Modelling of precipitation extremes using parametric and nonparametric methods with automated threshold selection

Jan Holešovský, Michal Fusek, and Jaroslav Michálek

***Abstract*—When dealing with extreme values estimation, the threshold models are often used. However, a proper threshold selection is one of the problems which have to be solved. In this paper, we concentrate on this issue in order to compare an automated threshold selection based on multiple-threshold model with double bootstrap technique based on semiparametric estimation. Both estimation procedures are compared using simulations. A case study is carried out to evaluate estimations of intensity-duration-frequency curves which represent commonly used hydrological tool. A special attention is also paid to the assessment of an impact of the series length on the estimation quality.**

***Keywords*—Extreme value, bootstrap, generalized Pareto distribution, moment estimator, peaks-over-threshold.**

I. Introduction

The interest in extreme hydrological events prediction has lately increased in Central Europe (see e.g. [1]). To ensure protection of people and property from the harmful effects of hydrological situations (see e.g. [2]), proper evaluations of the available data should be carried out. On that account, many authors have studied the impact of using modern statistical methods on evaluation of hydrological data (see e.g. [3], [4]). The most frequent statistical method used to evaluate rainfall data (especially for design and operation of urban drainage systems) is based on estimation of so-called intensity-duration-frequency (IDF) curves. The IDF curves describe a dependency between rainfall intensity and duration and give us information about return levels of a T-year event. The T-year return level z_T is a value which is exceeded on average once every T years. In some cases, it is more convenient to estimate the R-observation return level z_R which is a value which is exceeded on average once every R observations. In order to estimate return levels of hydrological events, extreme value theory is often applied.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (iid) random variables and let M_n be a transformed random variable defined as $M_n = \max\{X_1, \ldots, X_n\}$. Fisher and Tippett [5] showed that, given suitable sequences $a_n > 0$ and b_n, the only one nondegenerate limit distribution of properly normalized random variables $(M_n - b_n)/a_n$ arises in the form of generalized extreme value (GEV) distribution with cumulative distribution function (cdf)

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

where $\mu \in R$, $\sigma > 0$ and $\xi \in R$ are location, scale and shape parameters respectively, and $x_+ = \max(0, x)$. The shape parameter, also referred to as the extreme value index (EVI), plays a crucial role in relation to tail properties of the distribution, and needs to be properly estimated.

In practice, mostly when working with observations over time, a significant dependence structure is often present [6]. Since the estimation procedures, for example the commonly used maximum likelihood (ML) method, are based on iid observations, it is usually necessary to draw samples from a series that can be considered independent. As described, for example, in [7], two methods of drawing samples which can be considered independent are distinguished.

When the block maxima method is used, the sequence is divided into blocks of a given size, and the maxima from all the blocks form a new sequence with independent members (see e.g. [8]–[10]). As blocks get larger, the distribution of block maxima approaches the GEV distribution (see e.g. [9], [11]) with cdf (1). A disadvantage of the method is usually an arbitrary choice of block size. When blocks are large, a small sample size is obtained which usually leads to wide confidence intervals for parametric functions' estimates. In case of small blocks, dependence between the maxima still may be preserved. It was shown that it is reasonable to select a maximum value of the random variable in each year of the measured period (see e.g. [7]). The samples are then called annual maxima series. However, modelling of extremes using block maxima is often unsuitable for practical purposes, because it leads to an extensive reduction of information contained in data, since other extreme values besides the maxima in given blocks are omitted. This is problematic especially if only short time series are available. Therefore, the extreme value analysis is often based on threshold exceedances.

When the peaks-over-threshold (POT) approach is used, only values above a specified threshold are considered (see e.g. [9], [11]). The dependence in the series is removed by

This work was supported by the specific research project No. FCH/FSI-J-14-2439 at Brno University of Technology.

J. Holešovský is with the Institute of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896/2, Brno 61669, Czech Republic (corresponding author to provide phone: 00420-541-142-726; e-mail: honza.holesovsky@gmail.com).

M. Fusek is with the Department of Mathematics, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 2848/8, Brno 61600, Czech Republic (e-mail: fusekmi@feec.vutbr.cz).

J. Michálek is with the Department of Econometrics, Faculty of Military Leadership, University of Defence, Šumavská 4, Brno 66210, Czech Republic (e-mail: michalek@fme.vutbr.cz).

separating the series into independent events (see e.g. [8] for more details) from which only peaks are further considered.

Pickands [12] showed that given sequences u_n of thresholds increasing with n, the limiting distribution of threshold exceedances of a random variable X is the generalized Pareto (GP) distribution. In practice, a high enough threshold u is selected and kept fixed. The variable $Y = X - u$ conditioned by $X > u$ is modelled by the GP distribution with cdf

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)_+^{-1/\xi}, \tag{2}$$

where $\xi \neq 0$ and $\sigma_u > 0$ are shape and scale parameters, and $y = x - u > 0$. In correspondence to GEV distribution (1), the relation $\sigma_u = \sigma + \xi(u - \mu)$ holds (see e.g. [9], [13] for more details).

Although the GP-analysis allows us to make use of more observations than the GEV-analysis, selecting a proper threshold can be problematic. A threshold too low leads to desired smaller uncertainty in parameters' estimates, however, it provides a possibly insufficient approximation by the limiting GP distribution, and vice versa. Moreover, a suitable scheme for drawing independent samples from the original time series must also be determined. However, such a discussion is not a part of this paper, and further on we follow commonly used methods which will be briefly introduced in Sect. II.

A comprehensive review of possible threshold selection procedures was provided by Scarott and MacDonald [14], including several nowadays obsolete or rarely used methods. In practice, approaches combining some of the graphical methods with experiences in a specific scientific fields are often applied. It is obvious that such a complex, laborious and often subjective approach can be difficult to use when dealing with extensive data files. Therefore, one important question arises. Is it possible to identify the proper threshold automatically? Here we may notice the method lately proposed by Wadsworth and Tawn [15] being a one remarkable result in the field of automated threshold selection. However, the suggested approach simultaneously chooses both threshold and also the so-called runs declustering parameter in order to draw an independent sample from a given time series. Afterwards, Northrop and Coleman [16] introduced a multiple-threshold penultimate model based on piecewise constant shape parameter approximation. This modified approach which concentrates only on the proper threshold identification also reduced the computational complexity of the approach described in [15].

The purpose of this paper is to compare the method proposed by Northrop and Coleman [16] with a bootstrap-based methodology developed for an optimal sample fraction estimation [17], [18]. The properties of both approaches will be studied using simulations and the results will be used for estimation of the IDF curves which play an important role in hydrological risk assessment. The extreme value theory in combination with bootstrap is rather common and can also be found in financial applications (see e.g. [19]).

Unknown parameters of the probability distributions will be estimated using parametric and nonparametric methods. Specifically, the ML method and the method of moments which is often applied in hydrological applications will be used. The ML method is rather commonly used in many application areas, e.g. in biometric models of signal transduction process [20] or in measuring service quality [21]. In this paper, the ML theory will be implemented in a similar way. However, using the ML theory can also bring difficulties, because in order to obtain ML estimates, it is often necessary to find a solution of generally rather complicated nonlinear equations. On that account, many authors focus on numerically less demanding methods like method of moments or L-moments (see e.g. [22]) when looking for estimates of unknown parameters of particular distributions. Instead of moment estimator, Hill's estimator [23] can also be used. However, it suffers from tendency to provide biased results when only small sample is analyzed. Nevertheless, such difficulty can be overcome by application of wavelet analysis and kernel estimate of the tail distribution [24].

II. Sample Drawing and Proper Sample Fraction Identification

Continuous rainfall series with the time step of 1 minute from stations located in the South Moravian Region, Czech Republic, were used for the analysis. Since the observations cannot be considered independent, a preprocessing needs to be applied before carrying on with the analysis. The separating procedure involves identification of the independent rainfall events from which only maxima of mean intensities over different rainfall durations (5, 10, 15, 20, 30, 45, 60, 90, 120, 180, 240 and 360 minutes) are considered. For the purpose of this study, the methodology of Madsen et al. [25] was applied, and the events were separated by ceasing the rainfall for a period equal to the considered duration but at least one hour.

Since some of the stations were established only in the last decades, lengths of the available series vary between 11 and 41 years of records. We aim to study the impact of the chosen threshold estimation methodology on the IDF curves estimates with respect to the series length which can significantly influence the estimates [4], [8].

A. The Multiple-Threshold GP Model

The traditional diagnostic tool for a threshold identification consists in graphical inspection of parameter estimates and their stability. Should u_0 be a proper threshold to be selected, then parameter estimates (changing with threshold) $\widehat{\xi}(u)$ and $\widehat{\sigma}(u)$ would follow a constant and a linear trend for $u > u_0$ respectively [9]. Northrop and Coleman [16] proposed an automated method based on the property of the shape parameter. In principle, they proposed a discretized version of testing the null hypothesis $H : \xi(u) = \xi(u_0)$, for all $u \geq u_0$.

Let $u_1 < u_2 < \cdots < u_m$ denote an increasing threshold sequence, Y be an excess of u_1, and denote $v_i = u_i - u_1$, for $i = 1, \ldots, m$. The shape parameter is modelled as a piecewise constant function $\xi(y)$ with change-points at $v_i, i = 2, \ldots, m$, i.e.

$$\xi(y) = \begin{cases} \xi_i & u_i < y < u_{i+1}, i = 1, \ldots, m-1, \\ \xi_m & y > u_m. \end{cases} \tag{3}$$

In order to avoid discontinuity in density of Y, the scale parameter is considered piecewise linear with σ_1 on interval (u_1, u_2) (see [16] for more details). Considering the threshold u_1 first, the hypothesis whether a common GP model holds on all intervals $(u_i, u_{i+1}), i = 1, \ldots, m-1$, is tested. Rejection of the null hypothesis $\mathrm{H} : \xi_1 = \cdots = \xi_m$ attests to requirement of a higher threshold. Let $\widehat{\theta}_0$ denote the sequence of ML estimates of shape ξ_1 and scale σ_1 parameters estimated under the null hypothesis, and $\widehat{\theta}$ the sequence of ML estimates of σ_1 and $\xi_i, i = 1, \ldots, m$, estimated without additional constrains given by (3). Northrop and Coleman proposed two tests - score test and likelihood ratio test - which are based on the following test statistics

$$S = U^T(\widehat{\theta}_0)J^{-1}(\widehat{\theta}_0)U(\widehat{\theta}_0), \tag{4}$$

$$LR = 2\left[l(\widehat{\theta}) - l(\widehat{\theta}_0)\right], \tag{5}$$

where $U(\theta)$ is the score function, $J(\theta)$ is the expected Fisher information matrix, and $l(\theta)$ is the log-likelihood function. More details about derivation of the score function, the expected Fisher information matrix, and the log-likelihood function can be found in [16]. Provided that $\xi_m > -1/2$ [26] in each case the asymptotic null distribution of the statistics (4) and (5) is χ^2_{m-1}. The advantage of the test statistic (4) is that it requires only a fit of the null model at the threshold of interest. The calculation of the test statistic (5) is more difficult and time-consuming as it requires m shape parameters to be estimated.

Let us assume now that the lowest threshold considered is u_k. The null hypothesis to be tested is $\mathrm{H} : \xi_k = \cdots = \xi_m$ and the asymptotic null distribution is χ^2_{m-k}, $k = 1, \ldots, m-1$. The p-values associated with the test indicates whether a threshold higher than u_k is required.

Once a proper value u_0 is selected, the T-year return level z_T can be estimated from (2) as the $1 - 1/(n_x T)$ quantile of the GP distribution, where n_x denotes the number of observations per year, i.e.

$$\widehat{z}_T = u_0 + \frac{\widehat{\sigma}_u}{\widehat{\xi}}\left[\left(\widehat{\lambda}_u T n_x\right)^{\widehat{\xi}} - 1\right]. \tag{6}$$

The estimator $\widehat{\lambda}_u := P(X > u)$ can be calculated as a proportion of values exceeding the threshold u_0, i.e. $\widehat{\lambda}_u = n_u/n$, where n_u denotes the number of observations above the threshold u_0 and n is the sample size (the length of the rain series). The rest of the unknown parameters in (6) were replaced by their ML estimates. The asymptotic confidence intervals for the T-year return level $\widehat{z}_T$ can be obtained by delta method using the asymptotic normality of the ML estimates [9].

In case the R-observation return level has to be estimated, the quantity Tn_x in equation (6) is replaced by R.

B. Bootstrap-Based Optimal Sample Fraction Selection

A different approach to optimal sample fraction identification was studied by Draisma et al. [18]. The method is based on k largest order statistics which should in some sense balance the approximation by the GP distribution and properties of the EVI. Here, we focus on moment estimator (MoE) $\widehat{\xi}_M$ defined as

$$\widehat{\xi}_M(k) = M_n^{(1)}(k) + 1 - \frac{1}{2}\left[1 - \frac{\left(M_n^{(1)}(k)\right)^2}{M_n^{(2)}(k)}\right]^{-1}, \tag{7}$$

where

$$M_n^{(j)}(k) = \frac{1}{k}\sum_{i=0}^{k-1}\left(\log X_{(n-i)} - \log X_{(n-k)}\right)^j \tag{8}$$

is the j-th moment of the GP distribution, $k = 2, \ldots, n-1$, and $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ denote the order statistics. The choice of a proper k is accompanied by difficulties very similar to the choice of an optimal threshold in the POT analysis. When a high value of k is selected, we may expect higher estimation precision but a weak approximation by the GP distribution which may lead to a significant bias, and vice versa. The optimal value k_0 balancing the bias and the variance can be determined only if the underlying distribution is known. As shown for example in [13], the limiting distribution of $\widehat{\xi}_M$ depends on a second-order parameter which is problematic to estimate. Therefore, the double bootstrap methodology proposed in [18] and developed in [17], [27] is used.

Hereby, the optimal value k_0 is chosen to minimize the mean square error (MSE) of the MoE, i.e.

$$k_0 \in \operatorname{argmin}_k \mathrm{E}\left(\widehat{\xi}_M(k) - \xi\right)^2, \tag{9}$$

although only in the asymptotic sense. The unknown EVI is replaced by an auxiliary estimator $\widehat{\xi}_{AUX}(k)$ calculated as the third-moment estimator (see [18], page 5).

The bootstrap methodology is used to resample B times independently a sample $\mathbf{X}^* = (X_1^*, \ldots, X_m^*)$ of size $m < n$ from the underlying series $\mathbf{X} = (X_1, \ldots, X_n)$ of separated (independent) rainfall observations (see Sect. II). For each replication and for all $k = 2, \ldots, m$ the estimators $\widehat{\xi}^*_M(k)$ and $\widehat{\xi}^*_{AUX}(k)$ are evaluated. Finally, the MSE is replaced by its estimator

$$\mathrm{MSE}^*(n,k) = \frac{1}{B}\sum_{b=1}^{B}\left(\widehat{\xi}^*_M(k) - \widehat{\xi}^*_{AUX}(k)\right)^2, \tag{10}$$

$k = 2, \ldots, m$, and minimized with respect to k. Denote the optimal value k which minimizes (10) by $k_0^*(m)$. The key step is to apply the bootstrap methodology twice: Firstly, for $m := n_1$; secondly, for $m := n_2$. As derived in [17] or [18], we avoid estimation of the second-order parameter by setting $n_2 = \lceil (n_1)^2/n \rceil$. The optimal sample fraction k_0^* is then obtained using formula

$$\widehat{k}_0^* = \frac{(k_0^*(n_1))^2}{k_0^*(n_2)}. \tag{11}$$

From the theoretical point of view, n_1 should be smaller than n; however, simulations show better performance if n_1 is set as high as possible [17]. In order to achieve better estimation stability, the whole double bootstrap procedure is applied repeatedly (N times) leading to N estimates $\widehat{k}_{0,i}, i = 1, \ldots, N$, of optimal sample fraction (11). The estimator $\widehat{k}_0$

of k_0 is computed as median from all the bootstrapped values of $\widehat{k}_{0,i}$.

Once the optimal value $\widehat{k}_0$ is determined, the shape parameter is estimated by the MoE (7), and the scale parameter is estimated using formula (see Theorem 4.3.3 in [13])

$$\widehat{\sigma}_M = X_{(n-k)} M_n^{(1)} \left(1 - \widehat{\xi}_M + M_n^{(1)}\right). \tag{12}$$

Standard deviations of the estimated parameters are estimated on the basis of asymptotic normality [13]. The desired return level estimates are obtained from formula (6) where parameter estimates are replaced by the appropriate moment estimators and the threshold is replaced by $X_{(n-\widehat{k}_0)}$.

III. Simulation Study

In this section, properties of the methods for an optimal sample fraction identification, which were described in the previous section, will be studied and compared using simulations. The attention will be paid to a comparison of the double bootstrap approach with the multiple-threshold GP (MT-GP) model based only on the score statistic (4). As already mentioned, use of the statistic (5) requires the shape parameter to be estimated on all threshold subintervals which leads to enormous computational times. For this very reason, the methodology based on the test statistic (5) is applied only in the case study in the following section. Some restrictions had to be also adopted because of the computational time demands of the bootstrap-based methodology.

Input parameters for the simulations will be based on real data set of the studied rainfall series. Specifically, attention will be paid to rainfall series measured at two stations located in the highly urbanized area of Brno, the second largest city in the Czech Republic. The rainfall series from the Tuřany station contains 41 years of records and it is the longest series available. The rainfall series from the Jundrov station consists of only 11 years of records and it is the shortest series available. Independent samples were drawn from the rainfall series according to methodology described in [25] and their sizes vary between 1836 and 10791 for the Tuřany station, and between 618 and 3228 for the Jundrov station considering all rainfall durations. Therefore, in order to study the impact of the series length on the estimation quality, simulations will be carried out considering sample sizes in the range from 500 to 8000.

The underlying distribution was estimated from the independent samples and among other distributions with positive support, 4 distributions were shown to be the most appropriate for description of the rainfall data, specifically, log-normal, log-logistic, Weibull, and gamma distribution (see Fig. 1). In particular, Weibull distribution is often used in practice because of its flexibility (see e.g. [28], where Weibull distribution is fitted to the monthly wind speed frequency). Suitability of the probability distributions was tested using the goodness-of-fit tests (Pearson's χ^2 test, Kolmogorov-Smirnov test, and Anderson-Darling test).

In the simulation study, $M = 1000$ random samples of sizes 500, 1000, 2000, 3000, 5000, and 8000 from the given distributions were generated. First of all, an optimal

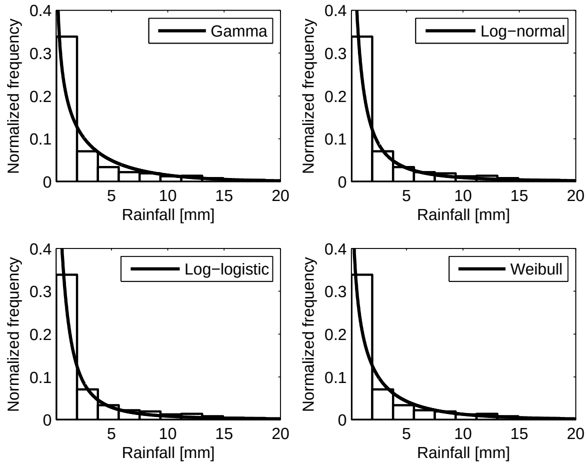


Fig. 1. Histograms of rainfall intensities at Jundrov station (180-minute rainfall duration) fitted by probability density function of the considered distributions. The estimated shape and scale parameters are 0.57 and 5.62 for gamma distribution, 0.09 and 1.56 for log-normal distribution, 0.08 and 0.94 for log-logistic distribution, and 2.40 and 0.68 for Weibull distribution.

threshold was selected by means of the MT-GP model using the test criterion (4). In each case, the optimal threshold was selected as the smallest value at which the null hypothesis is not rejected. Next, the double bootstrap methodology was applied with $B = 250$ sample replications. Such number of replications should be suitable for this methodology [29]. Since this method is computationally very demanding, the whole estimation procedure was repeated $N = 100$ times.

The threshold selection in the MT-GP model strongly depends on the level of discretization m. On one hand, as discussed in [16], a large value of m should be preferred taking into consideration a set of thresholds that approximates the range of possible thresholds. On the other hand, a balance must be found between the level of discretization and the variability of the estimated shape parameter in each interval (u_i, u_{i+1}), $i = 1, \ldots, m-1$. One would like to set m large in order to achieve a dense coverage of all possible thresholds. However, the large m leads to an increased uncertainty in parameter estimates and differences between the estimates are detected with a smaller probability. Moreover, if m is large, there can also arise convergence problems of the ML method. A more complex study could be carried out on this topic, however, here we follow Northrop and Coleman [16] and set $m = 10, 20, 40$ and possible threshold values between 0.5% and 90% sample quantiles.

In Fig. 2, there are boxplots visualizing the proportion of n_u (X, Y, and Z columns) and k_0 (B columns) to the particular sample sizes considering 4 above mentioned distributions. It can be seen that medians of k_0, i.e. optimal order statistics estimated using the double bootstrap methodology, are smaller than medians of n_u, i.e. estimated by the MT-GP model, in all the cases. An instability of the MT-GP methodology is visible for intermediate sample sizes of log-normal and large sample sizes of log-logistic distribution. In these cases the p-values of test criterion (4) fluctuate near the significance level of

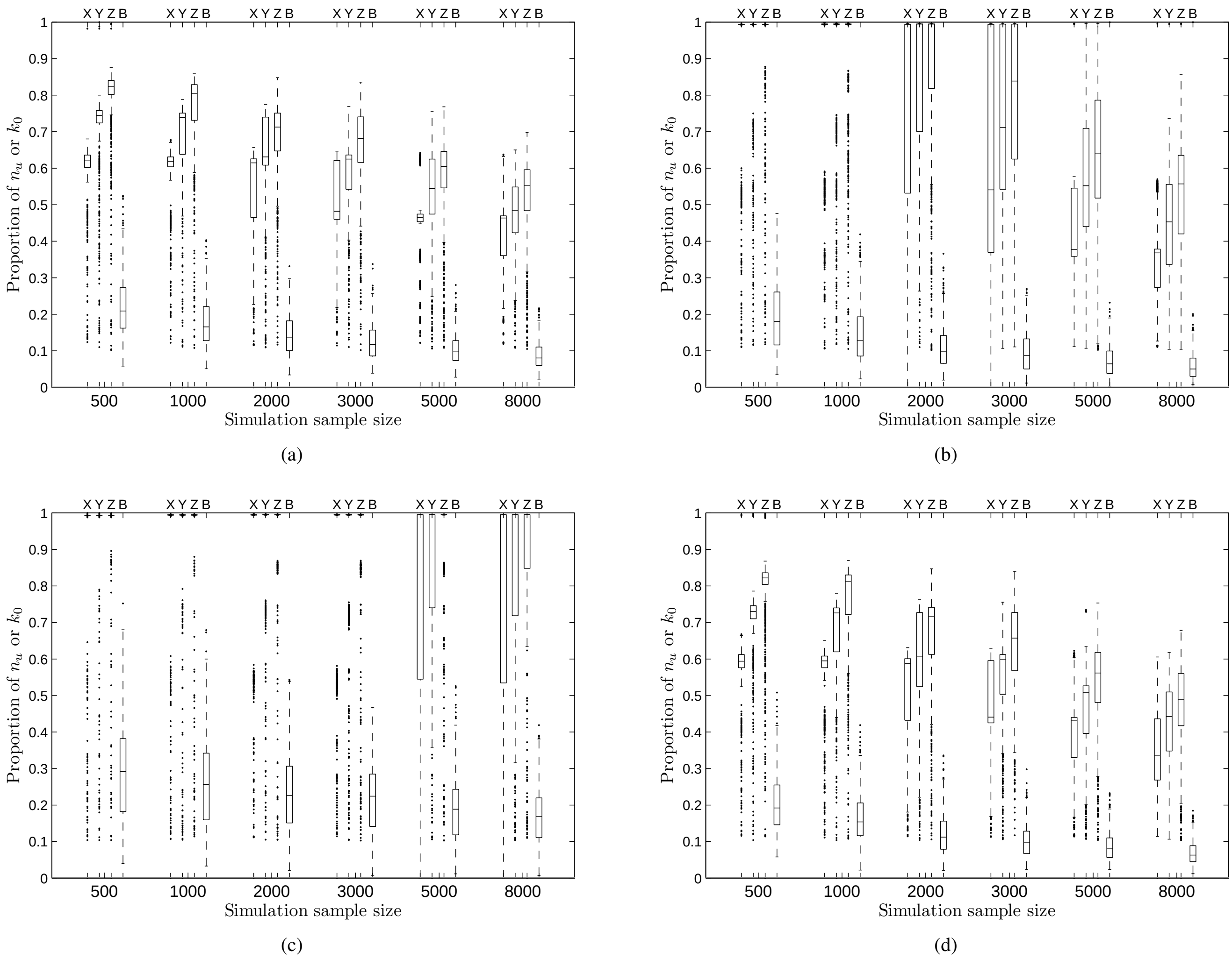


Fig. 2. Boxplots visualizing the proportions of n_u and k_0 to the particular sample sizes ($M = 1000$ repetitions); the MT-GP model with $m = 10$ (X), $m = 20$ (Y), $m = 40$ (Z), and double bootstrap methodology (B). Samples are drawn from (a) gamma, (b) log-normal, (c) log-logistic, and (d) Weibull distribution.

0.05 with unclear boundary. In case of small and intermediate sample sizes generated from the log-logistic distribution the proportions of observations n_u above the proper threshold u_0 remain stable near 1. Since the estimated shape parameter of the distribution is close to 1, in which case the log-logistic distribution coincides with the GP distribution, this behaviour was highly expected. In Fig. 2, the dependence of an optimal threshold identification on the level of discretization m is clearly visible. Specifically, the proportion of observations n_u above the proper threshold u_0 grows higher with the higher level of discretization m.

Moreover, a hybrid distribution with a positive support used by Northrop and Coleman [16] was also considered in the simulation study. This distribution's probability density function consists of a constant density up to 75% quantile and the GP density with shape parameter 0.1 from the 75% quantile upwards (not shown). Theoretical aspects imply the null hypothesis tested by the MT-GP model is true for the excesses of the 75% quantile. In this specific case, the double bootstrap methodology proved to act completely in opposite to the formerly obtained results, and the proportion of the k_0 largest order statistics approaches 1 as the sample size increases.

IV. Application Results

In this section, a behavior of the threshold selection procedures will be illustrated on Nidd River flow rates data. Moreover, attention will be also paid to estimation of the IDF curves considering rainfall series of various lengths as discussed in Sect. II.

A. Nidd River Flow Rates Example

This case study was previously closely investigated by Wadsworth and Tawn [15] and also by Northrop and Coleman [16]. Here we show the difference between the MT-GP and the double bootstrap approaches. In order to be able to compare the results with previous studies, the set of flow rates measured at Nidd River (Yorkshire, UK) [16] over the period 1934–1969 consisting of 154 observations exceeding $65\,\mathrm{m}^3\mathrm{s}^{-1}$ will be used. Since the data has already been preprocessed, the sample can be treated as independent. Fig. 3 shows shape and scale parameter stability plots with 95% pointwise confidence intervals. Considering the graphical methods for a threshold selection described in [9], the optimal threshold about $u_0 = 100\,\mathrm{m}^3\mathrm{s}^{-1}$ seems reasonable. It can be seen that the scale parameter roughly linearly depends on the threshold value and the shape parameter remains roughly constant above the optimal threshold u_0.

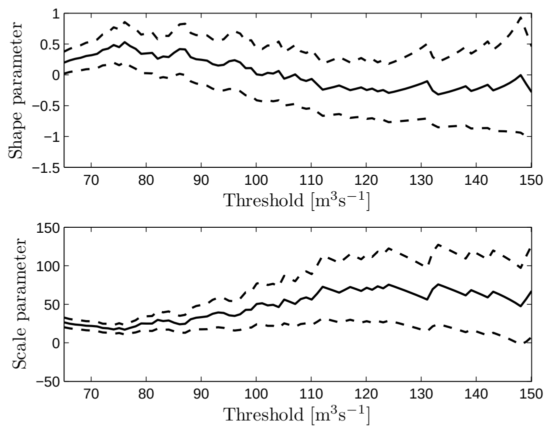


Fig. 3. The ML estimates of shape and scale parameters of the GP distribution considering various threshold values with 95% confidence intervals (dashed line).

Northrop and Coleman suggested the MT-GP model with $m = 10$ and thresholds between 65 and $120\,\mathrm{m}^3\mathrm{s}^{-1}$ in equidistant steps of $5\,\mathrm{m}^3\mathrm{s}^{-1}$. Both test criteria (4) and (5) indicate the threshold equal to $70\,\mathrm{m}^3\mathrm{s}^{-1}$ to be sufficient with $n_u = 138$. Wadsworth and Tawn, whose method includes tests based on goodness-of-fit of the GP distribution across various thresholds, suggested the threshold equal to $75\,\mathrm{m}^3\mathrm{s}^{-1}$. When the double bootstrap methodology introduced in Sect. II-B is used ($B = 250$ sample replications, $N = 1000$ repetitions) for estimation of a proper k_0, in contrast to the foregoing results, the number of optimal largest order statistics $\widehat{k}_0 = 149$ is estimated. This value corresponds to the threshold equal to about $67\,\mathrm{m}^3\mathrm{s}^{-1}$.

In Fig. 4, there are the estimated R-observation return levels. The heavy solid line visualizes the ML estimates obtained by use of the threshold model with $u = 70\,\mathrm{m}^3\mathrm{s}^{-1}$, the regular solid line denotes return levels estimated on the basis of moment estimators of the parameters with $\widehat{k}_0 = 149$. It can be seen that both estimates are rather similar. It is also clear that the threshold-based estimates lie in the 95% confidence bounds for the bootstrap-based curve. However, this is not the case if $u \geq 75\,\mathrm{m}^3\mathrm{s}^{-1}$. Moreover, if $u = 70\,\mathrm{m}^3\mathrm{s}^{-1}$ the confidence bounds are much narrower in case of the bootstrap-based curve.

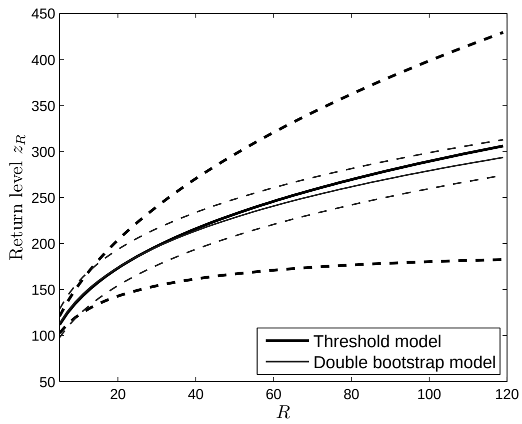


Fig. 4. R-observation return levels of Nidd River flow rates estimated using the threshold model with $u = 70\ \mathrm{m}^3\mathrm{s}^{-1}$ and the double bootstrap methodology with $k_0 = 149$; 95% confidence intervals are denoted by dashed lines.

B. IDF Curves Estimation

In the following case study, methods and results from the previous sections will be applied on real data set and used for estimation of the IDF curves. Specifically, results for stations with the longest (Tuřany, 41 years) and the shortest (Jundrov, 11 years) rainfall series will be presented.

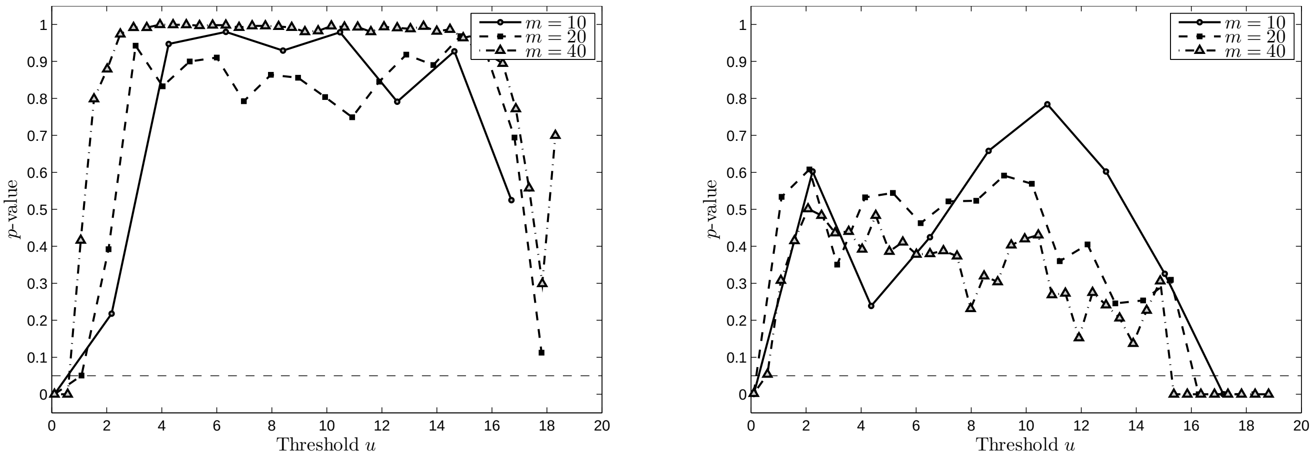
As discussed in Sect. III, the threshold identification in the MT-GP model may depend on the level of discretization, i.e. the value of m. In our case, we choose a defensive approach regarding several proposals in [16], and we set $m = 10, 20, 40$, and thresholds between 10% and 99.5% sample quantile for all stations and rainfall durations. For all values of m, the value $u_0(m)$ is selected as the lowest threshold that ensures the validity of the null hypothesis at the significance level of 0.05. Then the optimal threshold is chosen as $u_0 = \max_m(u_0(m))$. In Fig. 5, typical results obtained from testing the null hypothesis using the score test statistic at various thresholds are visualized, specifically, the p-values obtained from the limiting χ^2 distribution are plotted against the considered thresholds.

TABLE I
TUŘANY STATION: NUMBER OF OBSERVATIONS OBTAINED BY THE DOUBLE BOOTSTRAP ($\widehat{k}_0$) AND THE MT-GP MODEL WITH THE SCORE (n_u^S) AND THE LIKELIHOOD RATIO STATISTICS (n_u^{LR}).

Duration [min]	5	10	15	20	30	45	60	90	120	180	240	360
k_0	174	186	243	215	209	111	111	225	245	186	221	278
n_u^S	423	540	418	556	725	776	804	807	777	740	725	689
n_u^{LR}	487	471	499	498	479	496	507	535	516	500	481	482
k_0/n_u^S	0.41	0.34	0.58	0.39	0.29	0.14	0.14	0.29	0.32	0.25	0.30	0.40
k_0/n_u^{LR}	0.36	0.39	0.49	0.43	0.44	0.22	0.22	0.42	0.47	0.37	0.46	0.58

TABLE II
JUNDROV STATION: NUMBER OF OBSERVATIONS OBTAINED BY THE DOUBLE BOOTSTRAP ($\widehat{k}_0$) AND THE MT-GP MODEL WITH THE SCORE (n_u^S) AND THE LIKELIHOOD RATIO STATISTICS (n_u^{LR}).

Duration [min]	**5**	**10**	**15**	**20**	**30**	**45**	**60**	**90**	**120**	**180**	**240**	**360**
k_0	56	41	44	50	69	70	83	105	90	103	115	120
n_u^S	192	205	210	211	207	216	191	194	197	205	206	196
n_u^{LR}	157	147	150	150	145	149	147	146	140	141	134	131
k_0/n_u^S	0.29	0.20	0.21	0.24	0.33	0.32	0.43	0.54	0.46	0.50	0.56	0.61
k_0/n_u^{LR}	0.36	0.28	0.29	0.33	0.48	0.47	0.56	0.72	0.64	0.73	0.86	0.92



(a) (b)

Fig. 5. The calculated (using the score statistic) p-values at (a) Tuřany and (b) Jundrov station for various values of m. The horizontal dashed line denotes significance level of 0.05.

Further, ML estimates of parameters of the GP distribution were calculated for all determined thresholds. Their variances were established on the basis of asymptotic normality, and confidence intervals of return levels were calculated using the delta method [9].

Next, an optimal sample fraction using the double bootstrap methodology (Sect. II-B) was determined. We set $B = 250$ bootstrap replications and $N = 1000$ repetitions of the whole procedure. As it was mentioned before, it is convenient to set the value of n_1 as high as possible providing $n_1 < n$. In our study, we use $n_1 = \lceil n^{0.995} \rceil$ (see [17]). The optimal sample fractions estimated using the MT-GP model and the double bootstrap methodology are summarized in Tables I and II. Notice that use of the MT-GP model leads to lower 'thresholds' implicating larger sample fraction meant for the analysis.

Dependency of the parameter estimates on the selection of k largest order statistics is visualized in Fig. 6. On one hand, setting value k low provides better approximation of the limiting GP distribution. However, it can be seen that it is loaded with higher uncertainty. On the other hand, high k leads to a significant bias. The standard deviations of shape parameter estimates are presented in Table III. It can be seen that the moment estimate is mostly burdened with higher variability than those based on the MT-GP model. A similar behaviour was observed in case of the scale parameter.

On the basis of estimated parameters, return levels of 5, 10, 20, 50, and 100-year events were calculated. Similar results were obtained by both approaches in case of the Tuřany station (long series) and lower (< 30 minutes) rainfall durations. For higher rainfall durations, return levels estimated using the MT-GP approach overestimate the bootstrap-based estimates. In case of the Jundrov station (short series), the MT-GP approach based on the score statistic gives higher estimates than the bootstrap-based one for all rainfall durations. This mostly applies also if the likelihood ratio test is used. The bootstrap-based estimates provide narrower confidence intervals for the estimated IDF curves as documented in Fig. 7.

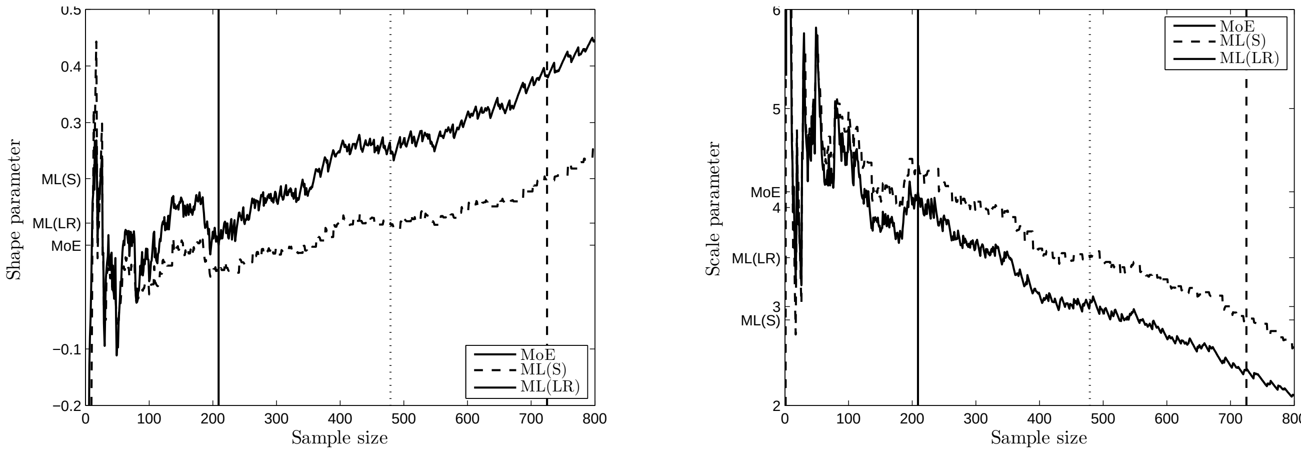
V. CONCLUSION

In this paper, two automated threshold selection procedures were introduced and compared using simulations. It was shown that use of the MT-GP model implicates larger sample fraction meant for the analysis in all the cases. However, in some cases, there is an instability in the optimal threshold selection procedure when using the MT-GP model. Moreover, the procedure based on the MT-GP model also crucially depends on the level of discretization m.

Both methods were used for real data analysis. First of all, Nidd River flow rates data were analyzed. In was shown that the return levels estimated using both methods are rather comparable. However, the confidence bounds are narrower when using double bootstrap methodology.

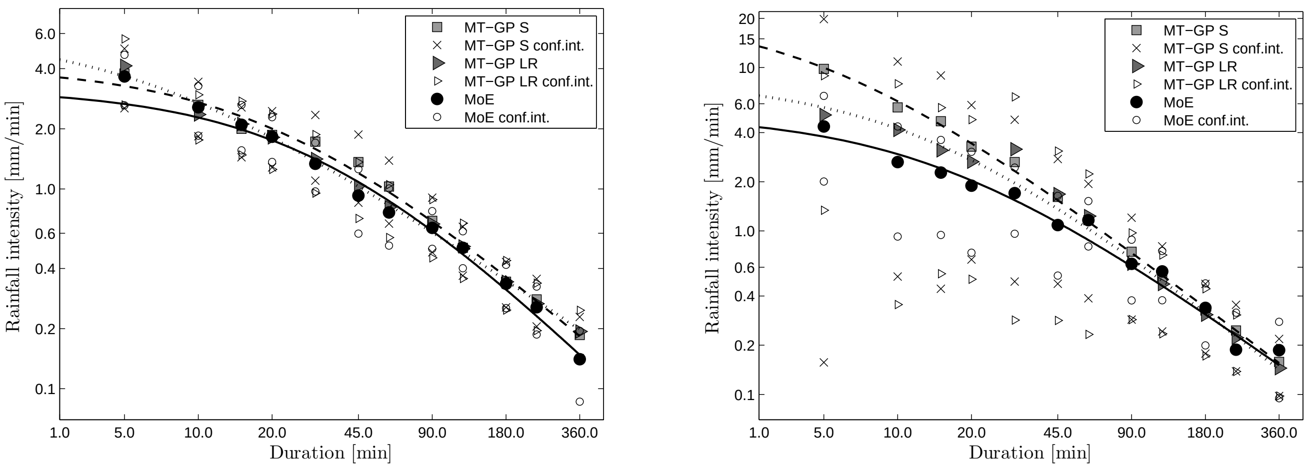
TABLE III
PROPORTIONS OF STANDARD DEVIATIONS OF SHAPE PARAMETERS ESTIMATED USING THE MOE AND THE MT-GP MODEL WITH THE SCORE (S) AND THE LIKELIHOOD RATIO (LR) TEST STATISTICS.

Duration [min]	**5**	**10**	**15**	**20**	**30**	**45**	**60**	**90**	**120**	**180**	**240**	**360**
Tuřany (MoE/MT-GP(S))	1.32	1.46	1.24	1.37	1.41	1.94	2.06	1.61	1.62	1.86	1.62	1.50
Tuřany (MoE/MT-GP(LR))	1.31	1.56	1.27	1.32	1.27	1.81	1.91	1.35	1.31	1.54	1.40	1.18
Jundrov (MoE/MT-GP(S))	1.15	1.47	1.45	1.47	1.26	1.36	1.25	1.12	1.31	1.30	1.18	1.17
Jundrov (MoE/MT-GP(LR))	1.15	1.18	1.24	1.20	0.90	1.07	0.94	1.00	1.15	1.15	1.11	1.10



(a) (b)

Fig. 6. Dependency of the moment (heavy curve) and the ML estimators (dashed curve) of (a) shape and (b) scale parameters on the sample fraction for 30-minutes rainfall at Tuřany station. The estimates obtained by the score and the likelihood ratio test statistics are denoted ML(S) and ML(LR), respectively. Vertical lines visualize the optimal values of k (solid) and n_u (dashed).



(a) (b)

Fig. 7. Estimated 100-year return levels at (a) Tuřany and (b) Jundrov stations with their 95 % confidence intervals; logarithmic scale on both axes. IDF curves were obtained using nonlinear regression: MT-GP(S) (dashed line), MT-GP(LR) (dotted line), and MoE (solid line).

Next, both methods were used for IDF curves estimation. It was shown that moment estimates of the GP distribution parameters are burdened with higher variability than those based on the MT-GP model. However, the bootstrap-based estimates provide narrower confidence intervals for the IDF curves. Readers can also get an idea about the width of the confidence intervals when dealing with estimation of long-time events in situations where only short rainfall series are available. Although the automation of proper sample fraction identifications through the double bootstrap technique is very comfortable, it is more demanding in terms of computing.

REFERENCES

[1] J. Daňhelka, J. Kubát, *Flash floods in the Czech Republic in June and July 2009.* Prague, Ministry of Environment of the Czech Republic, 2009 (in Czech).

[2] D. Butler, J. Davies, *Urban drainage.* 2nd edition, London, Spon press, 2004.

[3] H. Madsen, K. Arnbjerg-Nielsen, P.S. Mikkelsen, "Update of regional intensity-duration-frequency curves in Denmark: Tendency towards increased storm intensities,"' in *Atmos. Res.*, vol. 92, 2009, pp. 343–349.

[4] P. Willems, "Revision of urban drainage design rules after assessment of climate change impacts on precipitation extremes at Uccle, Belgium," in *J. Hydrol.*, vol. 496, 2013, pp. 166–177.

[5] R. A. Fisher, L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest members of a sample," in *Proc. of Cambridge Philosophical Society 24*, 1928, pp. 180–190.

[6] J. Holešovský, M. Fusek, J. Michálek, "Extreme value estimation for correlated observations," in *Proc. 20th MENDEL Conf.*, 2014, pp. 359–364.

[7] V. T. Chow, D. R. Maidment, L. W. Mays, *Applied hydrology.* 1st edition, New York, McGraw-Hill, 1988.

[8] A. Ben Zvi, "Rainfall intensity-duration-frequency curves relationships derived from large partial duration series," in *J. Hydrol.*, vol. 367, no. 1-2, 2009, pp. 104–114.

[9] S. G. Coles, *An Introduction to statistical modelling of extreme values.* London, Springer, 2001.

[10] H. Madsen, C. P. Pearson, D. Rosbjerg, "'Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events, 2. Regional modeling,"' in *Water Resour. Res.*, vol. 33, 1997, pp. 759–769.

[11] M. N. Khaliq, T. B. M. J. Ouarda, J.-C. Ondo, P. Gachon, B. Bobèe, "Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observation: A review," in *J. Hydrol.,* vol. 329, no. 3-4, 2006, pp. 534–552.

[12] J. Pickands "Statistical inference using extreme order statistics," in *Ann. Stat.,* vol 3, 1975, pp. 119–131.

[13] L. de Haan, A. Ferreira, *Extreme value theory: An introduction.* New York, Springer, 2006.

[14] C. Scarott, A. MacDonald, "A review of extreme value threshold estimation and uncertainty quantification," in *Statistical Journal*, vol. 10, no. 1, 2012, pp. 33-60.

[15] J. L. Wadsworth, J. A. Tawn, "Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling," in *J. R. Statist. Soc. B*, vol. 74, no. 3, 2012, pp. 543–567.

[16] P. J. Northrop, C. L. Coleman, "Improved threshold diagnostic plots for extreme value analyses," in *Extremes*, vol. 71, 2014, pp. 289–303.

[17] M. I. Gomes, O. Oliveira, "The Bootstrap methodology in statistics of extremes - Choice of the optimal sample fraction," in *Extremes*, vol. 4, no. 4, 2002, pp. 331–358.

[18] G. Draisma, L. de Haan, L. Peng, T. T. Pereira, "A Bootstrap-based method to achieve optimality in estimating the extreme-value index," in *Extremes*, vol. 2, no. 4, 1999, pp. 367–404.

[19] L. Homolka, "Extreme value approach for estimating value at risk metrics with respect to Basel II", in *International Journal of Mathematics and Computers in Simulation*, vol. 7, no. 2, 2013, pp. 171–178.

[20] D. Prathumwan, Y. Lenbury, P. Satiracoo and C. Rattanakul, "Euler-Maruyama approximation and maximum likelihood estimator for a stochastic differential equation model of the signal transduction process," in *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 6, no. 2, 2012, pp. 323–331.

[21] A. Andronov, N. Kolmakova and I. Yatskiv, "A quasi regression model for polytomous data and its application for measuring service quality," in *International Journal of Mathematics and Computers in Simulation*, vol. 4, no. 2, 2010, pp. 50–57.

[22] D. Bílková, "Lognormal distribution and using L-moment method for estimating its parameters," in *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 6, no. 1, 2012, pp. 30–44.

[23] B. M. Hill, "'A simple general approach to inference about the tail of a distribution,"' in *Ann. Statist.*, vol. 3, 1975, pp. 1163–1174.

[24] V. Niola, R. Oliviero, G. Quaremba, "A non-parametric method for the estimation of a GPD parameter based on wavelet theory". *Proceedings of the 9th WSEAS International Conference on Computers*, ICCOMP05. Stevens Point, Wisconsin, USA:WSEAS Press, 2005.

[25] H. Madsen, P. S. Mikkelsen, D. Rosbjerg, P. Harremöes, "Regional estimation of rainfall intensity-duration-frequency curves using generalized least square regressions of partial durations series," in *Water Resour. Res.*, vol. 38, no. 11, 2002, pp. 21-1–21-11.

[26] R. L. Smith, "Extreme value theory based on the r largest annual events," in *J. Hydrol.*, vol. 86, no. 1-2, 1986, pp. 27-43.

[27] F. Caeiro, M. I. Gomes, "Semi-parametric tail inference through probability-weighted moments," in *J. Stat. Plan. Infer.,* vol. 141, 2010, pp. 937–950.

[28] K. Philippopoulos, D. Deligiorgi, "Statistical simulation of wind speed in Athens, Greece based on Weibull and ARMA models," in *International Journal of Energy and Environment*, vol. 3, no. 4, 2009, pp. 151–158.

[29] B. Efron, R. Tibshirani, *An Introduction to the bootstrap.* Boca Ration, Chapman, 1994.