ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 25, NO. 3, 2008, 507–516

A Newly-Discovered GPD-GEV Relationship Together with Comparing Their Models of Extreme Precipitation in Summer

DING Yuguo*[1] (丁裕国), CHENG Bingyan[2] (程炳岩), and JIANG Zhihong[1] (江志红)

[1] *National Key Laboratory of Meteorological Diseases of Jiangsu, Nanjing University of Information Science and Technology, Nanjing* 210044

[2] *Climate Center of Chongqing City Meteorological Bureau, Chongqing* 401147

(Received 13 April 2007; revised 14 September 2007)

ABSTRACT

It has been theoretically proven that at a high threshold an approximate expression for a quantile of GEV (Generalized Extreme Values) distribution can be derived from GPD (Generalized Pareto Distribution). Afterwards, a quantile of extreme rainfall events in a certain return period is found using L-moment estimation and extreme rainfall events simulated by GPD and GEV, with all aspects of their results compared. Numerical simulations show that POT (Peaks Over Threshold)-based GPD is advantageous in its simple operation and subjected to practically no effect of the sample size of the primitive series, producing steady high-precision fittings in the whole field of values (including the high-end heavy tailed). In comparison, BM (Block Maximum)-based GEV is limited, to some extent, to the probability and quantile simulation, thereby showing that GPD is an extension of GEV, the former being of greater utility and higher significance to climate research compared to the latter.

Key words: Generalized Pareto Distribution, Generalized Extreme Value, daily rainfall, extreme precipitation, rainstorm

DOI: 10.1007/s00376-008-0507-5

1. Introduction

Recent modeling studies have reconfirmed that possible future changes of weather and climate extremes are derived by comparing models of natural climate variability and its future state with atmospheric constituents affected by human activity (e.g., ever-increased greenhouse gases, sulfate aerosols etc.), as reported by IPCC. Consequently, the possible regime of extreme climate/weather events has become a major concern in climate research (IPCC, 2001; Meehl et al., 2000; Huang, 2000; Christensen and Christensen, 2003; Ding et al., 2006).

Rapid advances in climate modeling capabilities have made reality-closer estimate of future mean climate change during the past decade. Current global coupled climate models have improved resolution (grid points typically at $2.5^{\circ} \times 2.5^{\circ}$), with more detailed and accurate land surface schemes, and dynamical sea ice formulations. Some have even higher resolution in the ocean near the equator (leading to improved simulations of El Niño). These changes are coupled with improved techniques to study climate changes and processes at smaller regional scales from GCM results through either embedding high-resolution regional models (with grid points every 50 km or so) in the global models or statistical downscaling techniques. However, direct imitation of extreme climate events and their changes remain rather difficult. The combination of a downscaling method under development with the theory on the distribution of extreme values used in statistical climate has advanced the research into the prediction of possible climate extreme values in a certain return period to follow (Jones et al., 1997; IPCC, 2001; Ding and Sun, 2006).

In general, the statistical simulation of the extreme

*Corresponding author: DING Yuguo, dygnim@jlonline.com

climate values must be given by using the extreme value distribution of each climatic element so that a statistical inference of the future climatic extreme value from above provided future climate projections scenario can be obtained (Katz et al., 2005; Meehl et al., 2000; Easterling et al., 2000). Thus, to determine a certain model for extreme value distribution is the foundation of diagnosis and prediction for extreme climate according to a projection of the future climate scenario (Katz et al., 2005; Meehl et al., 2000; Karl and Easterling, 1999) .

Jenkinson (1955) proposed the classical theory on extreme distribution advanced by Fisher and Tippett (1928), in which the three kinds of typical extremes distribution were changed into a three-parameter Generalized Extreme Values distribution (GEV) as a new model (Gumbel, 1958) for applied researches. GEV studies on meteorological extreme value distribution are now popular, especially type-I Gumbel model (Fisher and Tippett, 1928; Jenkinson, 1955; Gumbel, 1958). Pickands (1975) introduced Generalized Pareto Distribution (GPD) into hydro-meteorological research, which was improved by Hosking and Wallis (1987) and is now being diffused (Pickands, 1975, Davison and Smith, 1990; Abild et al., 1992; Guttman et al., 1993). GPD is capable of extracting the yearly maxima (or minima) above (below) a given threshold from a primitive sequence, e.g., of calendar year, a threshold that falls into a so-called POT (Peaks Over Threshold) sampling scheme (Davison and Smith, 1990) so that the needed number of years is greatly decreased and the number of samples is increased for extremes, thereby overcoming the major disadvantage of the BM (Block Maxima) or AM (Annual Maxima) sampling for GEV or Gumbel distribution (Katz et al., 2005; Davison and Smith, 1990). Since the mid 1960s research into GPD statistical theories and applications have been actively undertaken, leading to rapid advances and new results appearing one after another (Cunnane, 1973, 1979; Ashkar and Rousselle, 1987). They made a study on the problems of threshold choice and GPD extremes distribution.

In recent years, it is demonstrated theoretically that at high thresholds there is a close association of parameters between GEV and GPD, which has further improved a stochastic point-process theory proposed by Coles (2001) and Katz et al. (2005), independently. Due to the fact that GPD parameters (threshold, scale and shape) are of generalization, they bear unique relations to the logarithmic, Beta and Pareto distributions at different thresholds, thus making application even easier and more flexible. Currently, GPD finds a variety of uses in hydrology, achieving numerous advances in the distribution of rainstorms and flood water levels as well as the statistical inference for flood likelihood (Pickands, 1975). In contrast, little is reported regarding the utility in atmospheric sciences.

China is a country often hit by local rainstorms that tend to trigger floods. If an exceptionally heavy rain occurred in a return period of decades or even a hundred years, major rivers and lakes would be swollen, making for deluges (Zhai et al., 1999, Wang et al., 2002). But previous studies of rainstorm extremes predominantly made use of the classical Gumbel or GEV distribution (Jenkinson, 1955; Gumbel, 1958), and their indispensable prerequisite is that the maximum value sampling takes just one value per year, i.e., Annual Maximum (AM). In reality, the stochastic variability of a maximum is quite high in the same year, and for different regions having varied wetness, one maximum taken for each year is unjustifiable. For example, there may be more than one extreme rainfall events beyond a set warning level in some years and it is quite possible that just a single rainstorm is observed in others. As a result, the sampling of one maximum per year is likely to lose a considerable amount of useful information. However, for arid or semi-arid areas it is quite probable that practically no such events reach the critical value throughout the year and extraction, if made, is bound to include spurious information on the maximum, a demerit that is inherent in the classical extremes distribution (Pickands, 1975; Yao and Ding, 1990; Coles, 2001; Ding et al., 2004; Katz et al., 2005).

The purpose of the paper is to introduce the latest theory on GEV and GPD models and to further demonstrate theoretically their approximate relation in seeking the quantile at high thresholds. By use of observations we have demonstrated and assessed the merit and demerit of GEV-, Gumbel- and GPD-fitted extreme rainfall amounts and attempted to predict the quantile, whereon an optimal model of the extremes distribution is explored, with which to lay a foundation for better simulating and predicting the statistical features of extreme precipitation event in our country.

2. Linkage between GPD and GEV theories

GPD is essentially an simple primitive distribution model, designed specifically to describe probability distribution features of the whole dataset of observations beyond a given critical value (threshold), for example, floods above a given critical value level, rainfall over a given peak value for an hour, day, pentad, decade or month, temperature higher than a threshold and gust stronger than a set wind speed, and its distribution function is in the form

$$F(X) = 1 - \left[1 - k\left(\frac{x-\beta_1}{\alpha_1}\right)\right]^{1/k} ,$$
$$k > 0,\ \beta_1 \leqslant x \leqslant \beta_1 + \frac{\alpha_1}{k}$$
$$\text{or } k < 0 ,\ \beta_1 \leqslant x < \infty , \tag{1}$$

$$F(x) = 1 - \exp\left[-\left(\frac{x-\beta_1}{\alpha}\right)\right] ,\ k = 0 , \tag{2}$$

in which β_1 denotes the threshold, α_1 the scale parameter and k the shape parameter (linear type). If $y = x - \beta_1$ denotes the values of the variable X above the threshold β_1, we are allowed to rewrite the distribution function as $F(y)$. It is seen from Eq. (2) that for $k = 0$, the GPD can be simplified to a logarithmic distribution (Coles, 2001; Katz et al., 2005).

To infer the GPD quantile at a given probability it is necessary to determine the crossing rate $\lambda\ (> 0)$ for variable X over a certain threshold β_1. Theoretically, POT is generally assumed to experience the Poisson process before reaching exceedance that means to be related to a crossing rate over a given threshold, with the unbiased estimate as , where stands for the number of POT and for the number of years of records. Following the theory on the Poisson process, the crossing rates (i.e., exceedances) obey the Poisson distribution (Gumbel, 1958; Wang, 1991). From the size of samples with one year as a unit, the annual crossing rates above a certain threshold in years t can be given (Katz et al., 2005) as

$$\lambda_x = \lambda t[1 - F(x)] \tag{3}$$

where λ is a mean over multi-year crossing rates between the given value of an extreme event and the threshold β_1, viz., the expectation value of the number of POT each year. As a result, if the quantile X_T with a return period T (years) is assumed, then $\lambda_x (= 1)$ of Eq.(3) is simplified to a unit element 1 (unity). Now we can derive the solution to the quantile from Eqs. (1) and (3)

$$x_T = \beta_1 + \frac{\alpha_1}{k}\left[(1 - (\lambda T)^{-k}\right] ,\quad k \neq 0 , \tag{4}$$
$$x_T = \beta_1 + \alpha_1 \ln(\lambda T) ,\quad k = 0 , \tag{5}$$

in which λ denotes the yearly mean crossing rate, β_1 the given threshold, α_1 scale parameter, and k the shape parameter (to denote the distribution curve type). Then, the related GPD model and its quantile can be found by obtaining GPD parameters by use of the given estimating method. In a special case with $\lambda = 1$, i.e., the POT crossing rate appearing once a year, we simplify Eqs. (4)–(5) to (6)–(7), respectively.

$$x_T = \beta_1 + \frac{\alpha_1}{k}(1 - T^{-k}) ,\quad k \neq 0 , \tag{6}$$
$$x_T = \beta_1 + \alpha_1 \ln(T) ,\quad k = 0 . \tag{7}$$

On the other hand, according to GEV distribution theory proposed by Jenkinson (1955), its distribution function can be written as

$$F(x) = \exp\left\{-\left[1 - k\left(\frac{x-\beta}{\alpha}\right)\right]^{1/k}\right\} ,\quad k \neq 0 , \tag{8}$$

$$F(x) = \exp\left[-\exp\left(-\frac{x-\beta}{\alpha}\right)\right] ,\quad k = 0 , \tag{9}$$

Evidently, for $k = 0$, GEV falls into Tippett type I (i.e., Gumbel distribution), into its type II for $k < 0$ and into its type III (Weibull distributionm) for $k > 0$. Given a return period T, we have the relationship between the distribution function and T in the form

$$F(x_T) = 1 - \frac{1}{T} , \tag{10}$$

from which we obtain the associated quantile with a return period T according to Eqs. (8) and (9).

$$X_T = \beta + \frac{\alpha}{k}\left\{1 - \left[-\ln\left(1 - \frac{1}{T}\right)\right]^k\right\} ,\ k \neq 0 \tag{11}$$

$$X_T = \beta - \alpha \ln\left[-\ln\left(1 - \frac{1}{T}\right)\right] ,\quad k = 0 . \tag{12}$$

Coles (2001) and Katz et al. (2005) have proven, independently, that parameters of GEV and GPD are bound up with each other. With GEV parameters β, $\alpha > 0$ and k given for GPD, if we assume the threshold β_1 and scale parameter α_1, then the distribution function of GEV can be approximately written as that of GPD under a high enough threshold, leading to

$$F(x) = 1 - \left[1 - k\left(\frac{x-\beta_1}{\alpha_1}\right)\right]^{1/k} , \tag{13}$$

where the parameter k is constant for both, but the parameters between GPD and GEV have the following relations (Coles, 2001)

$$\alpha_1 = \alpha + k(\beta_1 - \beta) , \tag{14}$$
$$\beta_1 = \beta + \frac{\alpha}{k}(\lambda^{-k} - 1) , \tag{15}$$

In Eq. (15), λ denotes the exceedance of Poisson in a randomized point process, which is obtained from the data series by means of the Poisson probability model. Again, from Eq. (15) we have the relation

$$\ln \lambda = -\frac{1}{k} \ln\left[1 + k\frac{(\beta_1 - \beta)}{\alpha}\right] . \tag{16}$$

Obviously, substitution of Eq. (14) into Eq. (16) yields the general expressions of GEV parameters α and β

$$\ln\alpha = \ln\alpha_1 + k\ln\lambda\,. \tag{17}$$

Then, taking Eq. (15) into account, we find

$$\beta = \beta_1 - \frac{\alpha}{k}(\lambda^{-k} - 1)\,. \tag{18}$$

When GPD parameters α_1 and β_1 are given, we are allowed to calculate GEV parameters α and β by means of parameter λ, which is obtained from the data series using the Poisson probability model. It follows that their parameters are acquired by inter-conversion with constant k (refer to Coles, 2001). We can also see from Eq. (14) that with $\beta_1 \to \beta$, GPD and GEV are equivalent.

3. Approximate relation between GPD and GEV quantiles

We attempt to further prove the relationship between quantiles of GPD and GEV at higher thresholds.

In the logarithmic term on the right-hand side of the expression for GEV quantile of Eqs. (11) and (12) there is $p = 1/T$ of small probability in general, where T represents the return period, and by power series expansion, and neglecting the high-order terms, we have an approximate expression in the form

$$-\ln\left(1-\frac{1}{T}\right) \approx \frac{1}{T}\,, \quad \left|\frac{1}{T}\right| \leqslant 1\,. \tag{19}$$

Correspondingly, we arrive at the approximate expressions of Eqs. (11) and (12) as follows

$$X_T \approx \beta + \frac{\alpha}{k}[1-(T)^{-k}]\,, \tag{20}$$

and

$$X_T \approx \beta + \alpha\ln T\,. \tag{21}$$

Substituting Eqs. (14) and (15) into above equations, we have approximate expressions for GPD quantile

$$X_T = \beta_1 + \frac{\exp[\ln(\alpha_1\lambda^k)]}{k}(2-T^{-k}-\lambda^{-k})\,, \quad k \neq 0\,, \tag{22}$$

$$X_T = \beta_1 + \exp[\ln(\alpha_1\lambda^k)][\ln T - \frac{1}{k}(\lambda^{-k}-1)]\,, \quad k = 0\,. \tag{23}$$

It is evident that at a higher threshold that will lead to the annual crossing rate $\lambda \to 1$. Eqs. (22) and (23) are simplified, respectively, to

$$X_T = \beta_1 + \frac{\alpha_1}{k}(1-T^{-k})\,, \tag{24}$$

$$X_T = \beta_1 + \alpha_1(\ln T)\,, \tag{25}$$

indicating clearly that GPD parameters β_1 and α_1 are completely equivalent to GEV parameters β and α so that Eqs. (24) and (25) are the equivalent formulae to the approximate expressions (20) and (21), respectively.

The above approximate expressions show that in seeking GEV quantile, when the yearly crossing rate $\lambda \to 1$ and T is big, Eq. (11) is close to Eq. (20) or Eq. (24) for seeking the GPD quantile and Eq. (12) is close to Eq. (21) or Eq. (25) for seeking the GPD quantile. From the above we see that at higher thresholds as a special case GEV approximately represents GPD, the latter being the more generalized form of the former. It follows that GEV distribution can be viewed as a special case of GPD under certain circumstances, thereby demonstrating theoretically that when a high threshold is assumed, with the decrease in the yearly crossing rate, leading to $\lambda \to 1$, the solutions to the quantile of all types of GPD can be obtained by approximate calculation from those of all types of GEV distribution and vice versa, regardless of whether or not the shape parameter $k = 0$. In other words, both the distributions are exchangeable in use under a given condition. If $\lambda T = T^*$ is assumed in Eqs. (4) and (5), then we can give the related forms

$$x_T = \beta + \frac{\alpha}{k}(1-T^*)^{-k}\,, \quad k \neq 0\,, \tag{26}$$

$$x_T = \beta + \alpha\ln(T^*)\,, \quad k = 0\,. \tag{27}$$

Similarly, substituting the approximate expression (19) into Eqs. (26) and (27), we have

$$X_{T^*} \approx \beta + \frac{\alpha}{k}\left\{1-\left[-\ln\left(1-\frac{1}{T^*}\right)\right]^k\right\}\,, \quad k \neq 0\,, \tag{28}$$

$$X_{T^*} = \beta - \alpha\ln\left[-\ln\left(1-\frac{1}{T^*}\right)\right]\,, \quad k = 0\,. \tag{29}$$

which indicate that for GEV, the derived quantile in a return period T^* is just equivalent to the extreme obtained in the return interval $T = T^*/\lambda$ from GPD fitting because of $\lambda T = T^*$. Generally speaking, GPD sampling mode is POT, with the yearly crossing rate $\lambda \geqslant 1$ while for GEV its sampling mode is AM, so that the former is more generalized compared to the latter. The utility of Eqs. (28) and (29) lies in that the use of GPD fittings and the above-mentioned relations allows us to derive the GEV parameters and the quantile without specially calculating the counterparts of GEV.

4. An estimation method of parameters: L-moment estimation based on PWM

Following Hosking and Wallis (1987) we can obtain the L moment estimation formulae for GPD parameters as follows,

$$\mu_1 = \beta_1 + \frac{\alpha_1}{1+k} , \tag{30}$$

$$\mu_2 = \frac{\alpha_1}{(1+k)(2+k)} . \tag{31}$$

According to the relation between PWM (Probability Weighed Moment) and L moment we have the following

$$\mu_1 = b_0 , \tag{32}$$

$$\mu_2 = 2b_1 - b_0 , \tag{33}$$

where μ_1 and μ_2 are, respectively, the first- and second-order L moment, and

$$b_0 = \bar{x} , \tag{34}$$

$$b_1 = \sum_{j=1}^{n-1} \frac{(n-j)}{n(n-1)} x_j , \tag{35}$$

where x_j is a sequenced statistics of observations, viz., $x_n \leqslant x_{n-1} \leqslant \ldots \leqslant x_1$, suggesting a sequence that is formed by extremes above a given threshold arranged in a decreasing order. Finally we get general expressions for parameters α and k of GPD, viz.,

$$\hat{k} = \frac{b_0}{2b_1 - b_0} - \frac{\beta_1}{2b_1 - b_0} - 2 , \tag{36}$$

$$\alpha_1 = (b_0 - \beta_1)\left(1 + \frac{b_0}{2b_1 - b_0} - \frac{\beta_1}{2b_1 - b_0} - 2\right) , \tag{37}$$

wherein the threshold β_1 denotes a critical value given from a certain condition. For rainfall, a range of standard critical precipitation amounts can be defined for experiment or we can assume one, two and three SD (Standard Deviation) as the critical values for experiments.

5. Case study

5.1 *Data*

A set of samples for experiment consists of daily rainfall from May to September, 1953–2002, taken out of 10 representative stations, including the Qiqihar, Beijing, Xi'an, Chongqing, Nanjing, Hangzhou, and Guangzhou areas.

5.2 *Comparison between fitting accuracies*

The parameters of GPD, GEV and Gumbel distribution are estimated using the L moment method of Eqs.(30)–(37), indicating once more the advantages, e.g., the statistical method being simple and the estimated parameters having greater robustness as well as demonstrating by example that the GPD, GEV and Gumbel distributions show the highest, higher and lowest fitting accuracy, in order, for which three test indices are utilized, which are Kolmogoroff-Smirnoff statistic test (K-S), correlation coefficient (R) and mean square error (MSE) (see Hosking, 1990). Table 1 gives the fittings of the 10 representative stations for comparison. But as regards to the calculation of high-end quantile they are somewhat similar to each other, still in a decreasing order for the precision. Besides, N denotes the sample size of extreme values for each model in Table 1.

From the given examples it is demonstrated that in terms of fitting accuracy, GPD is higher than GEV and Gumbel, with GEV, in turn, superior to Gumbel.

Here we made use of three test criteria, as mentioned above. As shown in Table 1, on average, the biggest correlation coefficient between the Gumbel modeled and measured curves does not exceed 0.55, compared to 0.99 in the case of GPD. In other words, GPD shows the perfect fit between its theoretical and observed curves while Gumbel model gives a good fit only in part of the curves.

To investigate the GPD utility, we have calculated, separately, different sizes of samples and the related parameters of GPD at given thresholds and GEV, as well as Gumbel. Table 2 gives the fittings of these models for comparison in the Hangzhou area as an example. It is seen from Table 2 that the GPD parameters are steadier in all the cases because of their smaller standard deviation, suggesting that their fitt-

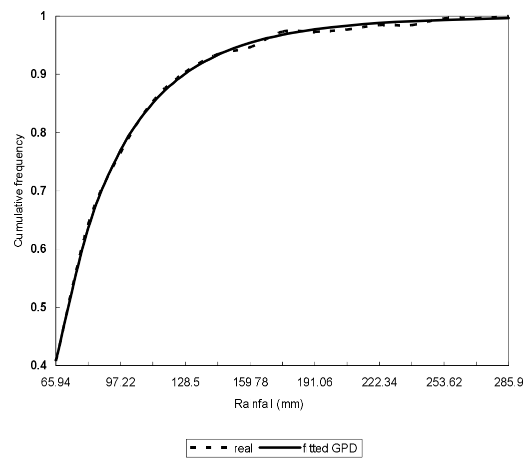

Fig. 1. GPD fitted curve of rainfall for Guangzhou area.

Table 1. Comparison among the fittings from GPD, GEV and Gumbel models.

Place	GPD				GEV				Gumbel			
	K-S	R	MSE	N	K-S	R	MSE	N	K-S	R	MSE	N
Qiqihar	0.07	0.99	0.01	31	0.23	0.99	0.03	50	0.94	0.52	0.15	50
Beijing	0.03	1	0	111	0.27	0.98	0.03	50	0.96	0.36	0.18	50
Xi'an	0.06	0.99	0.01	33	0.24	0.98	0.03	50	0.96	0.45	0.17	50
Zhengzhou	0.03	1	0	90	0.18	0.99	0.03	50	0.96	0.42	0.17	50
Yichang	0.07	0.99	0.01	135	0.23	0.98	0.03	50	0.92	0.48	0.15	50
Wuhan	0.04	1	0	197	0.31	0.97	0.04	50	0.94	0.41	0.17	50
Chongqing	0.05	0.99	0	134	0.34	0.96	0.04	50	0.94	0.41	0.17	50
Nanjing	0.04	1	0	143	0.24	0.99	0.03	50	0.92	0.55	0.14	50
Hangzhou	0.05	1	0	157	0.2	0.99	0.03	50	0.98	0.45	0.16	50
Guangzhou	0.01	1	0	260	0.25	0.98	0.03	50	0.92	0.52	0.14	50

Table 2. Comparison of stability fitted models (threshold 50.0 mm for GPD) in Hangzhou area.

Model	Parameter	N					Mean	S.D.
		10	20	30	40	50		
GPD	Shape k	−0.02	−0.06	−0.07	−0.07	−0.05	−0.05	0.02
	Scale α_1	21.37	20.53	20.12	20.73	21.4	20.83	0.49
Gumbel	Shape α_2	21.65	34.1	38.28	36.02	24.79	30.97	6.54
	LP β_2	5.14	14.66	28.13	47.89	73.85	33.93	24.59
GEV	Shape k	−0.69	0.33	0.05	0.33	0.04	0.01	0.37
	Scale α	16.89	30.37	38.86	40.26	25.16	30.31	8.70
	LP β	17.63	34.34	50.23	68.68	88.16	51.81	24.83

Note: S.D. stands for "standard deviation" and LP for "local parameter".

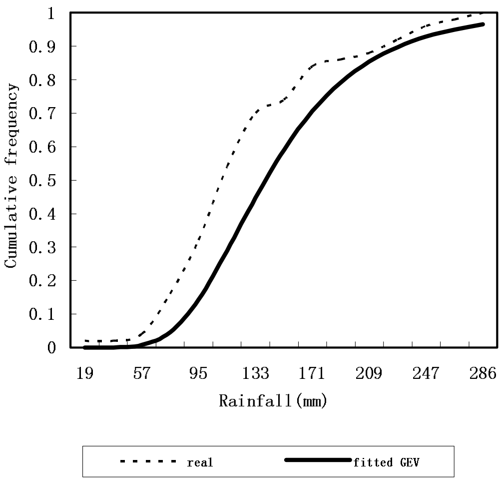

Fig. 2. As in Fig. 1 but for GEV fitted curve.

ings are subject to practically no effect of sample sizes, next being the GEV parameters fittings, and calculated Gumbel parameters show greater dependence upon the sample size, thus producing relatively big mean square error.

Now, we take the Guangzhou area for example. The fitting curves of the distribution models are presented in Fig. 1 for GPD and Fig. 2 for GEV, where we can readily visualize considerable difference among the flood-season rainfall fitting precisions from these models in the same size of samples and the same area (see below).

It is seen that the GPD is best suitable for fitting all frequency points, next being the GEV, with the Gumbel fittings relatively better just at a higher or upper boundary of cumulative frequency (figure not shown), whose residual part is, however, very poor. In other words, viewed from the entire course of fitting, GPD result is the best and the Gumbel is the worst among the three cases. However, viewed from predicting the high-end quantile in a given return period, i.e., the small-probability extreme event, their results may vary little but in comparison, GPD is the optimal.

5.3 *GEV Parameters used as GPD ones and vice versa for experiments*

To illustrate the inter-conversion we take the Beijing case for example, with the results depending on different thresholds given in Table 3, where we can determine at what thresholds the combinations are optimal. From Table 3 we see that at higher thresholds, the GEV fitted parameters can be used to approximately calculate GPD parameters by means of Eqs. (15) and (16). And with the constant shape parame-

Table 3. Relative errors (Re, %) of the estimated scale parameter and quantile (X_{100}) for a return period of 100-yr in the Beijing area.

Threshold	GPD	(GPD)	Re (%)	X_{100}	(X_{100})	Re (%)
62	25.31	29.52	0.17	210.41	186.00	0.12
64	26.53	29.34	0.11	210.41	192.30	0.09
66	26.11	28.63	0.10	210.41	190.45	0.09
68*	29.17	30.5	0.05	210.4	204.00	0.03
70	32.41	28.83	0.11	210.4	217.92	0.04
72	32.84	28.27	0.14	210.4	219.79	0.05
74*	29.00	28.40	0.02	210.4	203.85	0.03
76	27.26	27.44	0.01	210.41	196.80	0.06
78*	30.31	29.16	0.04	210.4	209.28	0.01
80*	29.50	29.43	0.002	210.4	206.17	0.02
82	26.37	27.73	0.05	210.4	194.02	0.08
84	23.12	26.0	0.12	210.4	181.57	0.13

Note: The asterisk-denoted thresholds represent <5% relative errors of the calculated scale parameter and quantile, and bracketed (GPD) designates the GPD scale parameter calculated by use of the GEV equivalent and bracketed (X_{100}) denotes the calculated upper-level quantile, respectively.

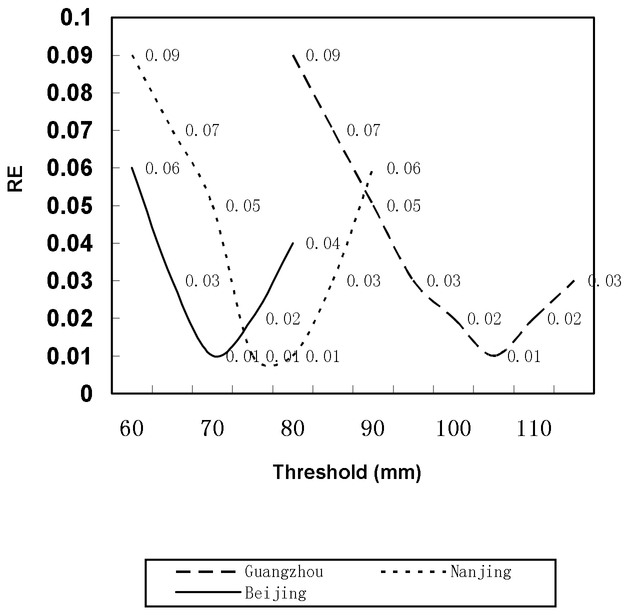

Fig. 3. Threshold-dependent error of approximately calculated quantile for the Guangzhou, Nanjing and Beijing areas.

ter (Coles, 2001), the relative error of the scale parameter generally does not exceed 17%, with the minimum error close to zero.

Also, from the table, we see an exceedingly small error of the calculated quantile at the upper part (the thick tail) of the distribution density curve and particularly at higher thresholds the relative error is generally less than 5%, with the minimum on the order of 1%. The asterisk-denoted entries in Table 3 show the relative errors of the estimated parameters and quantiles to be smaller than 5% so that the set thresholds are optimal. The results from the use of GEV fitted parameters instead of the GPD ones are quite comparable for the other chosen stations.

5.4 *Approximate calculation of ouantile*

Under a certain high threshold, the quantile for GPD can be approximately estimated from the parameters of fitted GEV, or otherwise, the quantile to GEV is still approximately estimated from the parameters of fitted GPD by using Eqs. (20) and (21) or equivalents to Eqs. (24) and (25) as well as Eqs. (26) and (27). If at higher thresholds parameters derived by GPD or GEV are substituted into the above formulae, the upper-part quantile can be obtained by means of either of them for the other. Tables 4 and 5 indicate the approximate calculations and the relative errors to the different thresholds, respectively, in the Beijing and Nanjing areas.

Figure 3 depicts, separately, the threshold-varying approximate error of the calculated quantile for the Guangzhou, Nanjing and Beijing areas, indicating that the optimal threshold for the quantile in a 100-yr return period depends on regional difference in rainfall climate features. It is also seen from the figure that mean rainfall differs greatly between the three city areas during the flood period, to which the optimal threshold may be related. From this unexpected discovery a new approach is found to further reveal a relationship between the original variable and GPD parameters.

5.5 *Selection of thresholds*

As stated earlier, for GPD it is required to infer an optimal threshold before seeking a quantile in a certain

Table 4. Approximate calculations of GPD quantile (mm) for the Beijing area at a range of thresholds (mm) at $N = 50$ years with relative error (Re) presented.

Thresholds	60	65	70	75	80
GPD	177.3	191.3	217.9	200.1	206.2
GPD*	167.2	185.9	219.1	204.2	215.7
T^*	64.9	82.0	104.2	116.3	138.9
λ	1.54	1.22	0.96	0.86	0.72
Re	0.06	0.03	0.01	0.02	0.04

Note: The GPD* represent the approximate GPD quantile and the T^* represent the equivalent return period of GEV, and λ is the yearly crossing rate.

Table 5. Approximate calculations of the GPD quantile (mm) for the Nanjing area at $N = 50$ years with the threshold in units of mm and relative error denoted as Re.

Threshold	60	65	70	75	80	85	90
GPD	212.1	206.6	181.3	230.7	224.9	239.0	241.7
GPD*	192.0	191.9	171.9	227.3	226.1	246.6	255.4
T^*	50	58.8	65.8	*90.9	*104.2	125	147.1
λ	2	1.7	1.52	*1.1	*0.96	0.8	0.68
Re	0.09	0.07	0.05	*0.01	*0.01	0.03	0.06

Note: The GPD* represent the approximate GPD quantile and the T^* represent the equivalent return period of GEV, and λ is the yearly crossing rate.

return period, viz., an extreme event of small probability. As indicated by numerous investigators, the choice of an optimal threshold is related to the sampling independence and the smallest interval between two extreme events, that is, to guarantee the interindependence of the selected extremes without involving their correlativity. However, it should be noted that the choice of a threshold is related to the purpose, so that the optimal threshold is to make minimal the error of the calculated quantile when examining the approximate estimate of the quantile or inferring it, and the mean square error of the calculated parameters is to be made minimal when considering the accuracy of the calculated model parameters.

6. Conclusions and review

6.1 *Comparison of the sampling schemes*

GEV distribution is the one of asymptotical extremes derived from primitive distribution when the sequence of an original variable is long enough to require a maximum in the interval of time, say, in a year, as the sample size, i.e., an extreme taken from the series on a yearly basis, which is known as Block Maxima (BM) or Annual Maxima (AM) (refer to Hosking et al., 1985; Katz et al., 2005). Obviously, just a few samples of extremes can be taken if the sequence is very short, thus leading to the fact that the number of the extremes is too small to meet the needs. For example, only 10 extremes are available in a 10-yr data series. Evidently, for stations or regions that have shorter periods of records the errors of calculated parameters would definitely be greater, resulting in credibility that is by no means idealized. In fact, it is often against the reality to take annual extremes from a large size as examples. Take the rainfall of a particular station, for example, whose climate may differ from others in that more than one daily rainfall maximum (beyond a critical value) may occur if the station is in a rainy area or year such that not only does annual maximum rainfall exceed a set critical value as an event of intense precipitation but the following one or two extremes may remain intense as well, which are often observed in many places. By contrast, in arid or semi-arid zones or in rainless years, even though the annual maximum may hardly reach the level of heavy rains (actually of moderate rainfall), it is obviously irrational to take one extreme in the year. Furthermore, sampling just one maximum one year will lose much useful information or involve useless information, a major shortcoming that is inherent in GEV and Gumbel distributions. In contrast, GPD has its most prominent merit just in that it extracts all maxima above a given level directly from a set of primitive data (as in many years), known as POT sampling method can lead to bigger sample size of extremes (as maxima) above a given threshold. Hence, the length of years is cut down a lot, but the size of extremes is increased, thus overcoming the demerit of sampling based on BM or AM for GEV and Gumbel distribution. This provides as full meaningful information for extreme value research as possible. As a result, the increased size of samples allows us to have a bigger size of effective samples even for stations and regions that keep shorter-length records, which would greatly improve fitting precision and augment the robustness of calculated parameters. In particular, GPD distribution is featured by its probability better described for the high-end thick tail part compared to GEV and Gumbel distributions.

6.2 *Comparison of fitting accuracies of the models*

The present study shows that GPD fittings have the highest accuracy so that it is most suitable for the study of rainfall extremes in a rainy season, thus superior to GEV and Gumbel models, the latter of which fails to completely match empirical distribution but provides some information on the quantile calculated for the high-end for reference. The GPD is generally subject to no effect of a size of samples because the volume of POT sampling extremes is larger than that of AM or BM for GEV and Gumbel so that it is especially suitable for stations having shorter length of samples.

6.3 *Conclusions*

(1) The expressions proposed for the high end quantile in a given return period using GPD for GEV and vice versa to further improve their theoretical inter-relationship and it is thus of higher utility. For instance, the use of GPD fitting and its relation to GEV parameters and quantile permits us to calculate those of GEV without the need to calculate them specifically.

(2) Using the method for approximately calculating a quantile presented in this paper we can approximately find the GEV quantile at a given return period with the aid of GPD parameters and vice versa. Experiments show that at Poisson exceedance $\lambda \rightarrow 1$ and in a large return period T (say, 50 or 100 years) the quantile to be calculated can be of rather high accuracy.

(3) From Fig. 3 we see that the mean rainfall varies vastly for the Guangzhou, Nanjing and Beijing areas and this may be directly associated with the set optimal threshold, an unexpected discovery of a new approach that helps reveal the relationship between the distribution of parameters of an original variable and GPD parameters. This problem will be dealt with in a separate paper.

Acknowledgements. This work is supported jointly by the National Natural Science Foundation of China (Grant No. 40675043) and Program of the Jiangsu Key Laboratory of Meteorological Disaster (Grant No. KLME050209).

REFERENCES

Abild, J., E. Y. Andersen, and D. Rosbjerg, 1992: The climate of extreme winds at the Great Belt, Denmark. *J. Wind Eng. Indus. Aerodynamic.*, 41–44, 521–532.

Ashkar, F., and J. Rousselle, 1987: Partial duration series modeling under the assumption of a Poissonian flood count. *J. Hydrol.*, **90**, 135–144.

Christensen, J. H., and O. B. Christensen, 2003: Severe summertime flooding in Europe. *Nature*, **421**, 805–806.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values.* Springer Verlag, London, UK., 205pp.

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

Cunnane, C., 1979: A note on the Poisson assumption in partial duration series models. *Water Resour. Res.*, **15**, 489–494.

Davison, A. C., and R. L. Smith, 1990: Models for exceedances over high thresholds. *J. Royal Statistical Society*, B, **52**, 393–442.

Ding, Y. H., and Y. Sun, 2006: Recent advances in climate change science. *Advances in Climate Change Research*, **2**(4), 161–167. (in Chinese)

Ding, Y. G., J. Liu, and Y. Zhang, 2004: Probability-weighed estimate-based modeling experiment with space/temporal distribution of extreme temperatures in China. *Chinese J. Atmos. Sci.*, **23**(2), 771–782. (in Chinese)

Ding, Y. H., J. Zhang, and Y. Song, 2006: Changes in weather and climate extreme events and their association with global warming. *Meteorological Monthly*, **28**, 3–7. (in Chinese)

Easterling, D. R., L. Evans, P. Y. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje，2000: Observed variability and trends in extreme climate events: A brief review. *Bull. Amer. Meteor. Soc.*, **81**(3), 417–425.

Fisher, R. A., and L. H. C. Tippett, 1928: Limiting forms of frequency distribution of the largest or smallest member of a sample. *Proceeding of Cambridge Philosophic Sociecy*, **24**, 180–190.

Gumbel, E. J., 1958: *Statistics of Extremes.* Columbia University Press, New York, 375pp.

Guttman, N. B., J. R. M. Hosking, and R. Wallis，1993: Regional precipitation quantile values for the continental United States computed from L moments. *J. Climate*, **6**, 2326–2340.

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

Hosking, J. R. M., 1990: L-moments analysis and estimation of distributions using combinations of order statistics. *Journal of the Royal Statical Society* (*B*),

52(1), 105–124.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood, 1985: Estimation of the generalized Extreme Value distribution by the method of Probability-Weighted Moments. *Technometries*, **27**, 251–261.

Huang, J. Y., 2000: The response of climatic jump in summer in North China to global warming. *Adv. Atmos. Sci.*, **17**(2), 184–192.

IPCC, 2001: *Climate Change in the Scientific Basis.* Cambridge Univercity Press, 881pp.

Jenkinson, A. F., 1955: The frequency distribution of the annual maximum (minimum) values of meteorological elements. *Quart. J. Roy Meteor. Soc.*, **18**, 158–171.

Jones, R. G., J. M. Murphy, M. Noguer, and A. B. Keen, 1997: Simulation of climate change over Europe using a nested regional climate model. II: Comparison of driving and regional model responses to a doubling of carbon dioxide. *Quart. J. Royal Meteor. Soc.*, **123**, 265–292.

Katz, R. W., G. S. Brush, and M. B. Parlange, 2005: Statistics of extremes: Modeling ecological disturbance. *Ecology*, **86**(5), 1124–1134.

Karl, T. R., and D. R. Easterling, 1999: Climate extremes: Selected review and future research directions. *Climatic Change*, **42**, 309–325.

Meehl, G. A., and Coauthors 2000: An introduction to trends in extreme weather and climate events: Observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. *Bull. Amer. Meteor. Soc.*, **81**(3), 413–416.

Pickands, J., 1975: Statistical inference using extreme order statistics. *Annual Statistics*, **3**, 119–131.

Wang, S. W., J. Cai, J. Zhu, and D. Gong, 2002: The interdecadal variations of annual precipitation in China. *Acta Meteorologica Sinica*, **60**(5), 637–640.

Wang, Q. J., 1991: The POT Model Described by the Generalized Pareto Distribution with Poisson Arrival Rate. *J. Hydrol.*, **129**, 263–280.

Yao, Z. S., and Y. G. Ding, 1990: *Climate Statistics.* China Meteorological Press, Beijing, 954pp. (in Chinese)

Zhai, P. M., A. J. Sun, F. M. Ren, X. Liu, B. Gao, and Q. Zhang, 1999: Changes of extreme events in China. *Climatic Change*, **42**, 203–218.