Thai Journal of Mathematics
Special Issue (2023) :Annual Meeting in Mathematics 2022
Pages 41 - 53

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

Application of Extreme Value Theory for Maximum Rainfall at Nakhon Ratchasima Province.

Nop Sopipan[1,*]

[1] *Mathematics Education, Faculty of Science and Technology, Nakhonratchasima Rajabhat University*
e-mail: nop.s@nrru.ac.th

Abstract This research aimed to apply Extreme Value Theory (EVT) to describe maximum rainfall events in Nakhon Ratchasima Province. We considered the Generalized Extreme Value (GEV) distribution and Generalized Pareto (GPD) distribution models for the highest monthly rainfall data in Nakhon Ratchasima Meteorological Station from January 1982 to December 2021. Our findings revealed that the Fréchet distribution is the most suitable model from the GEV family. Whereas the exponential distribution provided an optimal model for monthly rainfall data above the 235 mm. threshold, we found that an increase in the return period improved the return level. Comparing the return levels between GEV and GPD, our results show that GPD is higher than GEV. Our second goal was to investigate VaR to determine the maximum rainfall risk for the following year using the Block Maxima (BM) and Peak over Threshold (POT). Using VaR backtesting, the POT outperformed the BM because in-sample data had actual violations close to the expected number.

MSC: 62G32; 62P05; 62M20
Keywords: Generalized extreme value distribution (GEV), Generalized Pareto distribution (GPD), maximum rainfall, Value at risk.

Submission date: 31.05.2022 / Acceptance date: 15.12.2022

1. Introduction

As low-pressure systems move from the South China Sea through Vietnam and into northeastern Thailand, it rains primarily in Nakhon Ratchasima province. Rain typically blows through the Northeast 2-3 times per year, resulting in the appropriate amount of rain. However, it is still causing frequent flooding in Nakhon Ratchasima, as it has in the past, such as in 1984, 2007, 2010, and 2021, and has done so for the past ten years due to heavy rain in many areas. The number of consecutive rainy days exceeds the monthly average, particularly in Nakhon Ratchasima and the surrounding areas. Furthermore, the ability of cities and communities to drain water from upstream to downstream is insufficient to support flooding. The construction of several dams and the size of canals is not enough to drain the proper amount of water. The continuous development of cities

*Corresponding author.

Published by The Mathematical Association of Thailand.

and communities also results in the invasion of water areas. As a result, drainage efficiency is low, causing the water volume to be much higher than expected, resulting in damage to housing, the economy, society, and agriculture.

Rainfall forecasting has been done using time series and the decomposition method. Monthly rainfall forecasts for Nong Khai Province ([14]) and Nakhon Ratchasima Province ([9]) used exponential smoothing methods and Box-Jenkins techniques. This study examined all of the data models. In addition, most researchers considered maximum rainfall in rainfall forecasting using extreme values. Extreme Value analysis is a statistical tool that would use the maximum or minimum value. ([10]), including infrequent simulated events ([11]), such as maximum-minimum rainfall, maximum wind speed, maximum-minimum temperature, and so on. This extreme analysis employs two distribution models: the Generalized Extreme Value Distribution (GEV) and Generalized Pareto distribution (GPD), e.g., Panpharisa et al.,([13]) used a GEV to create a model of the highest rainfall in Thailand's upper north. Kaewmun ([1]) modeled the highest rainfall in the central and northeastern regions in addition to finding the return level of the highest rainfall in the return period. Rattanawan et al.,([16]) modeled the lower northeastern maximum rain using a GED and the highest rainfall's degree of recurrence during the year. Also Charin et al., ([3]), used a generalized extreme distribution to model monthly rainfall and a generalized Pareto distribution to model daily precipitation and calculate the minimum temperature change in winter and the summer peak temperature in the central Northeast. In addition, he used a generalized Pareto distribution to model the daily minimum temperature in winter and the daily maximum color temperature in summer.

We can estimate how often the extreme quantiles occur at a certain return level based on the extreme value theory. The return value is defined as a value that is expected to be equaled or exceeded on average once every interval of time (T) (with $T = \frac{1}{p}$ when p is a probability of extreme value distribution). However, it is difficult to calculate a 100% accuracy ($p = 1$) forecasting the return levels with $T = 1$ as the time horizon. For this reason, we apply method as a substitue of return levels with $T = 1$. The technique would be used to calculate the maximum expected rainfall over a given period. It is measured in months by a statistical confidence level known as the percentage of confidence (% of confidence level), such that we apply the Value at Risk (VaR). VaR is the most popular method of measuring financial risk in this study. It must measure the amount of rainfall the following year to predict the threat of heavy rain and flooding the following year. The greater the VaR, the more likely the event. Heavy rains may cause flooding, but it also means that rainfall is unlikely to exceed that VaR amount.

Therefore, we are interested in finding the return level and VaR as information to help plan and prepare for problems caused by changes in rainfall and weather conditions in Nakhon Ratchasima.

2. Value at risk (VaR).

Value at Risk (VaR) is mainly concerned with market risk. We define the VaR over the time horizon with probability α as ([15])

$$1 - \alpha = Pr[L > VaR] = 1 - F(VaR), \tag{2.1}$$

when L is the value of a stochastic variable with cumulative distribution function F. We will apply the VaR to determine the risk of the highest rainfall in this research in a unit of time given one year.

3. Extreme value theory

In this section, we want to find the distribution of extreme events, such as peak rainfall, that are not routine. There are two methods for determining a catastrophic event. First, the data is separated into sets. This set is called a blog. For example, our information divides into years, consisting of a data set within the year. The most important observations in each block are separate from each other. This method is called block maxima. In the second method, all data above certain limits called thresholds are extreme. The following presents more detail about both.

3.1. Block maxima (BM)

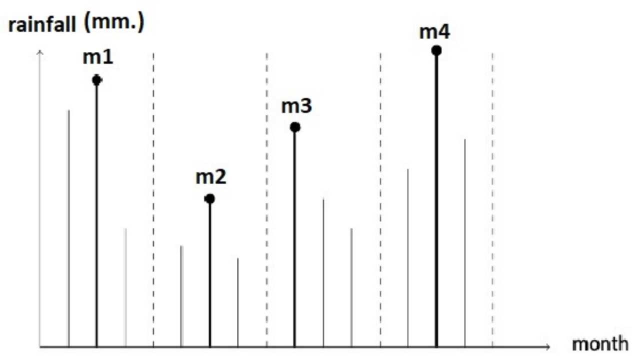

Figure 1. Example of the maximum observation in each set or block as an extreme event.

Let $X_1, X_2, ..., X_n$ be a sequence of iid random variables, where X_n has an unknown distribution function $F(x)$. Let $M_n = max\{X_1, X_2, ..., X_n\}$ such that M_n is the block maxima, see Figure 1. We need to find the distribution of M_n from now :

$$\begin{aligned} Pr(M_n \leq x) &= Pr(X_1 \leq x, X_2 \leq x, ..., X_n \leq x) \\ &= Pr(X_1 \leq x)Pr(X_2 \leq x)...Pr(X_n \leq x) \\ &= (F(x))^n. \end{aligned} \tag{3.1}$$

The Fisher-Tippett theorem states that M_n converges in distribution to the generalized extreme value (GEV) family. The GEV family is given by

$$H_{u,\sigma,\xi}(x) = \begin{cases} \exp(-(1+\xi(\frac{x-u}{\sigma}))^{-\frac{1}{\xi}}), & \xi \neq 0 \\ \exp(-\exp(-(\frac{x-u}{\sigma}))), & \xi = 0, \end{cases} \tag{3.2}$$

where u is the location parameter, σ is the scale parameter, and ξ is the shape parameter. If $\xi > 0$, then M_n follows the Fréchet distribution, when $\xi = 0$, M_n follows the Gumbel distribution, and has exponential distribution for $\xi < 0$.

Let $H(x) = p$ where p is the probability of GEV family. Thus, by inverting (3.2), we get

$$x_p = \begin{cases} u - \frac{\sigma}{\xi}(1-(-\ln p)^{-\xi}), & \xi \neq 0, \\ u - \sigma\ln(-\ln p), & \xi = 0. \end{cases} \tag{3.3}$$

We call x_p the quantile of the GEV.

The quantiles of the GEV distribution are of particular interest because of their interpretation as return levels : the value expected to be exceeded on average once every in T is a time horizon when $T = \frac{1}{p}$, where $1-p$ is the probability associated with the quantile. We seek x_p such that $H(x_p) = 1-p$, where H is as in (3.2). The associated return level x_p is

$$x_p = \begin{cases} u - \frac{\sigma}{\xi}(1 - (-ln(1-p))^{-\xi}), & \xi \neq 0, \\ u - \sigma \ln(-ln(1-p)), & \xi = 0. \end{cases} \quad (3.4)$$

Recall, $Pr(L > VaR_\alpha) = 1 - \alpha$, is equivalent to $Pr(L \leq VaR_\alpha) = \alpha$, can be written as $F(VaR_\alpha) = \alpha$. Hence, to find VaR_α as a quantile of F, it is impossible to find the quantile by simply inverting F because F is unknown. However, we now know that $Pr(VaR_\alpha) = Pr(M_n \leq x) = (F(x))^n$ and $Pr(M_n \leq x)$ can be approximated by H for large n, so that we get

$$p = H(VaR_\alpha) = Pr(M_n \leq VaR_\alpha) = (F(VaR_\alpha))^n = \alpha^n$$

Thus, if $p = \alpha^n$ we get

$$VaR_\alpha = \begin{cases} u - \frac{\sigma}{\xi}(1 - (-nln\alpha)^{-\xi}), & \xi \neq 0, \\ u - \sigma \ln(-nln\alpha), & \xi = 0, \end{cases} \quad (3.5)$$

which is the desired quantile ([15]).

3.2. Peaks-over-threshold (POT)

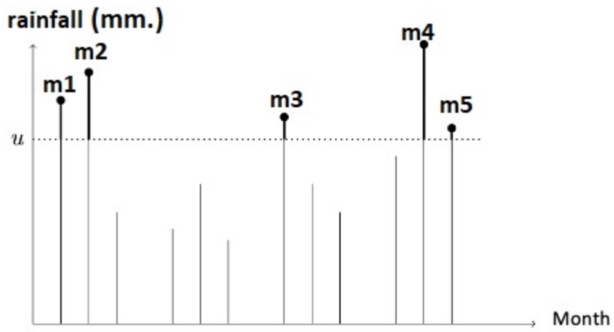

Figure 2. The observation above the threshold u as an extreme event in POT.

Let $X_1, X_2, ..., X_N$ be a sequence of iid random variables. In the POT method, we want to find $F(x)$ of all excess observations greater than threshold (u) see Figure 2. Let

$$F_u(x) = Pr(X - u \leq x)|X > u) = \frac{F(u+x) - F(u)}{1 - F(u)}. \quad (3.6)$$

However, $F_u(x)$ is unknown since $F(x)$ is unknown. We use the Pickands-Balkema-de Haan theorem: PBdH([12]) to find a (positive measurable) function σ that depends on the threshold u such that

$$\lim_{u \to x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi,\sigma}(x)| = 0,$$

where $F(x)$ is in the domain of attraction of $G_\xi, \xi \in \mathbb{R}$. So, $x_F \leq \infty$ is the right endpoint of $F(x)$ and $G(x)$ is the distribution function of General Pareto Distribution (GPD), defined as

$$G_{\xi,\sigma}(x) = \begin{cases} 1-(1+\xi x/\sigma)^{-1/\xi}, & \xi \neq 0, \\ 1-exp(-x/\sigma), & \xi = 0, \end{cases} \tag{3.7}$$

with shape parameter ξ and scale parameter σ. The $G(x)$ includes three distributions when $\xi = 0$, it is the Exponential distribution, when $\xi > 0$ it is the Ordinary Pareto distribution and, when $\xi < 0$, it is Pareto-II type distribution.

Thus, we can approximate $F_u(x)$ by $G_{\xi,\sigma}$ and (3.6) can written as

$$G_{\xi,\sigma}(x) = \frac{F(u+x)-F(u)}{1-F(u)}.$$

or

$$F(x+u) = G_{\xi,\sigma}(x)(1-F(u)) + F(u). \tag{3.8}$$

In equation (3.8), the $F(x+u)$ is stated, so we change variables from $(x+u)$ to x, such that

$$F(x) = G_{\xi,\sigma}(x-u)(1-F(u)) + F(u). \tag{3.9}$$

We still have to calculate $F(u) = Pr(X \leq u)$ Denote N as the total number of observations, and N_u, the number of observations that are greater than u. Then $F(u)$ can be approximated by

$$F(u) = Pr(X \leq u) = \frac{N-N_u}{N} = 1 - \frac{N_u}{N}. \tag{3.10}$$

Now, by (3.7) and (3.9), we have two cases to find $F(x)$. Firstly, $\xi \neq 0$, such that

$$F(x) = (1-(1-\frac{N_u}{N}))(1-(1+\xi\frac{x-u}{\sigma})^{-1/\xi}) + (1-\frac{N_u}{N}).$$

which can be simplified to

$$F(x) = 1 - \frac{N_u}{N}(1+\xi\frac{x-u}{\sigma})^{-1/\xi}. \tag{3.11}$$

Lastly, $\xi = 0$, we get

$$F(x) = 1 - \frac{N_u}{N}exp(-\frac{x-u}{\sigma}). \tag{3.12}$$

The probability of exceeding the threshold u called p is required. The value x_m that is exceeded on average once every N_u observations (return level) is

$$x_m = \begin{cases} u + \frac{\sigma}{\xi}((\frac{pN}{N_u})^{-\xi} - 1), & \xi \neq 0, \\ u + \sigma ln(\frac{pN}{N_u}), & \xi = 0. \end{cases} \tag{3.13}$$

Finding VaR, we use the quantile of (3.11), respectively for (3.12), for the confidence level α By setting $F(VaR_\alpha) = \alpha$ and solve VaR_α in (3.11) and (3.12), we get

$$VaR_\alpha = \begin{cases} u + \frac{\sigma}{\xi}((\frac{N_u}{N}(1-\alpha))^{-\xi} - 1), & \xi \neq 0, \\ u + \sigma ln(\frac{N_u}{N}(1-\alpha)), & \xi = 0. \end{cases} \tag{3.14}$$

We Choose the threshold, according to the PBdH theorem, the selection of u should be as high as possible. On the other hand, higher thresholds produce less intense observations. This results in a high variance for estimating distribution parameters. Therefore, a lower threshold is preferred. But that satisfies the finite hypothesis of the theorem quite well. Simulation studies by McNeil and Frey ([2]) concluded that the optimal level is when the observation of 10% is considered extreme. We use hill plots to help select the chosen threshold.

3.3. Backtesting

Backtesting VaR in this paper, we considered $VaR_{0.95}$ By (2.1), we have

$$Pr(L > VaR_{0.95}) = 0.05.$$

Therefore, we expect peak rainfall to exceed VaR 5% of the total time. We cannot expect the actual number of VaR violations to be precisely 5%. Therefore, the questions are how much average rainfall can deviate from the expected rainfall frequency and how much expected rainfall frequency is considered acceptable. We use the Kupiec test ([6]) to answer this question. We use the binomial distribution,

$$Pr(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k},$$

this means that the probability of success k from the n trial with a probability of success is equal to p. The cumulative distribution function is given by

$$Pr(X \leq k) = \sum_{i=1}^{k} \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}.$$

If the number of violations is less than expected, we calculate $Pr(X \leq k)$, where k is the number of violations. On the contrary, If the actual number of violations exceeds the expected number, $Pr(X \geq k)$ is calculated, while k is the absolute number of violations. Finally, these probabilities are compared with the selected significance level. The null hyperthesis was rejected if the probability is less than a significant degree. Otherwise, we used a two-tailed test following ([6]). We will find the confidence interval for the expected number of VaR violations by using the inverse of the binomial distribution ([8]).

4. Empirical methodology

4.1. Data

The monthly rainfall data set used in this study ranges from January 1982 to December 2021 (t =1,...,480 observations). The Nakhon Ratchasima Meteorological Station provided the data set. In the first step, we use data to determine return levels for two techniques. The second step divides the VaR data set into in-sample (R =360 observations) and out-of-sample (n =12 observations). Figure 3 depicts the rainfall plot.

Figure 3 shows no trend in monthly rainfall data from Nakhon Ratchasima Meteorological Station. The highest recorded maximum rainfall was 546.10 mm in September 2005. The heavy rain event in September 2017 came in second, with 451.0 mm of rain causing widespread flooding in Nakhon Ratchasima province and effectively halting the province's economy. We will use extreme value theory to analyze the rainfall, which includes two methods: the Block maxima technique with GEV and the Peak over threshold (POT) technique with GPD.

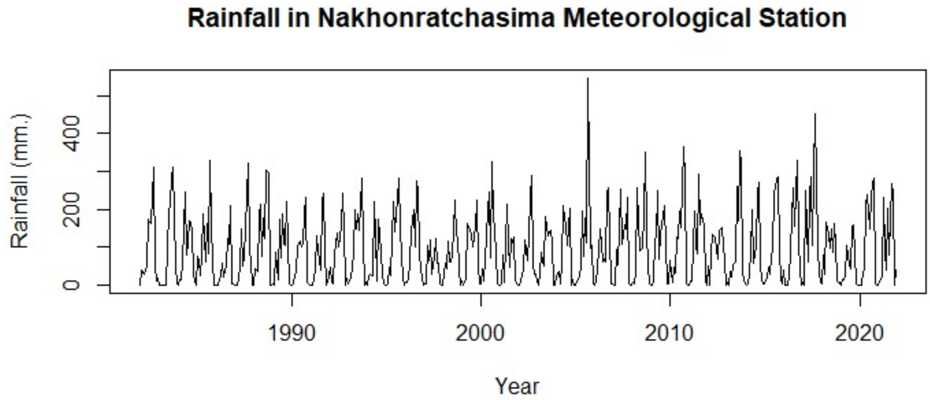

Figure 3. Monthly Rainfall at Nakhon Ratchasima Meteorological Station from 1982 to 2021.

Table 1. Generalized Extreme Value Parameter Estimates for maximum rainfall at Nakhon Ratchasima Meteorological Station.

Rainfall (mm.)	u	σ	ξ
Parameter Estimates	46.12	63.04	0.14
Parameter Standard Errors	1.94	1.77	0.02
Confidence Interval (CI) (95%)	(40.54, 52.18)	(58.03, 68.64)	(0.07, 0.21)
Anderson-Darling Statistics:	0.1671	P-Value: 0.9381	

Table 2. Generalized Extreme Value Return Level for maximum rainfall at Nakhon Ratchasima Meteorological Station.

Return Periods	2 Years	5 Years	10 Years	20 Years	100 Years
Return Levels	69.81	151.04	212.15	276.95	449.44
CI (95%)	(62.20 ,77.06)	(138.29 ,163.89)	(192.97, 232.41)	(247.72, 309.81)	(385.81, 530.64)

4.2. Generalized Extreme Value Modelling.

This section shows the results and discussion for modeling extreme rainfall in Nakhon Ratchasima province using GEV.

Table 1 shows the results of GEV modeling on extreme rainfall data in Nakhon Ratchasima province using the Block Maxima approach. The GEV parameters were estimated using the L-moments method, with a shape parameter greater than zero implying that the Fréchet distribution is the best model for the GEV family, as confirmed by Anderson-Darling Statistics of 0.1671 and P-Value of 0.9381.

Figure 4 depicts a GEV diagnostic diagram. The diagonal is close to the QQ plot point. This implies that the GEV distribution function is suitable. The density plot represents the relationship between the fitted GEV distribution function and the observed density. The return level plot in the final panel shows that the empirical return level is well-matched to the level from the fit distribution function.

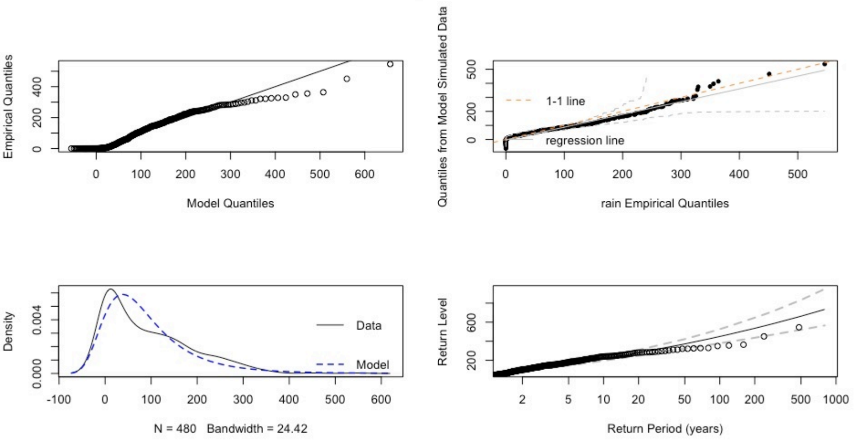

Figure 4. GEV Diagnostic Plots for maximum rainfall at Nakhon Ratchasima Meteorological Station.

Table 3. Generalized Pareto Distribution Parameter Estimates for maximum rainfall at Nakhon Ratchasima Meteorological Station.

Rainfall (mm.)	σ	ξ
Parameter Estimates	48.06	0.07
Parameter Standard Errors	7.34	0.11
Confidence Interval (CI) (95%)	(29.44, 73.48)	(-0.30, 0.35)
Anderson-Darling Statistics:	0.3467	P-Value: 0.4811

Table 2 shows the predicted maximum rainfall return level (in mm) and 95% confidence intervals for return periods of 2, 5, 10, 20, and 100 (in years). Table 2 shows that increasing the number of return periods leads to an increase in the number of return levels.

4.3. Generalized Pareto Distribution Modelling.

The threshold was determined using the hill plot in Figure 5. The threshold is chosen from the field containing the order statistics' relatively stable tail index. The tail index is long-lasting, we select the last area k=10% of 480 = 48. There are a total of 48 extreme data points with a threshold of 235 mm. where the Hill estimator is more stable. The mean excess function allows the establishment of the behavior of the distribution tail looking at the linear shape.

Table 3 shows the parameter estimates of the Generalized Pareto Distribution using L-moments estimation because there are 48 extreme data points. The confidence interval ξ has a value of zero. As a result, the exponential distribution very well fits the data, as confirmed by Anderson-Darling statistics of 0.3467 and a P-Value of 0.4811.

Figure 6 shows the model diagnostic plot for GPD for rainfall data in Nakhon Ratchasima province. This means that the GPD is an excellent match for the block maxima.

Table 4 shows the predicted maximum rainfall return level (in mm) and 95% confidence intervals for return periods of 10, 20, 50, and 100 (in years). The results show that increasing the return periods results in an increase in the return levels.

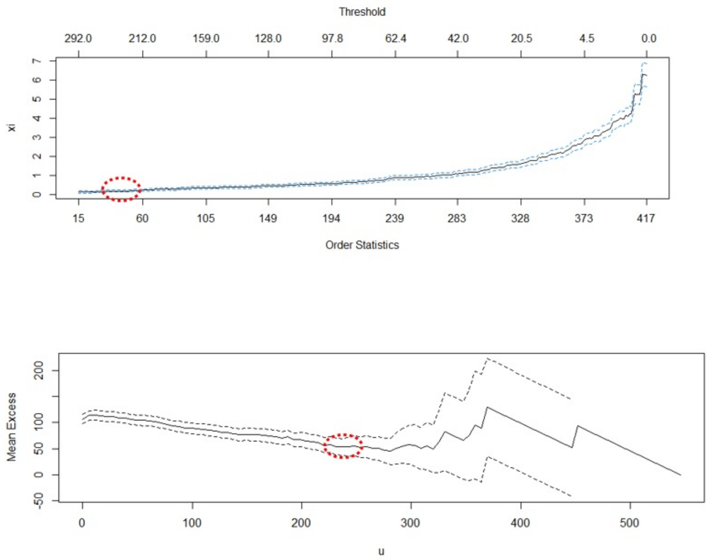

FIGURE 5. Hill Plot and Mean excess plot for maximum rainfall at Nakhon Ratchasima Meteorological Station.

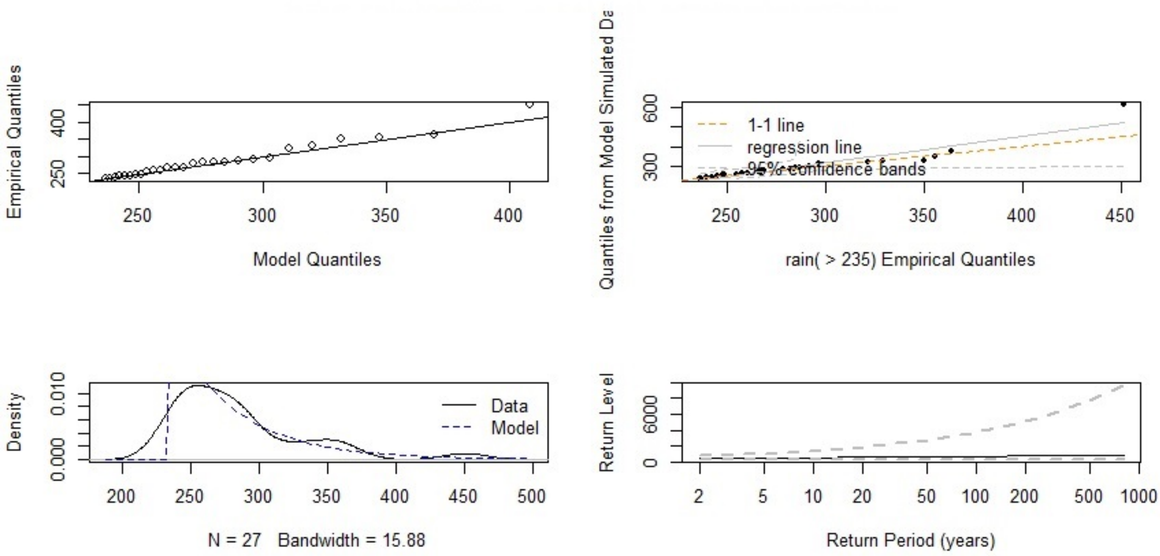

FIGURE 6. GPD Diagnostic Plots for maximum rainfall at Nakhon Ratchasima Meteorological Station.

When we compare the return levels for the GEV and GPD, we see that the GPD provides higher return levels and that the return levels provided are closer to reality for all return periods than the GEV.

Table 4. Generalized Pareto Distribution Return Level for maximum rainfall at Nakhon Ratchasima Meteorological Station.

Return Periods	2 Years	5 Years	10 Years	20 Years	100 Years
Return Levels	477.96	540.03	589.79	642.09	774.08
CI (95%)	(384.48, 621.53)	(401.17, 783.42)	(411.79, 943.16)	(420.46, 1130.32)	(436.10, 1816.68)

Table 5. In-sample and out-sample periods for the data.

in-sample	out-sample
1982-2011	2012
1983-2012	2013
1984-2013	2014
1985-2014	2015
1986-2015	2016
1987-2016	2017
1988-2017	2018
1989-2018	2019
1990-2019	2020
1991-2020	2021

4.4. VaR

The rainfall data are separated into in-sample and out-sample according to table 5, where the observations in the in-sample are 360 observations, and the out-sample is 12 observations.

The GEV and GPD distribution parameters calculate for each in-sample period. Finding the parameters (u, σ, ξ) of the GEV distribution, extracting the most significant rainfall each year, and then applying the L-moments to out-sample data.

For the GPD, the threshold has to decide at 235 mm. The observations are extreme, closer to the proposed 10% ([2]). Thus, the threshold was 235 mm. The L-moments were applied to rainfall greater than 235 mm.

The $VaR_{0.95}$ is calculated by (3.5) and (3.14) for out-sample period using the parameters of the corresponding in-sample period. The number of violations of VaR, i.e., how much rainfall is greater than VaR. In the last column, apply the Kupiec test to the number of violations for $\alpha = 0.95$, at the 95% level of confidence.

We have 30 monthly block maxima with n=12 observations in each block. The parameters (u, σ, ξ) for the GEV shown in Table 5. Estimates, VaR calculates by (3.5). Note that the estimates from 1982-2011 calculate the VaR of the year 2012, the VaR of 2013 from the estimate of 1983-2012, etc. The estimates of VaR are presented in Table 6. We have 360 observations in the in-sample periods. We expect $(0.05)(360) = 18$ violations of VaR under this period in-sample 1986-2015, 1987-2016, 1988-2017 and 1990-2019. Out-sample, we have 12 observations from 2012-2021. We expect $(0.05)(12) = 0.6$ violations of VaR under this period out-sample 2013-2017 and 2020.

We proceed with the results of the POT model. The frequency of extreme observations for each in and out-sample period is given in Table 5. We can note that the frequency

Table 6. L-moments of the parameters in the GEV and VaR for the years in-sample under the GEV distribution.

in-sample	u	σ	ξ	VaR_{95}	n violations in-sample	n violations out-sample
1982-2011	45.62	62.13	0.12	267.11	17	0
1983-2012	46.45	62.40	0.11	264.48	16	2
1984-2013	46.52	62.12	0.11	266.05	16	1
1985-2014	46.24	61.99	0.12	267.37	17	2
1986-2015	45.46	61.77	0.13	268.67	18	1
1987-2016	46.76	63.04	0.12	272.39	18	3
1988-2017	47.78	64.50	0.13	280.20	18	0
1989-2018	47.55	63.60	0.12	275.40	17	0
1990-2019	46.66	62.63	0.13	274.06	18	1
1991-2020	47.38	63.91	0.12	278.80	17	0

Table 7. L-moments of the parameters in the GPD distribution in-sample and frequency of threshold exceedances for each in-sample and out-sample period.

in-sample	σ	ξ	VaR_{95}	n of exceedances in-sample	n of exceedances out-sample
1982-2011	54.13	0.04	258.92	17	0
1983-2012	50.68	0.09	255.55	16	3
1984-2013	52.06	0.09	258.00	17	1
1985-2014	52.37	0.07	259.98	18	2
1986-2015	48.58	0.09	261.41	18	1
1987-2016	50.73	0.05	265.75	19	3
1988-2017	47.18	0.14	268.99	21	0
1989-2018	43.25	0.20	263.76	19	0
1990-2019	43.25	0.20	263.76	19	2
1991-2020	39.22	0.23	265.33	21	1

is more significant for the in-sample periods than for the out-sample periods. Further, the L-moments of the parameters of the GPD and estimates of VaR is then calculated by the same in-sample and out-sample periods as in the previous model. The results are presented in Table 7.

In total, we have 360 observations in the in-sample periods. Therefore we expect 18 violations of VaR under this period in the sample between 1985-2014 to 1991-2020. Out-sample, we have 12 observations in the period 2012-2021. Therefore we expect 0.6 violations of VaR under this period out-sample 2013-2017 and 2020-2021. Both models prove rather reasonable overall rainfall in the backtest, in which in-sample is By the way, in out-of-sample, GPD outperformed GEV.

5. Conclusion

This study demonstrates the use and significance of EVT in describing maximum rainfall events in Nakhon Ratchasima. For the highest monthly rainfall data in Nakhon Ratchasima from January 1982 to December 2021, the GEV and GPD models were used. L-moments estimation was used to estimate model parameters. The block maxima method corresponds to the maximum typical value distribution, whereas the peak over threshold method corresponds to the joint Pareto distribution.

According to our findings, the Fréchet distribution is the best model from the GEV family for monthly peak rainfall data. We discovered that, while the exponential distribution provided the best model for monthly rainfall data above 235 mm, increasing the return period increased the level of yield. When we compare GEV and GPD return levels, we find that GPD return levels are higher yearly than GEV. Model diagnostics demonstrate that the model is appropriate for modeling rainfall data.

The study will assist decision-makers in Nakhon Ratchasima in being aware of heavy rainfall events during the review period to help reduce crop infrastructure damage and make appropriate decisions to minimize crop infrastructure damage. This research will be useful in flood mitigation, early warning, management, preparedness, and response. Nakhon Ratchasima, on the other hand, is taking the right steps to create an environment conducive to climate change response. Future research could, however, simulate and predict severe rainfall in Nakhon Ratchasima across different regions. It is also a viable research strategy.

The second goal was to investigate the BM and POT of the most accurate VaR model for VaR estimation and backtesting. When VaR was backtested, we discovered that the POT method outperformed the BM method. BM performed better than POT in other studies, including K. Anderson ([7]), Cerovic and Karadzic ([5]), and Marineln et al. ([4]). Remember Embrechts et al.'s for choosing POT over BM ([12]).

Many other studies have concluded that POT is superior to BM; our findings are consistent with this study. This report, however, is based on a larger number of observations in Cerovic and Karadzic ([5]) and Marineln et al. ([4]). The benefits of POT over BM, as demonstrated by Embrechts et al. ([12]), clearly demonstrate that the POT does not discard useful information such as extremes as the number of observations increases.

Most studies sample extremes using the approach of block maxima, i.e., maximum daily, maximum monthly, and so on. The disadvantage of this method is that specific higher values from a block responsible for extreme events may be rejected. To overcome the disadvantage of block maxima, a Peak over Threshold approach can be used to sample extremes. Moreover, the discussion for a larger number of observations indicates that the method should work just as well as it did for Bekiros and Georgoutsos ([17]).

Furthermore, the frequency of the excess peak rainfall was higher during the sample period compared to the relevant off-sample period. Because the POT isolates severe losses more effectively than the BM (see [12]), this may indicate that risk is overestimated during the non-sampled period. This may explain why the POT underestimates risk while overestimating the BM. We can avoid this type of overestimation by using larger data sets. Combining the two models is another option. Divide the data into blocks and use the POT with the most observations extracted from each block.

As a result, when we use VaR to assess risk, we overestimate risk. Because of the optional POT model, VaR is superior for evaluating both models.

Acknowledgements

We would like to thank Nakhon Ratchasima Rajabhat University for funding this research, as well as the Northeastern Meteorological Department for providing data for analysis.

References

[1] A. Keawmun, P. Budchamruei, P. Busababodhin, Modeling Monthly Precipitation in Central Northeast of Thailand, KKU Science Journal,43(1) (2015), 137-147.

[2] A. McNeil, R. Frey, Estimation of Tail-Related Risk Measures for Heteroskedastic Financial Time Series: An Extreme Value Approach. Journal of Empirical Finance, 7, (2000), 271-300.

[3] B. Charin, W. Srisodaphol, P. Busababodhin, Modeling for the extreme temperature of the central northeast region of Thailand, Proceeding of International Conference on Applied Statistics 2014 (ICAS 2014), May 21-24, 2014, Pullman Khon Kaen Raja Orchid, Khon Kaen, Thailand, (2014), 67-72.

[4] C. Marinelli, S. D'addona, S.T.Rachev, A comparison of some univariate models for value-at-risk and expected shortfall', International Journal of Theoretical and Applied Finance 10(6) (2007), 1043–1075.

[5] J. Cerovic, V. Karadzic, Extreme value theory in emerging markets: evidence from the Montenegrin stock exchange, Economic Annals 60(206) (2005), 87– 116.

[6] J.C. Hull, Risk Management and Financial Institutions, fifth edition, Wiley. 2018.

[7] K. Anderson, A comparative study of VaR and ES using extreme value theory. Department of Economics, Lund University, 2020.

[8] K. Dowd, Measuring market risk, second edition, Wiley, 2005.

[9] N. Sopipan, Statistical model for forecasting rainfall of Nakhon Ratchasima Province. Naresuan University Journal, Year 23, Issue 2 (2015), 56-65.

[10] P. Busababodin, M. Siriboon, A. Kaewmun, Modeling of Extreme Precipitation in Upper Northeast of Thailand. Burapha Science Journal, 20(1) (2015), 106-117.

[11] P. Embrecht, C. Klüppelberg, T. Mikosch, Modelling extremal Events for insurance and finance.Springer-Verlag,Berlin, 1997.

[12] P. Embrechts, R. Frey, A.J. McNeil, Quantitative Risk Management Concepts, techniques and tools, Princeton University Press Princeton and Oxford, 2005.

[13] P. Khongthip, M. Khamkong, P. Bookamana, Modeling Annual Extreme Precipitation in upper Northern Region of Thailand. Burapha Science Journal, 18(1) (2013), 95-104.

[14] P. Kosolkul, Forecasting Monthly Rainfall of Nong Khai Province by Statistical Forecasting Techniques. University of the Thai Chamber of Commerce Journal,27(2) (2007), 138-150.

[15] R.S. Tsay, Analysis of Financial Time Series. John Wiley & Sons, 2005.

[16] S. Rattanawan, P. Busababodhin, B. Kumphon, Modeling for Extreme rainfall in lower Northeastern Thailand, Proceeding for The Asian Conference on Sustainability, Energy & the Environment (acsee2014), Osaka, Japan, 12-15June 2014: 162-171.

[17] S.D. Bekiros, D.A. Georgoutsos, Estimation of value at risk by extreme value and conventional methods: a comparative evaluation of their predictive performance, Journal of International Financial Markets, Institutions and Money 15(3) (2005), 209–228.