International Journal of Scientific and Innovative Mathematical Research (IJSIMR)
Volume 2, Issue 1, January - 2014, PP 23-36
ISSN 2347-307X (Print) & ISSN 2347-3142 (Online)
www.arcjournals.org

Extreme Value Modeling of Precipitation in Case Studies for China

Manuela Ender
Department of Mathematical Sciences
Xi'an Jiaotong Liverpool University
Suzhou,China
manuela.ender@xjtlu.edu.cn

Tong Ma
Department of Mathematical Sciences
Xi'an Jiaotong Liverpool University
Suzhou,China

Abstract: *This paper aims to model extreme rainfall events using 60 years of daily data based on extreme value theory for four cities in China. The purpose is to allow decision makers to make informed decisions and to avoid or at least reduce flood caused damage to life and property. Generalized extreme value distributions are used for fitting monthly and semiannual maxima according to the Block Maxima Approach. Rainfall exceeding an extremely high threshold is modelled by Generalized Pareto distributions .The thresholds are selected based on the analysis of three methods, Hill plot, mean excess plot and Standardized Precipitation Index. Finally, we estimate the parameters for both models and calculate return levels for five different return periods. Statistical tests for stationarity, KPSS and Man-Kendall tests support the study. The results show that GPD has a better fitting performance than GEV. Further, we can determine how often a flood occur in a certain city and during which season, rainy or dry season. For example, for Nanjing it can be neglected that a flood occurs during dry seasons. But for other cities, like Shantou, this might be rare events but with a higher frequency of occurrence.*

Keywords: *Extreme value theory, precipitation, generalized extreme value distribution, generalized Pareto distribution, return level, China*

1. Introduction

During the year 2013, China experienced severe floods caused by extreme precipitation. In August 2013 for example, the south provinces Guangdong, Guangxi and Fujian and the north eastern provinces Heilongjiang, Jilin and Liaoning were hit by heavy rain that caused flooding (RCSC, 2013). For the north eastern region, it was the worst flooding in 50 years (Hunt and Ke, 2013). Millions of residents were affected, many people died or went missing. Direct economic losses solely for Guangdong province are reported to be CNY 13 billion and for Heilongjiang province to be around CNY 7 billion where potential losses in future crop yields are not yet included (RCSC, 2013).

Beside these recent events, China is regularly affected by extreme rainfall that has enormous influence on social development and that threatens safety of human life and property (Feng et al., 2007). In 2008 for example, the Spring Festival was spoiled by heavy snowstorms in many provinces. The total costs were approximately CNY 80 billion which was mainly used for refugee settlements and maintenance of infrastructure (BBC, 2008). In 2011, a heavy rain brought much inconvenience for cars and pedestrians that soaked in Nanjing, Jiangsu province (CCTV, 2011). Most floods or mud-rock slides triggered by severe rainfall impair agricultural production, public facilities, buildings and transport (Yang, 2012).

Consequently, estimations of extreme rainfall events in China play a significant role in an efficient risk appraisal and for the reduction of losses of the economy. To know the statistics that a certain flood event will occur gives advice to decision makers to design efficient mitigating measures. This covers firstly the planning and construction of water management, sewerage systems, capacity of channels and river basins, etc. Second, it supports the decision how much and what kind of insurance against water damage should be bought. Finally, the knowledge of return levels of floods helps to inform the citizens that they are better prepared in a case of flooding (Overeem et al., 2008).

To contribute to the achievement of these goals, this paper focuses on the estimation of the most appropriate distribution of extreme precipitation in four Chinese cities, Nanjing, Shantou, Urumqi and Qiqihaer. The estimation is based on a universal principle, Extreme Value Theory (EVT). EVT contains two fundamental distributions, Generalized Extreme Value distribution (GEV) and Generalized Pareto distribution (GPD), which are both applied in this paper. Details for the GEV and GPD approach are given in Section 3. We elaborate for both models the return levels of certain flood events for 5, 10, 20, 50, and 100 years. As many regions in China have marked dry and wet periods during the year, we distinguish further between dry and wet seasons as risk mitigating measures may be different during each time of the year. The probability of occurrence of extreme rainfall might be underestimated by assuming that the model is stationary (Coles et al., 2003). Therefore, stationary tests are applied to determine whether non-stationarity influences the application of GEV and GPD models. To sum up, this paper contributes to the existing literature by a comprehensive statistical analysis of extreme precipitation events in four cities of China in the north, east, south and west. Based on most recent flood events in 2013, Shantou and Qiqihaer are included in particular to study them from a statistical perspective.

The next section of the paper gives a literature review of previous and similar work. In Section 3, the methodology of EVT, GEV and GPD is elaborated. Section 4 presents and discusses empirical results. Recommendations and limitations in Section 5 conclude the paper.

2. Literature Review

First insights in EVT are published by Fisher and Tippett in 1928. Significant contributions to the statistical modeling of extremes followed by Jenkinson 1955 for GEV, Balkema and de Haan 1974 and Pickands 1975 for GPD. In comparison to the long history of theoretical results of EVT, the empirical analysis of precipitation data using EVT is relatively new. In the last two decades, researchers all over the world applied either the GEV or the GPD approach to rainfall data from Europe (Miroslava, 1992; Bordi et al., 2007), America (Nadarajah, 2005), Oceania (Withers and Nadarajah, 2000; Li et al., 2005) or Asia (Nadarajah and Choi, 2007; McAleer et al, 2012).

For China in particular, several papers exist dealing with the analysis of extreme precipitation and its trends (Gemmer et al., 2004; Wang and Zhou, 2005; Liu et al., 2005; Zhai et al., 2005; Sun and Ao, 2013). For example, Wang and Zhou (2005) stress that an obvious increasing trend of extreme daily rainfall mainly took place in the east, southwest and northwest of China in summer months from 1961 to 2001. However, fewer studies apply EVT techniques for the estimation of the tail distribution and derivation of return levels. Feng et al. (2007) provide the first comprehensive study of GEV estimation and return levels based on Chinese data. The authors conclude that the highest return levels were found in the very Southern parts of China. In Eastern China, high return levels are reported in comparison to low return levels in the northwest. Studies using the GPD approach are conducted by Li (2013) for data from whole China or more locally based by Dong et al. (2011) for Yellow-Huaihe and Yangtze-Huaihe rivers basins and Jiang et al. (2009) for Eastern China. Jiang et al. (2009) state that the fitting performance of GPD is better than for GEV. Their reported return levels are in line with the findings of Feng et al. (2007) and decrease from the south to the north.

3. Methodology

3.1 Data Description

To cover locations in the north, east, south and west and in different climate zones of China, the four cities Qiqihaer (Heilongjiang province), Nanjing (Jiangsu province), Shantou (Guangdong province) and Urumchi (Xinjiang province) are chosen in the empirical study. Qiqihaer and Shantou are of particular interest given most recent flood events in 2013. All data starts from year 1951 to 2010 and are obtained from the National Meteorological Information Center. The daily precipitation amount is measured within 24 hours (8am on a particular day to 8am on next day). The climate of Qiqihaer is humid continental and monsoon-influenced. Nanjing is located in the subtropical monsoon climate zone. Shantou lies within the subtropical marine and monsoon climate zone. Urumchi belongs to the temperate zone with continental climate (Domroes and Peng, 1988).

The descriptive statistics of the daily rainfall datasets for the four cities are given in Table 1. According to their climate zones, Shantou is the city with most rain, followed by Nanjing and Qiqihaer. Urumchi is the driest city.

Table 1. *Descriptive statistics of daily rainfall datasets, 1951 to 2010 (unit: mm, n=21915)*

	Nanjing	Shantou	Urumchi	Qiqihaer
Minimum	0	0	0	0
Maximum	207.20	297.40	57.70	135.50
Mean	2.90	4.34	0.73	1.18
Median	0	0	0	0
Mode	0	0	0	0
Standard	9.73	14.80	2.76	5.02
Range	207.20	297.40	57.70	135.50

3.2 Generalized Extreme Value Distribution (GEV) and Block Maxima

The most classical model for extreme events is the Block Maxima approach. This model is appropriate when the maximum observations of each period or block with a predefined and fixed length are assembled from a large number of identically and independently distributed (iid) variables (McNeil, 1999). In this case, the asymptotic distribution of the maximum observations is exactly one of three well known distributions (Fisher and Tippett, 1928). The cumulative distribution function of these three distributions can be summarized by the GEV (Jenkinson, 1955) and is given by

$$H(x,\xi,\sigma,\mu)=\begin{cases} e^{-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}}, \xi\neq 0, \\ e^{-e^{-\frac{x-\mu}{\sigma}}}, \quad \xi=0, \end{cases} \tag{1}$$

where x are the extreme values from the blocks, and ξ, σ, μ are the shape, scale and location parameters, respectively. For $\xi = 0$, the Gumbel distribution is determined. For $\xi > 0$, we get the Frechet distribution with a fat-tail. In the case of $\xi < 0$, the Weibull distribution is obtained. The parameters ξ, σ, μ are estimated by Maximum likelihood estimation (MLE). When we assume that the variables are independent, the likelihood function is given by the product of the observations' densities.

3.3 Generalized Pareto Distribution (GPD) and Peak over Threshold (POT)

A second methods to analyze the distribution of extreme events is called Peak over Threshold (POT) method which considers the maximum variables exceeding a predetermined threshold. Given a threshold u, the distribution function of extreme values of X over u is,

$$F_u(y)=\Pr(X-u\leq y|X>u)=\frac{F(y+u)-F(u)}{1-F(u)}. \tag{2}$$

$F_u(y)$ represents the probability that the value of X exceeds the threshold u by at most amount y, where $y = X - u$. Balkema and de Haan (1974) and Pickands (1975) showed that the distribution $F_u(y)$ converges to GPD when the threshold is sufficiently high. The cumulative distribution function for GPD is,

$$H(x,\xi,\sigma,\mu)=\begin{cases} 1-\left[1+\xi\left(\frac{x-u}{\sigma}\right)\right]^{-\frac{1}{\xi}}, & \xi\neq 0 \\ 1-e^{-(x-u)/\sigma}, & \xi=0, \end{cases} \tag{3}$$

where x are the exceedances, ξ and σ are the shape and scale parameter, respectively. There are three types of GPD. For $\xi = 0$, we have an Exponential distribution with medium-size right tail. If $\xi < 0$, H(x) is an Ordinary Pareto distribution describing a positive long tail. In the case of $\xi > 0$, H(x) is a Pareto (II) type distribution for a positive short tail.

The estimation of parameters is possible with the method of probability weighted moments, the so called L-moments, or with MLE (Hosking et al., 1984). In this study, MLE is used. Before the parameter estimation, the initial step is to choose an appropriate threshold. We are faced with the problem to find a threshold that is sufficiently high to support the convergence. However, it should not be too high as otherwise the sample of these extreme high observations is very small. In other words, there is a tradeoff between bias and variance. We use three methods of threshold selection including Hill plot (Hill, 1975), sample mean excess plots, and Standardized Precipitation Index (SPI) (McKee et al., 1993).

The Hill plot specifies the relationship between the estimated tail index ξ_{hill} and either n or the threshold, where n is the number of exceedances and m is the sample size:

$$\xi_{hill} = \frac{1}{n-1}\sum_{i=1}^{n-1} \log X_{i,m} - \log X_{n,m}. \qquad (4)$$

A threshold is chosen from the Hill plot where the tail index starts to be stable (Hill, 1975).

The mean excess function is the expectation of each observation deducted by a fixed amount given that this observation is not smaller than that fixed amount:

$$e(u) = E(X - u | X > u) = \frac{\int_u^{\infty} (1 - F(u))\, du}{1 - F(u)} - x\ . \qquad (5)$$

Empirically, $e(u)$ is estimated by e_n based on the realizations of random variables, $x_1, x_2, \cdots, x_n$.

$$e_n(x) = \frac{\sum_{x_i > x} x_i}{N(i: x_i > x)} - x\ . \qquad (6)$$

The SPI is used to quantify extreme wetness and dryness per month in a location (McKee et al., 1993). McKee et al. (1993) state that the auto-correlation of individual observations is unobvious as the monthly precipitation is usually independent. Hence, the frequency distribution of monthly precipitation is estimated by a two-parameter Gamma distribution for every month of a year. As a next step, the empirical Gamma probability density distribution is transferred into the Normal distribution. Abnormality in the transformation quantifies the meaning of relatively wet and dry. Therefore, the SPI values can be converted into Z-indexes of the Standard Normal Distribution (Bordi et al., 2007). Based on Yuan and Zhou (2004), Table 2 lists the classification of scales for SPI and Z-index in China. For example, the extremely wet conditions can be identified by Z-values which are greater than 1.96.

To summarize, Hill-plot and sample mean excesses belong to the family of semi-parametric models, while SPI fits to non-parametric model. This paper applies all three methods to select thresholds.

Table. 2 *Classification of Scales for SPI and Z-Index*

Z-index	**SPI values**	**Class**
>1.96	>2.00	Extremely wet
1.44 to1.96	1.50 to 2.00	Moderately wet
0.84 to1.44	1.00 to 1.50	Slightly wet
-0.84 to 0.84	-1.00 to 1.00	Normal
-1.44 to -0.84	-1.50 to -1.00	Slightly dry
-1.96 to -1.44	-2.00 to -1.50	Moderately dry
<-1.96	<-2.00	Extremely dry

3.4 Extreme Quantile Estimation for GEV and GPD

A T-year return level has a probability of $100 \cdot \frac{1}{T}$ per cent to be exceeded once in a year. For example, rainfall of 170 mm or larger equally occurs every 60 years in a region. The return levels at different return periods can be evaluated through the quantile estimation of fitted GEV and GPD, that is,

$$U_T = F^{-1}\left(1 - {}^{1}/_{\lambda T}\right). \tag{7}$$

F^{-1} is the inverse of the GEV or GPD. For GPD, it is assumed that the number of exceedances n_u over the extremely large threshold u is approximately close to a Poisson distribution with parameter λ, which is also the rate of exceedances per year. Hence, λT is the number of exceedances in the return period of T years. The λ can be estimated by n_u/T_{data}, where T_{data} is the number of years with available data (Maraun, 2010).

3.5 Stationary Tests

Stationary tests including graphic examination, KPSS (Kwiatkowski, Philips, Schmidt and Skin) and non-parametric Mann-Kendall tests should be carried out since the assumption of stationarity is crucial for the application of GEV and GPD. The KPSS stationary test (Kwiatkowski, Phillips, Schmidt and Shin) judges whether the trend is stabilized around a constant, a linear line or non-stationary (Hasna and Chung, 2010). The test statistics are compared with critical values at different significant levels. Thus, the null (H_0) and alternative (H_1) hypotheses are:

H_0: Stationary around a constant or a linear trend,

H_1: The trend is non-stationary.

The Mann-Kendall (MK) test determines the existence of either an increasing or decreasing tendency in monthly and semi-annual extreme rainfall (Hasna and Chung, 2010). The p-value of the null hypothesis is used to determine the tendency of the observations. Thus, the null (H_0) and alternative (H_1) hypotheses are:

H_0: There is no trend,

H_1: There is an increasing/ decreasing trend.

4. Results and Discussion

4.1 Modeling using GEV

For GEV estimation, the Block Maxima of monthly and half-yearly rainfall are extracted for all four cities. Table 3 contains descriptive statistics for both periods.

Table. 3 *Descriptive statistics of monthly (n=720) and half-yearly (n=120) maximum, 1951 to 2010 (unit: mm)*

	Nanjing		Shantou		Urumchi		Qiqihaer	
	Monthly	Half-yearly	Monthly	Half-yearly	Monthly	Half-yearly	Monthly	Half-yearly
Minimum	0	24.30	0	24.60	0	6.60	0	8.80
Maximum	207.20	207.20	297.40	297.40	57.70	57.70	135.50	135.50
Mean	32.07	74.81	45.50	116.21	9.72	22.23	13.42	39.89
Median	24.00	65.20	30.60	103.25	7.30	20.95	5.50	33.65
Mode	12.80	40.20	0	75.00	4.00	18.60	0	30.30
Standard	28.91	37.15	46.33	53.28	8.37	10.14	17.69	21.77
Range	207.20	182.90	297.40	272.80	57.70	51.10	135.50	126.70

Table 4 gives the parameters of GEV as results of MLE fitted to monthly and half-yearly maxima. The first line of each parameter indicates the value and the standard errors (s.e.). The corresponding 95% confidence intervals (CI) are included in the second line. For monthly maxima, the shape parameter ξ is for all four cities positive and the CI do not include zero which supports the positive sign of ξ. This means that the fat-tailed Frechet distributions are obtained. For half-yearly maxima, zero is included in the CI. Therefore, a Gumbel distribution cannot be excluded. These findings are in line with Feng et al. (2007).

Table. 4 *GEV parameter estimates for monthly and half-yearly maxima*

	Nanjing		Shantou		Urumchi		Qiqihaer	
Parameters (95% CI)	Monthly / s.e.	Half-yearly / s.e.	Monthly / s.e.	Half-yearly / s.e.	Monthly / s.e.	Half-yearly / s.e.	Monthly / s.e.	Half-yearly / s.e.
ξ	0.319 / 0.036	0.171 / 0.089	0.497 / 0.054	0.035 / 0.073	0.285 / 0.038	0.074 / 0.074	1.245 / 0.067	0.127 / 0.083
(95% CI)	(0.250, 0.389)	(-0.005, 0.346)	(0.391, 0.603)	(-0.108, 0.178)	(0.211, 0.358)	(-0.070, 0.219)	(1.113, 1.377)	(-0.035, 0.288)
σ	14.478 / 0.534	24.510 / 2.082	21.647 / 0.973	40.135 / 3.081	4.550 / 0.167	7.248 / 0.565	3.906 / 0.263	14.693 / 1.203
(95% CI)	(13.469, 15.563)	(20.750,28.950)	(19.822, 23.641)	(34.529, 46.651)	(4.234, 4.889)	(6.221, 8.444)	(3.424, 4.457)	(12.515, 17.250)
μ	17.880 / 0.625	56.120 / 2.622	19.776 / 1.032	91.504 / 4.163	5.543 / 0.199	17.476 / 0.753	2.606 / 0.180	29.395 / 1.553
(95% CI)	(16.655, 19.104)	(50.981,61.259)	(17.753, 21.798)	(83.344, 99.664)	(5.154, 5.932)	(16.000, 18.952)	(2.254, 2.958)	(26.351, 32.439)

By checking density plots, cumulative distribution plots, probability plots and QQ plots respectively, the fitting performance of GEV can be analyzed. The fitted lines in the density plots have clear tails for all cities. The cumulative plots are well fitted to the empirical data for both monthly and half-yearly maxima. In terms of probability plots, the GEV fitted lines estimate the extreme values much more precise than the Normal distribution. For monthly and half-yearly observations, the probability fitting curves and QQ plots are congruent in all cases. Only a few highest points diverge from the fitted lines.

Table 5 lists the return level estimates at different return periods for monthly and half-yearly maxima. The 95% CI are included in brackets. The estimated return levels and CI increase with the increase of the return period. Compared with the lower bound of the confidence interval, the upper bound is likely to be further away from the predicted return level when the return period is longer. Recalling Table 3, the highest rainfall amount in Nanjing of the observed period from 1951 to 2010 was 207.2 mm. This value appears in CI at T=10 for monthly and at T=20 for half-yearly samples. This pattern is similar for Urumchi. For Shantou, the highest observation is also covered by T=20 for half-yearly samples. For monthly samples it is already included in T=5. The maximum value of Qiqihaer lies within the CI of T=50 for semiannual values. Return levels based on monthly maxima are not possible to compute.

According to China Meteorological Administration (2012), 24-hour rainfall amount exceeding 250 mm is considered as extreme rainfall that can cause floods in Southern China. For Nanjing, this value is contained in the CI of T=20 for maximum monthly observations and T=50 for semiannual maxima. In Shantou, the event of a floods occurs much more often as 250mm is included in the CI of T=5 for monthly and T=10 for half-yearly observations. Hence, floods might occur once in every 20 to 50 years in Nanjing and every 5 to 10 years in Shantou.

Table 5. *GEV return level estimates for monthly and half-yearly maxima*

Selection Period T (years)	5	10	20	50	100
Monthly Max Nanjing	139.438 (120.161, 166.147)	180.973 (151.041, 223.980)	232.660 (187.681, 299.699)	320.981 (247.536, 437.008)	407.135 (301.523, 578.915)

Half-yearly Max Nanjing	123.361 (109.749, 145.588)	150.920 (129.830, 190.498)	181.447 (149.689, 247.645)	227.373 (175.645, 348.148)	267.060 (195.140, 449.122)
Monthly Max Shantou	307.790 (242.743, 409.104)	445.022 (331.759, 631.786)	638.3502 (448.555, 970.388)	1020.600 (676.900, 1703.000)	Not computable
Half-yearly Max Shantou	185.452 (168.566, 210.172)	217.089 (194.011, 256.593)	248.900 (217.348, 309.097)	291.723 (245.484, 389.603)	324.8883 (264.927, 460.329)
Monthly Max Urumchi	40.713 (35.227, 48.393)	51.949 (43.543, 64.166)	65.608 (53.158, 84.387)	88.314 (68.240, 120.160)	109.878 (81.774, 156.250)
Half-yearly Max Urumchi	35.2310 (31.919, 40.163)	41.567 (36.916, 49.642)	48.121 (41.642, 60.700)	57.234 (47.523, 78.277)	64.521 (51.727, 94.274)
Half-yearly Max Qiqihaer	67.655 (60.116, 79.383)	82.383 (71.345, 102.623)	98.194 (82.228, 131.153)	121.141 (96.153, 179.253)	140.281 (106.432, 225.622)

4.2 Modeling using GPD

The choice of appropriate thresholds is based on the three methods presented in section 3.3. To examine the differences in dry and rainy seasons, the whole year's data is divided into two seasons: dry season from December to March and rainy season from April to November. We select one constant threshold for the whole year, and two seasonal thresholds for dry and rainy season solely.

The best indicated thresholds by all methods are finally chosen for the estimation of parameters. Fig. 1 displays the Hill plots for daily precipitation of the whole year for the city with most rain, Shantou and for Urumchi as the driest city. Hill plots for the other two cities follow a similar pattern. Two red dashed curves around the blue curve are the bounds of 95% CI. The curve for Shantou starts to be steady after around 100 mm for daily data, while the stability of the tail index emerges after 20 mm for Urumchi.

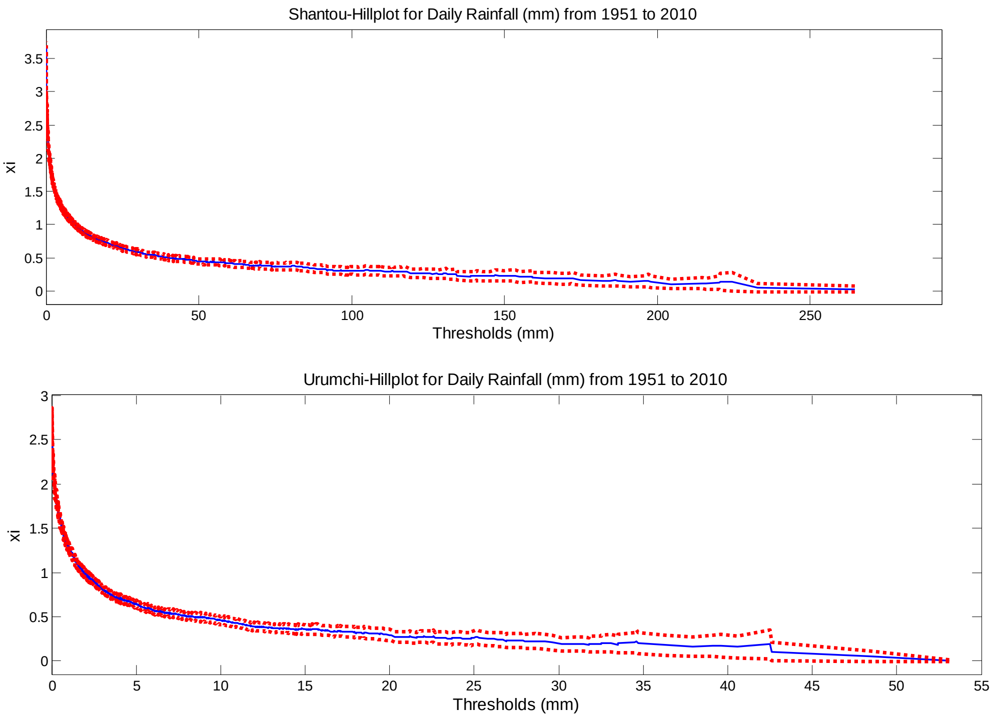

Figure 1. *Hill plots for daily precipitation for Shantou (above) and Urumchi (below)*

Second, the mean excess plots are plotted. The possible threshold is the point when the mean excess plot shows linearity, and estimated parameters look stable at different thresholds (Hasan and Chung, 2010). As an example, the sample mean excess plot for the constant threshold of Nanjing is presented in Fig. 2. In this case the constant threshold could be chosen as 60 mm. Mean excess plots for rainy and dry seasons for all cities can be drawn and analyzed in a similar way.

The third method incorporates the SPI. Fig. 3 exposes tamount of rainfall against Z-indexes for six years for Qiqihaer. The red straight line in each graph is the benchmark of 1.96. We can choose the level as a threshold when the Z-index starts to be greater than 1.96. As an average over the years, a constant threshold of 25 mm seems appropriate. The analysis like this is done for other cities and for rainy and dry seasons as well.

Table 6 gives an overview of the chosen threshold after completing the analysis of the three methods for all cities and all seasons.

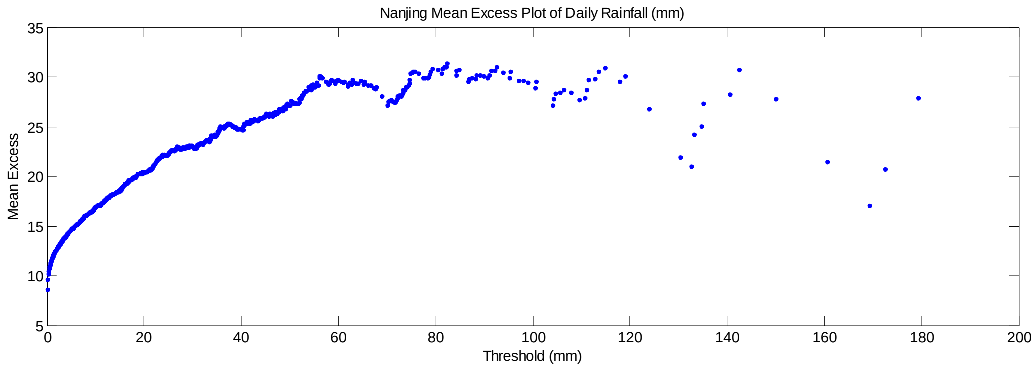

Figure 2. *Mean excess plot for daily precipitation for Nanjing*

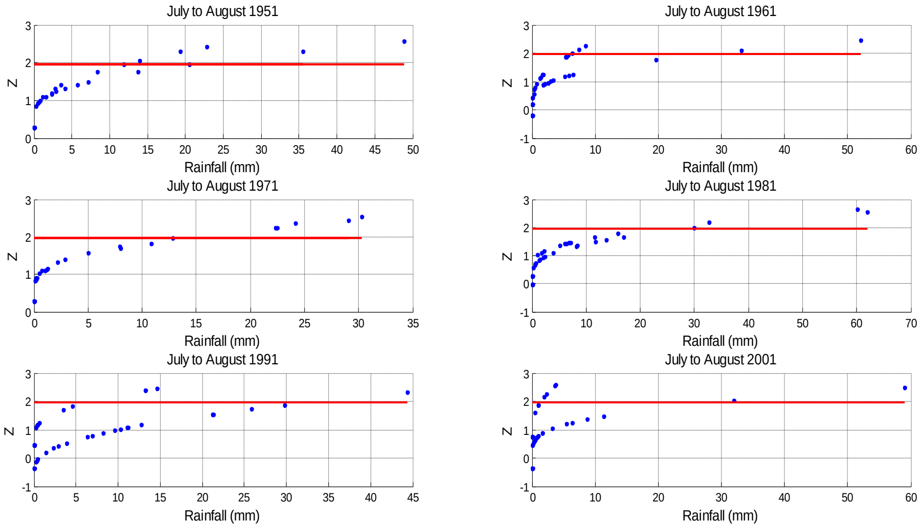

Figure 3. *Z-index against daily rainfall in July and August for six years for Qiqihaer*

Table 6. *Constant and seasonal thresholds*

Threshold (mm)	Nanjing	Shantou	Urumchi	Qiqihaer
constant	60 (n=124)	110 (n=78)	20 (n=90)	25 (n=222)
rainy season	90	150	20	45
dry season	30	60	10	10

After threshold selection, the parameters for GPD as given in equation (3) are estimated using MLE. The results are represented in Table 7. The exceedances of Nanjing have a long tail and follow the Ordinary Pareto distribution as the shape parameters for all seasons are negative. For Shantou, ξ is negative for constant and rainy threshold, but not for the dry season. Data from Urumchi follow the Ordinary Pareto distribution for rainy and dry season, but not for the whole year. In the case of Qiqihaer, a Pareto (II) type distribution was found in all cases. As all CI include the zero for the shape parameter, the Exponential distribution cannot be excluded.

Table 7. *GPD parameter estimates for constant and seasonal thresholds*

	Nanjing			**Shantou**		
Parameters (95% CI)	Constantover 60 / s.e.	Rainy over 90 / s.e.	Dry over 30 / s.e.	Constantover 110 / s.e.	Rainy over 150 / s.e.	Dry over 60 / s.e.
ξ	-0.052 / 0.095	-0.159 / 0.153	-0.101/ 0.108	-0.127 / 0.114	-0.232 / 0.203	0.238 / 0.189
(95% CI)	(-0.237, 0.134)	(-0.459, 0.141)	(-0.313, 0.110)	(-0.350, 0.097)	(-0.630, 0.166)	(-0.131, 0.608)
σ	31.052 / 4.049	34.746 / 7.416	9.495 / 1.449	50.443 / 8.072	54.692 / 14.983	21.269 / 5.078
(95% CI)	(24.048, 40.094)	(22.868, 52.792)	(7.040, 12.806)	(36.863, 69.026)	(31.969, 93.566)	(13.321, 33.959)

	Urumchi			**Qiqihaer**		
Parameters (95% CI)	Constant over 20 / s.e.	Rainy over 20/ s.e.	Dry over 10 / s.e.	Constant over 25 / s.e.	Rainy over 45 / s.e.	Dry over 10 / s.e.
ξ	0.002 / 0.113	-0.037 / 0.117	-0.118 / 0.095	0.095 / 0.076	0.032 / 0.139	0.372 / 0.311
(95% CI)	(-0.219, 0.223)	(-0.266, 0.193)	(-0.304, 0.069)	(-0.055, 0.244)	(-0.241, 0.304)	(-0.237, 0.981)
σ	7.907 / 1.220	8.765 / 1.420	4.908 / 0.665	13.078 / 1.326	15.895 / 3.079	3.764 / 1.310
(95% CI)	(5.843, 10.700)	(6.381, 12.040)	(3.764, 6.340)	(10.721, 15.952)	(10.874, 23.235)	(1.902, 7.447)

Similar to GEV, we examine how well the GPD model fits to the exceedances with probability distribution plots, fitted and empirical distribution plots, probability plots and QQ plots. We can conclude that the fitted probability density plots and the fitted cumulative plots match the empirical data consistently. Compared to GEV, the GPD probability plots are more appreciable for all cities, especially for higher values. From the QQ plots, the number of deviating observations for GPD is less than the number of observations modeled by GEV in all cases. To sum up, GPD is superior to GEV in terms of fitting which is in line with the literature (Jiang et al., 2009).

Table 8 describes the return level estimates at different return periods of daily exceedances over constant and seasonal thresholds. 95% CI are included in brackets. The pattern of the results is similar to GEV. The maximum values during the observation period never appear in the CI for dry seasons for Nanjing and Urumchi. For Shantou and Qiqihaer, the highest value of the past can be expected to happen even in dry seasons every 50 years.

Recalling the trigger amount of rain with 250 mm that can cause a flood in Southern China (CMA, 2012), we can conclude that a flood event does not happen during dry season in Nanjing. Even in rainy seasons, a flood above 250 mm should happen only once in 100 years. However, the upper bound of T=50 is already very close to this amount. For Shantou, we get a different picture. Here, the city experience a flood every 20 years in rainy seasons and every 50 years in dry seasons.

Table 8. *GPD return level estimates for constant and seasonal thresholds*

Selection Period	5	10	20	50	100
Constant Nanjing (60 mm)	128.328 (118.120, 142.419)	147.077 (133.909, 170.074)	165.170 (148.147, 201.586)	188.117 (164.424, 249.705)	204.770 (174.930, 291.557)
Rainy Nanjing (90 mm)	130.785 (120.245, 143.968)	149.330 (135.954, 168.371)	165.939 (150.176, 198.551)	185.265 (166.075, 248.597)	198.122 (175.882, 295.092)
Dry Nanjing (30 mm)	46.860 (43.908, 50.630)	52.072 (48.401, 58.100)	56.9300 (52.788, 66.726)	62.849 (56.978, 79.992)	66.975 (59.868, 91.571)
Constant Shantou (110 mm)	194.0472 (178.836, 213.038)	220.445 (201.863, 250.044)	244.621 (222.210, 292.282)	273.485 (244.695, 356.332)	293.197 (258.519, 411.417)
Rainy Shantou (150 mm)	192.070 (177.516, 212.280)	220.839 (200.222, 247.139)	245.334 (221.295, 289.245)	272.222 (244.958, 364.340)	289.084 (260.912, 437.436)
Dry Shantou (60 mm)	92.388 (82.513, 106.947)	114.232 (98.202, 145.591)	139.998 (116.803, 209.595)	181.290 (139.196, 361.692)	219.095 (156.720, 564.733)
Constant Urumchi (20 mm)	35.958 (33.031, 39.963)	41.4601 (39.5391, 43.8961)	46.968 (44.296, 50.777)	54.259 (50.168, 60.772)	59.782 (54.288, 69.058)
Rainy Urumchi (20 mm)	36.064 (34.618, 37.685)	41.661 (39.737, 44.061)	47.118 (44.513, 50.802)	54.121 (50.246, 60.348)	59.266 (54.138, 68.060)
Dry Urumchi (10 mm)	19.434 (18.760, 20.182)	21.964 (21.114, 23.008)	24.296 (23.208, 25.798)	27.101 (25.587, 29.436)	29.031 (27.121, 32.142)
Constant Qiqihaer (25 mm)	68.944 (62.709, 78.611)	81.291 (72.253, 97.540)	94.475 (81.618, 120.388)	113.284 (93.649, 157.913)	128.636 (102.475, 193.067)
Rainy Qiqihaer (45 mm)	69.788 (63.743, 77.935)	81.472 (73.038, 95.772)	93.415 (82.286, 119.709)	109.607 (93.368, 163.860)	122.169 (101.131, 209.891)
Dry Qiqihaer (10 mm)	14.949 (13.003, 17.821)	19.382 (16.032, 26.898)	25.119 (19.609, 47.758)	35.372 (24.564, 118.854)	45.815 (28.154, 251.963)

Comparing return levels estimated by GEV with the results of GPD, there are not significant differences at short periods (T=5 and T=10). The apparent differences occurs at the long return periods (after T=20). The confidence intervals obtained using GEV are higher compared to GPD. Especially, the upper limits of GPD are much less than the upper limits of GEV for the long return periods.

4.3 Stationary Tests

In this section, we present the results of the statistical tests introduced in section 3.5 for GEV and GPD results. Just from graphical inspection, there are no explicit evidences of trends and no changes in the pattern of variation in maximum precipitation. Further to graphic inspection, KPSS tests are carried out to test for trends around a constant or a deterministic linearity. The test statistics are revealed in Table 9. The critical values at 5% and 10% significant levels are 0.463

and 0.347 for trends around a constant, while 0.146 and 0.119 are for stationarity around a deterministic trend, respectively. When the test statistics are smaller than the critical values, it indicates insignificant evidence to reject the null hypotheses that there is stationarity around a constant or a linear trend. This is the case for data from Nanjing. The result for Shantou is similar with one exception for GPD with constant threshold. In the case of Urumchi, for all results from GEV the null hypotheses can be rejected and non-stationarity can be assumed. This continues for GPD and for stationarity around a linear trend which is rejected for constant and rainy season threshold. Qiqihaer accepts stationarity for GEV, but rejects it for GPD with constant and rainy season threshold. We can conclude that the assumption that data are stationary is appropriate for Nanjing and Shantou. But there might be problems for Urumchi and also for Qiqihaer in some cases which reveals limitations of the EVT approach.

Table 9. *KPSS test statistics for GEV - monthly and half-yearly maxima and for GPD – constant and seasonal thresholds; critical values: 5%: constant 0.463, linear 0.146, 10%: constant 0.347, linear 0.119*

	Nanjing		**Shantou**		**Urumchi**		**Qiqihaer**	
GEV	Stat. around constant	Stat. around linear trend	Stat. around constant	Stat. around linear trend	Stat. around constant	Stat. around linear trend	Stat. around constant	Stat. around linear trend
Month	0.124	0.018	0.032	0.024	1.576	0.253	0.124	0.068
Half-year	0.118	0.024	0.215	0.049	0.675	0.169	0.233	0.111
GPD								
Constant	0.263	0.025	0.494	0.064	0.154	0.154	0.484	0.250
Rainy	0.066	0.059	0.194	0.060	0.129	0.129	0.457	0.122
Dry	0.179	0.090	0.080	0.071	0.146	0.027	0.049	0.041

The p-values of MK test are showed in Table 10. When the p-value is larger than 0.05, the null hypotheses cannot be rejected that neither increasing nor decreasing trends exist in the exceedances. This is the case for almost all cities with only four exceptions, two for GPD with constant threshold from Nanjing and Shantou. The other two exceptions are from Urumchi with GEV where it is assumed that an increasing trend exists in the exceedances.

Table 10. *MK p-values (5%) for GEV - monthly and half-yearly maxima and for GPD – constant and seasonal thresholds*

	Nanjing		**Shantou**		**Urumchi**		**Qiqihaer**	
GEV	Test for positive trend	Test for negative trend	Test for positive trend	Test for negative trend	Test for positive trend	Test for negative trend	Test for positive trend	Test for negativ e trend
Month	0.193	0.807	0.247	0.753	<0.001	1.000	0.344	0.656
Half-year	0.191	0.809	0.632	0.358	0.009	0.991	0.243	0.757
GPD								
Constant	0.029	0.971	0.987	0.013	0.698	0.302	0.212	0.788
Rainy	0.326	0.674	0.859	0.151	0.736	0.264	0.118	0.882
Dry	0.898	0.102	0.558	0.450	0.138	0.862	0.825	0.175

5. CONCLUSION

This paper applied the Block Maxima model with GEV and the POT approach with GPD on 60 years of daily rainfall data in four cities in China, Nanjing, Shantou, Urumchi and Qiqihaer. This includes the estimation of parameters using MLE techniques and the calculation of return levels for different return periods. The purpose is to support decision makers in these regions with statistical knowledge about extreme precipitation that they can choose appropriate risk mitigating measures to reduce the damage caused by floods.

The results from the parameter estimation show that the GPD approach is the preferable model as it has a better goodness-of-fit performance. There are fewer deviations compared with GEV, especially in the QQ plots. However for both models, there are several small deviations for considerably large rainfall. The obtained return levels show that the return period of a flood event in Nanjing is 20 to 50 years for GEV and 50 to 100 years for GPD. This takes place during the rainy season from April to November, but it is highly unlikely to observe a flood in dry season from December to March. Shantou experiences a flood event caused by heavy rain every 5 to 10 years for GEV and every 10 to 20 years in rainy seasons for GPD. Every 50 years, it is further possible to see a flood during dry season. To sum up, extreme rainfall events can be predicted by the analysis of EVT. GEV and GPD are proper approaches to estimate return levels. This paper suggests using the GPD approach.

The limitations of this study are first, that it considers extreme rainfall without removing clusters. The classical GEV and GPD models assume that the observations are independent, but fail to take account of dependency in climate data in reality. Thus, further studies should use observations after declustering. Furthermore, GEV and GPD models need to be refined because of the existence of seasonality. Besides that, the spatial homogeneity of extreme rainfall for a region is valuable to test, which will be useful for choosing the best model from GEV and GPD for each city in a region. Instead of MLE, other estimation techniques such as L-moments should be considered for this purpose. As the first results of this paper are promising, future research is needed to overcome those limitations.

REFERENCES

[1] Balkema, A. A., and De Haan, L., Residual life time at great age. The Annals of Probability, pp. 792-804, (1974).

[2] BBC, BBC News, Retrieved from.: http://news.bbc.co.uk/2/hi/asia-pacific/7231622.stm, February 7, 2008.

[3] Bordi, I., Fraedrich, K., Petitta, M., and Sutera, A., Extreme value analysis of wet and dry periods in Sicily. Theoretical and applied climatology, 87(1-4), pp. 61-71, (2007).

[4] CMS, China Meteorology Administration, Retrieved From: http://www.cma.gov.cn/2011xzt/2012zhuant/20120928_1_1_1_1/2010052703/201212/t20121212_195616.html , 2012.

[5] Coles, S., Pericchi, L. R., and Sisson, S., A fully probabilistic approach to extreme rainfall modeling. Journal of Hydrology, *273*(1), pp. 35-50, (2003).

[6] Ding, Y., Cheng, B., and Jiang, Z., A newly-discovered GPD-GEV relationship together with comparing their models of extreme precipitation in summer. *Advances in Atmospheric Sciences*, *25*(3), pp. 507-516, (2008).

[7] Domroes, M., and Peng, G., *The climate of China*. Germany: Springer, 1988.

[8] Dong, Q., Chen, X., and Chen, T., Characteristics and Changes of Extreme Precipitation in the Yellow-Huaihe and Yangtze-Huaihe Rivers Basins, China. Journal of Climate, 24(14), pp. 3781-3795, (2011).

[9] Feng, S., Nadarajah, S., and Hu, Q., Modeling annual extreme precipitation in China using the generalized extreme value distribution. Journal of the Meteorological Society of Japan, 85(5), pp. 599-613, (2007).

[10] Fisher, R. A., and Tippett, L. H. C., Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society, 24(2), pp. 180-190, (1928).

[11] Gemmer, M., Becker, S., and Jiang, T., Observed monthly precipitation trends in China 1951–2002. Theoretical and applied climatology, 77(1-2), pp. 39-45, (2004).

[12] Hasan, H., and Chung, Y.W., Extreme Value Modeling and Precipitation of extreme rainfall: a case study of Penang. AIP Conf. Proc. 1309, pp. 372-393, (2010).

[13] Hill B. M., A simple general approach to inference about the tail of a distribution. Ann Stat 3, pp. 1163–1174, (1975).

[14] Hosking, J.R.M. Wallis, J.R., and Wood, E.F., Estimation of Generalized Extreme-value distribution by the method of probability weighted moments. Institution of Hydrology. 27(89), pp. 251-261, (1984).
[15] Hunt, K., and Ke, F., Floods leave at least 107 dead in China. CNN.com, Retrieved from: http://www.cnn.com/2013/08/20/world/asia/china-floods/, August 20, 2013.
[16] Jenkinson, A. F., The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quarterly Journal of the Royal Meteorological Society, 81(348), pp. 158-171, (1955).
[17] Jiang, Z. H., Ding, Y. G., Zhu, L. F., Zhang, J. L. and Zhu, L.H., Extreme Precipitation Experimentation over Eastern China Based on Generalized Pareto Distribution. Plateau Meteorology, 3(12), (2009).
[18] Li, N., Risk analysis on extreme precipitation events over China based on peaks over threshold model. Hydraulic Engineering, (2013).
[19] Li, Y., Cai, W., and Campbell, E. P., Statistical modeling of extreme rainfall in southwest Western Australia. Journal of Climate, 18(6), pp. 852-863, (2005).
[20] Liu, B., Xu, M., Henderson, M., and Qi, Y., Observed trends of precipitation amount, frequency, and intensity in China, 1960–2000. Journal of Geophysical Research: Atmospheres (1984–2012), 110(D8), (2005).
[21] Maraun, D., Extreme Value Theory. The Block Maxima Approach in a Nutshell. Invited presentation. Young researcher school on analyzing extreme events.5 to 8 Sep. 2010, Aussois, France.
[22] McAleer, M., Chu, L. F., and Wang, S. H., Statistical Modeling of Recent Changes in Extreme Rainfall in Taiwan. University of Canterbury, Department of Economics and Finance, (2012).
[23] McKee T. B., Doesken N. J., and Kleist J., The relationship of drought frequency and duration to time scales, Preprints,8th Conference on Applied Climatology, 17–22 January, Anaheim, CA, pp. 179–184, (1993).
[24] McNeil, A. J., Extreme value theory for risk managers. Departement Mathematik ETH Zentrum, (1999).
[25] Miroslava, U., The extreme value distribution of rainfall data at Belgrade, Yugoslavia. Atmósfera, 5(1), pp. 47-56, (1992).
[26] Nadarajah, S., Extremes of daily rainfall in West Central Florida. Climatic change, 69(2-3), pp. 325-342, (2005).
[27] Nadarajah, S., and Choi, D., Maximum daily rainfall in South Korea. Journal of Earth System Science, 116(4), pp. 311-320, (2007).
[28] Overeem, A., Buishand, A., and Holleman, I., Rainfall depth-duration-frequency curves and their uncertainties. Journal of Hydrology, 348(1), pp. 124-134, (2008).
[29] Pickands III, J. Statistical inference using extreme order statistics. the Annals of Statistics, pp. 119-131, (1975).
[30] RCSC, Information Bulletin – China: Floods and typhoons. No. 1. GLIDE No. TC-2013-000099-CHN. International Federation of Red Cross and Red Crescent Societies. August 23, 2013.
[31] Sun, J. Q., Ao J., Changes in precipitation and extreme precipitation in a warming environment in China. Chinese Science Bulletin, 58(12), pp. 1395-1401, (2013).
[32] Wang, M., Heavy rain wreaks havoc in E China's Nanjing. CCTV News, Retrieved from: http://english.cntv.cn/20110719/110856.shtml, July 7, 2011.
[33] Wang, Y., and Zhou, L., Observed trends in extreme precipitation events in China during 1961–2001 and the associated changes in large-scale circulation. Geophysical Research Letters, 32(9), (2005).
[34] Withers, C. S., and Nadarajah, S., Evidence of trend in return levels for daily rainfall in New Zealand. J. Hydrol.(NZ), 39(2), pp. 155-166, (2000).
[35] Yang, E., Extreme weather - heavy rainfall. Retrieved from: http://www.niwa.co.nz/natural-hazards/extreme-weather-heavy-rainfall , (2012).
[36] Yuan, W.P. and Zhou, G.S., Comparison between standardized precipitation and Z-index in China. Acta Phytoecologica Sinica. 28 (4), pp. 523-529, (2004).
[37] Zhai, P., Zhang, X., Wan, H., and Pan, X., Trends in total precipitation and frequency of daily precipitation extremes over China. Journal of Climate, 18(7), pp. 1096-1108, (2005).

AUTHORS' BIOGRAPHY

Manuela Ender works as a lecturer at Xi'an Jiaotong-Liverpool University in Suzhou China. She got her PhD from the University of Cologne, Germany and worked as research manager and management consultant for a software company that is ranked among the companies leading the market for financial software solutions in Germany. Her research interests lie in the area of financial derivatives and risk management.

Tong Ma holds a double degree in BEcon Financial Mathematics from the Xi'an Jiaotong-Liverpool Univeristy and in BSc in Financial Mathematics from the University of Liverpool. Her research interests lie in the area of mathematical finance, especially derivatives pricing and hedging.