Modeling for Extreme Rainfall in Lower Northeastern of Thailand

Saowanee Rattanawan, Mahasarakham University, Thailand
Piyapatr Busababodhin, Mahasarakham University, Thailand
Bungon Kumphon, Mahasarakham University, Thailand

The Asian Conference on Sustainability, Energy & the Environment 2014
Official Conference Proceedings
0684

Abstract
The purpose of this research is to model extreme rainfall in lower northeastern of Thailand by using the Generalized Extreme Value Distribution (GEV) and Generalized Pareto Distribution (GPD). The daily rainfall, which are obtained from the Meteorological Department of Thailand, during January 1982 to September 2013; 149,475 values, from eleven stations in the lower northeastern of Thailand are studied. The source code from analyzing the data was provided by R program with extreme function in extRemes library. It is able to directly model a data for each station by using GEV distribution and GPD with stationary process and also estimate the return levels for various return periods. The study finds that the best model is Fréchet distribution and Exponential distribution for GEV distribution and GPD, respectively. Since the Si Sa Ket Agromet station which is set at Muang district of Srisaket province has a highest return level from various return periods for both models, so it should be the first consideration for the preventing or reducing the severity of floods.

Keywords: generalized extreme value, generalized pareto distribution, return period and return level

iafor
The International Academic Forum
www.iafor.org

1. Introduction

The problem of flooding in Lower Northeast region of Thailand occurs almost every year and is likely to be more severe because of the world climate change. In August to October 2013, this area had big floods because of heavy rain; especially Srisaket, Surin and Ubon Ratchathani province (Meteorological Report, 2013). Model of Lower Northeast region rainfall can have significant value for resources planning and management, e.g. reservoir operations, agricultural practices, and flood emergency responses. In particular increased population stress on the Moon River basin, one of the key regions for Lower Northeast's socioeconomic well-being, is resulting in water quantity and quality problems. To mitigate this, effective planning and management of water resources is necessary. If amount of rainfall at each probabilities level and each return periods for this area can be calculated then the farmers will use both of them for choose suitable crops and managing the efficient crops system that appropriate in their any areas.

For a long period of time, phenomenological cases of extreme events have been studied. For example which is worth to be mentioned is a water levels record of Nile, which have recorded the lowest and highest water levels for over 5000 years in order to analyze hunger or disasters when the levels are too low or too high (Albeverio et al., 2005). As to the statistical method, Fisher and Tippett (1928) first explored extreme value theory then Gnedenko (1943) formalized extreme value distribution to which block maxima converges (Faranda et al., 2011). Over the last 50 years, extreme value theory has been used widely in applied sciences and various disciplines, such as physical, financial markets, insurance industry, environment, failure cases, and so on (e.g., Coles, 2001).

This paper mainly focuses on the analysis of daily rainfall because this area is the agricultural area, the major of framers need to used rainfall for planting. The extreme value theory are applied; GEV distribution and GPD. In Section 2 presents the definitions and theories related to extreme value theory, and estimation methods which are used for empirical analysis. Then, description of empirical analysis of data set is presented in section 3 and the estimation process and the results are shown in Section 4. Finally, a conclusion of the estimation results in daily maxima rainfall is shown in section 5.

2. Background

The theory states that under certain regularity conditions, if the maximum or minimum of random variable taken over suitably large blocks have a non-degenerate distribution, then that distribution must be the generalized extreme value (GEV) distribution. Similarly, for excesses over a suitably high threshold, analogous results state that their distribution is generalized Pareto distributions (GPD) (e.g., Coles, 2001).

2.1 The Generalized Extreme Value (GEV) Distribution

In the block maxima method is supposed to have observed the maximum value of some quantities over a number of 'blocks', a typical example being that a black is a month and the observed quantities may be some environmental quantity such as the rainfall data set at a specific location. In this paper, monthly rainfall in space and time can provide guidelines for crop scheduling, better cropping patterns and the planning and design of water resources development projects.

Suppose $X_1, X_2, ..., X_n$ be a sequence of independent variable with common distribution function $F(x)$, the maximum value of random variable $X_1, X_2, ..., X_n$ is $X_{(n)} = \max(X_1, X_2, ..., X_n)$. The cumulative distribution function (cdf.) of the GEV distribution is (Coles, 2001)

$$F(x) = \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, \tag{1}$$

and its probability density function (pdf.) is as follow,

where $1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0$, $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$.

The GEV distribution has three parameters; μ is location parameter which specifies the center of distribution, σ is scale parameter which determines the size of deviations of μ and ξ is shape parameter which shows how rapidly the upper tail decays.

The representation of Eq.(1) is combined single model which can lead to three types of non-degenerate distribution function families, i) Gumbel family which corresponds to case $\xi = 0$ i.e., GEV family with limits as $\xi \to \infty$, ii) Fréchet family which corresponds to case $\xi > 0$ of GEV family and iii) Weibull family which corresponds to case $\xi < 0$ of GEV family.

2.2 Generalized Pareto distribution (GPD)

In 1975, Pickands proposed "GPD" and it was gained wide acceptance in the EVT. It has all the flexible of exponential distribution and it was developed by Leadbetter et al. (1983). This method considers, instead of annual maxima, excess over a sufficiently high threshold in the time series (Mendes, 2010; Peter, 2011). Hence, the data set is enlarged to decrease the sampling uncertainty. As the meteorological variables tend to present successive dependent extreme values, the technique of de-clustering was applied, which considers successive extremes as belonging at the same event.

This method is supposed to have observed all values which are larger than some suitable threshold, for example all rainfall data in excess of 80 mm. These values are then assumed to follow the GPD function.

It is natural to regard as extreme events those of the x_i that exceed some high threshold, *u*. For large enough *u*, the distribution function of $x_i - u$, conditional on $x_i > u$, is approximately, (Coles, 2001)

$$H(y) = 1-\left(1+\frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{3}$$

defined on $\{y : y > 0, 1+\xi y/\tilde{\sigma} > 0\}$, $\tilde{\sigma} = \sigma + (u-\mu)$ and with scale parameter σ $(\sigma > 0)$ and shape parameter ξ $(-\infty < \xi < \infty)$. If $\xi > 0$ $(\xi < 0)$ then the GPD is simplified into the Pareto (Gamma) distribution. For $\xi \to 0$, GPD is simplified into the Exponential distribution. The family of distribution defined by (1) is called the generalized Pareto family. Denote by σ_u the value of the GPD scale parameter for a threshold of $u > u_0$, where $\sigma_u = \sigma_{u_0} + \xi(u - u_0)$, so that the scale parameter changes u unless $\xi = 0$. A modified scale is obtained by reparameterising the GPD scale parameter as $\sigma^* = \tilde{\sigma} - \xi u$. A threshold μ_0 is selected as the lowest value of u for which the estimates of σ^* and ξ remain near constant and the probability density function (cdf.) of GPD is

$$h(y) = 1 + \left[\xi\left(\frac{y-\mu}{\sigma}\right)\right]^{\left(-1/\xi\right)} \quad (4)$$

where $\sigma > 0$ and $-\infty < \xi < \infty$.

2.4 Return Levels

Return values contain two quantities: return period $1/p$ and return level (recurrence interval), z_p. For annual maxima as an example, return level is an estimated high value of annual maxima temperature which is expected to be exceeded in any year during return period $1/p$ with probability p where $0 < p < 1$.

2.4.1 Return Level of GEV

In term of quantiles, take $0 < p < 1$ and define

$$z_T = \mu - \frac{\sigma}{\xi}\left\{1 - \left[-\log(1-p)\right]^{-\xi}\right\}, \tag{8}$$

where $p = \frac{1}{T}$. The MLE of return levels can be found interval confidence 100(1 − α)% for $\hat{Z}_T$ as shown the equation,

$$I_{\hat{Z}_T} : \hat{Z}_T \pm Z_{\frac{\alpha}{2}}(\sqrt{V(\hat{Z}_T)}) \tag{9}$$

2.4.2 Return Level of GPD

Suppose GPD has σ and ξ parameter was an appropriate model for the data is higher than u, that note where $Y > u$, can be written the probability function as,

$$\Pr\{Y > y\} = \varsigma_u \left[1 + \xi\left(\frac{y-u}{\sigma}\right)\right]^{-1/\xi}, \tag{10}$$

where $\xi_u = \Pr\{Y > u\}$, return levels note that mean of exceedances over threshold u, for all m observation is as

$$\zeta_u \left[1 + \xi\left(\frac{y_m - u}{\sigma}\right)\right]^{-1/\xi} = \frac{1}{m} \tag{11}$$

Consider the change of equation is

$$\hat{y}_m = u + \frac{\sigma}{\xi}\left[(m\zeta_u)^{\xi} - 1\right] \quad , \quad \text{if } \xi \neq 0 \tag{12}$$

where $\hat{y}_m$ is return level of for all of m - observation, if number of observation n_y per year, where N is number of year calculating from $m = N \times n_y$. Therefore, return level of N year is

$$\hat{Y}_N = u + \frac{\sigma}{\xi}\left[(Nn_y\zeta_u)^{\xi} - 1\right], \text{if } \xi \neq 0 \tag{13}$$

and the interval confidence 100(1−α)% for $\hat{Y}_N$ is

$$I_{\hat{Y}_N} : \hat{Y}_N \pm Z_{\frac{\alpha}{2}}(\sqrt{V(\hat{Y}_N)}) \text{ (14)}$$

3. Research Methodology

3.1 Data

This study, the maximum rainfall data of lower northeast of Thailand are used to analyze. They are obtained from the Meteorological Department of Thailand during January 1, 1982 to September 30, 2013 for eleven meteorological stations; Chaiyaphum(CP), Chokchai(NM1), Nakhon Ratchasima(NM2), Pak Chong Agromet(NM3), Nang Rong(BR), Surin Agromet(SR1), Surin(SR2), Tha Tum(SR3), Si Sa Ket Agromet(SK), Ubon Ratchathani Agromet(UB1) and Ubon Ratchathani(UB2), which are located in six provinces; Chaiyaphum, Nakhon Ratchasima, Burirum, Si Sa Ket, Surin and Ubon Ratchathani. The locations and summary statistics of the corresponding data sets are showed in Table 1.

Table 1 Locations and some summary statistics of data

Location	Latitude	Longitude	Mean	Median	Max	Skewness
CP	15.48.0.0	102.2.0.0	11.49	4.60	162.5	5.248
NM1	14.43.8.0	102.10.7.0	9.72	4.00	147.5	5.764
NM2	14.58.5.9	102.5.9.7	10.09	3.80	129.7	5.404
NM3	14.38.38.0	101.19.55.0	9.42	3.90	145.9	5.022
BR	14.35.0.0	102.48.0.0	10.44	4.40	130.5	5.352
SR1	14.53.0.0	103.30.0.0	12.57	5.20	241.6	5.618
SR2	14.53.0.0	103.27.0.0	12.21	5.30	279.5	5.582
SR3	15.19.0.0	103.41.0.0	12.75	5.70	177.7	5.245
SK	15.0.0.0	104.3.0.0	13.69	6.15	263.4	6.059
UB1	15.14.20.9	105.1.24.6	12.96	5.60	254.3	5.038
UB2	15.15.0.0	104.52.0.0	13.79	6.40	173.1	4.952

The package “Extreme” in R program is used which is able to perform parametric inferential analysis of the GEV and GPD distribution for each location in the phenomena listed above.

Fig. 1 Geographical distribution of the eleven stations in lower northeast Thailand

3.2 Analysis of GEV and GPD

The analysis of GEV is formed of three steps; 1) to find the block-maxima to define the extreme rainfall as the maximum of monthly rainfalls within each year, 2) to find the estimates of parameters in the GEV distribution by using maximum likelihood estimation (MLE) method.

Respect to the analysis of GPD is also formed of two steps; 1) to find the extreme rainfall as the maximum of excess over threshold is used. The values of threshold and the number of excesses for each location are presented in Table 4. The GPD is fitted to the tails of daily maxima rainfall data using threshold around 46.6-63.3 mm. (fixed percentile method at 99^{th}) for each stations, 2) the MLE is used with R program. Notice that thresholds are selected similarity to the Meteorological Department of Thailand’s criterion for highest rainfall stage.

4. Results of Extreme Value Theory

All estimation and calculation in empirical analysis are implemented in R with package extRemes. First a GEV model will be studied in order to get an overall idea of the annual maxima, this is compared to the GPD method; then follows the analysis of rainfall (the quantities mentioned earlier: frequency, duration, and mean maxima).

4.1 Results of GEV

As introduced above, the estimation of parameter in GEV distribution could be done by MLE method. The results of GEV are shown in Table 2.

Table 2 Locations, scale, and shape parameters of GEV distributions

Locations	μ	σ	ξ
CP	16.55(13.83,19.27)	19.53 (17.24,21.82)	0.22(0.06,0.38)
NM1	16.09(13.67,18.51)	18.37(16.35,20.39)	0.22(0.08,0.36)
NM2	15.98(13.52,18.44)	18.33(16.27,20.39)	0.23(0.08,0.38)
NM3	10.49(8.16,12.82)	14.15(12.02,16.28)	0.49(0.24,0.74)
BR	17.44(14.87,20.01)	19.56(17.44,21.68)	0.20(0.06,0.34)
SR1	19.08(16.04,22.12)	22.32(19.75,24.87)	0.24(0.09,0.39)
SR2	18.39(15.36,21.42)	21.88(19.32,24.44)	0.23(0.07,0.39)
SR3	17.25(13.89,20.61)	21.53(18.58,24.48)	0.34(0.13,0.55)
SK	14.11(10.52,17.70)	19.19(15.85,22.53)	0.57(0.28,0.86)
UB1	16.83(13.36,20.30)	21.46(18.40,24.52)	0.33(0.11,0.55)
UB2	16.19(12.59,19.79)	21.20(17.96,24.44)	0.42(0.17,0.67)

Table 2 gives the details of location, scale, and shape parameters of GEV. Notice that Shape parameter (ξ) values is positive; this implies upper bounded distribution. The 95% confidence intervals are also on the positive side. These are indicated that data are best fitted by the Fréchet distribution at all stations.

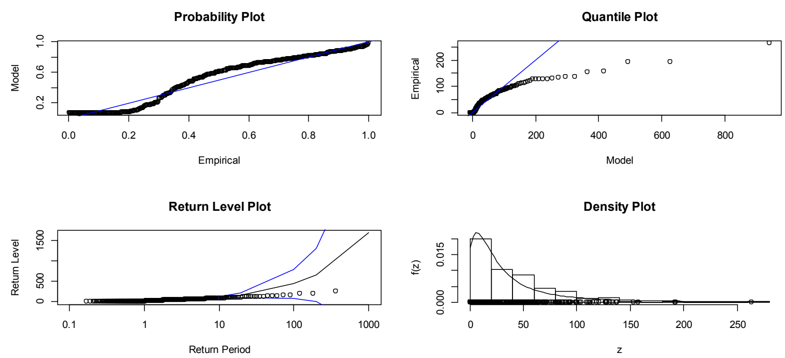


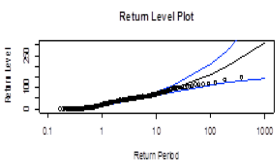
Fig.2 Diagnostic plots for GEV fit to the monthly maxima rainfall at at Si Sa Ket Agromet (SK)

Fig. 2 shows various diagnostic plots for the results of MLE of GEV fit at Si Sa Ket Agromet (SK) Station. The probability plot (P-P Plot) and quantile plot (Q-Q plot) appears approximately linear. This indicates the validity of the MLE fitting result of GEV distribution.

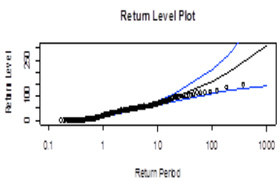
Table 3 Return levels, 95% confidence intervals associated with different return periods for GEV

Location	Return Periods			
	10-years	25-years	50-years	100-years
CP	73.60(65.69,85.72)	107.62(90.94,138.79)	137.93(110.54,194.17)	173.09(130.93,267.48)
NM1	69.43(62.21,79.97)	**101.00(86.34,126.52)**	**129.01(105.34,173.89)**	**161.39(125.27,235.26)**
NM2	69.66(62.28,80.56)	101.77(86.47,128.72)	130.42(105.47,178.27)	163.71(125.38,243.10)
NM3	**68.97(58.21,91.42)**	121.07(90.68,203.19)	178.98(120.72,363.87)	260.26(156.98,645.41)
BR	72.78(65.44,83.33)	104.49(89.87,119.57)	132.14(108.77,175.68)	163.63(128.33,234.38)
SR1	85.53(76.26,99.51)	126.11(106.70,161.47)	162.73(130.89,226.35)	205.68(156.49,312.51)
SR2	82.86(73.89,96.63)	121.72(102.92,157.38)	156.55(125.72,221.20)	197.15(149.66,306.17)
SR3	89.77(78.08,110.19)	141.15(112.71,202.45)	191.41(141.28,312.77)	254.60(172.64,477.28)
SK	**101.69(82.99,146.36)**	**188.38(133.14,365.01)**	**290.67(181.27,722.72)**	**441.88(270.00,969.99)**
UB1	88.89(77.22,110.91)	139.75(11.20,209.01)	189.39(139.08,329.23)	251.70(169.53,512.34)
UB2	96.08(81.74,125.72)	160.45(122.50,262.83)	228.04(158.05,448.63)	318.39(198.92,759.39)

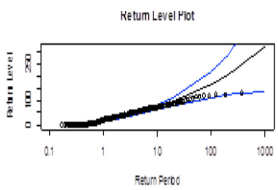
Table 3 gives the details of return levels and theirs 95% return levels confidence intervals of each location. For 10-years return period, the SK has maximum return levels but the NM3 has minimum return levels. For 25-, 50- and 100-years return period, the SK has maximum return levels but the NM1 has minimum return levels. It is clearly to show by the plots for various return periods as in Fig. 3.



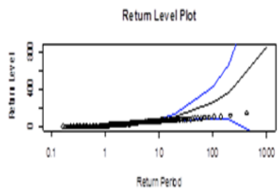
CP



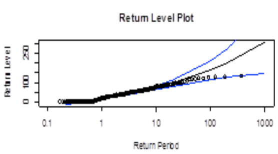
NM1



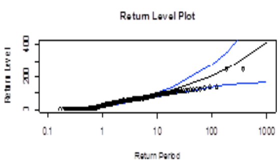
NM2



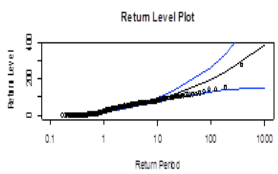
NM3



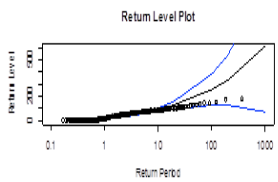
BR



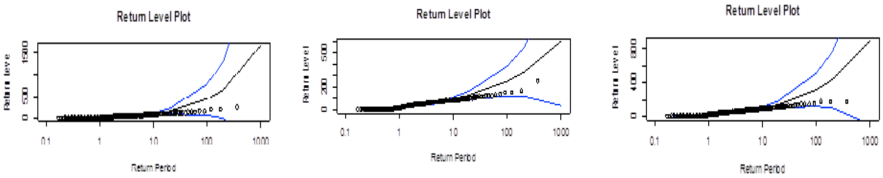
SR1



SR2



SR3



SK UB1 UB2

Fig. 3 Return level (mm.) of GEV distribution associated with 10-, 25-, 50-, and 100-year return period of each location.

4.2 Results of GPD

As to the MLE of return values estimation with GEV, method has been presented in 4.1. The corresponding results of GPD are shown in Table 4. And the MLE return level associated with a range of return period between 0 and 100 years is shown in Fig. 4, also.

Table 4 Scale and shape parameters of GPD distributions with the best fitting distribution

Location	**u (mm.)**	**No. of excesses**	σ	ξ
CP	51.6	112	18.92(14.02,23.82)	0.03(-0.15,0.21)
NM1	46.6	112	20.75(14.20,27.29)	0.03(-0.22,0.28)
NM2	49.2	113	20.91(14.93,26.89)	-0.06(-0.28,0.16)
NM3	50.1	113	19.56(15.09,24.03)	-0.10(-0.24,0.04)
BR	48.7	110	24.52(17.52,31.79)	-0.11(-0.35,0.13)
SR1	57.4	111	19.88(14.18,25.58)	0.16(-0.06,0.38)
SR2	60.8	113	20.99(15.66,26.32)	0.10(-0.08,0.28)
SR3	60.5	108	20.88(15.04,26.72)	0.06(-0.16,0.28)
SK	60.5	112	27.68(19.86,35.50)	0.09(-0.13,0.31)
UB1	62.1	112	17.78(12.64,22.92)	0.17(-0.05,0.39)
UB2	63.3	113	27.30(19.87,34.73)	-0.05(-0.25,0.15)

Table 4 shows the thresholds selected above with the excesses were fitted to the GPD with the number over the respective thresholds for each station. The parameters of distributions fit by the GPD model for excesses with the best fitting distribution are also shown. Shape parameter (ξ) values indicate that data were best fitted by the exponential distribution at all stations and theirs 95% confidence intervals are good agreement. As can be seen in Fig.5, most reanalysis data fall within the up to 25 years return periods suggesting that the strong rainfall.

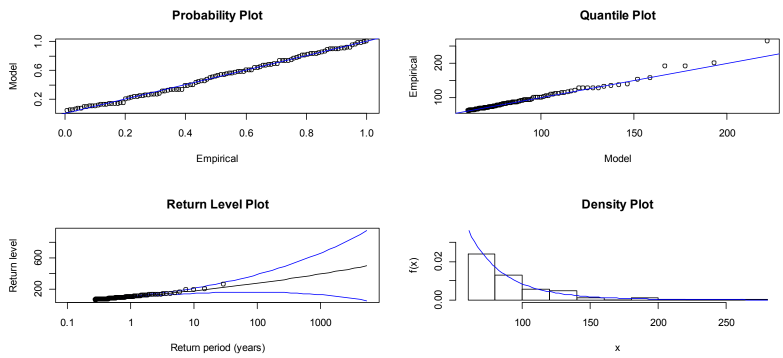


Fig.4 Diagnostic plots for GPD fit to the monthly maxima rainfall at Si Sa Ket Agromet (SK)

Fig. 4 shows various diagnostic plots for the results of MLE of GPD fit at Si Sa Ket Agromet (SK). The probability plot (P-P Plot) and quantile plot (Q-Q plot) appears approximately linear. This indicates the validity of the MLE fitting result of GPD. The corresponding estimate for such a for 10-, 25-, 50-, and 100- year return levels values and 95% confidence intervals are presented in Table 5.

Table 5 Return levels, 95% confidence intervals associated with different return periods for GPD

Location	**Return Periods**			
	10-years	**25-years**	**50-years**	**100-years**
CP	123.61(110.15,151.25)	143.43(123.86,192.37)	158.83(133.36,230.49)	174.59(142.11,275.91)
NM1	125.19(109.65,163.99)	146.71(122.61,220.48)	163.38(130.93,275.98)	180.41(138.15,345.51)
NM2	117.33(105.88,143.25)	132.64(116.23,178.12)	143.72(122.59,209.21)	154.38(127.86,244.98)
NM3	**107.20(98.58,122.24)**	**119.07(108.28,142.41)**	**127.34(114.71,158.59)**	**135.05(120.39,175.60)**
BR	122.55(111.35,149.07)	136.79(121.06,182.65)	146.63(126.65,211.62)	155.73(131.02,243.99)
SR1	156.65(134.74,210.04)	191.29(155.92,299.29)	221.04(171.85,394.99)	254.24(187.62,524.43)
SR2	147.62(129.19,185.99)	176.42(148.77,247.21)	200.04(163.34,307.12)	225.35(177.65,381.94)
SR3	143.98(127.57,181.06)	168.03(143.14,236.92)	187.08(153.84,291.11)	206.88(163.62,358.24)
SK	**177.52(153.04,233.76)**	**213.97(176.41,320.88)**	**243.60(193.07,408.04)**	**275.13(208.81,518.99)**
UB1	150.74(130.06,200.56)	183.77(149.87,285.15)	212.51(164.84,376.68)	244.94(179.76,501.53)
UB2	152.46(137.68,184.26)	172.54(151.79,227.36)	187.07(160.69,265.24)	201.07(168.24,308.28)

Table 5 gives the details of 95% return levels confidence intervals of each location. Notice that, the return levels of SK is higher than another location and the return levels of NM3 is lower than another location for all return period.

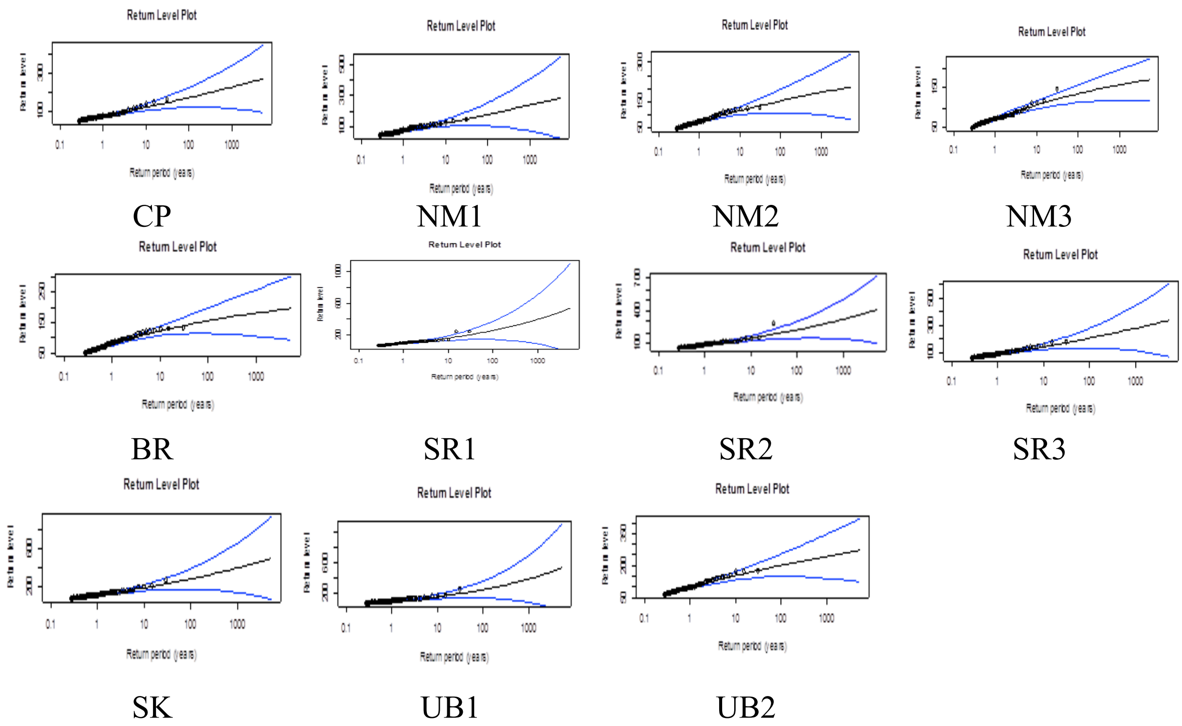


Fig. 4 Return level (mm.) of GPD distribution associated with 10-, 25-, 50-, and 100-year return period.

5. Conclusions

The extreme rainfall in Thailand's lower northeast region is occurred during the middle of May through the middle of October in every year. A statistical modeling of the annual maximum rainfall data of eleven locations in this region is developed by selecting from geographical origin of the main river sources, Moon River. By using the GEV distribution and GPD, the Fréchet distribution and the Exponential distribution is the best model selection which had an evidence of stationary for all stations, respectively. The return levels and the 95% confidence intervals associated with different return periods; 10, 25, 50 and 100 years, for GEV distribution and GPD are provided. Since the Si Sa Ket Agomet Station (SK) which is set at Muang District of Srisaket province has a highest return level with various return periods for both models, so it should be the first consideration for the preventing or reducing the severity of floods and other water-related natural disasters.

6. Acknowledgements

This paper is partially supported by Mahasarakham University, Mahasarakham, Thailand. The authors would like to thank the Meteorological Department of Thailand for the precipitation data.

7. References

Albeverio, S., Jentsch, A., and Kantz, H. Extreme Events in Nature and Society, Springer, 2005.

Coles, S. An introduction to statistical modeling of extreme values, Springer Verlag, London, 2001.

Faranda, D., Lucarini, V., Turchetti, G., and Vaietti, S. Numerical Convergence of the Block-Maxima Approach to the Generalized Extreme Value Distribution, Journal of Statistics and Physics, 145: 1156-1180, Springer, 2011.

Fisher RA, Tippett LH . Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proc Camb Phylos Soc, 24:180–190,1928.

Gnedenko B.V. Sur la distribution limite du terme maximum d'une serie aleatoire, Annals of Mathematics, 44, 423–453, 1943.

Katz RW, Parlange MB, Naveau P. Statistics of extremes in hydrology. Advanced Water Resource 25:1287–1304, 2002.

Leadbetter,M.R,Lindgren,G. and Rootzén,H. Extremes and Related propoties of random sequences and series, Springer verlag,New York, 1983.

Mendes D, Souza EP, Marengo JA, Mendes MCD. Climatology of extratropical cyclones over the South American-southern oceans sector. Theor Appl Climatol, 100:239–250, 2010.

Peter G. and Bahaj, AbuBakr S.A comparison of estimators for the generalised Pareto distribution, Ocean Engineering, 38, (11-12), 1338-1346, 2011.

Thai Meteorological Department. Climate of Thailand in year 2012[Internet]. [Cited 2013 Dec 22]. Available from www.tmd.go.th/programs%5CyearlySum.