arXiv:1802.02928v1 [stat.ME] 7 Feb 2018

Statistical tests for daily and total precipitation volumes to be abnormally extremal

V. Yu. Korolev[1], A. K. Gorshenin[2], K. P. Belyaev[3]

Abstract. In this paper, two approaches are proposed to the definition of abnormally extremal precipitation. These approaches are based on the negative binomial model for the distribution of duration of wet periods measured in days [19]. This model demonstrates excellent fit with real data and provides a theoretical base for the determination of asymptotic approximations to the distributions of the maximum daily precipitation volume within a wet period and of the total precipitation volume over a wet period. The asymptotic distribution of the maximum daily precipitation volume within a wet period turns out to be a tempered Snedecor–Fisher distribution whereas the total precipitation volume for a wet period turns out to be the gamma distribution. Both approximations appear to be very accurate. These asymptotic approximations are deduced using limit theorems for statistics constructed from samples with random sizes. The first approach to the definition (and determination) of abnormally extreme precipitation is based on the tempered Snedecor–Fisher distribution of the maximum daily precipitation. According to this approach, a daily precipitation volume is considered to be abnormally extremal, if it exceeeds a certain (pre-defined) quantile of the tempered Snedecor–Fisher distribution. The second approach is based on that the total precipitation volume for a wet period has the gamma distribution. Hence, the hypothesis that the total precipitation volume during a certain wet period is abnormally large can be formulated as the homogeneity hypothesis of a sample from the gamma distribution. Two equivalent tests are proposed for testing this hypothesis. One of them is based on the beta distribution whereas the second is based on the Snedecor–Fisher distribution. Both of these tests deal with the relative contribution of the total precipitation volume for a wet period to the considered set (sample) of successive wet periods. Within the second approach it is possible to introduce the notions of relatively abnormal and absolutely abnormal precipitation volumes. The results of the application of these tests to real data are presented yielding the conclusion that the intensity of wet periods with abnormally large precipitation volume increases.

Key words: wet period, total precipitation volume, negative binomial distribution, asymptotic approximation, extreme order statistic, random sample size, gamma distribution, Beta distribution, Snedecor–Fisher distribution, testing statistical hypotheses.

[1]Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Russia; Institute of Informatics Problems, Federal Research Center ¡¡Computer Science and Control¿¿ of Russian Academy of Sciences, Russia; Hangzhou Dianzi University, China; vkorolev@cs.msu.su

[2]Institute of Informatics Problems, Federal Research Center ¡¡Computer Science and Control¿¿ of Russian Academy of Sciences, Russia; agorshenin@frccsc.ru

[3]Institute of Oceanology of Russian Academy of Sciences, Russia; Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Russia; kosbel55@gmail.com

1 Introduction

1.1 Motivation and the structure of the paper

Estimates of regularities and trends in heavy and extreme daily precipitation are important for understanding climate variability and change at relatively small or medium time horizons. However, such estimates are much more uncertain compared to those derived for mean precipitation or total precipitation during a wet period. This uncertainty is due to that, first, estimates of heavy precipitation depend closely on the accuracy of the daily records; they are more sensitive to missing values [25, 26]. Second, uncertainties in the estimates of heavy and extreme precipitation are caused by the inadequacy of the mathematical models used for the corresponding calculations. Third, these uncertainties are boosted by the lack of reasonable means for the unambiguous (algorithmic) determination of extreme or abnormally heavy precipitation amplified by some statistical significance problems owing to the low occurrence of such events. As a consequence, continental-scale estimates of the variability and trends in heavy precipitation based on daily precipitation might generally agree qualitatively but may exhibit significant quantitative differences. In [27] a detailed review of this phenomenon is presented where it is noted that for the European continent, most results hint at a growing intensity of heavy precipitation over the last five decades.

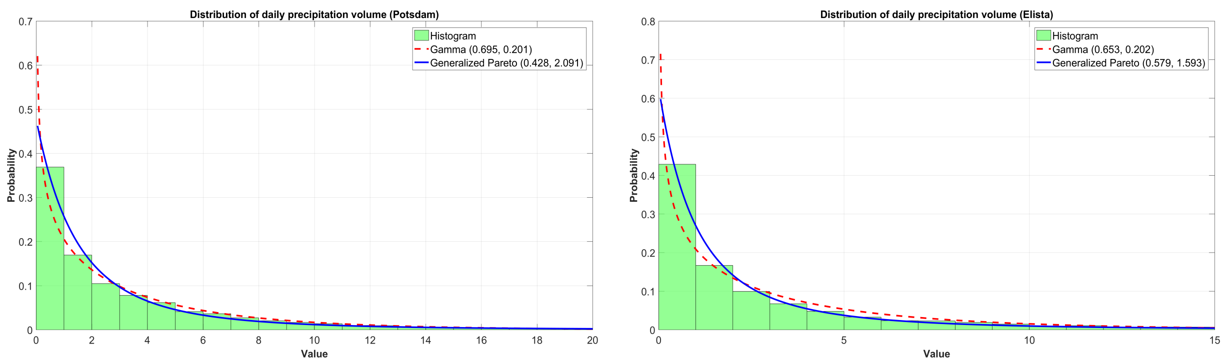
At the same time, the climate variability and trends at relatively large time horizons are of no less importance for long-range business, say, agricultural projects and forecasting of risks of water floods, dry spells and other natural disasters. In the present paper we propose a rather reasonable approach to the unambiguous (algorithmic) determination of extreme or abnormally heavy daily and total precipitation within a wet period. This approach is based on the negative binomial model for the duration of wet periods measured in days [19]. This model demonstrates excellent fit with the real data and provides a theoretical base for the determination of asymptotic approximations to the distributions of the maximum daily precipitation volume within a wet period and of the total precipitation volume for a wet period. The asymptotic distribution of the maximum daily precipitation volume within a wet period turns out to be a tempered Snedecor–Fisher distribution whereas the total precipitation volume for a wet period turns out to be the gamma distribution. Both approximations appear to be very accurate. These asymptotic approximations are deduced using limit theorems for statistics constructed from samples with random sizes.

In this paper, two approaches are proposed to the definition of abnormally extremal precipitation. The first approach to the definition (and determination) of abnormally heavy daily precipitation is based on the tempered Snedecor–Fisher distribution. The second approach is based on the assumption that the total precipitation volume over a wet period has the gamma distribution. This assumption is theoretically justified by a version of the law of large numbers for sums of a random number of random variables in which the number of summands has the negative binomial distribution and is empirically substantiated by the statistical analysis of real data. Hence, the hypothesis that the total precipitation volume during a certain wet period is abnormally large can be formulated as the homogeneity hypothesis of a sample from the gamma distribution. Two equivalent tests are proposed for testing this hypothesis. One of them is based on the beta distribution whereas the second is based on the Snedecor–Fisher distribution. Both of these tests deal with the relative contribution of the total precipitation volume for a wet period to the considered set (sample) of successive wet periods. Within the second approach it is possible to introduce the notions of relatively abnormal and absolutely abnormal precipitation volumes. The results of the application of these tests to real data are presented yielding the conclusion that the intensity of wet periods with abnormally large precipitation volume increases.

The proposed approaches are to a great extent devoid of the drawbacks mentioned above: first, estimates of total precipitation are weakly affected by the accuracy of the daily records and are less sensitive to missing values. Second, they are based on limit theorems of probability theorems that yield unambiguous asymptotic approximations which are used as adequate mathematical models. Third, these approaches provide unambiguous algorithms for the determination of extreme or abnormally heavy daily or total precipitation that do not involve statistical significance problems owing to the low occurrence of such (relatively rare) events.

Our approaches improve the one proposed in [26], where an estimate of the fractional contribution from the wettest days to the total was developed which is less hampered by the limited number of wet days. For this purpose, in [26] an assumption was enacted (yet, without any theoretical justification) that the statistical regularities in daily precipitation follow the gamma distribution and the parameters of the gamma distribution are estimated from the observations. This assumption made it possible to derive a theoretical distribution of the fractional contribution of any percentage of wet days to the total from the gamma distribution function.

However, a more thorough statistical analysis showed that, although being rather adequate and, in general, acceptable model, the gamma distribution is not the best model for statistical regularities in *daily* precipitation. For example, the analysis of meteorological data (daily precipitation volumes) registered registered during 60 years at two geographic points with very different climate: Potsdam (Brandenburg, Germany) with mild climate influenced by the closeness to the ocean with warm Gulfstream flow and Elista (Kalmykia, Russia) with radically continental climate convincingly suggests the Pareto-type model for the distribution of daily precipitation volumes, see Figure 3.3a, 3.3b. For comparison, on these figures there are also presented the graphs of the best gamma-densities which, nevertheless, fit the histograms in a noticeably worse way than the Pareto distributions.



a) b)

Fig. 1: The histogram of daily precipitation volumes in Potsdam (a) and Elista (b) and the fitted Pareto and gamma distributions.

The Pareto model for the daily precipitation volume together with the observation that the duration of a wet period has the negative binomial distribution makes it possible to propose a reasonable model for the distribution of the maximum daily precipitation within a wet period as an asymptotic approximation provided by the limit theorems for extreme order statistics in samples with random size. In Section 2 we give a strict derivation of such a model having the form of the tempered Snedecor–Fisher distribution (that is, the distribution of a positive power of a random variable with the Snedecor–Fisher distribution) and discuss its properties as well as some methods of statistical estimation of its parameters. This model makes it possible to propose the following approach to the definition (and determination) of an abnormally heavy daily precipitation volume. The grounds for this approach is an obvious observation that if $X_1, X_2, \ldots, X_N$ is a sample of N positive observations, then with finite (possibly, random) N, among X_i's there is *always* an extreme observation, say, X_1, such that $X_1 \geq X_i$, $i = 1, 2, \ldots, N$. Two cases are possible: (i) X_1 is a 'typical' observation and its extreme character is conditioned by purely stochastic circumstances (there *must* be an extreme observation within a finite homogeneous sample) and (ii) X_1 is abnormally large so that it is an 'outlier' and its extreme character is due to some exogenous factors. As it will be shown in Section 2, the distribution of X_1 in the 'typical' case (i) is the tempered Snedecor–Fisher distribution. Therefore, if X_1 exceeds a certain (pre-defined) quantile of the tempered Snedecor–Fisher distribution (say, of the orders 0.99, 0.995 or 0.999), then it is regarded as 'suspicious' to be an outlier, that is, to be abnormally large (the quantile orders specified above mean that it is pre-determined that one out of a hundred of maximum daily precipitations, one out of five hundred of maximum daily precipitations, or one out of a thousand of maximum daily precipitations is abnormally large, respectively).

Methodically, this approach is similar to the classical techniques of dealing with extreme observations [1]. The novelty of the method proposed in Section 2 is in a more accurate specification of the distribution of extreme daily precipitation. In applied problems dealing with extreme values there is a common tradition which, possibly, has already become a prejudice, that statistical regularities in the behavior of extreme values necessarily obey one of well-known three types of extreme value distributions. In general, this is certainly so, if the sample size is very large, that is, the time horizon under consideration is very wide. In other words, the models based on the extreme value distributions have *asymptotic* character. However, in real practice, when the sample size is finite and the extreme values of the process under consideration are studied on the time horizon of a moderate length, the classical extreme value distributions may turn out to be inadequate models. In these situations a more thorough analysis may generate other models which appear to be considerably more adequate. This is exactly the case discussed in the present paper. Here, within the first approach, along with the 'large' parameter, the expected sample size, one more 'small' parameter is introduced and new models are proposed as asymptotic approximations when the small parameter is infinitesimal. These models prove to be exceptionally accurate and demonstrate excellent fit with the observed data.

To construct another test for distinguishing between the cases (i) and (ii) mentioned above, in Section 3 we also strongly improve the results of [27] by giving theoretical grounds for the correct application of the gamma distribution as the model of statistical regularities of *total precipitation volume during a wet period*. These grounds are based on the negative binomial model for the distribution of the duration of a wet period. In turn, the adequacy of the negative binomial model has serious empirical and theoretical rationale the details of which are described below. With some caveats the gamma model can be also used for the *conditional* distribution of daily precipitation volumes. The proof of this result is based on the law of large numbers for random sums in which the number of summands has the negative binomial distribution. Hence, the hypothesis that the total precipitation volume during a certain wet period is abnormally large can be re-formulated as the homogeneity hypothesis of a sample from the gamma distribution. Two equivalent statistics are proposed for testing this hypothesis. The corresponding tests are scale-free and depend only on the easily estimated shape parameter of the negative binomial distribution and the time-scale parameter determining the denominator in the fractional contribution of a wet period under consideration. It is worth noting that within the second approach the test for a total precipitation volume during one wet period to be abnormally large can be applied to the observed time series in a moving mode. For this purpose a *window* (a set of successive observations) is determined. The observations within a window constitute the sample to be analyzed. Let m be the number of observation in the window (the sample size). As the window moves rightward, each fixed observation falls in exactly m successive windows. A fixed observation may be recognized as abnormally large within *each* of m windows containing this observation. In this case this observation will be called *absolutely abnormally large* with respect to a given time horizon (determined by the sample size m. Also, a fixed observation may be recognized as abnormally large within *at least one* of m windows containing this observation. In this case the observation will be called *relatively abnormally large* with respect to a given time horizon.

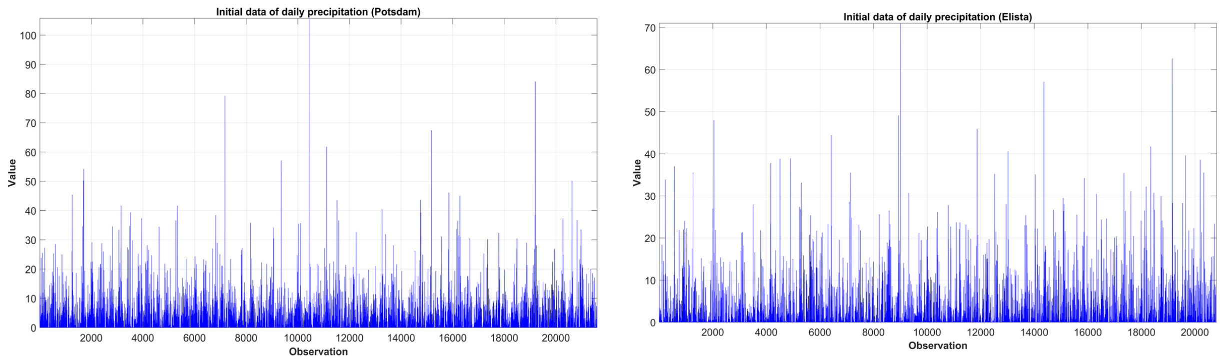
The preconditions and backgrounds of all the approaches as well as their peculiarities will also be discussed.

The paper is organized as follows. In Sect. 1.2 we discuss the negative binomial model for the distribution of the duration of a wet period measured in days. Notation, definitions and some mathematical preliminaries are presented in Sect. 1.3. In Section 2 we introduce the test for a *daily* precipitation volume to be abnormally large. In Sect. 2.1 an asymptotic approximation is proposed for the distribution of the maximum daily precipitation volume within a wet period. Some analytic properties of the obtained limit distribution are described. In particular, it is demonstrated that under certain conditions the limit distribution is mixed exponential and hence, is infinitely divisible. It is shown that under the same conditions the limit distribution can be represented as a scale mixture of stable or Weibull or Pareto or folded normal laws. The corresponding product representations for the limit random variable can be used for its computer simulation. Several methods for the statistical estimation of the parameters of this distribution are proposed in Sect. 2.2. Section 2.3 contains the results and discussion of fitting the distribution proposed in Section 2.1 to real data. The results of application of the test for a daily precipitation to be abnormally large based on the tempered Snedecor–Fisher distribution to real daily precipitation data are presented and discussed in Sect. 2.4. Section 3 deals with the test for a *total* precipitation volume over a wet period to be abnormally large based on testing the homogeneity hypothesis of a sample from the gamma distribution. Two equivalent tests are introduced in Sect. 3.1. In Sect. 3.2 the application of these tests to a time series in a moving mode is discussed and the notions of relatively abnormally large and absolutely abnormally large precipitation are introduced. The results of application of these tests to real daily precipitation data are presented and discussed in Sect. 3.3.

1.2 The negative binomial model for the duration of wet periods

In most papers dealing with the statistical analysis of meteorological data available to the authors, the suggested analytical models for the observed statistical regularities in precipitation are rather ideal and inadequate. For example, it is traditionally assumed that the duration of a wet period (the number of subsequent wet days) follows the geometric distribution (for example, see [27]) although the goodness-of-fit of this model is far from being admissible. Perhaps, this prejudice is based on the conventional interpretation of the geometric distribution in terms of the Bernoulli trials as the distribution of the number of subsequent wet days (¡¡successes¿¿) till the first dry day (¡¡failure¿¿). But the framework of Bernoulli trials assumes that the trials are independent whereas a thorough statistical analysis of precipitation data registered in different points demonstrates that the sequence of dry and wet days is not only independent, but it is also devoid of the Markov property so that the framework of Bernoulli trials is absolutely inadequate for analyzing meteorological data.

It turned out that the statistical regularities of the number of subsequent wet days can be very reliably modeled by the negative binomial distribution with the shape parameter less than one. For example, in [19] we analyzed meteorological data registered at two geographic points with very different climate: Potsdam (Brandenburg, Germany) with mild climate influenced by the closeness to the ocean with warm Gulfstream flow and Elista (Kalmykia, Russia) with radically continental climate. The initial data of daily precipitation in Potsdam and Elista are presented on Figures 2a and 2b, respectively. On these figures the horizontal axis is discrete time measured in days. The vertical axis is the daily precipitation volume measured in centimeters. In other words, the height of each "pin" on these figures is the precipitation volume registered at the corresponding day (at the corresponding point on the horizontal axis).



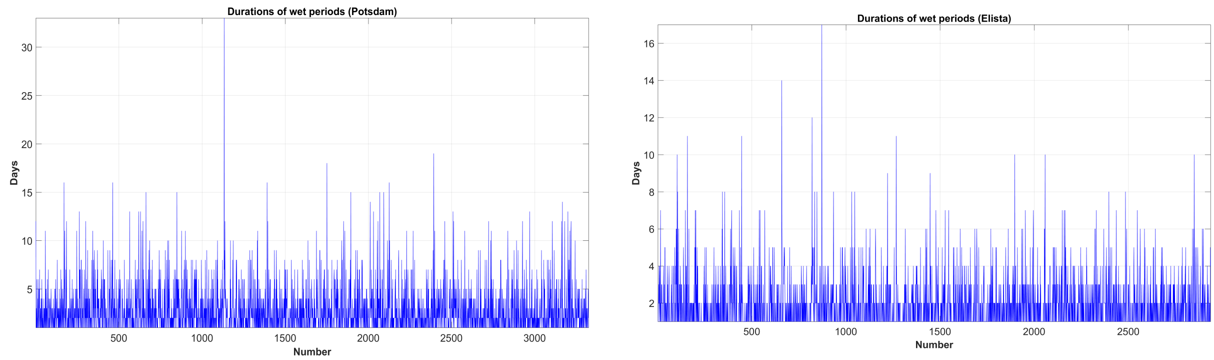
a) b)

Fig. 2: The initial data of daily precipitation in Potsdam (a) and Elista (b).

In order to analyze the statistical regularities of the duration of wet periods this data was rearranged as shown on Figures 3a and 3b.

On these figures the horizontal axis is the number of successive wet periods. It should be mentioned that directly before and after each wet period there is at least one dry day, that is, successive wet periods are separated by dry periods. On the vertical axis there lie the durations of wet periods. In other words, the height of each ¡¡pin¿¿ on these figures is the length of the corresponding wet period measured in days and the corresponding point on the horizontal axis is the number of the wet period.

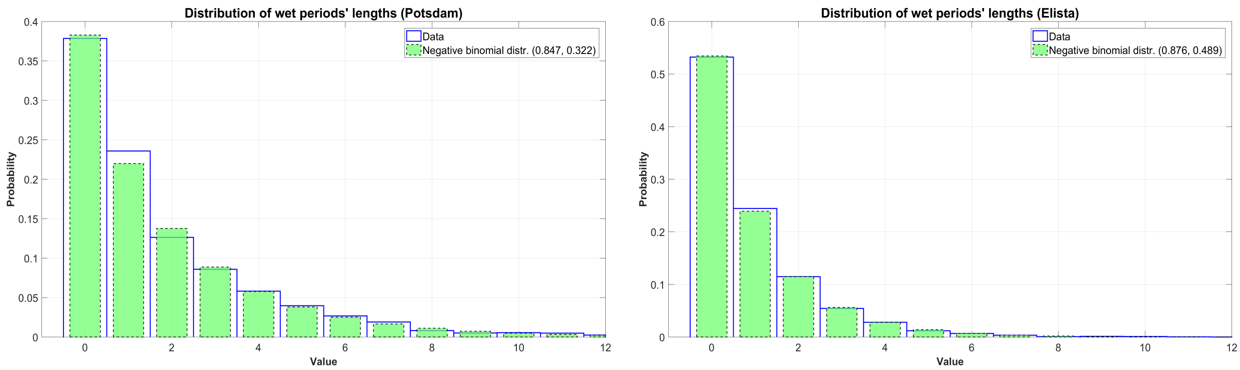
The samples of durations in both Potsdam and Elista were assumed homogeneous and independent. It was demonstrated that the fluctuations of the numbers of successive wet days with very high confidence fit the negative binomial distribution with shape parameter less than one (also see [4]). Figures 4a and 4b show the histograms constructed from the



a) b)

Fig. 3: The durations of wet periods in Potsdam (a) and Elista (b).

corresponding samples of duration periods and the fitted negative binomial distribution. In both cases the shape parameter r turned out to be less than one. For Potsdam $r = 0.876$, $p = 0.489$, for Elista $r = 0.847$, $p = 0.322$.



a) b)

Fig. 4: The histogram of durations of wet periods in Potsdam (a) and Elista (b) and the fitted negative binomial distribution.

In the same paper a schematic attempt was undertaken to explain this phenomenon by the fact that negative binomial distributions can be represented as mixed Poisson laws with mixing gamma-distributions. As is known, the Poisson distribution is the best model for the discrete stochastic chaos [7] by virtue of the universal principle of non-decrease of entropy in closed systems (see, e. g., [3, 12]) and the mixing distribution accumulates the statistical regularities in the influence of stochastic factors that can be assumed exogenous with respect to the local system under consideration.

In the paper [18] this explanation of the adequacy of the negative binomial model was concretized. For this purpose, the concept of a mixed geometric distribution introduced in [13] (also see [14, 15]) was used. In [18] it was demonstrated that any negative binomial distribution with shape parameter no greater than one is a mixed geometric distribution (this result is reproduced below as Theorem 1). Thereby, a ¡¡discrete¿¿ analog of a theorem due to L. Gleser [2] was proved. Gleser's theorem establishes that a gamma distribution with shape parameter no greater than one can be represented as a mixed exponential distribution.

The representation of a negative binomial distribution as a mixed geometric law can be interpreted in terms of the Bernoulli trials as follows. First, as a result of some ¡¡preliminary¿¿ experiment the value of some random variables (r.v.'s) taking values in $[0, 1]$ is determined which is then used as the probability of success in the sequence of Bernoulli trials in which the original ¡¡unconditional¿¿ r.v. with the negative binomial distribution is nothing else than the ¡¡conditionally¿¿ geometrically distributed r.v. having the sense of the number of trials up to the first failure. This makes it possible to assume that the sequence of wet/dry days is not independent, but is conditionally independent and the random probability of success is determined by some outer stochastic factors. As such, we can consider the seasonality or the type of the cause of a rainy period.

The negative binomial model for the distribution of the duration of wet periods makes it possible to obtain asymptotic approximations for important characteristics of precipitation such as the distribution of the total precipitation volume per wet period and the distribution of the maximum daily precipitation volume within a wet period. The first of these approximations was proposed in [18], where an analog of the law of large numbers for negative binomial random sums was presented stating that the limit distribution for these sums is the gamma distribution. The construction of the second approximation is the target of Section 3 of the present paper.

1.3 Notation, definitions and mathematical preliminaries

Although the main objects of our interest are the probability distributions, for convenience and brevity in what follows we will expound our results in terms of r.v.'s with the corresponding distributions assuming that all the r.v.'s under consideration are defined on one and the same probability space $(\Omega, \mathfrak{F}, \mathsf{P})$.

In the paper, conventional notation is used. The symbols $\stackrel{d}{=}$ and $\Longrightarrow$ denote the coincidence of distributions and convergence in distribution, respectively. The integer and fractional parts of a number z will be respectively denoted $[z]$ and $\{z\}$.

A r.v. having the gamma distribution with shape parameter $r > 0$ and scale parameter $\lambda > 0$ will be denoted $G_{r,\lambda}$,

$$\mathsf{P}(G_{r,\lambda} < x) = \int_0^x g(z; r, \lambda)dz, \quad \text{with} \quad g(x; r, \lambda) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \ x \geq 0,$$

where $\Gamma(r)$ is Euler's gamma-function, $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$, $r > 0$.

In these notation, obviously, $G_{1,1}$ is a r.v. with the standard exponential distribution: $\mathsf{P}(G_{1,1} < x) = \big[1 - e^{-x}\big]\mathbf{1}(x \geq 0)$ (here and in what follows $\mathbf{1}(A)$ is the indicator function of a set A).

The gamma distribution is a particular representative of the class of *generalized gamma distributions* (GG-distributions), which were first described in [23] as a special family of lifetime distributions containing both gamma distributions and Weibull distributions. A GG-distribution is the absolutely continuous distribution defined by the density

$$g^*(x; r, \gamma, \lambda) = \frac{|\gamma|\lambda^r}{\Gamma(r)} x^{\gamma r - 1} e^{-\lambda x^\gamma}, \qquad x \geq 0,$$

with $\gamma \in \mathbb{R}$, $\lambda > 0$, $r > 0$.

The properties of GG-distributions are described in [23, 24]. A r.v. with the density $g^*(x; r, \gamma, \lambda)$ will be denoted $G^*_{r,\gamma,\lambda}$. It can be easily made sure that

$$G^*_{r,\gamma,\lambda} \stackrel{d}{=} G^{1/\gamma}_{r,\lambda}. \tag{1}$$

For a r.v. with the Weibull distribution, a particular case of GG-distributions corresponding to the density $g^*(x;1,\gamma,1)$ and the distribution function (d.f.) $\big[1-e^{-x^\gamma}\big]\mathbf{1}(x\geq 0)$, we will use a special notation W_γ. Thus, $G_{1,1}\stackrel{d}{=}W_1$. It is easy to see that

$$W_1^{1/\gamma}\stackrel{d}{=}W_\gamma. \tag{2}$$

A r.v. with the standard normal d.f. $\Phi(x)$ will be denoted X,

$$\mathsf{P}(X<x)=\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-z^2/2}dz,\qquad x\in\mathbb{R}.$$

The distribution of the r.v. $|X|$, i. e. $\mathsf{P}(|X|<x)=2\Phi(x)-1$, $x\geq 0$, is called half-normal or folded normal.

The d.f. and the density of a strictly stable distribution with the characteristic exponent α and shape parameter θ defined by the characteristic function (ch.f.)

$$\mathfrak{f}_{\alpha,\theta}(t)=\exp\big\{-|t|^\alpha\exp\{-\tfrac{1}{2}i\pi\theta\alpha\mathrm{sign}t\}\big\},\qquad t\in\mathbb{R},$$

wheter $0<\alpha\leq 2$, $|\theta|\leq\min\{1,\frac{2}{\alpha}-1\}$, will be respectively denoted $F_{\alpha,\theta}(x)$ and $f_{\alpha,\theta}(x)$ (see, e. g., [28]). A r.v. with the d.f. $F_{\alpha,\theta}(x)$ will be denoted $S_{\alpha,\theta}$.

To symmetric strictly stable distributions there correspond the value $\theta=0$ and the ch.f. $\mathfrak{f}_{\alpha,0}(t)=e^{-|t|^\alpha}$, $t\in\mathbb{R}$. To one-sided strictly stable distributions concentrated on the nonnegative halfline there correspond the values $\theta=1$ and $0<\alpha\leq 1$. The pairs $\alpha=1$, $\theta=\pm1$ correspond to the distributions degenerate in ±1, respectively. All the rest strictly stable distributions are absolutely continuous. Stable densities cannot be explicitly represented via elementary functions with four exceptions: the normal distribution ($\alpha=2$, $\theta=0$), the Cauchy distribution ($\alpha=1$, $\theta=0$), the Lévy distribution ($\alpha=\frac{1}{2}$, $\theta=1$) and the distribution symmetric to the Lévy law ($\alpha=\frac{1}{2}$, $\theta=-1$).

In [22, 16, 17] it was proved that if $\alpha\in(0,1)$ and the i.i.d. r.v.'s $S_{\alpha,1}$ and $S'_{\alpha,1}$ have the same strictly stable distribution, then the density $v_\alpha(x)$ of the r.v. $R_\alpha=S_{\alpha,1}/S'_{\alpha,1}$ has the form

$$v_\alpha(x)=\frac{\sin(\pi\alpha)x^{\alpha-1}}{\pi[1+x^{2\alpha}+2x^\alpha\cos(\pi\alpha)]},\qquad x>0. \tag{3}$$

A r.v. $N_{r,p}$ is said to have the *negative binomial distribution* with parameters $r>0$ (¡¡shape¿¿) and $p\in(0,1)$ (¡¡success probability¿¿), if

$$\mathsf{P}(N_{r,p}=k)=\frac{\Gamma(r+k)}{k!\Gamma(r)}\cdot p^r(1-p)^k,\qquad k=0,1,2,...$$

A particular case of the negative binomial distribution corresponding to the value $r=1$ is the *geometric distribution*. Let $p\in(0,1)$ and let $N_{1,p}$ be the r.v. having the geometric distribution with parameter p:

$$\mathsf{P}(N_{1,p}=k)=p(1-p)^k,\qquad k=0,1,2,...$$

This means that for any $m\in\mathbb{N}$

$$\mathsf{P}(N_{1,p}\geq m)=\sum\nolimits_{k=m}^{\infty}p(1-p)^k=(1-p)^m.$$

Let Y be a r.v. taking values in the interval $(0,1)$. Moreover, let for all $p \in (0,1)$ the r.v. Y and the geometrically distributed r.v. $N_{1,p}$ be independent. Let $V = N_{1,Y}$, that is, $V(\omega) = N_{1,Y(\omega)}(\omega)$ for any $\omega \in \Omega$. The distribution

$$\mathsf{P}(V \geq m) = \int_0^1 (1-y)^m d\mathsf{P}(Y < y), \quad m \in \mathbb{N},$$

of the r.v. V will be called *Y-mixed geometric* [13].

It is well known that the negative binomial distribution is a mixed Poisson distribution with the gamma mixing distribution [5] (also see [12]): for any $r > 0$, $p \in (0,1)$ and $k \in \{0\} \bigcup \mathbb{N}$ we have

$$\frac{\Gamma(r+k)}{k!\Gamma(r)} \cdot p^r(1-p)^k = \frac{1}{k!}\int_0^\infty e^{-z}z^k g(z;r,\mu)dz, \tag{4}$$

where $\mu = p/(1-p)$.

Based on representation (4), in [18] it was proved that any negative binomial distribution with the shape parameter no greater than one is a mixed geometric distribution. Namely, the following statement was proved that gives an analytic explanation of the validity of the negative binomial model for the duration of wet periods measured in days (see the Introduction).

THEOREM 1 [18]. *The negative binomial distribution with parameters $r \in (0,1)$ and $p \in (0,1)$ is a mixed geometric distribution: for any $k \in \{0\} \bigcup \mathbb{N}$*

$$\frac{\Gamma(r+k)}{k!\Gamma(r)} \cdot p^r(1-p)^k = \int_\mu^\infty \Big(\frac{z}{z+1}\Big)\Big(1-\frac{z}{z+1}\Big)^k p(z;r,\mu)dz = \int_p^1 y(1-y)^k h(y;r,p)dy,$$

where $\mu = p/(1-p)$ and the probability densities $p(z;r,\mu)$ and $h(y;r,p)$ have the forms

$$p(z;r,\mu) = \frac{\mu^r}{\Gamma(1-r)\Gamma(r)} \cdot \frac{\mathbf{1}(z \geq \mu)}{(z-\mu)^r z},$$

$$h(y;r,p) = \frac{p^r}{\Gamma(1-r)\Gamma(r)} \cdot \frac{(1-y)^{r-1}\mathbf{1}(p < y < 1)}{y(y-p)^r}.$$

Furthermore, if $G_{r,1}$ and $G_{1-r,1}$ are independent gamma-distributed r.v.'s, $\mu > 0$, $p \in (0,1)$, then the density $p(z;r,\mu)$ corresponds to the r.v.

$$Z_{r,\mu} = \frac{\mu(G_{r,1} + G_{1-r,1})}{G_{r,1}} \tag{5}$$

and the density $h(y;r,p)$ corresponds to the r.v.

$$Y_{r,p} = \frac{p(G_{r,1} + G_{1-r,1})}{G_{r,1} + pG_{1-r,1}}.$$

2 The test for a daily precipitation volume to be abnormally large based on the tempered Snedecor–Fisher distribution

2.1 The tempered Snedecor–Fisher distribution as an asymptotic approximation to the maximum daily precipitation volume within a wet period

Following [20, 21], in this section we will determine the probability distribution of extremal daily precipitation within a wet period.

Let $F(x)$ be a d.f., $a \in \mathbb{R}$. Denote $\mathrm{rext}(F) = \sup\{x : F(x) < 1\}$, $F^{-1}(a) = \inf\{x : F(x) \geq a\}$.

In [20, 21] the following statement was proved using Lemma 2 from [10] and some results of the papers dealing with the asymptotic theory of extreme order statistics constructed from samples with random sizes that have mixed Poisson distribution (in particular, Theorem 3.1 of [11]).

THEOREM 2 [20, 21]. *Let $n \in \mathbb{N}$, $\lambda > 0$, $q \in (0,1)$ and let N_{r,p_n} be a r.v. with the negative binomial distribution with parameters $r > 0$ and $p_n = \min\{q, \lambda/n\}$. Let $X_1, X_2, \ldots$ be i.i.d. r.v:s with a common d.f. $F(x)$. Assume that* $\mathrm{rext}(F) = \infty$ *and there exists a number $\gamma > 0$ such that for each $x > 0$*

$$\lim_{y\to\infty} \frac{1 - F(xy)}{1 - F(y)} = x^{-\gamma}. \tag{6}$$

Then

$$\lim_{n\to\infty} \sup_{x\geq 0} \left| \mathsf{P}\left(\frac{\max\{X_1, \ldots, X_{N_{r,p_n}}\}}{F^{-1}(1 - \frac{1}{n})} < x \right) - F(x; r, \lambda, \gamma) \right| = 0,$$

where

$$F(x; r, \lambda, \gamma) = \left(\frac{\lambda x^{\gamma}}{1 + \lambda x^{\gamma}} \right)^{r}, \quad x \geq 0.$$

As we have seen in Sect. 1.2, the Pareto model is rather adequate for the distribution function F of the daily precipitation volume. It can be easily verified that the Pareto distribution function satisfies condition (6). Therefore, we can conclude that the theoretical conditions of Theorem 2 are in good correspondence with the empirically observed data.

It is worth noting that the limit distribution with the power-type decrease of the tail was obtained in Theorem 2 as a scale mixture of the Fréchet distribution (the type II extreme value distribution) in which the mixing law is the gamma distribution. Namely, since the Fréchet d.f. $e^{-x^{-\gamma}}$ with $\gamma > 0$ corresponds to the r.v. W_{γ}^{-1}, it is easy to make sure that the d.f. $F(x; r, \lambda, \gamma)$ corresponds to the r.v. $M_{r,\gamma,\lambda} \equiv G_{r,\lambda}^{1/\gamma} W_{\gamma}^{-1}$, where the multipliers on the right-hand side are independent. From (1) and (2) it follows that

$$M_{r,\gamma,\lambda} \stackrel{d}{=} \left(\frac{G_{r,\lambda}}{W_1} \right)^{1/\gamma} \stackrel{d}{=} \frac{G^{*}_{r,\gamma,\lambda}}{W_{\gamma}} \tag{7}$$

where in each term the multipliers are independent. Consider the r.v. $G_{r,\lambda}/W_1$ in (7) in more detail. We have

$$\frac{G_{r,\lambda}}{W_1} \stackrel{d}{=} \frac{G_{r,\lambda}}{G_{1,1}} \stackrel{d}{=} \frac{G_{r,1}}{\lambda G_{1,1}} \stackrel{d}{=} \frac{Q_{r,1}}{\lambda r},$$

where $Q_{r,1}$ is the r.v. having the Snedecor–Fisher distribution with parameters r, 1 ('degrees of freedom') defined by the Lebesgue density

$$f_{r,1}(x) = \frac{r^{r+1}x^{r-1}}{(1+rx)^{r+1}}, \quad x \geq 0,$$

(see, e. g., [6], Section 27).

So,

$$M_{r,\gamma,\lambda} \stackrel{d}{=} \Big(\frac{Q_{r,1}}{\lambda r}\Big)^{1/\gamma}, \tag{8}$$

that is, the distribution of the r.v. $M_{r,\gamma,\lambda}$ up to a non-random scale factor coincides with that of the positive power of a r.v. with the Snedecor–Fisher distribution. In other words, the distribution function $F(x;r,\lambda,\gamma)$ up to a power transformation of the argument x coincides with the Snedecor–Fisher distribution function. In statistics, distributions with arguments subjected to the power transformation are conventionally called *tempered*. Therefore, we have serious reason to call the distribution $F(x;r,\lambda,\gamma)$ *tempered Snedecor–Fisher distribution.*

The statement of theorem 2 can be re-formulated as

$$\frac{\max\{X_1,\ldots,X_{N_{r,p_n}}\}}{F^{-1}(1-\frac{1}{n})} \Longrightarrow M_{r,\gamma,\lambda} \equiv \frac{G_{r,\lambda}^{1/\gamma}}{W_\gamma} \stackrel{d}{=} \Big(\frac{Q_{r,1}}{\lambda r}\Big)^{1/\gamma} \quad (n\to\infty). \tag{9}$$

The density of the limit distribution $F(x;r,\gamma,\lambda)$ of the extreme daily precipitation within a wet period has the form

$$p(x;r,\gamma,\lambda) = \frac{r\gamma\lambda^r x^{\gamma r-1}}{(1+\lambda x^\gamma)^{r+1}} = \frac{\gamma r\lambda^r}{x^{1+\gamma}(\lambda+x^{-\gamma})^{r+1}}, \quad x>0. \tag{10}$$

Some properties of the distribution of the r.v. $M_{r,\gamma,\lambda}$ were discussed in [21]. In particular, it was shown there that $p(x;r,\gamma,\lambda) = O(x^{-1-\gamma})$ as $x\to\infty$. Therefore $\mathsf{E}M_{r,\gamma,\lambda}^\delta < \infty$ only if $\delta<\gamma$. Moreover, it is possible to deduce explicit expressions for the moments of the r.v. $M_{r,\gamma,\lambda}$.

THEOREM 3 [21]. *Let* $0<\delta<\gamma<\infty$. *Then*

$$\mathsf{E}M_{r,\gamma,\lambda}^\delta = \frac{\Gamma\big(r+\frac{\delta}{\gamma}\big)\Gamma\big(1-\frac{\delta}{\gamma}\big)}{\lambda^{\delta/\gamma}\Gamma(r)}.$$

THEOREM 4 [21]. *Let* $r\in(0,1]$, $\gamma\in(0,1]$, $\lambda>0$. *Then the following product representations are valid*:

$$M_{r,\gamma,\lambda} \stackrel{d}{=} \frac{G_{r,\lambda}^{1/\gamma}S_{\gamma,1}}{W_1}, \tag{11}$$

$$M_{r,\gamma,\lambda} \stackrel{d}{=} \frac{W_\gamma}{W'_\gamma}\cdot\frac{1}{Z_{r,\lambda}^{1/\gamma}} \stackrel{d}{=} W_1\cdot\frac{R_\gamma}{W'_1 Z_{r,\lambda}^{1/\gamma}} \stackrel{d}{=} \frac{\Pi R_\gamma}{Z_{r,\lambda}^{1/\gamma}} \stackrel{d}{=} \frac{|X|\sqrt{2W_1}R_\gamma}{W'_1 Z_{r,\lambda}^{1/\gamma}}, \tag{12}$$

where $W_\gamma \stackrel{d}{=} W'_\gamma$, $W_1 \stackrel{d}{=} W'_1$, *the r.v.* R_γ *has the density* (3), *the r.v.* Π *has the Pareto distribution*: $\mathsf{P}(\Pi>x) = (x+1)^{-1}$, $x\geq 0$, *and in each term the involved r.v.:s are independent.*

With the account of the relation $R_\gamma \stackrel{d}{=} R_\gamma^{-1}$, from (12) we obtain the following statement.

Corollary 1 [21]. *Let $r \in (0,1]$, $\gamma \in (0,1]$, $\lambda > 0$. Then the d.f. $F(x;r,\gamma,\lambda)$ is mixed exponential*:

$$1 - F(x;r,\gamma,\lambda) = \int_0^\infty e^{-ux} dA(u), \quad x \geq 0,$$

where

$$A(u) = \mathsf{P}\big(W_1 R_\gamma Z_{r,\lambda}^{1/\gamma} < u\big), \quad u \geq 0,$$

and all the involved r.v:s are independent.

Remark 1. It is worth noting that the mixing distribution in Corollary 1 is mixed exponential itself.

Corollary 2 [21]. *Let $r \in (0,1]$, $\gamma \in (0,1]$, $\lambda > 0$. Then the d.f. $F(x;r,\gamma,\lambda)$ is infinitely divisible.*

Theorem 3 states that the limit distribution in Theorem 2 can be represented as a scale mixture of exponential or stable or Weibull or Pareto or folded normal laws. The corresponding product representations for the r.v. $M_{r,\gamma,\lambda}$ can be used for its computer simulation.

In practice, the asymptotic approximation $F(x;r,\lambda,\gamma)$ for the distribution of the extreme daily precipitation within a wet period proposed by Theorem 2 is adequate, if the ¡¡success probability¿¿ is small enough, that is, if on the average the wet periods are long enough.

It should be mentioned that the same mathematical reasoning can be used for the determination of the asymptotic distribution of the maximum daily precipitation within m wet periods with arbitrary finite $m \in \mathbb{N}$. Indeed, fix arbitrary positive $r_1,\ldots,r_m$ and $p \in (0,1)$. Let $N_{r_1,p}^{(1)},\ldots,N_{r_m,p}^{(m)}$ be independent random variables having the negative binomial distributions with parameters r_j, p, $j = 1,\ldots,m$, respectively. By the consideration of characteristic functions it can be easily verified that

$$N_{r_1,p}^{(1)} + \ldots + N_{r_m,p}^{(m)} \stackrel{d}{=} N_{r,p}, \tag{13}$$

where $r = r_1 + \ldots + r_m$. If all r_j coincide, then $r = mr_1$ and in accordance with theorem 2 and (13) the asymptotic distribution of the maximum daily precipitation within m wet periods has the form

$$F^{(m)}(x;r,\lambda,\gamma) = F(x;mr_1,\lambda,\gamma) = \left(\frac{\lambda x^\gamma}{1+\lambda x^\gamma}\right)^{mr_1}, \quad x \geq 0.$$

And if now m infinitely increases and simultaneously λ changes as $\lambda = cm$, $c \in (0,\infty)$, then, obviously,

$$\lim_{m\to\infty} F^{(m)}(x;r,\lambda,\gamma) = \lim_{m\to\infty} F(x;mr_1,cm,\gamma) = \lim_{n\to\infty}\left(1 - \frac{1}{1+cmx^\gamma}\right)^{mr_1} = e^{-\mu x^{-\gamma}}$$

with $\mu = (cr_1)^{-1}$, that is, the distribution function $F^{(m)}(x;r,\lambda,\gamma)$ of the maximum daily precipitation within m wet periods turns into the classical Fréchet (inverse Weibull) distribution (the type II distribution of extreme values).

2.2 Statistical estimation of the parameters r, λ and γ of the tempered Snedecor–Fisher distribution

Some methods of statistical estimation of the parameters r, λ and γ of the tempered Snedecor–Fisher distribution were proposed in [20, 21]. Here we briefly recall some ideas underlying the algorithms proposed in [20, 21] and give the corresponding formulas for practical computation.

From (10) it can be seen that the realization of the maximum likelihood method for the estimation of the parameters r, λ and γ inevitably assumes the necessity of numerical solution of a system of transcendental equations by iterative procedures without any guarantee that the resulting maximum is global. The closeness of the initial approximation to the true maximum likelihood point in the three-dimensional parameter set might give a hope that the terminal extreme point found by the numerical algorithm is global.

For rough estimation of the parameters, the following considerably simpler method can be used. The resulting rough estimates can be used as a starting point for the 'full' maximum likelihood algorithm mentioned above in order to ensure the closeness of the initial approximation to the true solution. The rough method is based on that the quantiles of the d.f. $F(x;r,\lambda,\gamma)$ can be written out explicitly. Namely, the quantile $x(\epsilon;r,\lambda,\gamma)$ of the d.f. $F(x;r,\lambda,\gamma)$ of order $\epsilon\in(0,1)$, that is, the solution of the equation $F(x;r,\lambda,\gamma)=\epsilon$ with respect to x, obviously has the form

$$x(\epsilon;r,\lambda,\gamma)=\left(\frac{\epsilon^{1/r}}{\lambda-\lambda\epsilon^{1/r}}\right)^{1/\gamma}.$$

Let at our disposal there be observations $\{X_{i,j}\}$, $i=1,\ldots,m$, $j=1,\ldots,m_i$, where i is the number of a wet period (the number of a sequence of rainy days), j is the number of a day in the wet sequence, m_i is the length of the ith wet sequence (the number of rainy days in the ith wet period), m is the total number of wet sequences, $X_{i,j}$ is the precipitation volume on the jth day of the ith wet sequence. Construct the sample $X_1^*,\ldots,X_m^*$ as

$$X_k^*=\max\{X_{k,1},\ldots,X_{k,m_k}\},\quad k=1,\ldots,m. \tag{14}$$

Let $X^*_{(1)},\ldots,X^*_{(m)}$ be order statistics constructed from the sample $X_1^*,\ldots,X_m^*$. Since we have three unknown parameters r, λ and γ, fix three numbers $0<p_1<p_2<p_3<1$ and construct the system of equations

$$X^*_{([mp_k])}=\left(\frac{p_k^{1/r}}{\lambda-\lambda p_k^{1/r}}\right)^{1/\gamma},\quad k=1,2,3$$

(here the symbol $[a]$ denotes the integer part of a number a).

This system can be solved by standard techniques. For example, first, the number $s\equiv\frac{1}{r}$ can be found numerically as the solution of the equation

$$Cs=\log\frac{1-p_3^s}{1-p_1^s}\log\frac{X^*_{([mp_1])}}{X^*_{([mp_2])}}-\log\frac{1-p_2^s}{1-p_1^s}\log\frac{X^*_{([mp_1])}}{X^*_{([mp_3])}},$$

where

$$C=\log\frac{X^*_{([mp_1])}}{X^*_{([mp_3])}}\log\frac{p_1}{p_2}-\log\frac{X^*_{([mp_1])}}{X^*_{([mp_2])}}\log\frac{p_1}{p_3}.$$

Having obtained the value of s, one can then find the values of γ and λ explicitly:

$$\widetilde{\gamma}=\frac{s(\log p_1-\log p_3)+\log(1-p_3^s)-\log(1-p_1^s)}{\log X^*_{([mp_1])}-\log X^*_{([mp_3])}}, \tag{15}$$

$$\widetilde{\lambda} = \frac{p_2^s}{(1-p_2^s)(X^*_{([mp_2])})^{\gamma}}. \tag{16}$$

As p_1, p_2 and p_3 one may take, say, $p_k = \frac{k}{4}$, $k = 1, 2, 3$. Or it is possible to fix a $\tau \in (0, \frac{1}{4})$, set $p_1 = \tau$, $p_2 = \frac{1}{2}$, $p_3 = 1 - \tau$ and choose a τ that provides the best fit of the estimated model d.f. $F(x; r, \lambda, \gamma)$ with the empirical d.f.

If the parameter r is known (for example, it is estimated beforehand while solving the problem of fitting the negative binomial distribution to the empirical data on the lengths of wet periods), then the parameters λ and γ can be estimated directly by formulas (15) and (16).

With known r, more accurate estimates of the parameters λ and γ can be also found by minimizing the discrepancy between the empirical and model d.f:s by the least squares techniques. Namely, this approach leads to the estimators

$$\widehat{\gamma} = \frac{(m-1)\sum_{i=1}^{m-1} c_i \log X^*_{(i)} - \sum_{i=1}^{m-1} \log X^*_{(i)} \sum_{i=1}^{m-1} c_i}{(m-1)\sum_{i=1}^{m-1} (\log X^*_{(i)})^2 - (\sum_{i=1}^{m-1} \log X^*_{(i)})^2}, \tag{17}$$

$$\widehat{\lambda} = \exp\left\{\frac{1}{m-1}\Big(\sum_{i=1}^{m-1} c_i - \widehat{\gamma}\sum_{i=1}^{m-1} \log X^*_{(i)}\Big)\right\}, \tag{18}$$

where

$$c_i = \log \frac{i^{1/r}}{m^{1/r} - i^{1/r}}.$$

So, the following algorithm can be recommended for the practical obtaining of the parameters r, λ and γ of the tempered Snedecor–Fisher distribution.

First, obtain the estimate of r when fitting the negative binomial distribution to the sample of observed durations of wet periods, say, by the maximum likelihood method or by the method of moments. The second method is, possibly, less accurate, but is considerably simpler. Indeed, let $D_1, \ldots, D_m$ be the initial sample of durations of m wet periods measured in days. Then the method of moments yields the following simple estimate for r:

$$\widetilde{r} = \frac{(\overline{D})^2}{S_D^2},$$

where

$$\overline{D} = \frac{1}{m}\sum_{j=1}^{m} D_j, \qquad S_D^2 = \frac{1}{m-1}\sum_{j=1}^{m}\big(D_j - \overline{D}\big)^2.$$

In practice, any estimation procedure can be used that is built in standard statistical program tools of fitting distributions, e. g., in Matlab, etc.

Second, obtain the estimates of γ and λ by formulas (17) and (18).

An example of practical application of this algorithm with r estimated by the Matlab tools is discussed in the next section.

2.3 The statistical fitting of the tempered Snedecor–Fisher distribution model to real data

In this section we present the results of statistical estimation of the distribution of extremal daily precipitation within a wet period by the methods described in the preceding section.

As the data, we use the observations of daily precipitation in Potsdam and Elista from 1950 to 2009, presented on Fig. 1.2.

First of all, notice that the Pareto distribution of daily precipitation volumes (see Figure 4) satisfies condition (6). Therefore, the asymptotic approximation provided by Theorem 2 can be applied to the statistical analysis of the real data.

The numerical results of estimation of the parameters are presented in Tables 1 and 2. In these table the first column indicates the censoring threshold: since the tempered Snedecor–Fisher distribution is an asymptotic model which is assumed to be more adequate with small ¡¡success probability¿¿, the estimates were constructed from differently censored samples which contain only those wet periods whose duration is no less than the specified threshold. The second column contains the correspondingly censored sample size. The third and fourth columns contain the sup-norm discrepancy between the empirical and fitted tempered Snedecor–Fisher distribution for two types of estimators (quantile and least squares) described above. The rest columns contain the corresponding values of the parameters estimated by these two methods. According to Tables 1 and 2, the best accuracy is attained when the censoring threshold equals 3 days for Elista and 5-6 days for Potsdam.

Table 1: Potsdam ($r = 0.847$)

Minimum duration	Sample size	Discrepancy, quantile metod: (15), (16)	Discrepancy, LS method: (17), (18)	$\widetilde{\lambda}$ quantile method	$\widehat{\lambda}$ LS method	$\widetilde{\gamma}$ quantile method	$\widehat{\gamma}$ LS method
1	3323	0.09	0.092	0.169	0.211	1.177	1.29
2	2066	0.045	0.065	0.0381	0.0538	1.76	1.709
3	1282	0.031	0.041	0.01	0.013	2.261	2.189
4	862	0.026	0.027	0.00487	0.00454	2.449	2.523
6	384	0.025	0.026	0.0016	0.0012	2.822	2.948
8	163	0.04	0.045	0.0007	0.0005	3.174	3.253
10	73	0.041	0.042	0.0003	0.0003	3.389	3.352
15	12	0.13	0.09	0.0014	0.0009	2.667	2.972

Table 2: Elista ($r = 0.876$)

Minimum duration	Sample size	Discrepancy, quantile metod: (15), (16)	Discrepancy, LS method: (17), (18)	$\widetilde{\lambda}$ quantile method	$\widehat{\lambda}$ LS method	$\widetilde{\gamma}$ quantile method	$\widehat{\gamma}$ LS method
1	2937	0.06	0.06	0.361	0.347	1.057	1.266
2	1374	0.049	0.055	0.108	0.1	1.42	1.576
3	656	0.041	0.045	0.0454	0.0377	1.706	1.898
4	319	0.051	0.06	0.0231	0.0272	1.899	1.94
6	77	0.07	0.075	0.0178	0.0144	2.017	2.186
7	42	0.15	0.01	0.0201	0.0206	1.974	2.184
8	22	0.12	0.14	0.0143	0.0355	2.003	1.769
10	10	0.17	0.16	0.0137	0.0377	2.154	1.798

The values of the parameter r coincide with those of the corresponding negative binomial distribution (see the Introduction): $r = 0.847$ for Elista and $r = 0.876$ for Potsdam.

Figures 5, 6, 7 and 8 illustrate the approximation of the empirical d.f. by the model d.f. $F(x; r, \gamma, \lambda)$ with γ and λ estimated by the 'rough' formulas (15) and (16) as well as by the least squares formulas (17) and (18). To illustrate the asymptotic character of the approximation $F(x; r, \gamma, \lambda)$ we consider a sort of censoring in which the censoring threshold is the minimum length of the wet periods which varies from 1 day (full sample) to 10 (Elista) and 15 (Potsdam) days.

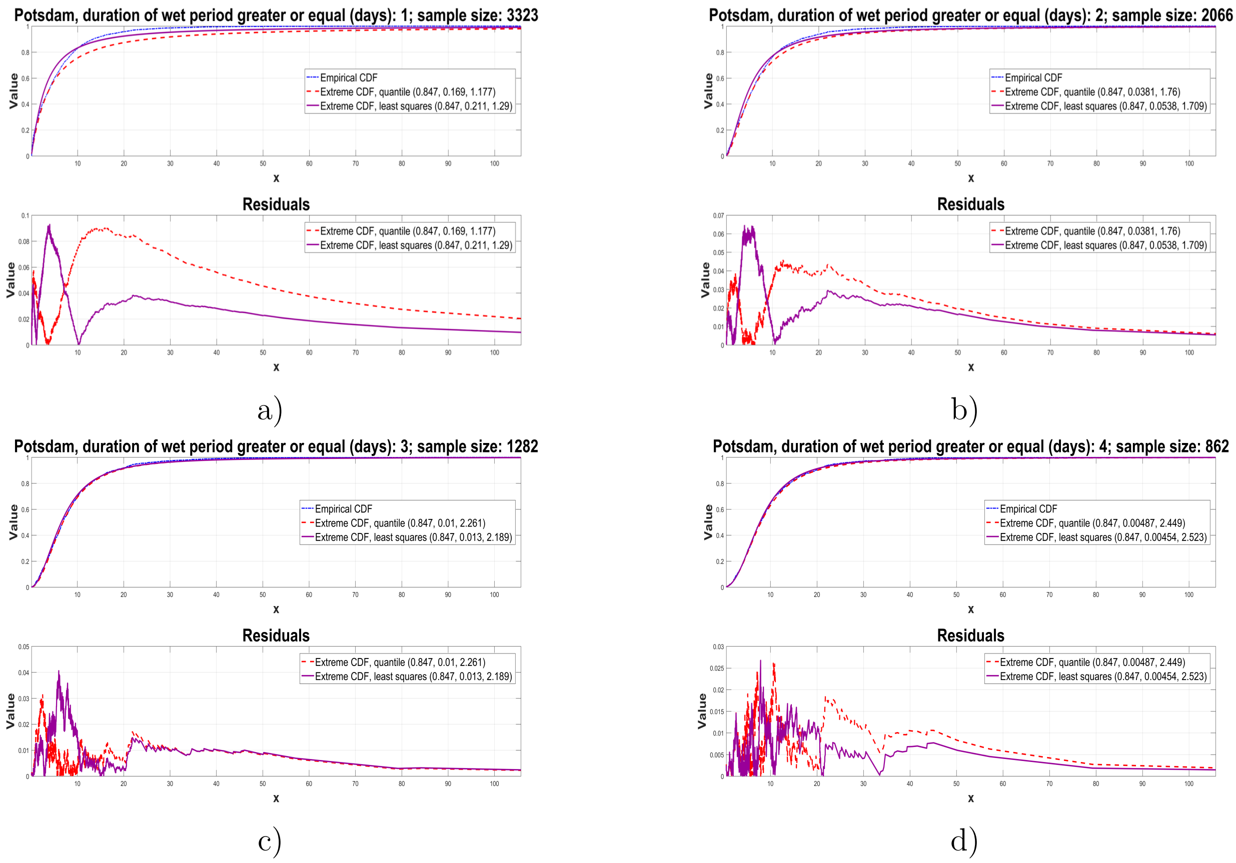


Fig. 5: Fitting the tempered Snedecor–Fisher distribution to the empirical data of maximum daily precipitation within a wet period in Potsdam with duration of wet periods no less than: a) one; b) two; c) three; d) four days

For each censoring threshold $h = \min_i m_i$ the sample is formed according to the rule (14). For each value of the threshold on the upper graph there are

- the empirical d.f. (continuous line);
- the d.f. $F(x; , r, \gamma, \lambda)$ with γ and λ estimated by the 'rough' formulas (15) and (16) (dash line);
- the d.f. $F(x; , r, \gamma, \lambda)$ with γ and λ estimated by the least squares formulas (17) and (18) (dotted line).

On the lower graph there is the discrepancy (the uniform distance) between the empirical d.f. and the fitted model d.f. The types of lines correspond to those on the upper graph.

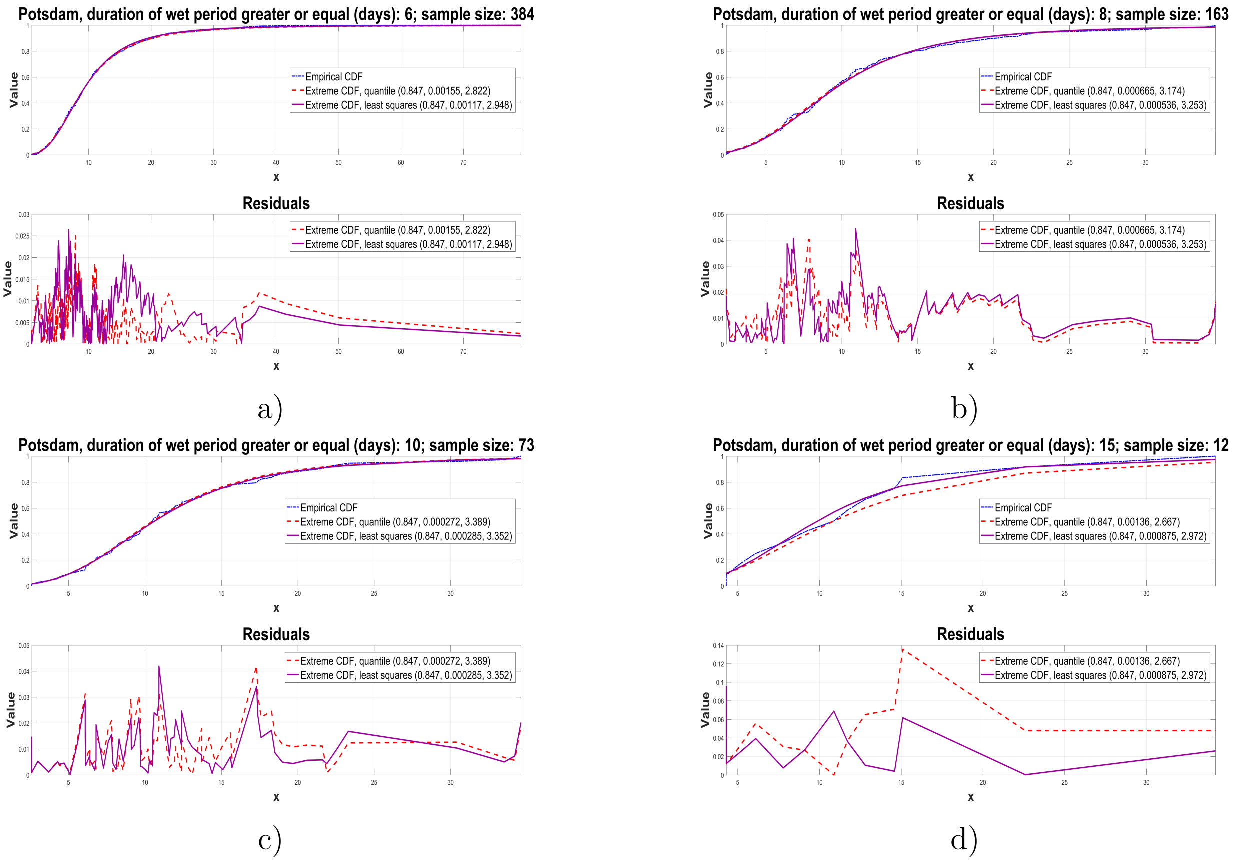


Fig. 6: Fitting the tempered Snedecor–Fisher distribution to the empirical data of maximum daily precipitation within a wet period in Elista with duration of wet periods no less than: a) six; b) eight; c) ten; d) fifteen days

First of all, from Figures 5–8 it is seen that the asymptotic model $F(x; r, \gamma, \lambda)$ provides very good approximation to the real statistical regularities in the behavior of extremal daily precipitation within a wet period. As this is so, the least squares formulas (17) and (18) yield more accurate estimates for the parameters of the model d.f.

It should be also noted that these figures illustrate the dependence of the accuracy of the approximation on the censoring threshold and the censored sample size. The approximation is satisfactory even if the censoring threshold h is greater or equal to three and the censored sample size is grater than 150. As this is so, the influence of the threshold h on the accuracy is more noticeable than that of the sample size.

2.4 The statistical analysis of real data

The approach to the determination of an abnormally heavy daily precipitation is methodically similar to the classical techniques of dealing with extreme observations [1]. The novelty of the proposed method is in an accurate specification of the mathematical model of the distribution of extreme daily precipitation which turned out to be the tempered Snedecor–Fisher distribution.

The algorithm of determination of an abnormally heavy daily precipitation is as follows. First, the parameters of the distribution function $F(x; r, \lambda, \gamma)$ are estimated from the historical data. Second, a small positive number ε is fixed. Third, the $(1 - \varepsilon)$-quantile

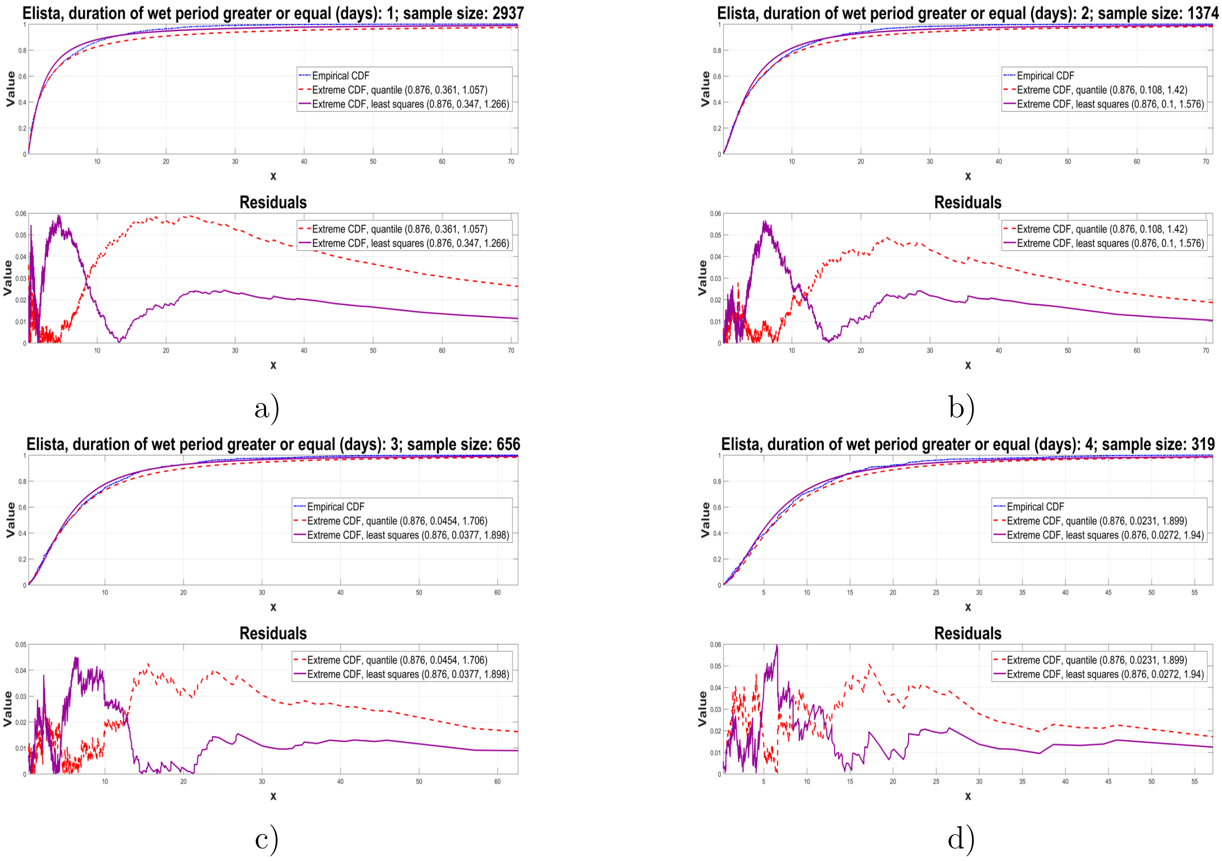


Fig. 7: Fitting the tempered Snedecor–Fisher distribution to the empirical data of maximum daily precipitation within a wet period in Elista with duration of wet periods no less than: a) one; b) two; c) three; d) four days

$\tau(1-\varepsilon; r, \lambda, \gamma)$ of the distribution function $F(x; r, \lambda, \gamma)$ is calculated.

If the maximum value, say, X of the daily precipitation volume within some wet period exceeds $\tau(1-\varepsilon; r, \lambda, \gamma)$, then X is regarded as 'suspicious' to be an outlier, that is, to be abnormally large.

It is easy to see that the the probability of the error of the first kind (occurring in the case where a 'regularly large' maximum value is erroneously recognized as an abnormally large outlier) for this test is approximately equal to ε.

The application of this test to real data is illustrated by Figures 9 and 10. On these figures the lower horizontal line corresponds to the threshold equal to the quantile of the fitted tempered Snedecor–Fisher distribution of order 0.9. The middle and upper lines correspond to the quantiles of orders 0.95 and 0.99 respectively.

Figure 9 contains all data. For the sake of vividness, on Figure 10 only one, maximum, daily precipitation is exposed for each wet period. From Figure 10 it is seen that during 58 years (from 1950 to 2007) in Potsdam there were 13 wet periods containing abnormally heavy maximum daily precipitation volumes (at 99% threshold) and 69 wet periods containing abnormally heavy maximum daily precipitation volumes (at 95% threshold). Other maxima were 'regular'. During the same period in Elista there were only 2 wet periods containing abnormally heavy maximum daily precipitation volumes (at 99% threshold) and 40 wet periods containing abnormally heavy maximum daily precipitation volumes (at 95% threshold). Other maxima were 'regular'. The proportion of abnormal maxima exceeding

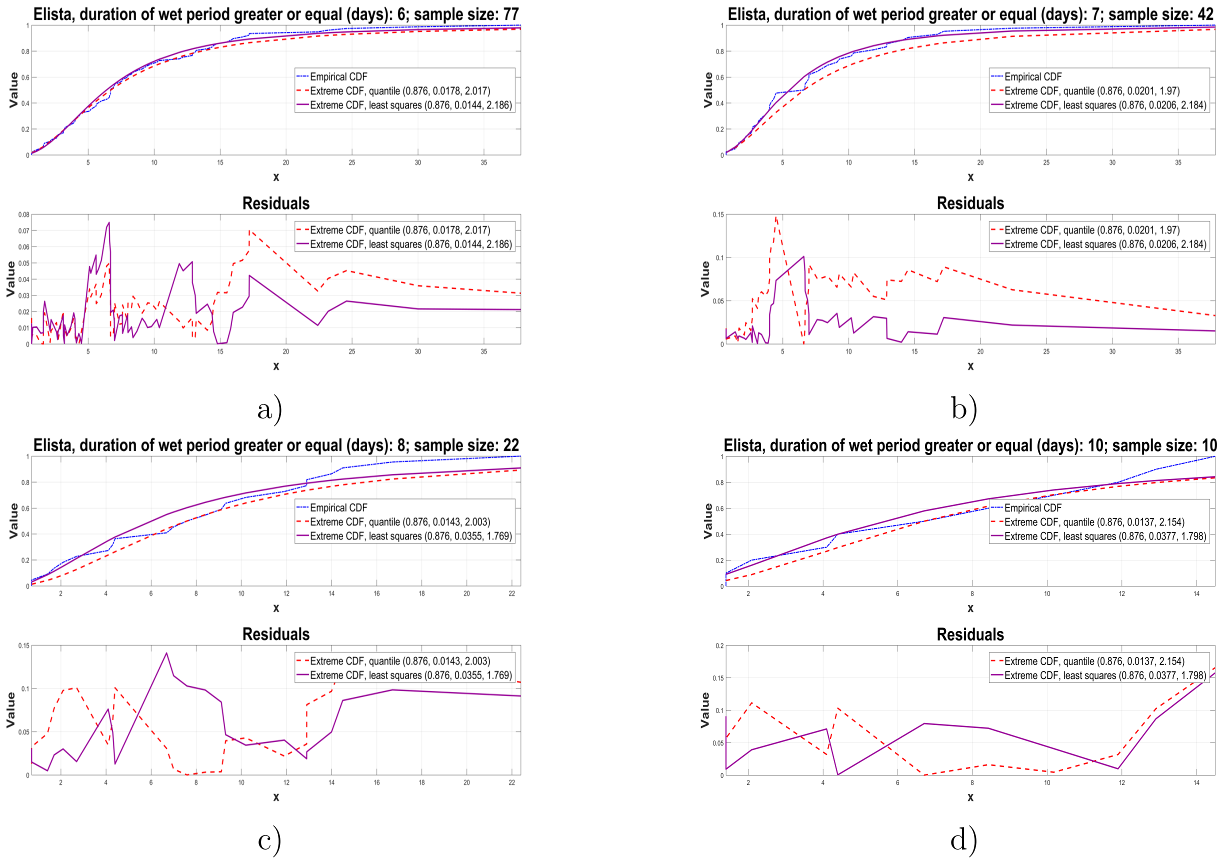


Fig. 8: Fitting the tempered Snedecor–Fisher distribution to the empirical data of maximum daily precipitation within a wet period in Elista with duration of wet periods no less than: a) six; b) seven; c) eight; d) ten days

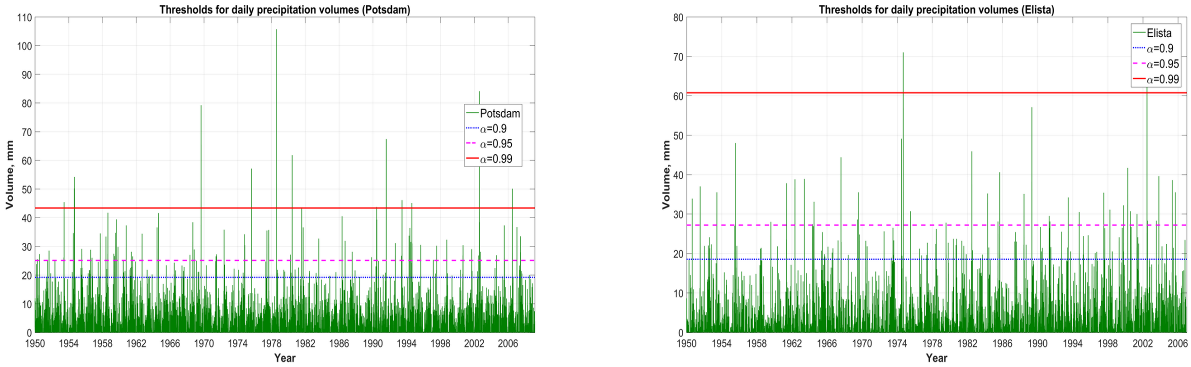
99% and 95% thresholds in Potsdam is quite adequate (the latter is approximately five times greater than the former) whereas in Elista this proportion is noticeably different. Perhaps, this can be explained by the fact that, for Elista, heavy rains are rare events.

3 The tests for a total precipitation volume to be abnormally extremal based on the homogeneity test of a sample from the gamma distribution

3.1 The tests based on the beta and Snedecor–Fisher distributions

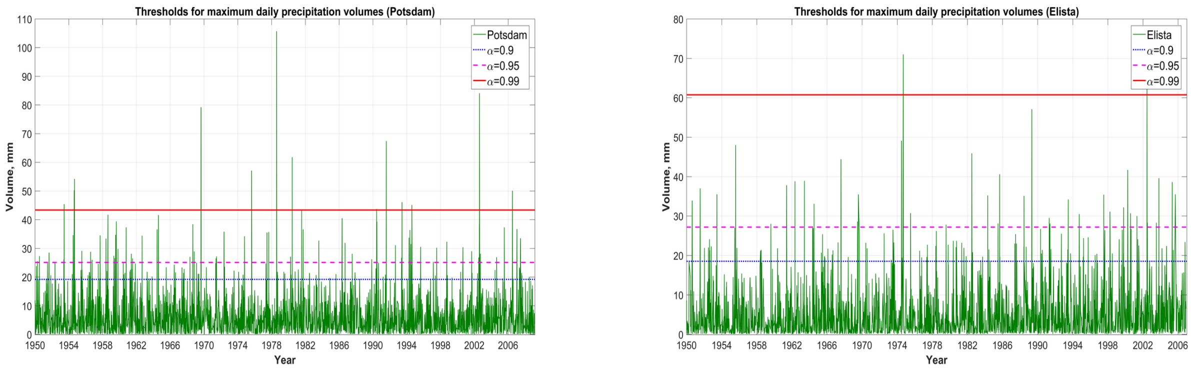
Here we will propose some algorithms of testing the hypotheses that a *total precipitation volume during a wet period* is abnormally extremal within a certain time horizon. Moreover, our approach makes it possible to consider relatively abnormally extremal volumes and absolutely abnormally extremal volumes for a given time horizon.

Let $m \in \mathbb{N}$ and $G_{r,\mu}^{(1)}, G_{r,\mu}^{(2)}, \ldots, G_{r,\mu}^{(m)}$ – be independent r.v.'s having the same gamma distribution with shape parameter $r > 0$ and scale parameter $\mu > 0$. In [26] it was suggested



a) b)

Fig. 9: Testing maximum daily precipitation within a wet period for abnormal heaviness: a) Potsdam; b) Elista, all data.



a) b)

Fig. 10: Testing maximum daily precipitation within a wet period for abnormal heaviness: a) Potsdam; b) Elista, data containing only maximum daily precipitation for every wet period.

to use the distribution of the ratio

$$R^* = \frac{G^{(1)}_{r,\mu}}{G^{(1)}_{r,\mu} + G^{(2)}_{r,\mu} + \ldots + G^{(m)}_{r,\mu}} \stackrel{d}{=} \frac{G^{(1)}_{r,1}}{G^{(1)}_{r,1} + G^{(2)}_{r,1} + \ldots + G^{(m)}_{r,1}} \tag{19}$$

as a heuristic model of the distribution of the extremely large precipitation volume based on the assumption that fluctuations of *daily* precipitation follow the gamma distribution. As we have seen in Sect. 1.2, the gamma model for the distribution of daily precipitation volume is less adequate than the Pareto model. Here we will modify the technique proposed in [26] and make it more adequate and justified.

For this purpose we will use the following auxiliary result. Consider a sequence of r.v.'s $W_1, W_2, \ldots$ Let $N_1, N_2, \ldots$ be natural-valued r.v.'s such that for every $n \in \mathbb{N}$ the r.v. N_n is independent of the sequence $W_1, W_2, \ldots$ In the following statement the convergence is meant as $n \to \infty$.

THEOREM 3 [8, 9]. *Assume that there exist an infinitely increasing* (*convergent to zero*) *sequence of positive numbers* $\{b_n\}_{n\geq 1}$ *and a r.v.* W *such that*

$$b_n^{-1} W_n \Longrightarrow W.$$

If there exist an infinitely increasing (convergent to zero) sequence of positive numbers $\{d_n\}_{n\geq 1}$ *and a r.v.* N *such that*

$$d_n^{-1} b_{N_n} \Longrightarrow N, \tag{20}$$

then

$$d_n^{-1} W_{N_n} \Longrightarrow W \cdot N, \tag{21}$$

where the r.v.'s on the right-hand side of (21) *are independent. If, in addition,* $N_n \longrightarrow \infty$ *in probability and the family of scale mixtures of the d.f. of the r.v.* W *is identifiable, then condition* (20) *is not only sufficient for* (21)*, but is necessary as well.*

Let $X_1, X_2, \ldots$ be daily precipitation volumes on wet days. For $k \in \mathbb{N}$ denote $S_k = X_1 + \ldots + X_k$. The statistical analysis of the observed data shows that the average daily precipitation volume on wet days is finite:

$$\frac{1}{n}\sum_{j=1}^{n} X_j \Longrightarrow a \in (0, \infty). \tag{22}$$

Figure 9 illustrates the stabilization of the cumulative averages of daily precipitation volumes as n grows in Potsdam (continuous line) and Elista (dash line), and thus, the practical validity of assumption (22).

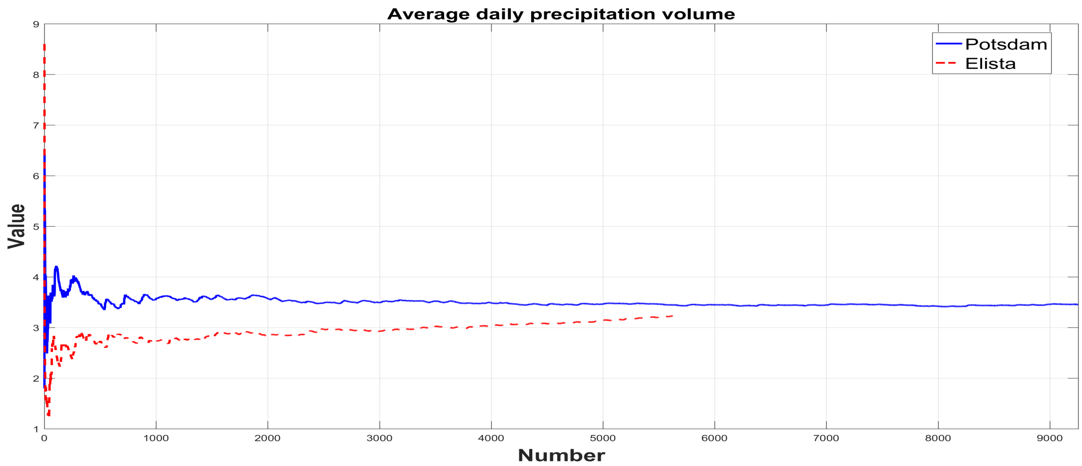


Fig. 11: stabilization of the cumulative averages of daily precipitation volumes as n grows in Potsdam (continuous line) and Elista (dash line).

It should be emphasized that in (22) we *do not* assume that $X_1, X_2, \ldots$ are independent.

Let $r > 0$, $\mu > 0$, $q \in (0, 1)$, $n \in \mathbb{N}$. Let the r.v. N_{r,p_n} have the negative binomial distribution with parameters r and $p_n = \min\{q, \mu/n\}$. Using the properties of characteristic functions it is easy to make sure that

$$n^{-1} N_{r,p_n} \Longrightarrow G_{r,\mu} \stackrel{d}{=} \tfrac{1}{\mu} G_{r,1} \tag{23}$$

as $n \to \infty$.

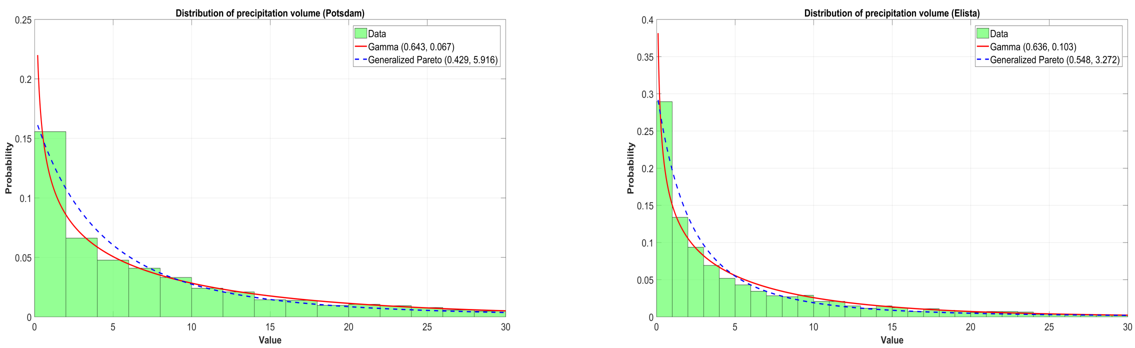
From (23) and theorem 3 we obtain the following analog of the law of large numbers for negative binomial random sums.

Theorem 4. *Assume that the daily precipitation volumes on wet days $X_1, X_2, \ldots$ satisfy condition* (22). *Let the numbers $r > 0$, $q \in (0,1)$ and $\mu > 0$ be arbitrary. For each $n \in \mathbb{N}$, let the r.v. N_{r,p_n} have the negative binomial distribution with parameters r and $p_n = \min\{q, \mu/n\}$. Assume that the r.v.'s N_{r,p_n} are independent of the sequence $X_1, X_2, \ldots$ Then*

$$n^{-1}S_{N_{r,p_n}} \Longrightarrow aG_{r,\mu} \stackrel{d}{=} \tfrac{a}{\mu}G_{r,1}$$

as $n \to \infty$.

Therefore, with the account of the excellent fit of the negative binomial model for the duration of a wet period (see Fig. 4), with rather small p_n, the gamma distribution can be regarded as an adequate and theoretically well-based model for the total precipitation volume during a (long enough) wet period. This theoretical conclusion based on the negative binomial model for the distribution of duration of a wet period is vividly illustrated by the empirical data as shown on Figure 3.1 where the histograms of total precipitation volumes in Potsdam (a) and Elista (b) and the fitted gamma distributions are shown. For comparison, the densities of the best generalized Pareto distributions are also presented. It can be seen that even the best fitted Pareto distributions demonstrate worse fit than the gamma distribution.



a) b)

Fig. 12: The histograms of total precipitation volumes in Potsdam (a) and Elista (b) and the fitted gamma and generalized Pareto distributions.

Let $m \in \mathbb{N}$ and $G_{r,\mu}^{(1)}, G_{r,\mu}^{(2)}, \ldots, G_{r,\mu}^{(m)}$ be independent r.v.'s having the same gamma distribution with parameters $r > 0$ and $\mu > 0$.

Consider the relative contribution of the r.v. $G_{r,\mu}^{(1)}$ to the sum $G_{r,\mu}^{(1)} + G_{r,\mu}^{(2)} + \ldots + G_{r,\mu}^{(m)}$:

$$R = \frac{G_{r,\mu}^{(1)}}{G_{r,\mu}^{(1)} + G_{r,\mu}^{(2)} + \ldots + G_{r,\mu}^{(m)}} \stackrel{d}{=} \frac{G_{r,1}^{(1)}}{G_{r,1}^{(1)} + G_{r,1}^{(2)} + \ldots + G_{r,1}^{(m)}} \stackrel{d}{=}$$

$$\stackrel{d}{=} \left(1 + \frac{1}{G_{r,1}^{(1)}}(G_{r,1}^{(2)} + \ldots + G_{r,1}^{(m)})\right)^{-1} \stackrel{d}{=} \left(1 + \frac{G_{(m-1)r,1}}{G_{r,1}}\right)^{-1}, \tag{24}$$

where the gamma-distributed r.v.'s on the right hand side are independent. So, the r.v. R characterizes the relative precipitation volume for one (long enough) wet period with respect to the total precipitation volume registered for m wet periods.

The distribution of the r.v. R is completely determined by the distribution of the ratio of two *independent* gamma-distributed r.v.'s. To find the latter, denote $k = (m-1)r$ and obtain

$$\frac{G_{k,1}}{G_{r,1}} = \frac{k}{r}\cdot\left(\frac{r}{k}\cdot\frac{G_{k,1}}{G_{r,1}}\right) \stackrel{d}{=} \frac{k}{r}\cdot Q_{k,r},$$

where $Q_{k,r}$ is the r.v. having the Snedecor–Fisher distribution determined for $k>0$, $r>0$ by the Lebesgue density

$$f_{k,r}(x) = \frac{\Gamma(k+r)}{\Gamma(k)\Gamma(r)}\left(\frac{k}{r}\right)^{k}\frac{x^{k-1}}{(1+\frac{k}{r}x)^{k+r}}, \quad x\geq 0, \tag{25}$$

(as is known, $Q_{k,r} \stackrel{d}{=} rG_{k,1}(kG_{r,1})^{-1}$, where the r.v.'s $G_{k,1}$ and $G_{r,1}$ are independent (see, e. g., [6], p. 32)). It should be noted that the particular value of the scale parameter is insignificant. For convenience, it is assumed equal to one.

So, $R \stackrel{d}{=} \big(1+\frac{k}{r}Q_{k,r}\big)^{-1}$, and, as is easily made sure by standard calculation using (25), the distribution of the r.v. R is determined by the density

$$p(x;k,r) = \frac{\Gamma(k+r)}{\Gamma(r)\Gamma(k)}(1-x)^{k-1}x^{r-1}, \quad 0\leq x\leq 1,$$

that is, it is the beta distribution with parameters $k=(m-1)r$ and r.

Then the test for the homogeneity of an independent sample of size m consisting of the gamma-distributed observations of total precipitation volumes during m wet periods with known γ based on the r.v. R looks as follows. Let $V_1,\ldots,V_m$ be the total precipitation volumes during m wet periods and, moreover, $V_1\geq V_j$ for all $j\geq 2$. Calculate the quantity

$$SR = \frac{V_1}{V_1+\ldots+V_m}$$

(SR means ¡¡Sample R¿¿). From what was said above it follows that under the hypothesis H_0: ¡¡the precipitation volume V_1 under consideration *is not* abnormally large¿¿ the r.v. SR has the beta distribution with parameters $k=(m-1)r$ and r. Let $\varepsilon\in(0,1)$ be a small number, $\beta_{k,r}(1-\varepsilon)$ be the $(1-\varepsilon)$-quantile of the beta distribution with parameters $k=(m-1)r$ and r. If $SR>\beta_{k,r}(1-\varepsilon)$, then the hypothesis H_0 must be rejected, that is, the volume V_1 of precipitation during one wet period must be regarded as abnormally large. Moreover, the probability of erroneous rejection of H_0 is equal to ε.

Instead of R, the quantity

$$R_0 = \frac{(m-1)G^{(1)}_{r,\mu}}{G^{(2)}_{r,\mu}+\ldots+G^{(m)}_{r,\mu}} \stackrel{d}{=} \frac{k}{r}\frac{G_{r,\mu}}{G_{k,\mu}} \stackrel{d}{=} \frac{k}{r}\frac{G_{r,1}}{G_{k,1}} \stackrel{d}{=} Q_{r,k}$$

can be considered. Then, as is easily seen, the r.v.'s R and R_0 are related by the one-to-one correspondence

$$R = \frac{R_0}{m-1+R_0} \quad\text{or}\quad R_0 = \frac{(m-1)R}{1-R},$$

so that the homogeneity test for a sample from the gamma distribution equivalent to the one described above and, correspondingly, the test for a precipitation volume during a wet period to be abnormally large, can be based on the r.v. R_0 which has the Snedecor–Fisher distribution with parameters r and $k=(m-1)r$.

Namely, again let $V_1, \ldots, V_m$ be the total precipitation volumes during m wet periods and, moreover, $V_1 \geq V_j$ for all $j \geq 2$. Calculate the quantity

$$SR_0 = \frac{(m-1)V_1}{V_2 + \ldots + V_m}$$

(SR_0 means ¡¡Sample R_0¿¿). From what was said above it follows that under the hypothesis H_0: ¡¡the precipitation volume V_1 under consideration *is not* abnormally large¿¿ the r.v. SR has the Snedecor–Fisher distribution with parameters r $k = (m-1)r$. Let $\varepsilon \in (0,1)$ be a small number, $q_{r,k}(1-\varepsilon)$ be the $(1-\varepsilon)$-quantile of the Snedecor–Fisher distribution with parameters r $k = (m-1)r$. If $SR_0 > q_{r,k}(1-\varepsilon)$, then the hypothesis H_0 must be rejected, that is, the volume V_1 of precipitation during one wet period must be regarded as abnormally large. Moreover, the probability of erroneous rejection of H_0 is equal to ε.

Let l be a natural number, $1 \leq l < m$. It is worth noting that, unlike the test based on the statistic R, the test based on R_0 can be modified for testing the hypothesis H_0': ¡¡the precipitation volumes $V_{i_1}, V_{i_2}, \ldots, V_{i_l}$ *do not* make an abnormally large cumulative contribution to the total precipitation volume $V_1 + \ldots + V_m$¿¿. For this purpose denote

$$T_l = V_{i_1} + V_{i_2} + \ldots + V_{i_l}, \qquad T = V_1 + V_2 + \ldots + V_m$$

and consider the quantity

$$SR_0' = \frac{(m-l)T_l}{l(T - T_l)}.$$

In the same way as it was done above, it is easy to make sure that

$$SR_0' \stackrel{d}{=} \frac{(m-l)G_{lr,l}}{lG_{(m-l)r,1}} \stackrel{d}{=} Q_{lr,(m-l)r}.$$

Let $\varepsilon \in (0,1)$ be a small number, $q_{lr,(m-1)r}(1-\varepsilon)$ be the $(1-\varepsilon)$-quantile of the Snedecor–Fisher distribution with parameters lr $k = (m-l)r$. If $SR_0' > q_{lr,(m-l)r}(1-\varepsilon)$, then the hypothesis H_0' must be rejected, that is, the cumulative contribution of the precipitation volumes $V_{i_1}, V_{i_2}, \ldots, V_{i_l}$ into the total precipitation volume $V_1 + \ldots + V_m$ must be regarded as abnormally large. Moreover, the probability of erroneous rejection of H_0' is equal to ε.

The examples of application of the test for a total precipitation volume within a wet period to be abnormally large will be discussed in Section 3.3.

3.2 The application of the tests to the statistical analysis of time series. Relative and absolute abnormality

In this section we present the results of the application of the test SR_0 to the analysis of the time series of daily precipitation observed in Potsdam and Elista from 1950 to 2009.

First of all it should be emphasized that the parameter m of the Snedecor–Fisher distribution of the test statistic SR_0 is tightly connected with the time horizon, the abnormality of precipitation within which is studied. Indeed, the average duration of a wet/dry period (or, which is the same, the average distance between the first days of successive wet periods) in Potsdam turns out to be $5.804 \approx 6$ days. So, one observation of a total precipitation during a wet period, on the average, corresponds to approximately 6 days. This means, that, for example, the value $m = 5$ corresponds to approximately one month on the time axis, the value $m = 15$ corresponds to approximately 3 months (a season), the value $m = 60$ corresponds to approximately one year.

Second, it is important that the test for a total precipitation volume during one wet period to be abnormally large can be applied to the observed time series in a moving mode. For this purpose a *window* (a set of successive observations) should be determined. The number of observations in this set, say, m, is called the *window width*. The observations within a window constitute the sample to be analyzed. After the test has been performed for a given position of the window, the window moves rightward by one observation so that the leftmost observation at the previous position of the window is excluded from the sample and the observation next to the rightmost observation is added to the sample. The test is performed once more and so on. It is clear that each fixed observation falls in exactly m successive windows. Two cases are possible: (i) the fixed observation is recognized as abnormally large within *each* of m windows containing this observation and (ii) the fixed observation is recognized as abnormally large within *at least one* of m windows containing this observation. In the case (i) the observation will be called *absolutely abnormally large* with respect to a given time horizon (approximately equal to $m \cdot 5.804 \approx 6m$ days). In the case (ii) the observation will be called *relatively abnormally large* with respect to a given time horizon.

Of course, these definitions admit intermediate cases where the observation is recognized as abnormally large for $q \cdot m$ windows with $q \in [\frac{1}{m}, 1]$.

3.3 The statistical analysis of real data

The results of the application of the test for a total precipitation volume during one wet period to be abnormally large based on SR_0 in the moving mode are shown on Figures 13 – 16 (Potsdam) and 17 – 20 (Elista) for different time horizons. It is seen that at relatively small time horizons the test yields non-trivial and unobvious conclusions. However, as the time horizon increases, the results of the test become more expected. At small time horizons there are some big precipitation volumes that are not recognized as abnormal. At large time horizons there are almost no ‘regular’ big precipitation volumes at significance level $\alpha = 0.05$ whereas at the smaller significance level $\alpha = 0.01$ there are some ‘regular’ big precipitation volumes which are thus not recognized as abnormal.

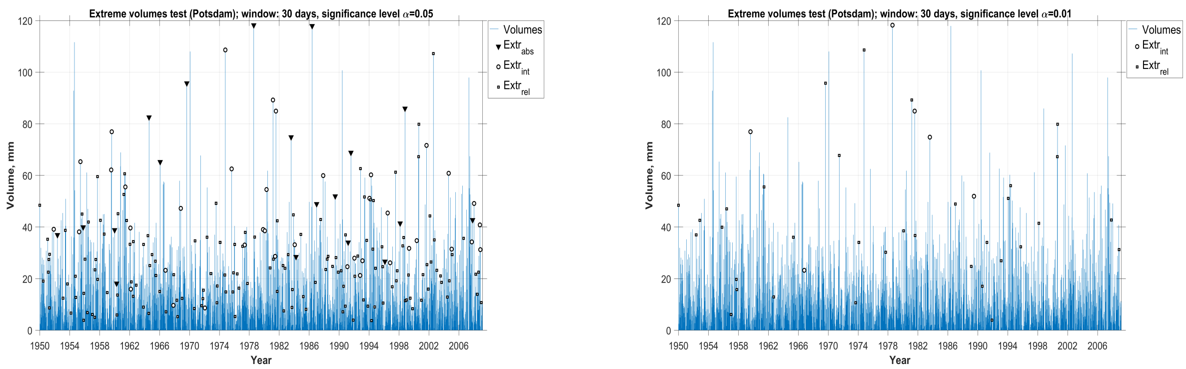


Fig. 13: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Potsdam, time horizon = 30 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

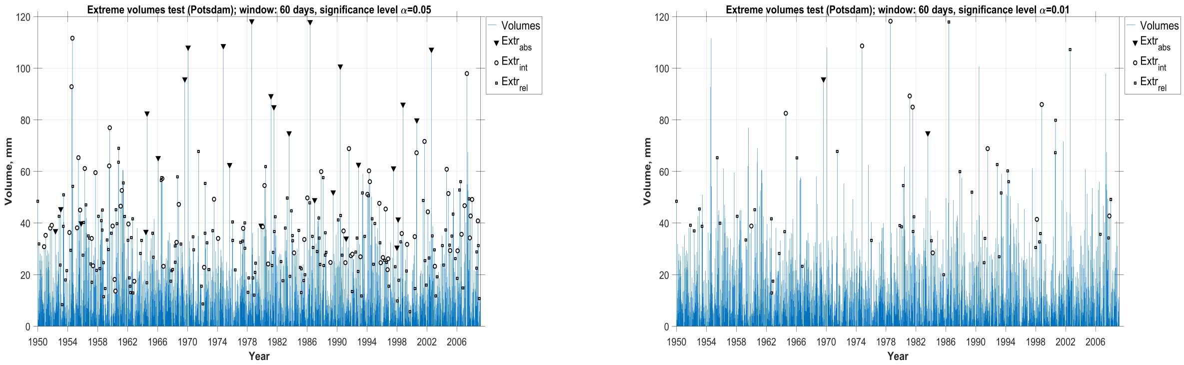


Fig. 14: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Potsdam, time horizon = 60 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

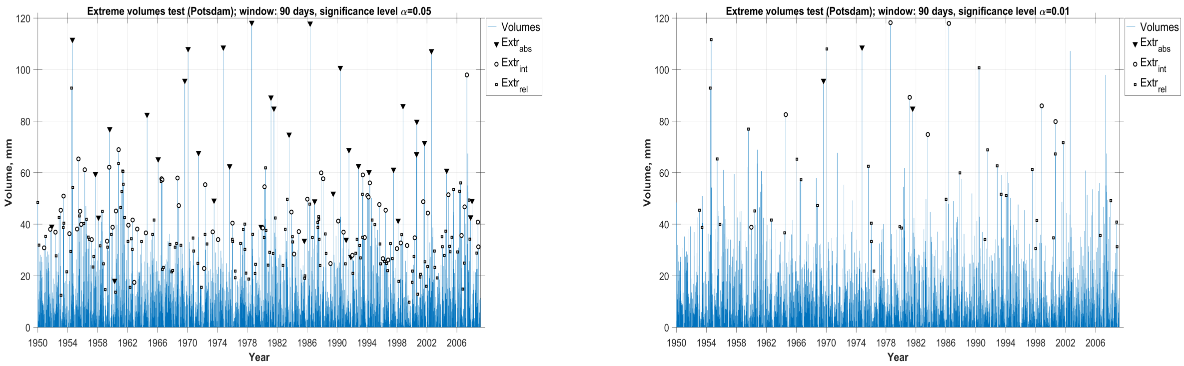


Fig. 15: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Potsdam, time horizon = 90 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

Acknowledgements

The research was supported by the Russian Foundation for Basic Research (project 17-07-00851) and the Ministry of Education and Science of the Russian Federation (No. 538.2018.5).

References

[1] *Embrechts P., Klüppelberg K., Mikosch T.* Modeling Extremal Events. – Berlin–New York: Springer, 1998.

[2] *Gleser L. J.* The gamma distribution as a mixture of exponential distributions // American Statistician, 1989. Vol. 43. P. 115–117.

[3] *Gnedenko B. V., Korolev V. Yu.* Random Summation: Limit Theorems and Applications. – Boca Raton: CRC Press, 1996.

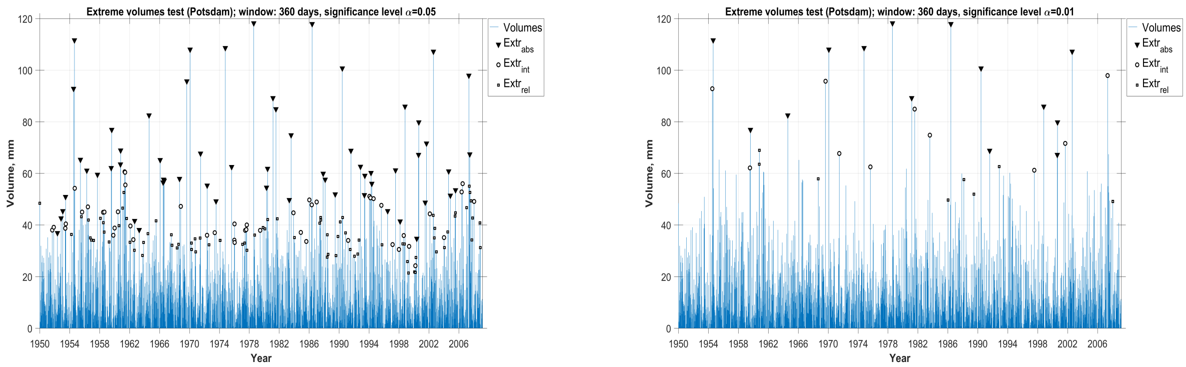


Fig. 16: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Potsdam, time horizon = 360 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

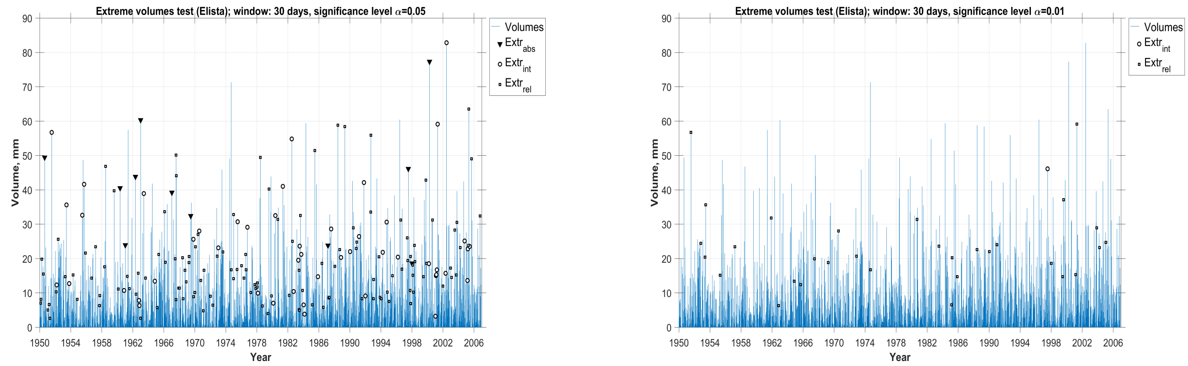


Fig. 17: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Elista, time horizon = 30 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

[4] *Gorhenin A. K.* On some mathematical and program methods for the construction of structure models of information flows // Informatics and Its Applications, 2017. Vol. 11. No. 1. P. 58–68.

[5] *Greenwood M., Yule G. U.* An inquiry into the nature of frequency-distributions of multiple happenings, etc. // J. Roy. Statist. Soc., 1920. Vol. 83. P. 255–279.

[6] *Johnson, N. L., Kotz S., Balakrishnan N.* Continuous Univariate Distributions, Vol. 2 (2nd Edition). – New York: Wiley, 1995.

[7] *Kingman J. F. C.* Poisson processes. – Oxford: Clarendon Press, 1993.

[8] *Korolev V. Yu.* Convergence of random sequences with independent random indexes. I // Theory of Probability and its Applications, 1994. Vol. 39. No. 2. P. 313–333.

[9] *Korolev V. Yu.* Convergence of random sequences with independent random indexes. II // Theory of Probability and iits Applications, 1995. Vol. 40. No. 4. P. 770–772.

[10] *Korolev V. Yu.* On convergence of distributions of compound Cox processes to stable laws // Theory of Probability and its Applications, 1999. Vol. 43. Vol. 4. P. 644–650.

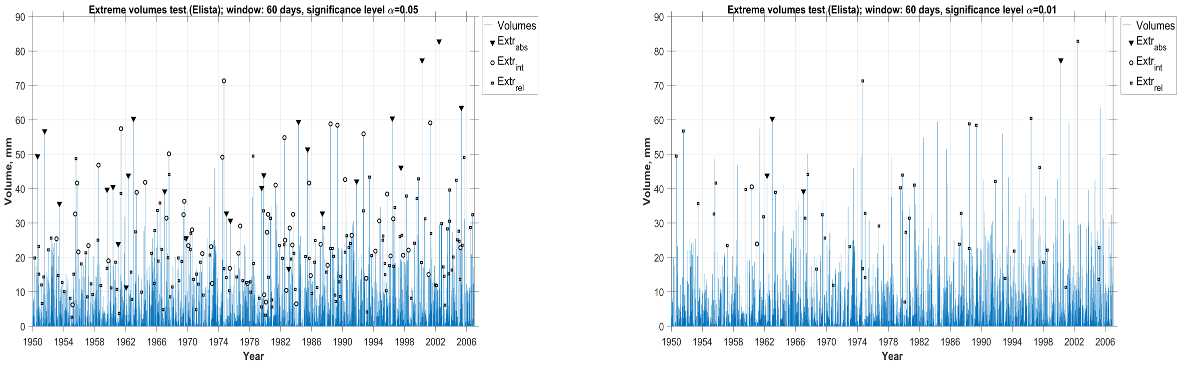


Fig. 18: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Elista, time horizon = 60 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

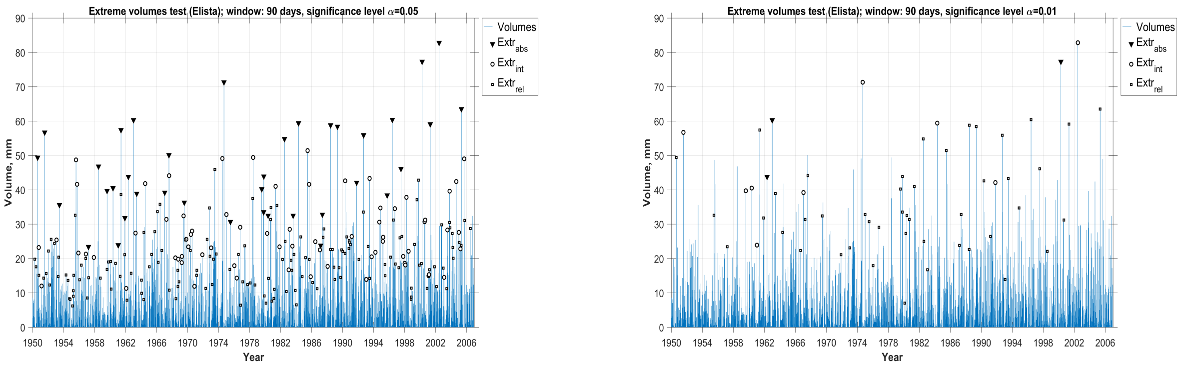


Fig. 19: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Elista, time horizon = 90 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

[11] *Korolev V. Yu., Sokolov I. A.* Mathematical Models of Inhomogeneous Flows of Extremal Events. – Moscow: Torus-Press, 2008.

[12] *Korolev V. Yu., Bening V. E., Shorgin S. Ya.* Mathematical Foundations of Risk Theory. 2nd Edition. – Moscow: FIZMATLIT, 2011.

[13] *Korolev V. Yu.* Limit distributions for doubly stochastically rarefied renewal processes and their properties // Theory Probab. Appl., 2016. Vol. 61. No. 4. P. 753–773.

[14] *Korolev V. Yu., Korchagin A. Yu., Zeifman A. I.* Poisson theorem for the scheme of Bernoulli trials with random probability of success and a discrete analog of the Weibull distribution // Informatics and Its Applications, 2016. Vol. 10. No. 4. P. 11–20.

[15] *Korolev V. Yu., Korchagin A. Yu., Zeifman A. I.* On doubly stochastic rarefaction of renewal processes // AIP Conference Proceedings, 2017. Vol. 1863. No. 090010.

[16] *Korolev V. Yu., Zeifman A. I.* A note on mixture representations for the Linnik and Mittag-Leffler distributions and their applications // Journal of Mathematical Sciences, 2017. Vol. 218. No. 3. P. 314–327.

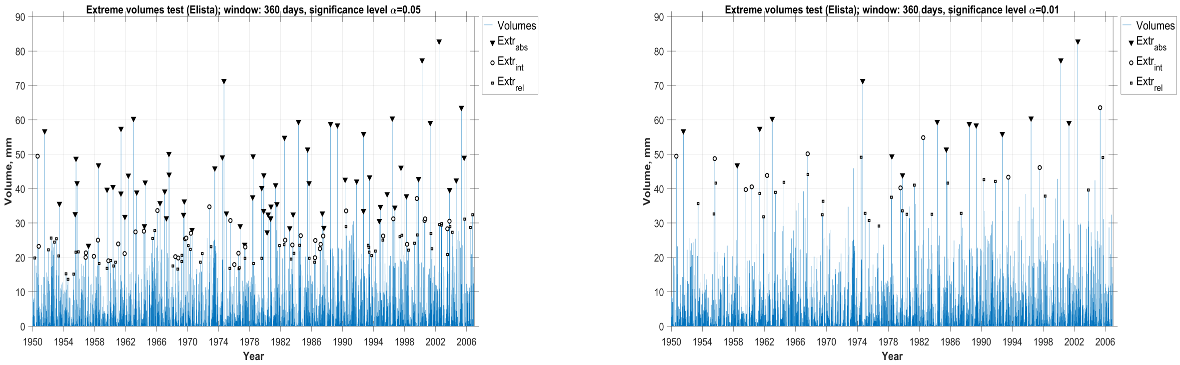


Fig. 20: Absolutely (triangles), relatively (squares) and intermediate (circles) abnormal precipitation volumes, Elista, time horizon = 360 days, significance levels $\alpha = 0.05$ (left) and $\alpha = 0.01$ (right).

[17] *Korolev V. Yu., Zeifman A. I.* Convergence of statistics constructed from samples with random sizes to the Linnik and Mittag-Leffler distributions and their generalizations // Journal of Korean Statistical Society, 2017. Vol. 46. No. 2. P. 161–181. Available online 25 July 2016. Also available on arXiv:1602.02480v1 [math.PR].

[18] *Korolev V. Yu.* Analogs of the Gleser theorem for negative binomial and generalized gamma distributions and some their applications // Informatics and Its Applications, 2017. Vol. 11. No. 3. P. 2–17.

[19] *Korolev V. Yu., Gorshenin A. K., Gulev S. K., Belyaev K. P., Grusho A. A.* Statistical Analysis of Precipitation Events // AIP Conference Proceedings, 2017. Vol. 1863. No. 090011 (4 p.). Also available on arXiv:1705.11055 [math.PR], 31 May, 2017.

[20] *Korolev V. Yu., Gorshenin A. K.* The probability distribution of extreme precipitation // Doklady Earth Sciences, 2017. Vol. 477. No. 2. P. 1461–1466.

[21] *Korolev V. Yu., Gorshenin A. K.* On the asymptotic approximation to the probability distribution of extremal precipitation // arXiv:1706.00308 [math.PR], 04 June, 2017.

[22] *Kotz S., Ostrovskii I. V.* A mixture representation of the Linnik distribution // Statistics and Probability Letters, 1996. Vol. 26. P. 61–64.

[23] *Stacy E. W.* A generalization of the gamma distribution // Annals of Mathematical Statistics, 1962. Vol. 33. P. 1187–1192.

[24] *Zaks L. M., Korolev V. Yu.* Generalized variance gamma distributions as limit laws for random sums // Informatics and Its Applications, 2013. Vol. 7. No. 1. P. 105–115.

[25] *Zolina O., Simmer C., Kapala A., Gulev S. K.* On the robustness of the estimates of centennial-scale variability in heavy precipitation from station data over Europe // Geophys. Res. Lett., 2005. Vol. 32. L14707; DOI:10.1029/2005GL023231.

[26] *Zolina O., Simmer C., Belyaev K., Kapala A., Gulev S.* Improving estimates of heavy and extreme precipitation using daily records from European rain gauges // Journal of Applied Meteorology, 2009. Vol. 10. P. 701–716.

[27] *Zolina O., Simmer C., Belyaev K., Gulev S., Koltermann P.* Changes in the duration of European wet and dry spells during the last 60 years // Journal of Climate, 2013. Vol. 26. P. 2022–2047.

[28] *Zolotarev V. M.* One-Dimensional Stable Distributions. – Providence, R.I.: American Mathematical Society, 1986.