in Zin et al. (2009) and Zin and Jemain (2010).

The GPD, which is a two-parameter distribution consisting of special cases of the standard Pareto, the exponential and the Pareto type-II distributions, is the limiting distribution of the excess over a threshold as the threshold approaches the endpoint of the variable. Compared to the AMS which considers only the data of maximum or minimum, the PDS considers all data exceeding a predetermined threshold. Generally, the PDS method involves three main steps; the first is to choose an appropriate threshold, the second is to estimate the GPD parameters and the final is to estimate the extreme quantiles. In PDS, the fitting of threshold exceedances requires the assumption of Poisson recurrence process and the GPD for exceedance data. One of the main issues in fitting the GPD on a selected large values of a random variable is that the exact location of the upper region or the exact threshold value, is required. Therefore, an analyst must compromise between two contradictory strategies, using a lower tail closer to the central data which provides more data but introduces bias towards the central values or using a higher tail which provides less data but leads to larger variance in the estimated parameters. Several methods have been proposed as guidelines for selecting the threshold value, but there are no unified method which has been generally agreed upon. However, practice-oriented guidelines on threshold selection have been proposed by Begueria (2005), Lang et al. (1999), Madsen et al. (1997) and Rasmussen et al. (1994).

The estimation of GPD parameters for extreme data are generally implemented using two main approaches, parametric and non-parametric. Most hydrological studies favoured the fully parametric approach (Coles et al. 2003; Lana et al. 2006) and in particular, the most widely used parametric estimation models are the maximum likelihood (ML), the method of moments (MOM) and the probability weighted moments (PWM). In particular, Hosking and Wallis (1987) compared the performances of the ML, the MOM and the PWM estimators and found that the MOM and the PWM were more reliable than the ML. However, Ashkar and Tatsambon (2007) proved that the ML estimator is consistent for either small or large sample sizes due to the implementation of numerical algorithm introduced by Davison (1984). Besides the ML, the MOM and the PWM, the L-Moments (LMOM) (Hosking 1990), the Least square (Moharram et al. 1993), the maximum penalized likelihood (Coles & Dixon 1999), the TL-moments (TLMOM) (Elamir & Seheult 2003) and the minimum density power divergence (Juarez & Schucany 2004) can also be used as alternatives for estimating the GPD parameters.

This paper aims to estimate the GPD parameters and the *T*-year return levels of extreme rainfall events using the PDS method based on the hourly rainfall data of five stations in Peninsular Malaysia. In particular, the parameter estimations were performed using five methods of the parametric approach namely the MOM, the PWM, the LMOM, the TLMOM and the ML and the performance of the *T*-year return level of each estimation method is analyzed based on the RMSE measure obtained from Monte Carlo simulation. Finally, the weighted average model, a model which assigns the inverse variance of several methods as weights, is proposed in this study to estimate the *T*-year return level. This paper contributes to the hydrological literatures in terms of three main elements. Firstly, this study suggests the use of hourly rainfall data as an alternative to provide a more detailed and valuable information for the analysis of extreme rainfall events. Secondly, this study applies five methods of parametric approach for estimating the GPD parameters and the *T*-year return level. Finally, this study proposes the weighted average model, a model that assigns the inverse variance of several methods as weights, for the estimation of the *T*-year return level.

METHOD

STUDY AREA

The hourly rainfall data from five stations in Peninsular Malaysia were analyzed in this paper, where the selected stations are Alor Setar which represents the NorthWest region, Pekan which represents the East region, Johor Bahru which represents the SouthWest region, Ampang which represent the West region and Chanis which represent the Center region of Peninsular Malaysia. The locations of the selected stations are shown in Figure 1, whereas the details of the selected stations are shown in Table 1. Since the percentage of missing data, which is less than 10%, is quite small, the missing values are ignored in the analysis. The wet days (the rainfall amount more than 1 mm) are extracted from the hourly rainfall data.

TABLE 1. List of stations

Raingauge station	Latitude (N)	Longitude (E)	Number of recorded years	Available years	% missing	Maximum hourly rainfall
Alor Setar	6° 07’	100° 23’	39	1970-2008	3	80.5
Pekan	3° 30’	103° 25’	39	1970-2008	5	90.9
Johor Bahru	1° 28’	103° 45’	39	1970-2008	4	98.0
Ampang	3° 09’	101° 45’	39	1970-2008	1	90.8
Chanis	2° 49’	102° 55’	29	1980-2008	7	83.5

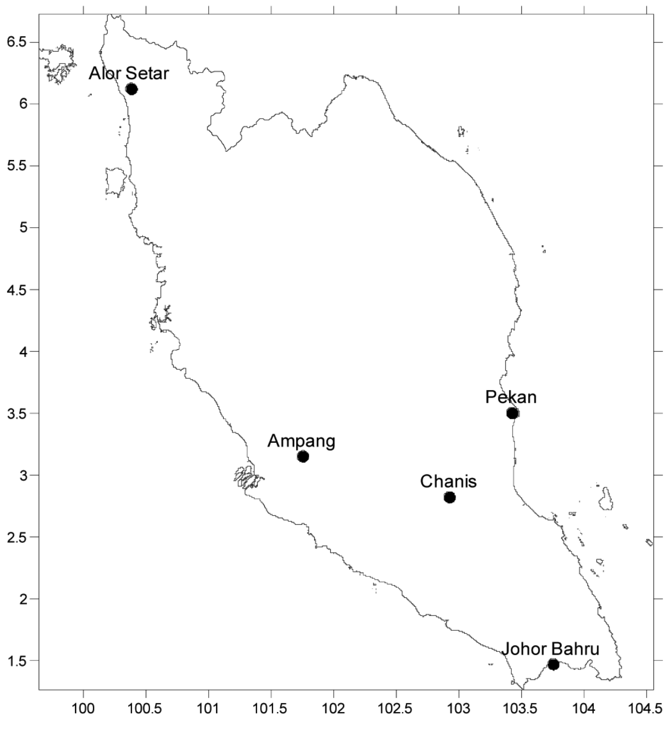

FIGURE 1. Location of stations

THRESHOLD SELECTION

Let $x_1, x_2, \ldots, x_n$ be a series of independent observations of a random variable X with unknown distribution function $F_x(x)$. For modeling the upper tail of $F_x(x)$, consider k exceedances of X over a specified threshold u, and let y_1, $y_2, \ldots, y_k$ be the excesses or peaks so that $y_i = x_i - u$. Based on the EVT, the conditional distribution of excesses, $Y = [(X-u) | X > u]$, follows the GPD as $u \to \infty$ (or u tends towards infinity) with distribution function (Balkema & de Haan 1974; Pickands 1975),

$$F_Y(y) = \begin{cases} 1 - \left(1 - \dfrac{\xi y}{\sigma}\right)^{1/\xi} & \xi \neq 0 \\ 1 - \exp\left(-\dfrac{y}{\sigma}\right) & \xi = 0 \end{cases}, \tag{1}$$

where σ is the scale parameter, ξ is the shape parameter, the range of y is $0 < y < \infty$ for $\xi \leq 0$ and $0 < y < \dfrac{\sigma}{\xi}$ for $\xi > 0$.

One of the main issues in fitting the GPD on a selected excess data is that the optimal threshold is required. To select an optimal threshold, an analyst must compromise between two contradictory strategies, using a lower tail closer to the central data which provides more data but introduces bias towards the central values or using a higher tail which provides less data but leads to larger variance in the estimated parameters. In this study, we use three plots, namely the mean frequency, the mean excess and the threshold choice, as our selection criteria for choosing an optimal threshold. For the mean frequency plot, the mean of annual number of exceedances above a set of threshold value is calculated and the threshold value is chosen between [1.2,5] to fulfill the independence condition of Poisson process (Begueria 2005). For the mean excess plot, the threshold value is chosen from the domain where the mean excess, $e(u) = E[(X-u) | X > u]$, is a linear function of the threshold level (Davison & Smith 1990). Finally, for the threshold choice plot, the GPD is fitted to a range of thresholds and the stability of parameter estimates for the shape parameter, ξ, and the modified scale parameter $\tilde{\sigma} = \sigma + \xi u$ are checked. The threshold value was chosen from the domain where the estimates are approximately constant above the threshold (Coles 2001).

Once the appropriate threshold has been selected, the interdependencies or the serial correlations of the excess data are analyzed using declustering technique. In particular, we will check the exceedances time series and reject the second peak if it occurs within 120 h (or 5 days) from the first peak (USWRC 1976).

ESTIMATION METHODS

In this study, the estimation of parameters are performed using the MOM, the PWM, the LMOM, the TLMOM and the ML methods. Let $y_1, y_2, \ldots, y_k$ be the sample of excess of random variable $Y = [(X-u) | X > u]$, $y_{1:k} \leq y_{2:k} \leq \ldots \leq y_{k:k}$ be the order statistics of the sample, $\bar{y}$ be the sample mean and s^2 be the sample variance.

Moments (MOM). MOM estimates of ξ and σ can be easily obtained by utilizing the mean and variance of the exceedances. The estimates of ξ and σ are given by $\hat{\xi} = \frac{1}{2}\left(\frac{\bar{y}^2}{s^2} - 1\right)$ and $\hat{\sigma} = \frac{1}{2}\bar{y}\left(\frac{\bar{y}^2}{s^2} + 1\right)$.

Probability Weighted Moments (PWM). Based on Hosking and Wallis (1987), the general expression for the rth order of PWM of GPD is given by:

$$M_r = \frac{u}{1+r} + \frac{\sigma}{(1+r)(1+r+\xi)}.$$

The estimate of the rth order of PWM is given by:

$$\hat{M}_r = \sum_{i=1}^{k} (1-P_i)^2 y_{i:k} / k,$$

with plotting position $P_i = \dfrac{i - 0.35}{k}$. The estimates of ξ and σ can be obtained using $\hat{\xi} = \dfrac{\hat{M}_0}{\hat{M}_0 - 2\hat{M}_1} - 2$ and $\hat{\sigma} = \dfrac{2\hat{M}_0\hat{M}_1}{\hat{M}_0 - 2\hat{M}_1}$.

L-Moments (LMOM). L-moments are linear combinations of order statistics. Therefore, even though the theory of L-moments is parallel to the theory of PWM, the L-moments are less affected by sampling variability and more robust to outliers (Hosking 1990). The L-moments of order r can be defined as:

$$\lambda_r = \frac{1}{r}\sum_{i=0}^{r-1}(-1)^i \binom{r-1}{i} E\left(Y_{r-i:r}\right),$$

for $r = 1, 2, \ldots$. By using the first and the second samples of L-moments, $\hat{\lambda}_1 = \dfrac{\hat{\sigma}}{1+\hat{\xi}}$ and $\hat{\lambda}_2 = \dfrac{\hat{\sigma}}{(1+\hat{\xi})(2+\hat{\xi})}$, the estimates of GPD parameters are $\hat{\xi} = \dfrac{\hat{\lambda}_1}{\hat{\lambda}_2} - 2$ and $\hat{\sigma} = \hat{\lambda}_1\left(1+\hat{\xi}\right)$.

Trimmed L-Moments (TLMOM). Trimmed L-moments are the natural generalization of L-moments that do not require the existence of the mean of the underlying distribution (Elamir & Seheult 2003). Since the TL-moments assign zero weight to extreme observations, they are more robust to outliers compared to the LMOM. The TL-moments of order r is defined as:

$$\lambda_r^{(t_1,t_2)} = \frac{1}{r}\sum_{i=0}^{r-1}(-1)^i \binom{r-1}{i} E\left(Y_{r+t_1} - i : r + t_1 + t_2\right),$$

where t_1 and t_2 are the amount of trimming (We focus on the symmetric case $t_1 = t_2 = t$). By using the first, the second and the third samples of TL-moments, $\hat{\lambda}_1^{(1)} = \dfrac{\hat{\sigma}\left(5+\hat{\xi}\right)}{\left(2+\hat{\xi}\right)\left(3+\hat{\xi}\right)}$, $\hat{\lambda}_2^{(1)} = \dfrac{6\hat{\sigma}}{\left(2+\hat{\xi}\right)\left(3+\hat{\xi}\right)\left(4+\hat{\xi}\right)}$

and $\hat{\lambda}_3^{(1)} = \frac{60\hat{\sigma}\left(1-\hat{\xi}\right)}{9\left(2+\hat{\xi}\right)\left(3+\hat{\xi}\right)\left(4+\hat{\xi}\right)\left(5+\hat{\xi}\right)}$, the estimates of GPD parameters are $\hat{\xi} = \frac{10-45\hat{\tau}_3^{(1)}}{9\hat{\tau}_3^{(1)}+10}$ and $\hat{\sigma} = \frac{\left(\hat{\xi}+2\right)\left(\hat{\xi}+3\right)}{\left(\hat{\xi}+5\right)}\hat{\lambda}_1^{(1)}$,

where $\hat{\tau}_3^{(1)} = \hat{\lambda}_3^{(1)} / \hat{\lambda}_2^{(1)}$.

Maximum Likelihood (ML). From the GPD distribution function in equation (1), the probability distribution function of GPD for $\xi \neq 0$ is:

$$f_Y(y_{i:k}) = \frac{1}{\sigma}\left(1-\frac{\xi y_{i:k}}{\sigma}\right)^{\frac{1}{\xi}-1}.$$

Therefore, the log-likelihood is:

$$l\left(y_{i:k};\hat{\sigma},\hat{\xi}\right) = -k\ln\hat{\sigma} + \left(\frac{1}{\hat{\xi}}-1\right)\sum_{i=1}^{k}\ln\left(1+\frac{\hat{\xi}y_{i:k}}{\hat{\sigma}}\right),$$

and the ML estimates of the shape, ξ, and the scale, σ, parameters can be obtained by maximizing the log likelihood. Numerical techniques such as the Newton-Raphson method can be applied to obtain the parameter estimates.

T-YEAR RETURN LEVEL

The tail distribution of X is useful for researchers of environmental and financial risk management studies, as it can be used to predict high or extreme quantiles. Since the distribution of $X < u$ is unknown and the threshold chosen is sufficiently large to assume the Poisson recurrence process, the yearly rate of exceedances, which is also known as the crossing rate, λ, is the number of exceedances over the number of recorded years, i.e. $\lambda = k/t$, and hence, the T-year exceedance Y_T is defined as $(1 - 1/\lambda T)$ quantile in the distribution of the exceedances. Inverting equation (1) and substituting $y = x - u$, the expression of the estimated T-year return level $\hat{X}_T$ is obtained as:

$$\hat{X}_T = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(1-\left(\frac{1}{\lambda T}\right)^{\hat{\xi}}\right). \qquad (2)$$

In this study, the return levels for $T = 5, 10, 25, 50, 100, 200$ and 500 years were calculated based on the parameters estimated from five methods (MOM, PWM, LMOM, TLMOM and ML). The performances of all methods are analyzed based on the RMSE obtained from Monte Carlo simulation (with 500 simulations). The RMSE is calculated as:

$$RMSE = \frac{E\left[\left(\hat{X}_T^* - \hat{X}_T\right)^2\right]^{0.5}}{\hat{X}_T}, \qquad (3)$$

where $\hat{X}_T^*$ is the simulated return levels. The 95% confidence bounds are also obtained from Monte Carlo simulation using:

$$mean\left(\hat{X}_T^*\right) \pm 1.96\sqrt{Var\left(\hat{X}_T^*\right)}. \qquad (4)$$

Since the size of simulation numbers (500 simulations) is large, the confidence bound in (4) can be considered as an appropriate asymptotic measure.

WEIGHTED AVERAGE RETURN LEVEL

We suggest the weighted average model for the estimation of T-year return level. The advantage of using this model is that several methods can be considered in estimating the return level, whereby each method is weighted by the inverse variance, indicating that the method with higher variance has lesser weight and vice versa. Intuitively, the best method for each station may differ, due to the uniqueness of data and the differences in the assumptions of each method. The proposed model is a suitable alternative if there is no single method which is uniformly best for all stations and if the analyst wants to take into accounts more than one method to predict the return level. The weighted average return level can be calculated as:

$$\hat{\theta} = \sum_{i=1}^{n} w_i\hat{\theta}_i, \qquad (5)$$

where $w_i = \left[\frac{1}{Var\left(\hat{\theta}_i\right)}\right] \Big/ \left[\sum_{i=1}^{n}\frac{1}{Var\left(\hat{\theta}_i\right)}\right]$ and n is the number of methods. The 95% confidence intervals are obtained through $\hat{\theta} \pm 1.96\sqrt{Var\left(\hat{\theta}\right)}$ with $Var\left(\hat{\theta}\right) = \sum_{i=1}^{n} w_i^2 Var\left(\hat{\theta}_i\right)$.

RESULTS AND DISCUSSION

In this study, optimal threshold was selected based on the mean frequency, the mean excess and the threshold choice plots which were constructed for the rainfall amounts ranging from the 95th to the 99.9th percentiles. As an example, Figure 2 shows the mean excess plot (in hours), the mean frequency plot (in hours) and threshold choice plot of Alor Setar station. The threshold was chosen from [1.2,5] of the mean frequency, the domain where the mean excess is linear and the domain where the ML estimates were approximately constant. The same plots were constructed and the same criteria were applied for choosing the optimal threshold of other stations.

Table 2 provides the optimal threshold, the mean frequency, the mean excess and the parameter estimates. The results indicate that the optimal thresholds of all stations were equivalent to the 99th percentile. After the selection of an appropriate threshold, the serial correlations of the excess data were checked using declustering technique. Descriptive summary for the excess data is provided in Table 3.

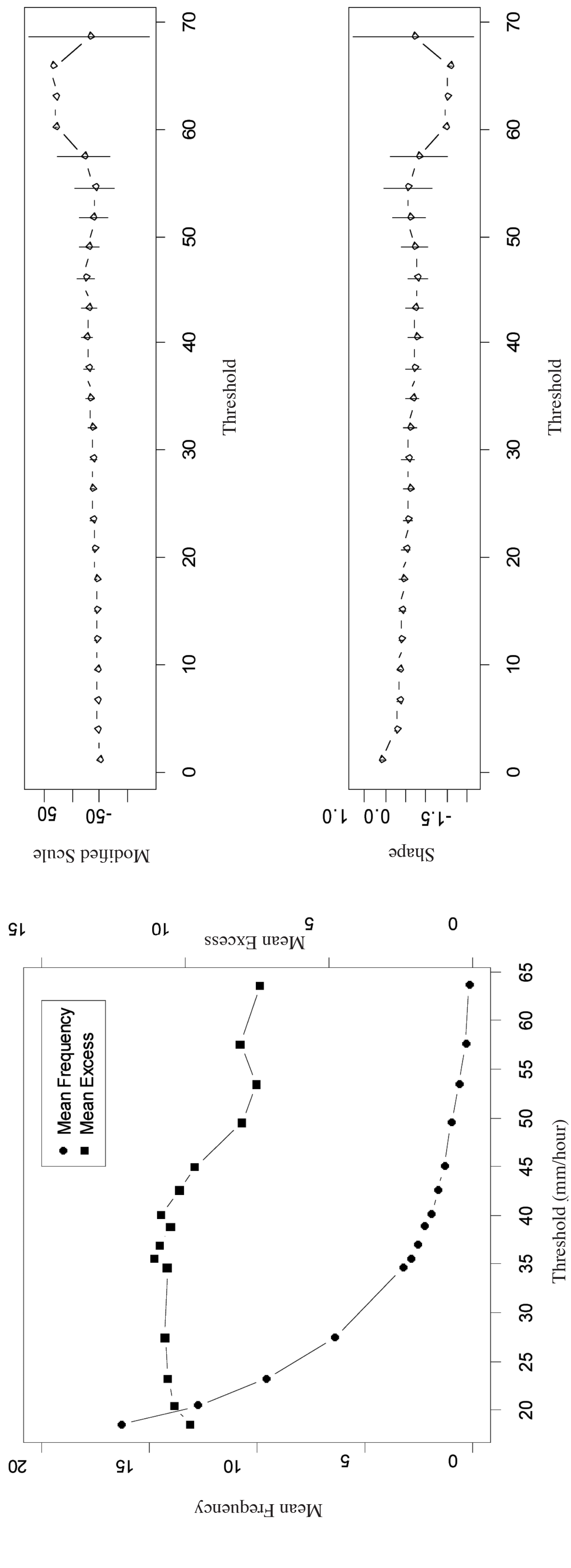

FIGURE 2. Mean frequency plot, mean excess plot and threshold choice plot for Alor Setar station

TABLE 2. List of selected thresholds

Raingauge station	Threshold value	Mean frequency	Mean excess	Scale estimated (standard error)	Shape estimated (standard error)	Number of exceedances (after declustering)
Alor Setar	34.60	3.26	10.62	12.95 (1.61)	- 0.18 (0.09)	112
Pekan	37.29	3.46	11.13	11.50 (1.47)	- 0.03 (0.09)	129
Johor Bahru	40.64	3.44	10.65	10.39 (1.46)	0.03 (0.11)	114
Ampang	46.35	3.18	11.30	12.16 (1.78)	- 0.08 (0.12)	112
Chanis	37.00	2.79	10.58	12.16 (1.93)	- 0.10 (0.11)	74

TABLE 3. Descriptive summary for exceedance data

Raingauge station	$\bar{y}$	s^2	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_1^{(1)}$	$\hat{\lambda}_2^{(1)}$	$\hat{\lambda}_3^{(1)}$
Alor Setar	11.38	95.97	11.38	5.29	10.02	2.86	0.51
Pekan	11.23	122.50	11.23	5.70	9.35	2.89	0.68
Johor Bahru	11.29	127.48	11.29	5.91	9.47	3.16	0.65
Ampang	11.96	124.40	11.96	6.03	10.34	3.31	0.62
Chanis	11.22	105.36	11.22	5.34	9.54	2.65	0.57

The exceedance data were fitted using the MOM, PWM, LMOM, TLMOM and ML methods and the parameter estimates are provided in Table 4. The results indicated that the parameter estimates are similar, showing only minimal deviations between different methods. The ML method provides converged solutions for all stations, indicating that the fitting procedures for the exceedance hourly data did not encounter any numerical problems. The shape parameters for all stations were nearly zero, implying possible utilization of the exponential distribution.

The *T*-year return levels for all methods were calculated using (2) and the results are displayed in Table 5. The performance of each estimation method was evaluated based on the RMSE measure obtained from Monte Carlo simulation (with 500 simulations), where a smaller RMSE indicated a better performance. From the overall results, the percentage of ML estimators that produced the lowest RMSE was 55.7%, followed by the MOM estimators (44.3%). This result was expected due to the asymptotic properties of the ML estimators.

Finally, the weighted *T*-year return levels based on the best two methods, the ML and the MOM, were calculated using (5) and the 95% confidence intervals of the weighted *T*-year return levels were obtained. The results are displayed in Table 6 and the plots are shown in Figure 3. As expected, the longer the years, the higher the return levels and the larger the confidence bounds. The 100-year return period is of particular interest as it is the standard measure utilized in infrastructure design.

CONCLUSION

In this paper, we estimated the GPD parameters of extreme rainfall events using the PDS method based on the hourly rainfall data of five stations in Peninsular Malaysia. In particular, the GPD parameters were estimated using five methods namely the MOM, the PWM, the LMOM, the TLMOM and the ML. Based on the results, the parameter estimates for all methods were similar, showing only minimal deviations between different methods.

In addition, from this research we predicted the *T*-year return level of each estimation method and analyzed the performance of each method based on the RMSE measure obtained from Monte Carlo simulation. The overall results

TABLE 4. Parameter estimates for all methods

Method	Alor Setar		Pekan		Johor Bahru		Ampang		Chanis	
	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Shape	Scale	Shape
MOM	13.36	0.174	11.39	0.015	11.28	-0.003	12.87	0.075	12.32	0.098
PWM	13.16	0.157	10.96	-0.024	10.34	-0.084	11.85	-0.010	12.47	0.111
LMOM	13.08	0.150	10.89	-0.030	10.26	-0.091	11.76	-0.017	12.35	0.101
TLMOM	13.36	0.174	10.85	-0.052	11.82	0.062	13.53	0.141	11.66	0.028
ML	13.73	0.205	11.50	0.024	11.36	0.006	13.67	0.138	12.52	0.115

TABLE 5. T-year return levels and simulated RMSE for all methods

Alor Setar	MOM				PWM				LMOM				TLMOM				ML			
T	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%
5	63.08	0.04	58.26	67.81	63.24	0.04	58.31	67.95	63.33	0.04	58.24	68.34	63.08	0.05	57.09	69.76	62.79	0.03	58.18	66.47
10	68.57	0.04	62.47	74.09	68.93	0.05	61.68	75.64	69.11	0.05	61.75	76.59	68.57	0.08	58.63	79.35	67.93	0.04	62.14	72.89
25	74.88	0.06	65.97	82.82	75.56	0.07	64.67	86.48	75.88	0.08	65.18	87.57	74.88	0.11	60.68	91.70	73.70	0.05	65.08	80.03
50	79.02	0.07	67.51	90.14	79.98	0.09	65.45	93.59	80.42	0.10	65.97	96.14	79.02	0.15	57.64	103.47	77.39	0.06	66.97	84.85
100	82.70	0.09	68.54	96.80	83.94	0.11	66.55	102.13	84.51	0.11	67.07	104.53	82.70	0.18	57.36	115.09	80.60	0.08	67.50	90.39
200	85.96	0.10	69.37	101.70	87.50	0.13	66.30	110.89	88.20	0.14	66.86	113.29	85.96	0.24	57.66	130.33	83.38	0.08	68.37	94.44
500	89.71	0.11	70.34	109.58	91.64	0.16	62.50	120.95	92.53	0.17	65.00	125.40	89.71	0.29	47.63	145.50	86.50	0.11	66.16	102.60
Pekan	**MOM**				**PWM**				**LMOM**				**TLMOM**				**ML**			
T	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%
5	68.59	0.05	61.75	74.76	69.08	0.05	62.23	75.62	69.17	0.05	62.03	75.92	70.27	0.07	61.03	79.76	68.49	0.05	61.33	75.11
10	76.12	0.06	66.43	84.87	77.27	0.07	66.55	87.44	77.46	0.07	66.63	88.49	79.18	0.09	64.93	94.20	75.88	0.06	65.73	84.01
25	85.96	0.08	71.24	99.49	88.29	0.10	70.88	104.52	88.69	0.11	70.38	107.62	91.46	0.15	65.66	119.69	85.45	0.09	69.78	99.28
50	93.31	0.10	75.27	110.62	96.78	0.13	72.40	121.68	97.39	0.14	72.40	124.08	101.14	0.19	64.52	138.70	92.56	0.12	70.42	113.15
100	100.58	0.12	74.79	121.64	105.42	0.17	70.92	140.55	106.27	0.17	72.48	143.92	111.18	0.27	58.53	175.60	99.54	0.14	71.58	125.25
200	107.78	0.14	77.03	138.19	114.20	0.18	72.89	154.54	115.33	0.20	72.38	162.63	121.58	0.35	48.61	213.14	106.41	0.16	70.90	137.45
500	117.18	0.19	76.02	161.62	126.02	0.25	68.20	191.97	127.61	0.28	64.66	203.41	135.92	0.49	26.72	280.11	115.32	0.19	70.84	157.83
Johor Bahru	**MOM**				**PWM**				**LMOM**				**TLMOM**				**ML**			
T	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%
5	70.89	0.05	64.26	77.11	71.73	0.05	64.11	79.15	71.79	0.05	64.11	79.01	74.00	0.06	65.56	82.27	70.84	0.04	64.53	76.91
10	78.72	0.06	69.60	87.08	80.98	0.07	69.12	92.40	81.16	0.07	69.32	92.63	81.75	0.08	70.05	95.24	78.56	0.06	69.02	87.44
25	89.07	0.08	74.22	101.23	94.05	0.11	73.31	113.46	94.49	0.11	74.72	115.04	91.50	0.12	70.53	111.91	88.72	0.09	72.91	102.15
50	96.90	0.10	77.32	113.12	104.63	0.14	77.36	133.19	105.34	0.14	77.40	134.47	98.52	0.16	71.17	131.09	96.36	0.11	74.54	116.74
100	104.73	0.12	80.13	127.29	115.85	0.17	78.02	156.02	116.90	0.19	75.66	160.98	105.24	0.22	63.92	153.93	103.98	0.14	73.80	131.48
200	112.56	0.14	79.59	142.36	127.74	0.21	76.21	178.99	129.21	0.22	73.91	184.11	111.67	0.23	63.45	165.35	111.56	0.16	75.50	144.38
500	122.92	0.18	80.37	165.14	144.56	0.26	72.75	220.61	146.72	0.30	69.11	237.30	119.76	0.36	46.26	211.27	121.53	0.21	71.62	170.66

(continued)

TABLE 5. *Continue*

Ampang	MOM				PWM				LMOM				TLMOM				ML			
T	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%
5	77.43	0.04	71.31	82.67	78.32	0.04	71.34	84.95	78.38	0.04	71.43	85.13	80.48	0.05	72.70	88.06	76.81	0.03	71.26	80.95
10	84.54	0.05	76.10	91.92	86.78	0.06	76.48	97.48	86.95	0.06	76.52	97.74	87.45	0.06	76.46	98.45	83.06	0.04	75.64	89.16
25	93.40	0.06	80.93	103.87	98.04	0.09	81.27	114.50	98.44	0.09	81.00	115.36	95.68	0.10	77.34	114.56	90.46	0.06	79.29	99.22
50	99.71	0.08	84.03	114.76	106.63	0.11	82.95	127.44	107.24	0.12	83.18	132.69	101.24	0.13	76.96	128.71	95.47	0.07	81.79	107.20
100	105.70	0.10	84.26	126.52	115.27	0.14	83.17	146.52	116.15	0.15	85.36	151.78	106.28	0.17	75.61	143.47	100.03	0.08	82.93	114.61
200	111.39	0.11	86.50	134.82	123.97	0.18	81.01	168.75	125.16	0.19	82.38	174.58	110.85	0.20	73.09	158.26	104.16	0.10	82.42	121.53
500	118.47	0.15	84.11	152.28	135.56	0.22	82.29	196.10	137.24	0.24	79.10	207.94	116.24	0.25	66.97	178.66	109.06	0.11	83.09	130.61
Chanis	MOM				PWM				LMOM				TLMOM				ML			
T	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%	$\hat{X}_T$	RMSE	LB 95%	UB 95%
5	64.76	0.04	59.75	69.53	64.65	0.04	59.19	70.01	64.72	0.04	59.78	69.88	63.88	0.05	57.38	70.17	64.63	0.04	59.59	69.25
10	71.19	0.04	64.82	76.78	70.92	0.05	63.65	77.90	71.11	0.05	63.79	78.28	70.88	0.07	60.93	80.94	70.85	0.05	63.62	76.65
25	79.05	0.06	68.51	87.81	78.51	0.08	66.73	89.83	78.91	0.07	68.13	89.81	79.91	0.12	62.42	98.57	78.36	0.06	67.87	87.28
50	84.54	0.08	71.01	97.54	83.75	0.09	69.22	97.97	84.34	0.09	70.08	99.58	86.60	0.16	62.06	115.91	83.53	0.08	69.84	94.98
100	89.68	0.10	71.97	106.41	88.61	0.12	68.77	108.98	89.41	0.11	70.11	110.30	93.15	0.21	58.09	134.84	88.31	0.09	71.43	102.68
200	94.47	0.12	71.88	115.72	93.11	0.15	68.04	121.17	94.13	0.15	68.22	122.86	99.57	0.28	53.29	159.14	92.72	0.11	71.09	109.54
500	100.33	0.13	73.20	124.73	98.55	0.17	68.39	132.64	99.89	0.19	66.65	139.92	107.88	0.35	43.81	187.51	98.03	0.13	71.26	120.91

TABLE 6. Weighted T-year return levels and their associated 95% confidence intervals

T (Years)	Alor Setar			Pekan			Johor Bahru			Ampang			Chanis		
	$\hat{X}_T$	LB 95%	UB 95%	$\hat{X}_T$	LB 95%	UB 95%	$\hat{X}_T$	LB 95%	UB 95%	$\hat{X}_T$	LB 95%	UB 95%	$\hat{X}_T$	LB 95%	UB 95%
5	62.91	59.50	65.77	68.54	63.51	72.97	70.86	66.25	75.17	77.07	72.79	80.16	64.49	61.10	67.96
10	68.22	63.92	71.82	76.00	68.76	81.75	78.64	71.95	84.63	83.68	77.94	88.22	71.03	66.09	74.90
25	74.22	67.78	78.95	85.72	74.76	95.18	88.91	77.71	97.55	91.72	83.09	98.12	78.70	71.03	84.71
50	78.02	70.02	84.04	93.01	78.85	106.09	96.67	81.75	109.04	97.19	86.69	106.27	84.01	74.17	92.42
100	81.43	71.53	89.32	100.13	80.66	115.95	104.43	85.03	121.53	102.07	88.48	113.82	88.93	76.45	99.59
200	84.40	72.89	93.18	107.15	83.52	128.55	112.11	87.31	133.70	107.02	90.21	120.61	93.48	77.38	106.28
500	87.99	73.62	100.31	116.26	86.10	147.12	122.33	89.87	154.28	112.14	91.07	130.05	99.14	79.60	115.35

(LB 95%: 95% Lower bound; UB 95%: 95% Upper bound)

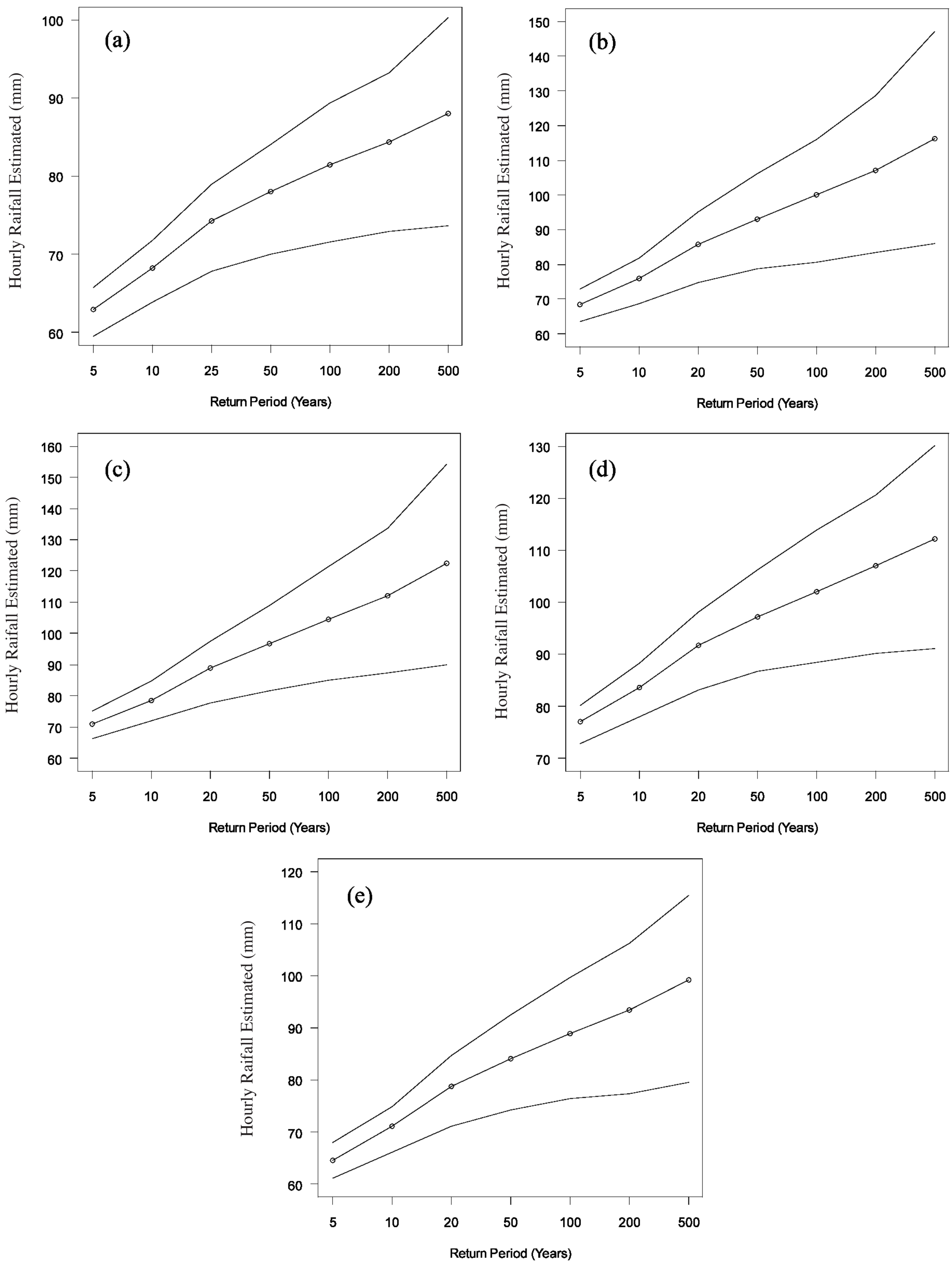

FIGURE 3. Plots of weighted return levels and 95% confidence intervals: (a) Alor Setar, (b) Pekan, (c) Johor Bahru, (d) Ampang and (e) Chanis

imply that the ML was the best method, followed by the MOM.

Finally, we also suggested the weighted average model, a model which assigns the inverse variance of several methods as weights, to estimate the *T*-year return level. Based on the results, the weighted average model was a suitable alternative for predicting the *T*-year return level if there was no single method which was uniformly best for all stations and if the analyst wants to take into accounts more than one method to predict the return level. In this study, the weighted *T*-year return level utilizes the ML and MOM methods.

For application purposes, the prediction of extreme rainfall events for several return periods provide important and valuable information for the management and the planning of water resources, especially for proper drainage systems, reservoirs and ground surface waters and for utilizations in agricultural sectors and socio-economic activities. The information can be used to facilitate the governments and other related parties in prioritizing water resources in their efforts to reduce or control the risks of large losses.

As mentioned previously, the estimation of GPD parameters for extreme data are generally implemented using two main approaches, parametric and non-parametric. This study has focused on the estimation methods from the fully parametric approach. In the last few years however, there has been a shift from the parametric approach for statistics of extremes towards a semi-parametric approach for the estimation of the right-tail distribution. In fact, several studies in semi-parametric approach have provided interesting and meaningful results, especially in determining the optimal threshold, which resulted in better estimates and inferences. The semi-parametric approach of the PDS method will be pursued in our future studies.

ACKNOWLEDGEMENTS

The authors express their thanks to the Drainage and Irrigation Department for providing the hourly rainfall data. The authors are also grateful to the reviewers for providing valuable comments that help to improve this manuscript. This research was funded by Universiti Kebangsaan Malaysia (UKM-ST-06-FRGS0181-2010).

REFERENCES

Ashkar, F. & Tatsambon, C.N. 2007. Revisiting some estimation methods for the generalized Pareto distribution. *Journal of Hydrology* 346: 136-143.

Balkema, A. & de Haan, L. 1974. Residual life time at great age. *The Annals of Probability* 2: 792-804.

Begueria, S. 2005. Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value. *Journal of Hydrology* 303: 215-230.

Coles, S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.

Coles, S. & Dixon, M.J. 1999. Likelihood-based inference for extreme value models. *Extremes* 2 (1): 5-23.

Coles, S., Pericchi, L.R. & Sisson, S. 2003. A fully probabilistic approach to extreme rainfall modelling. *Journal of Hydrology* 273: 35-50.

Cunnane, C. 1973. A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology* 18: 257-271.

Davison, A.C. 1984. Modelling excesses over high thresholds, with an application. In: *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira. Dordrecht: Reidel. pp. 461-482.

Davison, A.C. & Smith, R.L. 1990. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society* 52 (3): 393-442.

Deni, S.M., Suhaila, J., Zin, W.Z.W. & Jemain, A.A. 2009. Trends of wet spells over Peninsular Malaysia during monsoon seasons. *Sains Malaysiana* 38(2): 133-142.

Elamir, E.A. & Seheult, A.H. 2003. Trimmed L-moments. *Computational Statistics & Data Analysis* 43: 299-314.

Floris, M., D'Alpaos, A., Squarzoni, C., Genevois, R. & Marani, M. 2010. Recent changes in rainfall characteristics and their influence on thresholds for debris flow triggering in the Dolomitic area of Cortina d'Ampezzo, North-Eastern Italian Alps. *Natural Hazards Earth System* 10: 571-580.

Harasawa, H. & Nishioka, S.E. 2003. *Climate Change on Japan*. Tokyo: KokonShoin Publications.

Harris, R.I. 1999. Improvements to the 'method of independent storms'. *J. Wind Eng. Ind. Aerodyn*. 80: 1-30.

Hosking, J.R.M. & Wallis, J.R. 1987. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29: 339-349.

Hosking, J.R.M. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistical Society* 52: 105-124.

IPCC 2001. The Scientific Basis. *Contribution of Working Group I to the Second Assessment Report of the Intergovernment Panel on Climate Change*. New York: Cambridge University Press.

IPCC 2007. Climate Change 2007– *The Physical Basis*. United Kingdom: Cambridge University Press.

Juárez, S.F. & Schucany, W.R. 2004. Robust and efficient estimation of the generalized Pareto distribution. *Extremes* 7: 237-251.

Katz, R., Parlange, M. & Naveau, P. 2002. Statistics of extremes in hydrology. *Advance Water Resources* 25: 1287-1304.

Kouchak, A.A. & Nasrollahi, N. 2010. Semi-parametric and parametric inference of extreme value models for rainfall data. *Water Resources Management* 24: 1229-1249.

Lana, X., Burgueno, A., Martinez, M.D. & Serra, C. 2006. Statistical distribution and sampling strategies for the analysis of extreme dry spells in Catalonia (NE Spain). *Journal of Hydrology* 324: 94-114.

Lang, M., Ouarda, T.B.M.J. & Bobee, B. 1999. Towards operational guidelines for over-threshold modelling. *Journal of Hydrology* 225: 103-117.

Li, Y., Cai, W. & Campbell, E.P. 2005. Statistical modelling of extreme rainfall in southwest Western Australia. *Journal of Climate* 18: 852-863.

Madsen, H., Rasmussen, P.F. & Rosbjerg, D. 1997. Comparison of annual maximum series and partial duration series for modelling extreme hydrologic events, 1. At-site modelling. *Water Resources Research* 33: 747-757.

Moharram, S.H., Gosain, A.K. & Kapoor, P.N. 1993. A comparative study for the estimators of the generalized Pareto distribution. *Journal of Hydrology* 150: 169-185.

Pandey, M.D., Van Gelder, P.H.A.J.M. & Vrijling, J.K. 2003. Bootstrap simulations for evaluating the uncertainty associated with peaks-over-threshold estimates of extreme wind velocity. *Environmetrics* 14: 27-43.

Pickands, J. 1975. Statistical inference using extreme order statistics. *Annals of Statistics* 3: 119-130.

Rasmussen, P.F., Ashkar, F., Rosbjerg, D. & Bobee, B. 1994. The POT method for flood estimation: A review. *Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Extreme Values: Floods and Droughts* 1: 15-26.

Smith, R.L. 1986. Extreme value theory based on the *r* largest annual events. *Journal of Hydrology* 86: 27-43.

Smith, R.L. 2001. Extreme value statistics in meteorology and environment. *Environmental Statistics* 8: 300-357.

Todorovic, P. 1978. Stochastic models of floods. *Water Resources Research* 20: 914-920.

USWRC 1976. Guidelines for determining flood flow frequency. Washington, DC: United States Water Resources Council, Bulletin 17, *Hydrl. Comm.* 73pp.

Zin, W.Z.W., Jemain, A.A., Ibrahim, K., Suhaila, J. & Deni, S.M. 2009. A comparative study of extreme rainfall in Peninsular Malaysia: With reference to partial duration and annual extreme series. *Sains Malaysiana* 38(5): 751-760.

Zin, W.Z.W. & Jemain, A.A. 2010. Statistical distributions of extreme dry spell in Peninsular Malaysia. *Theoretical & Applied Climatology* 102(3-4): 253-264.

Zin, W.Z.W., Suhaila, J., Deni, S.M. & Jemain, A.A. 2010. Recent changes in extreme rainfall events in Peninsular Malaysia: 1971-2005. *Theoretical & Applied Climatology* 99(34): 303-314.

School of Mathematical Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: lingshinyie@gmail.com

Received: 30 September 2011
Accepted: 29 May 2012