

Article

Probability Models and Statistical Tests for Extreme Precipitation Based on Generalized Negative Binomial Distributions

Victor Korolev ^{1,2,3,4}  and Andrey Gorshenin ^{1,2,3,*} 

¹ Moscow Center for Fundamental and Applied Mathematics, Lomonosov Moscow State University, 119991 Moscow, Russia; vkorolev@cs.msu.ru

² Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, 119991 Moscow, Russia

³ Federal Research Center “Computer Science and Control” of the Russian Academy of Sciences, 119333 Moscow, Russia

⁴ Department of Mathematics, School of Science, Hangzhou Dianzi University, Hangzhou 310018, China

* Correspondence: agorshenin@frcscs.ru

Received: 4 April 2020; Accepted: 14 April 2020; Published: 16 April 2020



Abstract: Mathematical models are proposed for statistical regularities of maximum daily precipitation within a wet period and total precipitation volume per wet period. The proposed models are based on the generalized negative binomial (GNB) distribution of the duration of a wet period. The GNB distribution is a mixed Poisson distribution, the mixing distribution being generalized gamma (GG). The GNB distribution demonstrates excellent fit with real data of durations of wet periods measured in days. By means of limit theorems for statistics constructed from samples with random sizes having the GNB distribution, asymptotic approximations are proposed for the distributions of maximum daily precipitation volume within a wet period and total precipitation volume for a wet period. It is shown that the exponent power parameter in the mixing GG distribution matches slow global climate trends. The bounds for the accuracy of the proposed approximations are presented. Several tests for daily precipitation, total precipitation volume and precipitation intensities to be abnormally extremal are proposed and compared to the traditional PoT-method. The results of the application of this test to real data are presented.

Keywords: precipitation; limit theorems; statistical test; generalized negative binomial distribution; generalized gamma distribution; asymptotic approximations; extreme order statistics; random sample size

MSC: 60F05; 62G30; 62E20; 62P12; 65C20

1. Introduction

In this paper, we continue the research we started in [1,2]. We develop the mathematical models for statistical regularities in precipitation proposed in the papers mentioned above. We consider the models for the statistical regularities in the duration of a wet period, maximum daily precipitation within a wet period and total precipitation volume per wet period. The base for the models is the generalized negative binomial (GNB) introduced in the recent paper [3]. The GNB distribution is a mixed Poisson distribution, the mixing distribution being generalized gamma (GG). The results of fitting the GNB distribution to real data are presented and demonstrate excellent concordance of the GNB model with the empirical distribution of the duration of wet periods measured in days. Based on this GNB model, asymptotic approximations are proposed for 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 scale mixture of the gamma distribution in which the scale factor has the Weibull distribution, whereas the asymptotic approximation for the total precipitation volume for a wet period turns out to be the GG distribution. These asymptotic approximations are deduced using limit theorems for statistics constructed from samples with random sizes having the GNB distribution. The bounds for the accuracy of the proposed approximations are discussed theoretically and illustrated statistically. The proposed approximations appear to be very accurate. Based on these models, two approaches are proposed to the definition of abnormally extremal precipitation. These approaches can be regarded as a further development of those proposed in [2].

The importance of the problem of modeling statistical regularities in extreme precipitation is indisputable. Understanding climate variability and trends at relatively large time horizons is of crucial importance for long-range business, say, agricultural projects and forecasting of risks of water floods, dry spells and other natural disasters. Modeling regularities and trends in heavy and extreme daily precipitation is important for understanding climate variability and change at relatively small or medium time horizons. However, these models are much more uncertain as compared to those derived for mean precipitation or total precipitation during a wet period. In [4], a detailed review of this phenomenon is presented and it is noted that, at least for the European continent, most results hint at a growing intensity of heavy precipitation over the last decades.

In [2], we proposed a rather reasonable approach to the unambiguous (algorithmic) determination of extreme or abnormally heavy total precipitation for a wet period. This approach was based on the NB model for the duration of wet periods measured in days, and, as a consequence, on the distribution of the total precipitation volume during a wet period. This approach has some advantages. First, estimates of the parameters of the total precipitation are weakly affected by the accuracy of the daily records and are less sensitive to missing values. Second, the corresponding mathematical models are theoretically based on limit theorems of probability theorems that yield unambiguous asymptotic approximations, which are used as adequate mathematical models. Third, this approach gives an unambiguous algorithm for the determination of extreme or abnormally heavy total precipitation that does not involve statistical significance problems owing to the low occurrence of such (relatively rare) events.

The problem of the construction of a statistical test for the precipitation volume to be abnormally large can be mathematically formalized as follows. Let $m \geq 2$ be a natural number and consider a sample of m positive observations X_1, X_2, \dots, X_m . With finite m , among X_i 's there is always an extreme observation, say, X_1 , such that $X_1 \geq X_i$, $i = 1, 2, \dots, m$. 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.

To construct a test for distinguishing between these two cases for abnormally extreme daily precipitation, we use the fact that the distribution of the maximum daily precipitation per wet period is a tempered scale mixture of the gamma distribution in which the scale factor has the Weibull distribution. According to this model, a daily precipitation volume is considered to be abnormally extremal if it exceeds a certain (pre-defined) quantile of this distribution.

As regards testing for anomalous extremeness of total precipitation volume during a wet period, we use the GG distribution as the model of statistical regularities of its behavior. The theoretical grounds for this model are provided by the law of large numbers for random sums in which the number of summands has the GNB distribution. It turns out that, as compared to the ordinary negative binomial (NB) model (see [2]), the additional exponent power parameter in the corresponding GG distribution matches slow global climate trends. 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 GG 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. These tests are scale-free and depend only on the easily estimated shape parameter of the GNB distribution and the time-scale parameter determining the denominator in the fractional contribution of a wet period under consideration. The tests appeared to be applicable not only to total precipitation volumes over wet periods but also to the precipitation intensities (the ratios of total precipitation volumes per wet periods to the durations of the corresponding wet periods measured in days).

2. Generalized Negative Binomial Model for the Duration of Wet Periods

The main results of this paper strongly rely on the GNB model, a wide and flexible family of discrete distributions that are mixed Poisson laws with the mixing GG distribution. Namely, we say that a random variable $N_{r,\gamma,\mu}$ ($r > 0$, $\gamma \in \mathbb{R}$ and $\mu > 0$) has the generalized negative binomial distribution, if

$$\mathbb{P}(N_{r,\gamma,\mu} = k) = \frac{1}{k!} \int_0^\infty e^{-z} z^k g^*(z; r, \gamma, \mu) dz, \quad k = 0, 1, 2, \dots, \quad (1)$$

where $g^*(z; r, \gamma, \mu)$ is the density of GG distribution:

$$g^*(x; r, \gamma, \mu) = \frac{|\gamma| \mu^r}{\Gamma(r)} x^{r-1} e^{-\mu x^\gamma}, \quad x \geq 0, \quad (2)$$

with $\gamma \in \mathbb{R}$, $\mu > 0$, $r > 0$. The GNB distributions seem to be very promising in the statistical description of many real phenomena, being very convenient and almost universal models.

It is necessary to explain why this combination of the mixed and mixing distributions is considered. First of all, the Poisson kernel is used as mixed for the following reasons. Pure Poisson processes can be regarded as the best models of stationary (time-homogeneous) chaotic flows of events [5]. Recall that the attractiveness of a Poisson process as a model of homogeneous discrete stochastic chaos is due to at least two circumstances. First, Poisson processes are point processes characterized by the time intervals between successive points that are independent random variables (r.v.'s) with one and the same exponential distribution, and, as is well known, the exponential distribution possesses the maximum differential entropy among all absolutely continuous distributions concentrated on the nonnegative half-line with finite expectations, whereas the entropy is a natural and convenient measure of uncertainty. Second, the points forming the Poisson process are uniformly distributed along the time axis in the sense that for any finite time interval $[t_1, t_2]$, $t_1 < t_2$, the conditional joint distribution of the points of the Poisson process that fall into the interval $[t_1, t_2]$ under the condition that the number of such points is fixed and equals, say, n , coincides with the joint distribution of the order statistics constructed from an independent sample of size n from the uniform distribution on $[t_1, t_2]$, whereas the uniform distribution possesses the maximum differential entropy among all absolutely continuous distributions concentrated on finite intervals and very well corresponds to the conventional impression of an absolutely unpredictable random variable (see, e.g., [5,6]). However, in actual practice, as a rule, the parameters of the chaotic stochastic processes are influenced by poorly predictable «extrinsic» factors, which can be regarded as stochastic so that most reasonable probabilistic models of non-stationary (time-non-homogeneous) chaotic point processes are doubly stochastic Poisson processes, also called Cox processes (see, e.g., [5,7,8]). These processes are defined as Poisson processes with stochastic intensities. Such processes proved to be adequate models in insurance [5,7,8], financial mathematics [9], physics [10] and many other fields. Their one-dimensional distributions are mixed Poisson.

In order to have a flexible model of a mixing distribution that is “responsible” for the description of statistical regularities of the manifestation of external stochastic factors, we suggest to use the GG

distributions defined by the density (2). The class of GG distributions was first described as a unitary family in 1962 by E. Stacy [11] as the class of probability distributions simultaneously containing both Weibull and gamma distributions. The family of GG distributions contains practically all the most popular absolutely continuous distributions concentrated on the non-negative half-line. In particular, the family of GG distributions contains:

- The gamma distribution ($\gamma = 1$) and its special cases
 - The exponential distribution ($\gamma = 1, r = 1$),
 - The Erlang distribution ($\gamma = 1, r \in \mathbb{N}$),
 - The chi-square distribution ($\gamma = 1, \mu = \frac{1}{2}$);
- The Nakagami distribution ($\gamma = 2$);
- The half-normal (folded normal) distribution (the distribution of the maximum of a standard Wiener process on the interval $[0, 1]$) ($\gamma = 2, r = \frac{1}{2}$);
- The Rayleigh distribution ($\gamma = 2, r = 1$);
- The chi-distribution ($\gamma = 2, \mu = 1/\sqrt{2}$);
- The Maxwell distribution (the distribution of the absolute values of the velocities of molecules in a dilute gas) ($\gamma = 2, r = \frac{3}{2}$);
- The Weibull–Gnedenko distribution (the extreme value distribution of type III) ($r = 1, \gamma > 0$);
- The (folded) exponential power distribution ($\gamma > 0, r = \frac{1}{\gamma}$);
- The inverse gamma distribution ($\gamma = -1$) and its special case
 - The Lévy distribution (the one-sided stable distribution with the characteristic exponent $\frac{1}{2}$ – the distribution of the first hit time of the unit level by the Brownian motion) ($\gamma = -1, r = \frac{1}{2}$);
- The Fréchet distribution (the extreme value distribution of type II) ($r = 1, \gamma < 0$)

and other laws. The limit point of the class of GG-distributions is

- The log-normal distribution ($r \rightarrow \infty$).

GG distributions are widely applied in many practical problems. There are dozens of papers dealing with the application of GG-distributions as models of regularities observed in practice. Apparently, the popularity of GG-distributions is due to the fact that most of them can serve as adequate asymptotic approximations, since all the representatives of the class of GG-distributions listed above appear as limit laws in various limit theorems of probability theory in rather simple limit schemes. Below we will formulate a general limit theorem (an analog of the law of large numbers) for random sums of independent r.v.'s in which the GG-distributions are limit laws. It is worth noting that the GG distribution and its limit cases give a general form of the exponential distribution of rank 1 for the scale parameter.

In [1], the data registered in so climatically different points as Potsdam (Brandenburg, Germany) and Elista (Kalmykia, Russia) was analyzed, and it was demonstrated that the fluctuations of the numbers of successive wet days with very high confidence fit the NB distribution with shape parameters $r = 0.847$ and $r = 0.876$, respectively. In the same paper, a schematic attempt was undertaken to explain this phenomenon by the fact that NB distributions can be represented as mixed Poisson laws with mixing gamma distributions whereas, as it already has been mentioned, the Poisson distribution is the best model for the discrete stochastic chaos and the mixing distribution accumulates the stochastic influence of factors that can be assumed exogenous with respect to the local system under consideration.

The NB distributions are special cases of the GNB distributions. This family of discrete distributions is very wide and embraces Poisson distributions (as limit points corresponding to a degenerate mixing distribution), NB (Polya) distributions including geometric distributions

(corresponding to the gamma mixing distribution, see [12]), Sichel distributions (corresponding to the inverse gamma mixing distributions, see [13,14]), Weibull–Poisson distributions (corresponding to the Weibull mixing distributions, see [15]) and many other types supplying descriptive statistics with many flexible models. More examples of mixed Poisson laws can be found in [8,16].

It is quite natural to expect that, having introduced one more free parameter into the pure negative binomial model, namely, the power parameter in the exponent of the original gamma mixing distribution, instead of the negative binomial model one might obtain a more flexible GNB model that provides an even better fit with the statistical data of the durations of wet days. The analysis of the real data shows that this is indeed so.

In Figures 1 and 2 there are the histograms constructed from real data of 3323 wet periods in Potsdam and 2937 wet periods in Elista. On the same pictures, there are the graphs of the fitted NB distribution (that is, the GNB distribution with $\gamma = 1$) and the fitted GNB distribution with additionally adjusted scale and power parameters. For vividness, in the GNB model, the value of the shape parameter r was taken the same as that obtained for the NB model and equal to 0.876 for Elista and 0.847 for Potsdam. For the “fine tuning” of the GNB models with these fixed values of r , the minimization of the ℓ_1 -norm of the difference between the histogram and the fitted GNB model was used. In Appendix A, the Algorithm A1 of for the computation of GNB probabilities by the minimization of the ℓ_1 , ℓ_2 and ℓ_∞ -norms of the difference between the histogram and the fitted GNB model is presented.

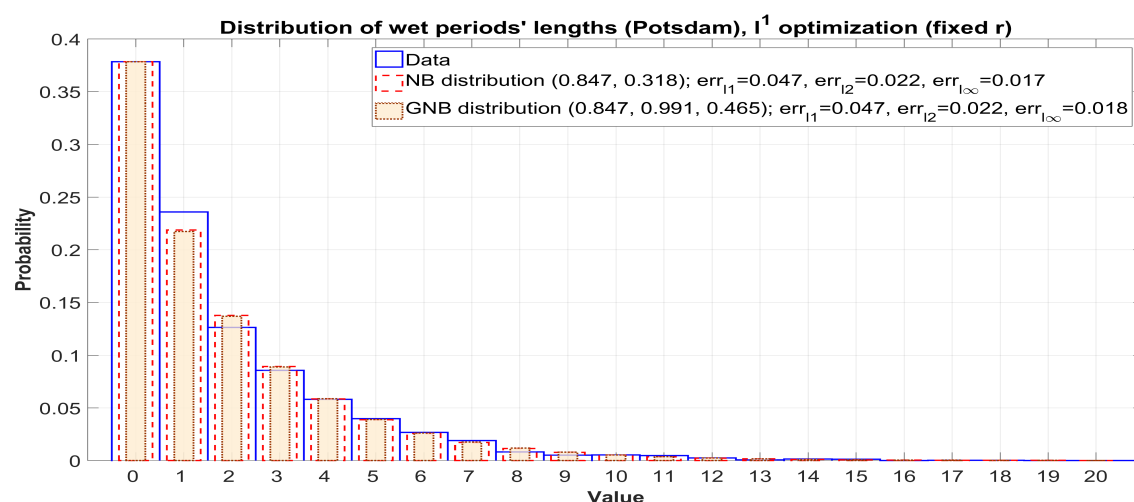


Figure 1. The histograms constructed from real data of 3320 wet periods in Potsdam and the fitted negative binomial (NB) and generalized negative binomial (GNB) models, ℓ_1 -distance minimization.

The analytic and asymptotic properties of the GNB distributions were studied in [3]. In particular, it was shown in that paper that the GNB distribution with shape parameter and exponent power parameter less than one is actually mixed geometric. The mixed geometric distributions were introduced and studied in [17] (also see [15,18]). A mixed geometric distribution can be interpreted in terms of the Bernoulli trials as follows. First, as a result of some “preliminary” experiment the value of some r.v. 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” mixed Poisson r.v. 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 wet period. So, since the GG-distribution is a more general and, hence, a more flexible model than the “pure” gamma

distribution, there arises a hope that the GNB distribution could provide an even better goodness of fit to the statistical regularities in the duration of wet periods than the “pure” NB binomial distribution.

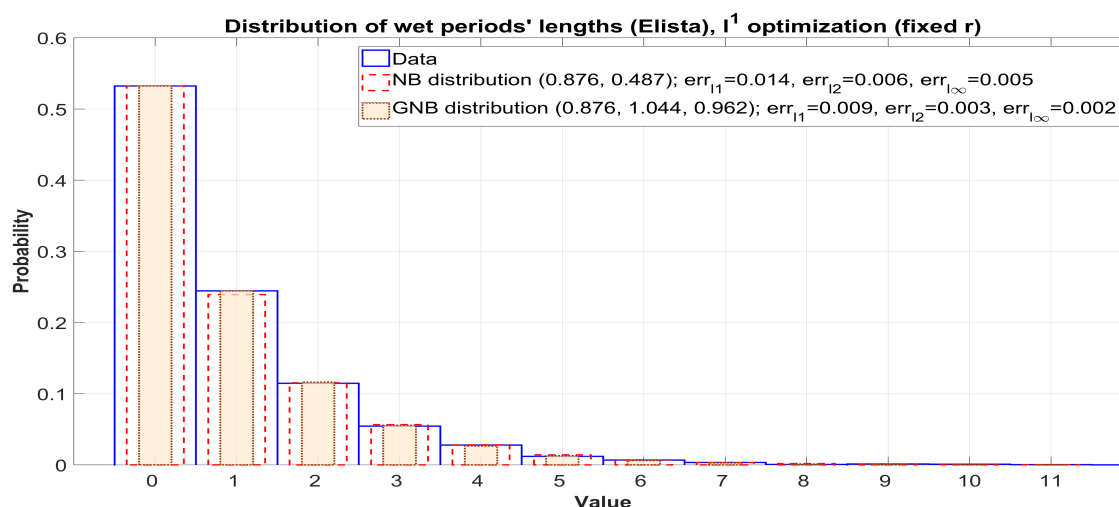


Figure 2. The histograms constructed from real data of 2937 wet periods in Elista and the fitted NB and GNB models, ℓ_1 -distance minimization.

3. Notation, Definitions and Mathematical Preliminaries

In the paper, conventional notation is used. The symbols $\stackrel{d}{=}$ and \implies denote the coincidence of distributions and convergence in distribution, respectively.

In what follows, for brevity and convenience, the results will be presented in terms of r.v.'s with the corresponding distributions. It will be assumed that all the r.v.'s are defined on the same probability space $(\Omega, \mathfrak{F}, P)$.

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

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

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

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

A GG-distribution is the absolutely continuous distribution defined by the density (Equation (2)). The distribution function (d.f.) corresponding to the density $g^*(x; r, \gamma, \mu)$ will be denoted $F^*(x; r, \gamma, \mu)$.

The properties of GG-distributions are described in [11,19]. An r.v. with the density $g^*(x; r, \gamma, \mu)$ will be denoted $\overline{G}_{r,\gamma,\mu}$. It can be easily made sure that

$$\overline{G}_{r,\gamma,\mu} \stackrel{d}{=} G_{r,\mu}^{1/\gamma}, \quad (3)$$

and hence,

$$(\overline{G}_{r,\gamma,\mu})^\gamma \stackrel{d}{=} G_{r,\mu} \quad (4)$$

For convenience, for an r.v. with the Weibull distribution, a particular case of GG-distributions corresponding to the density $g^*(x; 1, \gamma, 1)$ and the d.f. $[1 - e^{-x^\gamma}] \mathbf{1}(x \geq 0)$ with $\gamma > 0$, we will use a special notation W_α , that is, $W_\gamma \stackrel{d}{=} \overline{G}_{1,\gamma,1}$. Thus, $G_{1,1} \stackrel{d}{=} W_1$. The density $g^*(x; 1, \alpha, 1)$ with $\alpha < 0$ defines the Fréchet or inverse Weibull distribution. It is easy to see that

$$W_1^{1/\gamma} \stackrel{d}{=} W_\gamma. \quad (5)$$

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

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

A particular case of the NB 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 :

$$P(N_{1,p} = k) = p(1-p)^k, \quad k = 0, 1, 2, \dots$$

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

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

Let Y be an 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 $M = N_{1,Y}$, that is, $M(\omega) = N_{1,Y(\omega)}(\omega)$ for any $\omega \in \Omega$. The distribution

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

of the r.v. M will be called Y -mixed geometric [17].

It is well known that the negative binomial distribution is a mixed Poisson distribution with the gamma mixing distribution [12] (also see [20]): for any $r > 0$, $p \in (0, 1)$ and $k \in \{0\} \cup \mathbb{N}$ we have

$$\frac{\Gamma(r+k)}{k! \Gamma(r)} \cdot p^r (1-p)^k = \frac{1}{k!} \int_0^{\infty} e^{-z} z^k g(z; r, \mu) dz, \quad (6)$$

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

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.)

$$f(t; \alpha, \theta) = \exp \left\{ -|t|^\alpha \exp \left\{ -\frac{1}{2} i \pi \theta \alpha \operatorname{sign} t \right\} \right\}, \quad t \in \mathbb{R},$$

where $0 < \alpha \leq 2$, $|\theta| \leq \min\{1, \frac{2}{\alpha} - 1\}$, will be respectively denoted $F(x; \alpha, \theta)$ and $f(x; \alpha, \theta)$ (see, e.g., [21]). An r.v. with the d.f. $F(x; \alpha, \theta)$ will be denoted $S_{\alpha, \theta}$.

To symmetric, strictly stable distributions, there corresponds the value $\theta = 0$. To one-sided strictly stable distributions concentrated on the nonnegative halfline, there correspond the values $\theta = 1$ and $0 < \alpha \leq 1$. The pairs $\alpha = 1$, $\theta = \pm 1$ correspond to the distributions degenerate in ± 1 , respectively. All the other strictly stable distributions are absolutely continuous. Stable densities cannot explicitly be 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 ($\alpha = \frac{1}{2}$, $\theta = -1$). Expressions of stable densities in terms of the Fox functions (generalized Meijer G-functions) can be found in [22,23].

The standard normal d.f. will be denoted $\Phi(x)$,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy, \quad y \in \mathbb{R}.$$

An r.v. with the d.f. $\Phi(x)$ will be denoted X . The folded or half-normal distribution is the distribution of the r.v. $|X|$. It can be easily verified that $S_{2,0} \stackrel{d}{=} \sqrt{2}X$.

In [24,25], it was demonstrated that if $\gamma \in (0, 1]$, then

$$W_\gamma \stackrel{d}{=} W_1 \cdot S_{\gamma,1}^{-1} \quad (7)$$

with the r.v.'s on the right-hand side being independent.

For $r \in (0, 1)$ let $G_{r,1}$ and $G_{1-r,1}$ be independent gamma-distributed r.v.'s. Let $\mu > 0$. Introduce the r.v.

$$Z_{r,\mu} = \frac{\mu(G_{r,1} + G_{1-r,1})}{G_{r,1}} \stackrel{d}{=} \mu Z_{r,1} \stackrel{d}{=} \mu \left(1 + \frac{1-r}{r} Q_{1-r,r}\right), \quad (8)$$

where $Q_{1-r,r}$ is the r.v. with the Snedecor–Fisher distribution defined by the probability density

$$q(x; 1-r, r) = \frac{(1-r)^{1-r} r^r}{\Gamma(1-r)\Gamma(r)} \cdot \frac{1}{x^r [r + (1-r)x]}, \quad x \geq 0. \quad (9)$$

In the paper [26], it was shown that any gamma distribution with shape parameter no greater than one is mixed exponential. For convenience, we formulate this result as the following lemma.

Lemma 1 ([26]). *The density of a gamma distribution $g(x; r, \mu)$ with $0 < r < 1$ can be represented as*

$$g(x; r, \mu) = \int_0^\infty z e^{-zx} p(z; r, \mu) dz,$$

where

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

is the density of the r.v. $Z_{r,\mu}$ introduced above. In other words, if $0 < r < 1$, then

$$G_{r,\mu} \stackrel{d}{=} \frac{W_1}{Z_{r,\mu}}, \quad (10)$$

where the random variables W_1 and $Z_{r,\mu}$ are independent. Moreover, a gamma distribution with shape parameter $r > 1$ cannot be represented as a mixed exponential distribution.

Let $r > 0$, $\gamma \in \mathbb{R}$ and $\mu > 0$. Let the r.v. $N_{r,\gamma,\mu}$ have the GNB distribution. Its d.f. will be denoted $F_{GNB}(x; r, \gamma, \mu)$.

Along with the arguments given above in favor of the adequacy of the GNB models for the duration of wet periods based on their definition as mixed Poisson distributions, this effect can also be explained (at least in part) by their one more important property of being mixed geometric formulated as the following theorem.

Theorem 1 ([3]). *If $r \in (0, 1]$, $\gamma \in (0, 1]$ and $\mu > 0$, then a GNB distribution is a $Y_{r,\gamma,\mu}$ -mixed geometric distribution:*

$$P(N_{r,\gamma,\mu} = k) = \int_0^1 y(1-y)^k dP(Y_{r,\gamma,\mu} < y), \quad k = 0, 1, 2, \dots, \quad (11)$$

where

$$Y_{r,\gamma,\mu} \stackrel{d}{=} \frac{S_{\gamma,1} Z_{r,\mu}^{1/\gamma}}{1 + S_{\gamma,1} Z_{r,\mu}^{1/\gamma}} \stackrel{d}{=} \frac{\mu^{1/\gamma} S_{\gamma,1} (G_{r,1} + G_{1-r,1})^{1/\gamma}}{G_{r,1}^{1/\gamma} + \mu^{1/\gamma} S_{\gamma,1} (G_{r,1} + G_{1-r,1})^{1/\gamma}}, \quad (12)$$

where the r.v.'s $S_{\gamma,1}$ and $Z_{r,\mu}$ or $S_{\gamma,1}$, $G_{r,1}$ and $G_{1-r,1}$ are independent.

4. The Asymptotic Approximation to the Probability Distribution of Extremal Daily Precipitation within a Wet Period

In this section, the probability distribution of extremal daily precipitation within a wet period will be deduced as an asymptotic approximation. We will require some auxiliary statements formulated as lemmas.

The following asymptotic property of the GNB distribution will play the fundamental role in the construction of asymptotic approximations to the distributions of extreme daily precipitation within a wet period and the total precipitation volume per wet period and the corresponding statistical tests for precipitation to be abnormally heavy.

Lemma 2 ([3]). For $r > 0$, $\gamma \in \mathbb{R}$, $\mu > 0$ let $N_{r,\gamma,\mu}$ be an r.v. with the GNB distribution. We have

$$\mu^{1/\gamma} N_{r,\gamma,\mu} \Rightarrow \bar{G}_{r,\gamma,1} \stackrel{d}{=} G_{r,1}^{1/\gamma} \quad (13)$$

as $\mu \rightarrow 0$. If, moreover, $r \in (0, 1]$ and $\gamma \in (0, 1]$, then the limit law can be represented as

$$\bar{G}_{r,\gamma,1} \stackrel{d}{=} \frac{W_1}{S_{\gamma,1} Z_{r,1}^{1/\gamma}} \stackrel{d}{=} \frac{W_1^{1/\gamma}}{Z_{r,1}^{1/\gamma}} \stackrel{d}{=} \left(\frac{W_1 G_{r,1}}{G_{r,1} + G_{1-r,1}} \right)^{1/\gamma} \stackrel{d}{=} W_1^{1/\gamma} \cdot \left(1 + \frac{1-r}{r} Q_{1-r,r} \right)^{-1/\gamma}, \quad (14)$$

where the r.v.'s W_1 , $S_{\gamma,1}$ and $Z_{r,1}$ are independent as well as the r.v.'s W_1 and $Z_{r,1}$, or the r.v.'s W_1 , $G_{r,1}$ and $G_{1-r,1}$, and the r.v. $Q_{1-r,r}$ has the Snedecor–Fisher distribution with parameters $1 - r$ and r , see Equation (9).

Let $\mu > 0$, $\gamma > 0$. Instead of an infinitesimal parameter μ , in order to construct asymptotic approximations with “large” sample size, introduce an auxiliary “infinitely large” parameter $n \in \mathbb{N}$ and assume that $\mu = \mu_n = \mu n^{-\gamma}$. It can be easily made sure that in this case

$$\bar{G}_{r,\gamma,\mu/n^\gamma} \stackrel{d}{=} n \bar{G}_{r,\gamma,\mu}. \quad (15)$$

Then for $r > 0$, $\mu > 0$ for any $n \in \mathbb{N}$, we have

$$n^{-1} \bar{G}_{r,\gamma,\mu/n^\gamma} \stackrel{d}{=} \bar{G}_{r,\gamma,\mu} \stackrel{d}{=} \lambda^{-1/\gamma} \bar{G}_{r,\gamma,1} \stackrel{d}{=} \lambda^{-1/\gamma} G_{r,1}^{1/\gamma}. \quad (16)$$

The standard Poisson process (the Poisson process with unit intensity) will be denoted $P(t)$, $t \geq 0$.

Lemma 3 ([27]). Let $\Lambda_1, \Lambda_2, \dots$ be a sequence of positive r.v.'s such that for any $n \in \mathbb{N}$ the r.v. Λ_n is independent of the standard Poisson process $P(t)$, $t \geq 0$. The convergence

$$n^{-1} P(\Lambda_n) \Rightarrow \Lambda$$

as $n \rightarrow \infty$ to some nonnegative r.v. Λ takes place if and only if

$$n^{-1} \Lambda_n \Rightarrow \Lambda, \quad n \rightarrow \infty. \quad (17)$$

Lemma 3 can be regarded as a special case of the following result. Consider a sequence of r.v.'s W_1, W_2, \dots . Let N_1, N_2, \dots 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, \dots . In the following statement, the convergence is meant as $n \rightarrow \infty$.

Lemma 4 ([28,29]). Assume that there exists an infinitely increasing (convergent to zero) sequence of positive numbers $\{b_n\}_{n \geq 1}$ and an r.v. W such that

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

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

$$d_n^{-1} b_{N_n} \implies N, \quad (18)$$

then

$$d_n^{-1} W_{N_n} \implies W \cdot N, \quad (19)$$

where the r.v.'s on the right-hand side of Equation (19) are independent. If, in addition, $N_n \rightarrow \infty$ in probability and the family of scale mixtures of the d.f. of the r.v. W is identifiable, then Condition (18) is not only sufficient for Equation (19), but is necessary as well.

Consider a sequence of independent identically distributed (i.i.d.) r.v.'s X_1, X_2, \dots . Let N_1, N_2, \dots be a sequence of natural-valued r.v.'s such that for each $n \in \mathbb{N}$ the r.v. N_n is independent of the sequence X_1, X_2, \dots . Denote $M_n = \max\{X_1, \dots, X_{N_n}\}$.

Lemma 5 ([30]). Let $\Lambda_1, \Lambda_2, \dots$ be a sequence of positive r.v.'s such that for each $n \in \mathbb{N}$ the r.v. Λ_n is independent of the Poisson process $P(t)$, $t \geq 0$. Let $N_n = P(\Lambda_n)$. Assume that there exists a nonnegative r.v. Λ such that Convergence (17) takes place. Let X_1, X_2, \dots be i.i.d. r.v.'s a common d.f. $F(x)$. Assume also that $\sup\{x : F(x) < 1\} = \infty$ and there exists a number $\alpha > 0$ such that for each $x > 0$

$$\lim_{y \rightarrow \infty} \frac{1 - F(xy)}{1 - F(y)} = x^{-\alpha}. \quad (20)$$

Then

$$\lim_{n \rightarrow \infty} \sup_{x \geq 0} \left| \mathbb{P} \left(\frac{M_n}{F^{-1}(1 - \frac{1}{n})} < x \right) - \int_0^\infty e^{-zx^{-\alpha}} d\mathbb{P}(\Lambda < z) \right| = 0.$$

Now we turn to the main results of this section. The principal role in our reasoning will be played by Lemma 5. In order to justify its applicability, we need to make sure that the daily precipitation volumes satisfy Condition (20). A thorough statistical analysis shows that, although being rather adequate and, in general, acceptable model, the traditional gamma distribution (used, e.g., in [4]) is not the best model for statistical regularities in daily precipitation. The analysis of meteorological data (daily precipitation volumes) registered over 60 years at two geographic points with a very different climate: Potsdam (Brandenburg, Germany) with a mild climate influenced by the closeness to the ocean with warm Gulfstream flow and Elista (Kalmykia, Russia) with a radically continental climate convincingly suggests the Pareto-type model for the distribution of daily precipitation volumes, see Figures 3 and 4. For comparison, on these figures, the graphs of the best gamma-densities there are also presented. It can be seen that the gamma model fits the histograms in a noticeably worse way than the Pareto distribution.

Theorem 2. Let $n \in \mathbb{N}$, $\gamma > 0$, $\mu > 0$ and let N_{r,γ,μ_n} be an r.v. with the GNB distribution with parameters $r > 0$, $\gamma > 0$ and $\mu_n = \mu/n^\gamma$. Let X_1, X_2, \dots be i.i.d. r.v.'s with a common d.f. $F(x)$. Assume that $\text{rex}(F) = \infty$ and there exists a number $\alpha > 0$ such that Relation (20) holds for any $x > 0$. Then

$$\lim_{n \rightarrow \infty} \sup_{x \geq 0} \left| \mathbb{P} \left(\frac{\max\{X_1, \dots, X_{N_{r,\gamma,\mu_n}}\}}{F^{-1}(1 - \frac{1}{n})} < x \right) - F(x; r, \alpha, \gamma, \mu) \right| = 0,$$

where

$$F(x; r, \alpha, \gamma, \mu) = \int_0^\infty e^{-\lambda x^{-\alpha}} g^*(\lambda; r, \gamma, \mu) d\lambda \equiv \mathbb{P}(M_{r,\alpha,\gamma,\mu} < x), \quad x \in \mathbb{R}.$$

The limit r.v. $M_{r,\alpha,\gamma,\mu}$ admits the following product representations:

$$M_{r,\alpha,\gamma,\mu} \stackrel{d}{=} \frac{\bar{G}_{r,\alpha,\gamma,\mu}}{W_\alpha} \stackrel{d}{=} \left(\frac{\bar{G}_{r,\gamma,\mu}}{W_1} \right)^{1/\alpha} \stackrel{d}{=} \mu^{-1/\alpha\gamma} \left(\frac{G_{r,1}}{W_\gamma} \right)^{1/\alpha\gamma} \quad (21)$$

and in each term, the involved random variables are independent.

Proof. By definition, the GNB distribution is a mixed Poisson distribution with the GG mixing distribution. So, $N_{r,\gamma,\mu_n} \stackrel{d}{=} P(\bar{G}_{r,\gamma,\mu_n})$. Therefore, from Equation (16), Lemma 3 with $\Lambda_n = \bar{G}_{r,\gamma,\mu_n}$ and Lemma 5 with the account of the absolute continuity of the limit distribution it immediately follows that

$$\lim_{n \rightarrow \infty} \sup_{x \geq 0} \left| P\left(\frac{\max\{X_1, \dots, X_{N_{r,\gamma,\mu_n}}\}}{F^{-1}(1 - \frac{1}{n})} < x\right) - \int_0^\infty e^{-zx^{-\alpha}} g^*(z; r, \gamma, \mu) dz \right| = 0.$$

Since the Fréchet (inverse Weibull) d.f. $e^{-x^{-\alpha}}$ with $\alpha > 0$ corresponds to the r.v. W_α^{-1} , it is easy to make sure

$$F(x; r, \alpha, \gamma, \mu) \equiv \int_0^\infty e^{-zx^{-\alpha}} g^*(z; r, \gamma, \mu) dz = P\left(\frac{\bar{G}_{r,\gamma,\mu}^{1/\alpha}}{W_\alpha} < x\right).$$

Moreover, using relation $\bar{G}_{r,\gamma,\mu} \stackrel{d}{=} G_{r,\mu}^{1/\gamma}$, it is easy to see that

$$\frac{\bar{G}_{r,\gamma,\mu}^{1/\alpha}}{W_\alpha} \stackrel{d}{=} \frac{\bar{G}_{r,\alpha\gamma,\mu}}{W_\alpha} \stackrel{d}{=} \left(\frac{\bar{G}_{r,\gamma,\mu}}{W_1}\right)^{1/\alpha} \stackrel{d}{=} \mu^{-1/\alpha\gamma} \left(\frac{G_{r,1}}{W_\gamma}\right)^{1/\alpha\gamma},$$

where in each term the involved random variables are independent. The theorem is proved. \square

If $\gamma = 1$, then the limit distribution $F(x; r, \alpha, 1, \mu)$ corresponds to the results of [31,32].

Theorem 3. The distribution of the r.v. $M_{r,\alpha,\gamma,\mu}$ admits the following representations.

- (i) If $r \in (0, 1]$, it is the scale mixture of the distribution of the ratio of two independent Weibull-distributed r.v.'s:

$$M_{r,\alpha,\gamma,\mu} \stackrel{d}{=} (\mu Z_{r,1})^{-1/\alpha\gamma} \cdot \frac{W_{\alpha\gamma}}{W_\gamma},$$

where all the involved random variables are independent and the r.v. $Z_{r,1}$ is defined in Equation (8).

- (ii) If $\gamma \in (0, 1]$, it is the scale mixture of the tempered Snedecor–Fisher distribution with parameters r and 1:

$$M_{r,\alpha,\gamma,\mu} \stackrel{d}{=} \left(\frac{S_{\gamma,1}}{\mu r} \cdot Q_{r,1}\right)^{1/\alpha\gamma},$$

where $S_{\gamma,1}$ is a positive strictly stable r.v. with characteristic exponent γ independent of the r.v. $Q_{r,1}$ with the Snedecor–Fisher distribution in Equation (9) with parameters r and 1.

- (iii) If $\gamma \in (0, 1]$ and $r \in (0, 1]$, it is the scale mixture of the Pareto laws:

$$M_{r,\alpha,\gamma,\mu} \stackrel{d}{=} \Pi_\alpha (S_{\gamma,1} Z_{r,1}^{1/\gamma})^{-1/\alpha},$$

where $\mathbb{P}(\Pi_\alpha > x) = (x^\alpha + 1)^{-1}$, $x \geq 0$.

- (iv) If $r \in (0, 1]$ and $\alpha\gamma \in (0, 1]$, it is the scale mixture of the folded normal laws:

$$M_{r,\alpha,\gamma,\mu} \stackrel{d}{=} |X| \cdot \frac{\sqrt{2W_1}}{\mu^{1/\alpha\gamma} W_\alpha S_{\alpha\gamma,1} Z_{r,1}^{1/\alpha\gamma}},$$

where all the involved r.v.'s are independent.

Proof. To prove (i) it suffices to consider the rightmost term in Equation (21), apply relations $W_1^{1/\gamma} \stackrel{d}{=} W_\gamma$ and $G_{r,\mu} \stackrel{d}{=} \frac{W_1}{Z_{r,\mu}}$ (here $0 < r < 1$ and the r.v.'s W_1 and $Z_{r,\mu}$ are independent (for details, see Lemma 1)).

To prove (ii) it suffices to transform the rightmost term in (21) with the account of representation in Equation (7) and use the definition of the Snedecor–Fisher distribution as the distribution of the ratio of two independent gamma-distributed r.v.'s (see, e.g., Section 27 in [33]).

To prove (iii) it suffices to transform the second term in Equation (21) with the account of Equation (14) and notice that the distribution of the ratio of two independent exponentially distributed r.v.'s coincides with that of the random variable Π_1 .

To prove (iv) it suffices to transform the second term in (21) with the account of (14) and notice that $W_1 \stackrel{d}{=} |X|\sqrt{2W_1}$ with the r.v.'s on the right-hand side being independent (see, e.g., [25]). The theorem is proved. \square

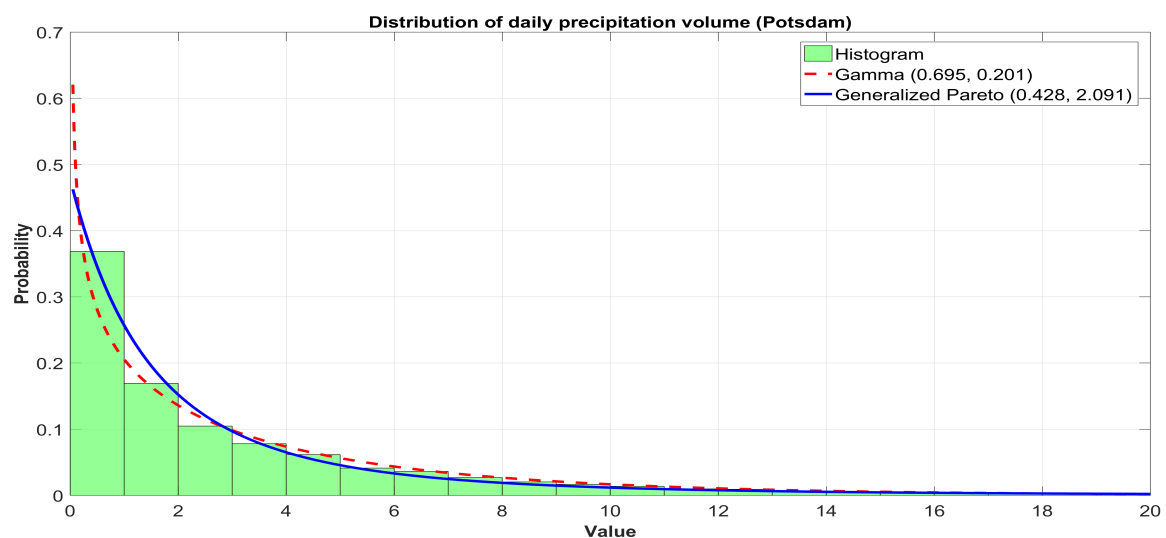


Figure 3. The histogram of daily precipitation volumes in Potsdam and the fitted Pareto and gamma distributions.

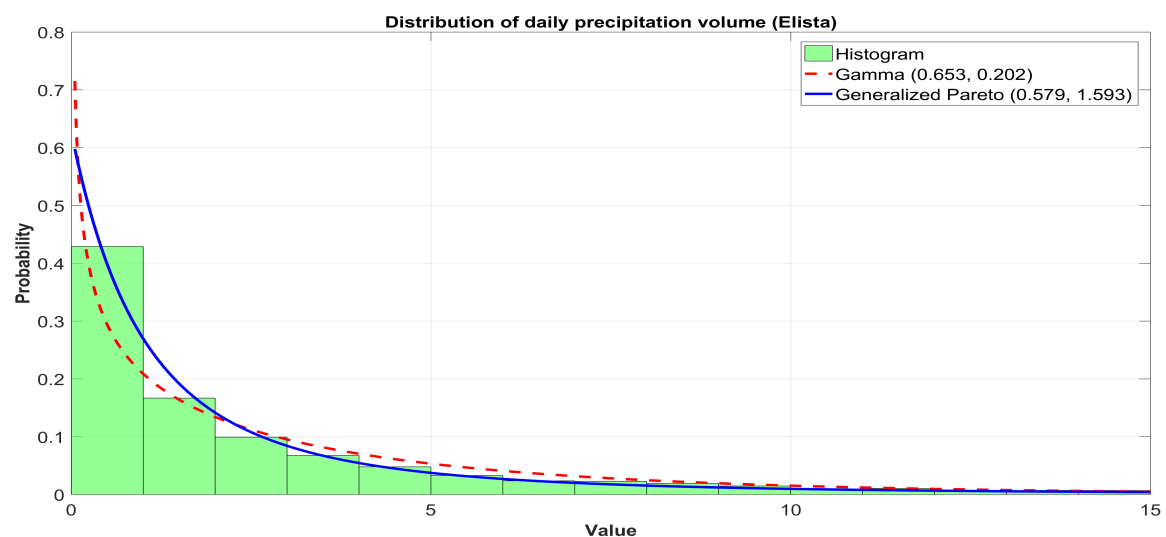


Figure 4. The histogram of daily precipitation volumes in Elista and the fitted Pareto and gamma distributions.

The product representations for the random value $M_{r,\alpha,\gamma,\mu}$ established in Theorem 3 can be used for computer simulation.

Theorem 4. If $r \in (0, 1]$, $\mu > 0$ and $\alpha\gamma \in (0, 1]$, then the d.f. $F(x; r, \alpha, \gamma, \mu)$ is mixed exponential:

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

where $A(u) = P(\mu^{1/\alpha\gamma} W_\alpha S_{\alpha\gamma,1} Z_{r,1}^{1/\alpha\gamma} < u)$, $u \geq 0$, and all the involved r.v.'s are independent.

Proof. To prove this statement, it suffices to transform the second term in Equation (21) with the account of Equation (14) and obtain

$$M_{r,\alpha,\gamma,\mu} \stackrel{d}{=} \frac{W_1}{\mu^{1/\alpha\gamma} W_\alpha S_{\alpha\gamma,1} Z_{r,1}^{1/\alpha\gamma}}.$$

□

Corollary 1. Let $r \in (0, 1]$, $\alpha\gamma \in (0, 1]$, $\mu > 0$. Then the distribution function $F(x; r, \alpha, \gamma, \mu)$ is infinitely divisible.

Proof. This statement immediately follows from Theorem 3 and the result of [34] stating that the product of two independent non-negative r.v.'s is infinitely divisible, if one of the two is exponentially distributed. □

It is possible to deduce explicit expressions for the moments of the r.v. $M_{r,\alpha,\gamma,\mu}$.

Theorem 5. Let $0 < \delta < \alpha$. Then

$$\mathbb{E}M_{r,\alpha,\gamma,\mu}^\delta = \frac{\Gamma(r + \frac{\delta}{\alpha\gamma})\Gamma(1 - \frac{\delta}{\alpha})}{\mu^{\delta/\alpha\gamma}\Gamma(r)}.$$

Proof. From Equation (14) it follows that $\mathbb{E}M_{r,\alpha,\gamma,\mu}^\delta = \mu^{-\delta/\alpha\gamma} \mathbb{E}G_{r,1}^{\delta/\alpha\gamma} \cdot \mathbb{E}W_1^{-\delta/\alpha}$. It is easy to verify that $\mathbb{E}G_{r,1}^{\delta/\alpha\gamma} = \Gamma(r + \frac{\delta}{\alpha\gamma})/\Gamma(r)$, $\mathbb{E}W_1^{-\delta/\alpha} = \Gamma(1 - \frac{\delta}{\alpha})$. Hence follows the desired result. □

Consider the bounds for the rate of convergence in Theorem 2. For this purpose, we will use one more auxiliary statement.

Lemma 6. Let $\lambda > 0$, X_1, X_2, \dots be i.i.d. r.v.'s with a common d.f. $F(x)$, $P(t)$ be the standard Poisson process independent of X_1, X_2, \dots . Assume that there exists a d.f. $H(x)$ such that for any $x \in \mathbb{R}$

$$\lim_{n \rightarrow \infty} P\left(\frac{1}{F^{-1}(1 - \frac{1}{n})} \max_{1 \leq k \leq n} X_k < x\right) = H(x). \quad (22)$$

Then for any $n \in \mathbb{N}$

$$\left| P\left(\frac{1}{F^{-1}(1 - \frac{1}{n})} \max_{1 \leq k \leq P(n\lambda)} X_k < x\right) - H^\lambda(x) \right| \leq |n[1 - F(xF^{-1}(1 - \frac{1}{n}))] - \log H(x)| \lambda H^\lambda(x).$$

Proof. This statement is a special case of Corollary 2 in [35]. □

Theorem 6. Let $n \in \mathbb{N}$, $\gamma > 0$, $\mu > 0$ and let N_{r,γ,μ_n} be an r.v. with the GNB distribution with parameters $r > 0$, $\gamma > 0$ and $\mu_n = \mu/n^\gamma$. Let X_1, X_2, \dots be i.i.d. r.v.'s with the common Pareto d.f.

$$F(x) = 1 - \frac{c}{ax^\alpha + c}, \quad x \geq 0,$$

with $a, c, \alpha > 0$,

$$F(x; r, \alpha, \gamma, \mu) = \int_0^\infty e^{-\lambda x^{-\alpha}} g^*(\lambda; r, \gamma, \mu) d\lambda, \quad x \in \mathbb{R}.$$

Then for any $x \in \mathbb{R}$

$$\begin{aligned} & \left| P\left(\left[\frac{a}{c(n-1)}\right]^{1/\gamma} \max_{1 \leq k \leq N_{r,\gamma,\mu_n}} X_k < x\right) - F(x; r, \alpha, \gamma, \mu) \right| \leq \\ & \leq \left| \frac{x^\alpha - 1}{x^\alpha(n-1) + 1} \right| \cdot \int_0^\infty \lambda e^{-\lambda x^{-\alpha}} g^*(\lambda; r, \gamma, \mu) d\lambda \leq \left| \frac{x^\alpha - 1}{x^\alpha(n-1) + 1} \right| \cdot \frac{\Gamma(r + \frac{1}{\gamma})}{\mu^{1/\gamma} \Gamma(r)}. \end{aligned}$$

Proof. First of all, check Condition (20). We have

$$\frac{1 - F(xy)}{1 - F(y)} = \frac{ay^\alpha + c}{ax^\alpha y^\alpha + c} \rightarrow x^{-\alpha}$$

as $y \rightarrow \infty$, that is, Condition (20) holds implying Equation (22) with $H(x) = e^{-x^{-\alpha}}$ in accordance with the classical theory of extremes (see, e.g., [36]). Second, note that in the case under consideration $F^{-1}(1 - \frac{1}{n}) = [\frac{c(n-1)}{a}]^{1/\alpha}$ so that $F(xF^{-1}(1 - \frac{1}{n})) = 1 - [x^\alpha(n-1) + 1]^{-1}$ and

$$n \left[1 - F\left(xF^{-1}\left(1 - \frac{1}{n}\right)\right) \right] - \log H(x) = \frac{n}{x^\alpha(n-1) + 1} - \frac{1}{x^\alpha} = \frac{x^\alpha - 1}{x^\alpha(n-1) + 1}.$$

Third, from Equation (15) it follows that $N_{r,\gamma,\mu/n^\gamma} \stackrel{d}{=} P(n\bar{G}_{r,\gamma,\mu})$ with independent $P(t)$ and $\bar{G}_{r,\gamma,\mu}$. Therefore, by Lemma 6 we have

$$\begin{aligned} & \left| P\left(\left[\frac{a}{c(n-1)}\right]^{1/\gamma} \max_{1 \leq k \leq N_{r,\gamma,\mu_n}} X_k < x\right) - F(x; r, \alpha, \gamma, \mu) \right| \leq \\ & \leq \int_0^\infty \left| P\left(\left[\frac{a}{c(n-1)}\right]^{1/\gamma} \max_{1 \leq k \leq P(n\lambda)} X_k < x\right) - e^{-\lambda x^{-\alpha}} \right| g^*(\lambda; r, \gamma, \mu) d\lambda \leq \\ & \leq \left| \frac{x^\alpha - 1}{x^\alpha(n-1) + 1} \right| \cdot \int_0^\infty \lambda e^{-\lambda x^{-\alpha}} g^*(\lambda; r, \gamma, \mu) d\lambda \leq \left| \frac{x^\alpha - 1}{x^\alpha(n-1) + 1} \right| \cdot E\bar{G}_{r,\gamma,\mu} = \\ & = \left| \frac{x^\alpha - 1}{x^\alpha(n-1) + 1} \right| \cdot \frac{\Gamma(r + \frac{1}{\gamma})}{\mu^{1/\gamma} \Gamma(r)}. \end{aligned}$$

The theorem is proved. \square

Actually, Theorem 6 states that the rate of convergence in Theorem 2 is $O(\mu_n^{1/\gamma})$ as $\mu_n \rightarrow 0$.

The results of this section serve as a theoretical base for the construction of a test for abnormally extreme daily precipitation. The distribution of the maximum daily precipitation per wet period can be assumed to be a tempered scale mixture of the gamma distribution in which the scale factor has the Weibull distribution. According to the typical construction of a test, a daily precipitation volume is considered to be abnormally extremal, if it exceeds a certain (pre-defined) quantile of this distribution. A detailed description of this test and algorithm of estimation of the parameters of the distribution mentioned above deserve a separate study as well as its application to real data.

5. The Asymptotic Approximation to the Probability Distribution of Total Precipitation over a Wet Period. Generalized Rényi Theorem for Gnb Random Sums

As far ago as in the 1950s, being interested in modeling rare events, A. Rényi studied rarefaction of renewal point processes and proved his famous theorem on convergence of rarefied renewal processes to the Poisson process [37,38]. The Rényi theorem states that the distribution of a geometric sum

(i.e., a sum of a random number of i.i.d. r.v.'s in which the number of summands is a r.v. with the geometric distribution independent of the summands) normalized by its expectation converges to the exponential law as the expectation of the sum infinitely increases. The normalization of a sum by its expectation is typical for laws of large numbers. Therefore, the Rényi theorem can be regarded as the law of large numbers for geometric sums. A general law of large numbers for random sums of independent identically distributed (i.i.d.) random variables (r.v.'s) was proved in [28]. It was demonstrated there that the distribution of a random sum normalized by its expectation converges to some distribution, if and only if the distribution of the random index (the number of summands) converges to the same distribution (up to a scale parameter) under the same normalization. In [3] the law of large numbers for GNB random sums was proved. However, a direct application of this result to modeling the probability distribution of total precipitation over a wet period is hampered by the following very interesting practical observation.

One might have expected that successive daily precipitation volumes X_1, X_2, \dots satisfy the classical law of large numbers, that is, the arithmetic mean $\frac{1}{n}(X_1 + \dots + X_n)$ converges to some number a almost surely as n infinitely grows, as it was done in [2]. However, a thorough analysis of real data shows that this not quite so. In Figure 5, there are the graphs of the averaged daily precipitation volumes in Potsdam and Elista demonstrating the slowly decreasing trend for Potsdam and slowly increasing trend for Elista.

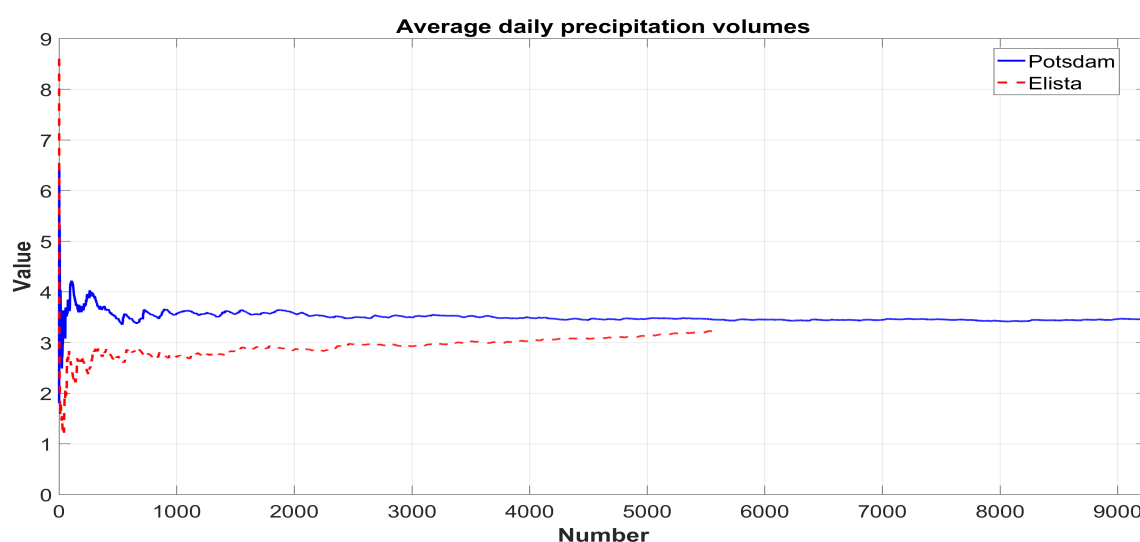


Figure 5. Stabilization of the cumulative averages of daily precipitation volumes as n grows in Potsdam (continuous line) and Elista (dash line).

This means that, in order to match the stabilization of the averages at some level a , it is required to normalize the sum $X_1 + \dots + X_n$ not by n , but by a somewhat more complicated function of n that can match the influence of slow global trends. As such, a function of n , consider a power function n^β with $\beta > 0$ and assume that not necessarily i.i.d. r.v.'s X_1, X_2, \dots satisfy the condition

$$\frac{1}{n^\beta} \sum_{j=1}^n X_j \implies a \in (0, \infty) \quad (23)$$

as $n \rightarrow \infty$. The parameters a and β can be rather reliably estimated by the least squares technique.

Let X_1, X_2, \dots, X_n be the observed values of successive nonzero daily precipitation volumes, $n \in \mathbb{N}$ be the total number of available observations. For a natural $k = 1, \dots, n$ denote $s_k = X_1 + \dots + X_k$.

If Condition (23) holds, then for k large enough ($1 \leq m \leq k \leq n$), the following estimates of the parameters a and β in Relation (23) can be used:

$$\tilde{a} = \exp \left\{ \frac{\sum_{k=m}^n \log s_k \cdot \sum_{k=m}^n (\log k)^2 - \sum_{k=m}^n \log k \cdot \sum_{k=m}^n (\log k \cdot \log s_k)}{(n-m+1) \sum_{k=m}^n (\log k)^2 - (\sum_{k=m}^n \log k)^2} \right\}, \quad (24)$$

$$\tilde{\beta} = \frac{\sum_{k=m}^n \log s_k - (n-m+1) \log \tilde{a}}{\sum_{k=m}^n \log k}. \quad (25)$$

Indeed, if Condition (23) holds, the following approximate equality can be written:

$$\frac{T_k}{k^{\tilde{\beta}}} \approx a \Leftrightarrow -\beta \log k + \log T_k \approx \log a.$$

Therefore, the estimates of the parameters a and β can be found as the solution of the least squares problem

$$\sum_{k=m}^n (\log T_k - \beta \log k - \log a)^2 \rightarrow \min_{\beta, \log a}.$$

This solution can be found explicitly and has the form

$$\begin{aligned} \widetilde{\log a} &= \frac{\sum_{k=m}^n \log T_k \cdot \sum_{k=m}^n (\log k)^2 - \sum_{k=m}^n \log k \cdot \sum_{k=m}^n (\log k \cdot \log T_k)}{(n-m+1) \sum_{k=m}^n (\log k)^2 - (\sum_{k=m}^n \log k)^2}, \\ \tilde{\beta} &= \frac{\sum_{k=m}^n \log T_k - (n-m+1) \widetilde{\log a}}{\sum_{k=m}^n \log k}, \end{aligned}$$

that leads to Formulas (24) and (25). This least squares method for estimation of a and β is realized by Algorithm A2 (see Appendix A).

The application of Equation (23) to real data from Potsdam and Elista with a and β estimated by Equations (24) and (25) is illustrated in Figure 6. It can be seen that the cumulative averages stabilize at the level $a = 4.087$ with $\beta = 0.981$ for Potsdam and at the level $a = 0.96$ with $\beta = 1.146$ for Elista.

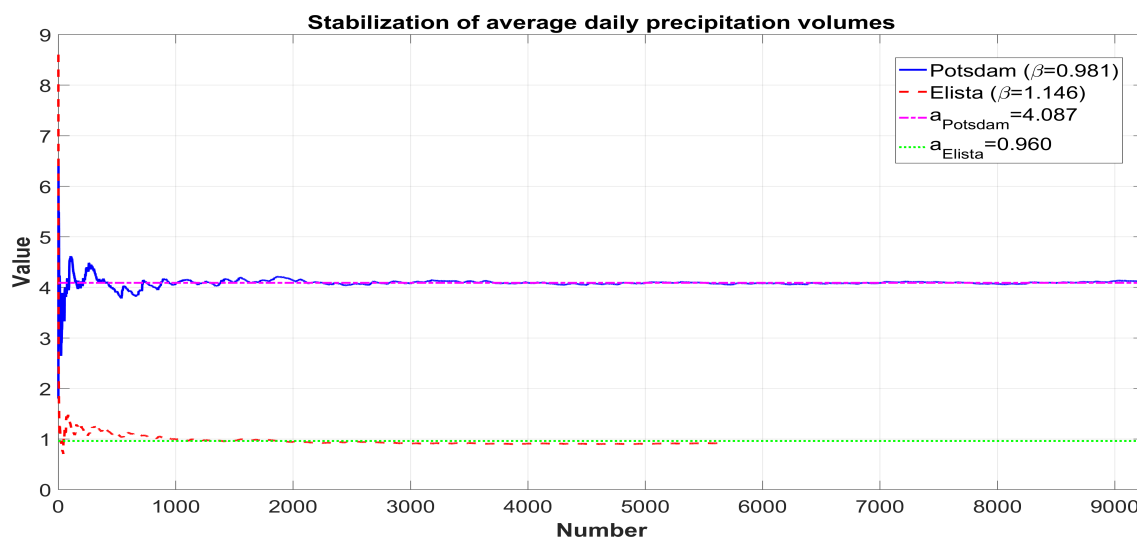


Figure 6. Stabilization of the cumulative averages of daily precipitation volumes as n grows with $\beta = 1.139$ for Potsdam (solid line) and with $\beta = 0.981$ for Elista (dashed line).

So, to construct the asymptotic approximation to the probability distribution of total precipitation over a wet period, we should prove a generalized Rényi theorem for GNB random sums improving an

analogous statement proved in [3]. It must be especially noted that in the following theorem, the r.v.'s X_1, X_2, \dots are not assumed to be i.i.d.

Theorem 7. Assume that the nonzero daily precipitation volumes X_1, X_2, \dots satisfy Condition (23) with some $\beta > 0$ and $a > 0$. Let the numbers $r > 0$, γ and $\mu > 0$ be arbitrary. For each $n \in \mathbb{N}$, let the r.v. N_{r,γ,μ_n} have the GNB distribution with parameters r , γ and $\mu_n = \mu/n^\gamma$. Assume that for each $n \in \mathbb{N}$ the r.v. N_{r,γ,μ_n} is independent of the sequence X_1, X_2, \dots . Then

$$\frac{a\mu^{\beta/\gamma}}{n^\beta} \sum_{j=1}^{N_{r,\gamma,\mu/n^\gamma}} X_j \Rightarrow \overline{G}_{r,\gamma/\beta,1} \stackrel{d}{=} G_{r,1}^{\beta/\gamma}$$

as $n \rightarrow \infty$.

Proof. The proof is based on Lemma 4 and Equation (16). From Equation (16) it follows that

$$\frac{\mu^{1/\gamma}}{n} \cdot N_{r,\gamma,\mu/n^\gamma} \Rightarrow \overline{G}_{r,\gamma,1} \quad (26)$$

as $n \rightarrow \infty$. By virtue of Condition (23), in Lemma 4 let $b_n = n^\beta/a$. As N_n in Lemma 4 take $N_{r,\gamma,\mu/n^\gamma}$. Then $b_{N_n} = \frac{1}{a} N_{r,\gamma,\mu/n^\gamma}^\beta$. From Equation (26) it follows that, as $n \rightarrow \infty$,

$$\frac{1}{a} N_{r,\gamma,\mu/n^\gamma}^\beta \cdot \frac{\mu^{\beta/\gamma}}{n^\beta} \Rightarrow \frac{1}{a} \overline{G}_{r,\gamma,1}^\beta \stackrel{d}{=} \frac{1}{a} \overline{G}_{r,\gamma/\beta,1} \stackrel{d}{=} \frac{1}{a} G_{r,1}^{\beta/\gamma}. \quad (27)$$

Therefore, as d_n we can take $d_n = n^\beta/\mu^{\beta/\gamma}$. So, using Equation (27) in the role of Equation (18) in Lemma 4, we obtain Equation (19) in the form

$$\frac{\mu^{\beta/\gamma}}{n^\beta} \sum_{j=1}^{N_{r,\gamma,\mu/n^\gamma}} X_j \Rightarrow \frac{1}{a} \overline{G}_{r,\gamma/\beta,1} \stackrel{d}{=} \frac{1}{a} G_{r,1}^{\beta/\gamma}, \quad (28)$$

whence follows the desired result. The theorem is proved. \square

Theorem 7 presents a good tool for the account of the parameters β and γ characterizing the deviation from traditional NB and arithmetic mean models due to the influence of possible (slow) global trends. If in Theorem 7 $r = \gamma = \beta = 1$, then we obtain a version of the Rényi theorem [39] generalized to non-identically distributed and not necessarily independent summands. If in Theorem 7 $\beta = 1$, then we obtain the law of large numbers for GNB random sums (see [3]). If in Theorem 7 $\gamma = 1$, then we obtain the law of large numbers for NB random sums modified for the case $\beta \neq 1$.

Therefore, if daily precipitation volumes X_1, X_2, \dots (of course, being non-identically distributed and not independent), with the account of the excellent fit of the GNB model for the duration of a wet period (see Figure 1), with rather small μ , the GG distribution can be regarded as an adequate and theoretically well-based model for the total precipitation volume over a (long enough) wet period.

As regards the bounds for the rate of convergence in Theorem 7, consider a special case of $\beta = 1$ and i.i.d. X_1, X_2, \dots

As a measure of the distance between probability distributions, consider the ζ -metric proposed by V. M. Zolotarev in [40,41] (also see [42], p. 44). Let $s > 0$. There exists a unique representation of the number s as $s = m + \alpha$ where m is an integer and $0 < \alpha \leq 1$. By \mathcal{F}_s we denote the set of all real-valued bounded functions f on \mathbb{R} that are m times differentiable and $|f^{(m)}(x) - f^{(m)}(y)| \leq |x - y|^\alpha$. Let X and Y be two r.v.'s in which the distribution functions will be denoted $F_X(x)$ and $F_Y(x)$, respectively.

The ζ -metric $\zeta_s(X, Y) \equiv \zeta_s(F_X, F_Y)$ in the space of probability distributions is defined by the equality $\zeta_s(X, Y) = \sup \{ |E(f(X) - f(Y))| : f \in \mathcal{F}_s \}$. In particular,

$$\zeta_1(X, Y) = \int_{\mathbb{R}} |F_X(x) - F_Y(x)| dx.$$

In [43], it was shown that in the case $\beta = 1$ and i.i.d. X_1, X_2, \dots for $1 \leq s \leq 2$ we have

$$\zeta_s \left(\frac{\mu^{1/\gamma}}{na} \sum_{j=1}^{N_{r,\gamma,\mu/n^\gamma}} X_j, \bar{G}_{r,\gamma,1} \right) \leq \frac{(EX_1^2)^{s/2}}{n^{s/2} |EX_1|^s} \cdot \frac{\Gamma(1+\gamma)\Gamma(r + \frac{s}{2\gamma})}{\Gamma(1+s)\Gamma(r)}.$$

In particular,

$$\zeta_2 \left(\frac{\mu^{1/\gamma}}{na} \sum_{j=1}^{N_{r,\gamma,\mu/n^\gamma}} X_j, \bar{G}_{r,\gamma,\mu} \right) \leq \frac{EX_1^2}{2n(EX_1)^2} \cdot \frac{\Gamma(r + \frac{1}{\gamma})}{\Gamma(r)}.$$

The results presented above justify the GG models for the probability distribution of total precipitation volume over a wet period improving the models considered in [2]. Statistical tests for the detection of anomalously extreme total volumes will be considered below.

6. Statistical Tests for Anomalously Extreme Total Precipitation Volumes

Now we turn to the construction of the tests for the total precipitation volume during a wet period to be abnormally large.

In what follows, based on the results of the preceding section, we will assume that the total precipitation volume during a wet period has the GG distribution with some parameters $r > 0$, $\gamma > 0$ and $\mu > 0$.

Let $m \in \mathbb{N}$ and $\bar{G}_{r,\gamma,\mu}^{(1)}, \bar{G}_{r,\gamma,\mu}^{(2)}, \dots, \bar{G}_{r,\gamma,\mu}^{(m)}$ be independent r.v.'s having the same GG distribution with parameters $r > 0$, γ and $\mu > 0$. Also, let $G_{r,\mu}^{(1)}, G_{r,\mu}^{(2)}, \dots, G_{r,\mu}^{(m)}$ be i.i.d. r.v.'s having the same gamma distribution with parameters $r > 0$ and $\mu > 0$.

The base for the first step in the construction of the desired test is the following obvious conclusion: if the r.v.'s $\bar{G}_{r,\gamma,\mu}^{(1)}, \bar{G}_{r,\gamma,\mu}^{(2)}, \dots, \bar{G}_{r,\gamma,\mu}^{(m)}$ are identically distributed (that is, the sample $\bar{G}_{r,\gamma,\mu}^{(1)}, \bar{G}_{r,\gamma,\mu}^{(2)}, \dots, \bar{G}_{r,\gamma,\mu}^{(m)}$ is homogeneous), then the r.v.'s $(\bar{G}_{r,\gamma,\mu}^{(1)})^\gamma, (\bar{G}_{r,\gamma,\mu}^{(2)})^\gamma, \dots, (\bar{G}_{r,\gamma,\mu}^{(m)})^\gamma$ are also identically distributed (that is, the sample $(\bar{G}_{r,\gamma,\mu}^{(1)})^\gamma, (\bar{G}_{r,\gamma,\mu}^{(2)})^\gamma, \dots, (\bar{G}_{r,\gamma,\mu}^{(m)})^\gamma$ is homogeneous).

Consider the relative contribution of the r.v. $(\bar{G}_{r,\gamma,\mu}^{(1)})^\gamma$ to the sum $(\bar{G}_{r,\gamma,\mu}^{(1)})^\gamma + (\bar{G}_{r,\gamma,\mu}^{(2)})^\gamma + \dots + (\bar{G}_{r,\gamma,\mu}^{(m)})^\gamma$:

$$R = \frac{(\bar{G}_{r,\gamma,\mu}^{(1)})^\gamma}{(\bar{G}_{r,\gamma,\mu}^{(1)})^\gamma + (\bar{G}_{r,\gamma,\mu}^{(2)})^\gamma + \dots + (\bar{G}_{r,\gamma,\mu}^{(m)})^\gamma}. \quad (29)$$

From Equation (4), it obviously follows that

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

(see Equation(29)).

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.

Note that

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

where the gamma-distributed r.v.'s on the right hand side are independent. The distribution of the r.v. R was described in [2] where it was demonstrated that $R \stackrel{d}{=} (1 + \frac{k}{r} Q_{k,r})^{-1}$ where $Q_{k,r}$ is the r.v. having the Snedecor–Fisher distribution determined for $k > 0, r > 0$ by the Lebesgue density

$$f_{k,r}(x) = \frac{\Gamma(k+r)}{\Gamma(k)\Gamma(r)} \left(\frac{k}{r}\right)^k \frac{x^{k-1}}{(1 + \frac{k}{r}x)^{k+r}}, \quad x \geq 0. \quad (30)$$

It should be noted that the particular value of the scale parameter is insignificant. For convenience, it is assumed equal to one. It can be easily made sure by standard calculation using Equation (30), 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 GG 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, \dots, 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

$$SR = \frac{V_1^\gamma}{V_1^\gamma + \dots + V_m^\gamma}$$

(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

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

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 GG 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, \dots, 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

$$SR_{GG} = \frac{(m-1)V_1^\gamma}{V_2^\gamma + \dots + V_m^\gamma}. \quad (31)$$

(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 and $k = (m-1)r$. 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

and $k = (m - 1)r$. 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 $V_{i_1}, V_{i_2}, \dots, V_{i_l}$ do not make an abnormally large cumulative contribution to the total precipitation volume $V_1 + \dots + V_m$ ». For this purpose denote

$$T_l^\gamma = V_{i_1}^\gamma + V_{i_2}^\gamma + \dots + V_{i_l}^\gamma, \quad T^\gamma = V_1^\gamma + V_2^\gamma + \dots + V_m^\gamma$$

and consider the quantity

$$SR'_0 = \frac{(m-l)T_l^\gamma}{l(T^\gamma - T_l^\gamma)}.$$

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, $q_{lr,(m-l)r}(1 - \varepsilon)$ be the $(1 - \varepsilon)$ -quantile of the Snedecor–Fisher distribution with parameters lr and $k = (m - l)r$. 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 $V_{i_1}, V_{i_2}, \dots, V_{i_l}$ into the total precipitation volume $V_1 + \dots + V_m$ must be regarded as abnormally large. Moreover, the probability of erroneous rejection of H'_0 is equal to ε .

7. Comparison of Tests for Anomalously Extreme Precipitation Volumes Based on Gamma and Gg Distributions

In this section, the results of the application of the test based on the statistic R in Equation (29) to the analysis of the time series of daily precipitation observed in Potsdam and Elista from 1950 to 2007 are considered and compared with similar results for the case of gamma distributed total precipitation volumes during wet periods [2].

The results of the application of the tests for a total precipitation volume during one wet period to be abnormally large based on GG and gamma models in the moving mode are shown in Figures 7 and 8 (Potsdam) and Figures 9 and 10 (Elista).

If m is the window width (the number of observations in a moving window). A fixed sample point falls in exactly m windows. One of the following cases can occur for a fixed observation:

- Absolute (abs) extreme, if at all m windows it is recognized as abnormally extreme;
- Intermediate (int) extreme, if it is recognized as abnormally extreme for at least half of windows containing it;
- Relative (rel) extreme, if it is recognized as abnormally extreme for at least one window;
- Not extremal, if it is not recognized as abnormally extreme for all windows.

Algorithm A3 (see Appendix A) realizes the method based on the statistic R described above.

For the sake of vividness on these figures the time horizon equals 90 and 360 days and the significance level α of the tests is 0.01. The absolutely, intermediate and relatively abnormal precipitation volumes are marked by downward-pointing triangles, circles and squares, respectively, for the test based on the gamma model, whereas the corresponding test based on the statistic R based on the GG distribution are marked by upward-pointing triangles, diamonds and right-pointing triangles, respectively. It is worth noting that MATLAB's notations are used here for these markers.

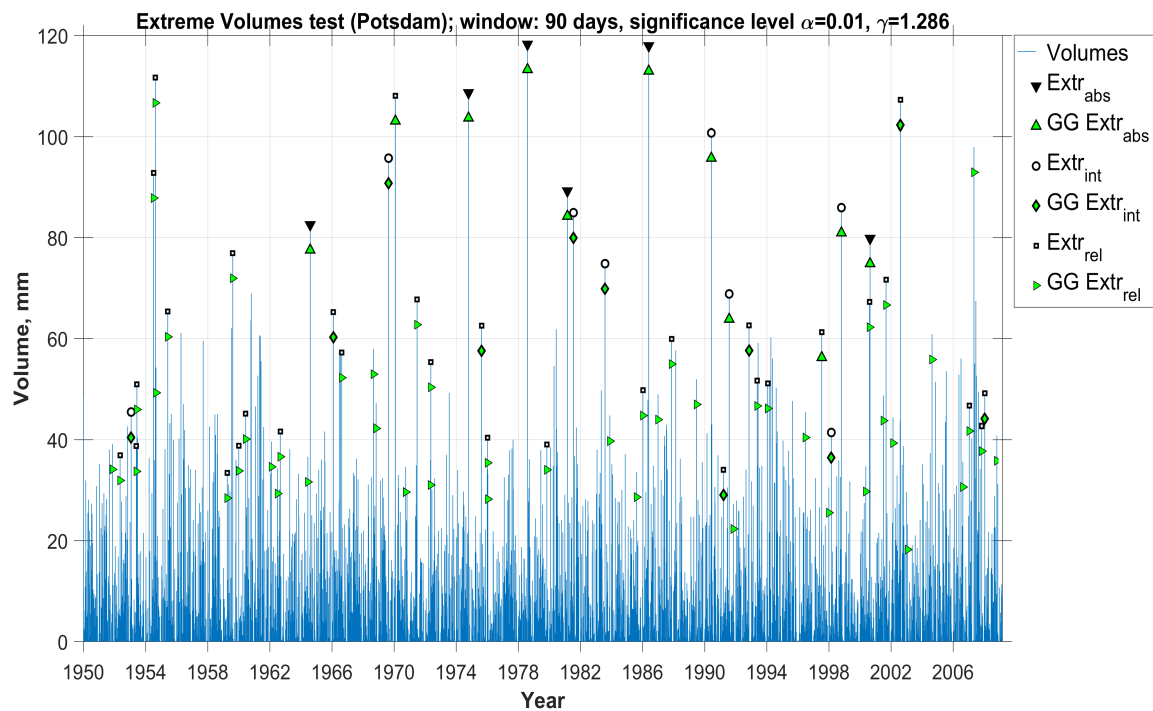


Figure 7. Abnormal precipitation volumes (Potsdam, 90 days).

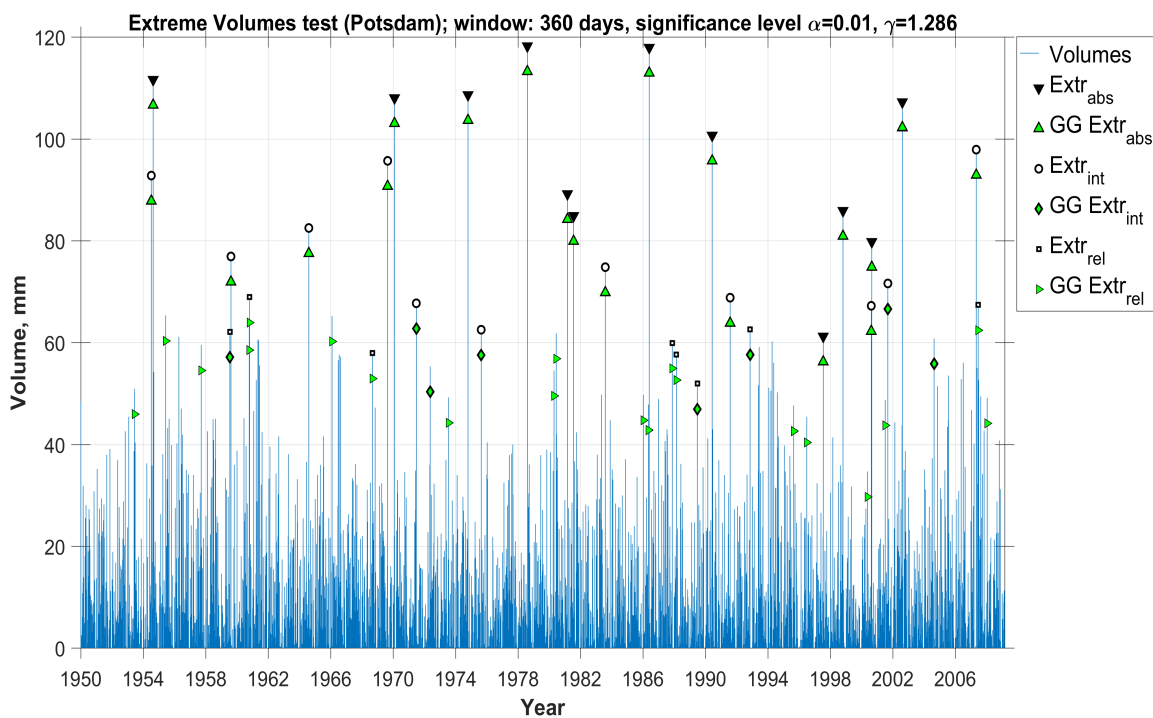


Figure 8. Abnormal precipitation volumes (Potsdam, 360 days).

Figures 7–10 demonstrate non-trivial values of the parameter γ , that is, $\gamma \neq 1$. For Potsdam $\gamma = 1.286$, whereas for Elista γ equals 1.279. At the same time, the results of the two methods are quite close, although the approach based on the GG distribution demonstrates a higher quality of determining potentially extreme observations. The same conclusions are valid for smaller window sizes.

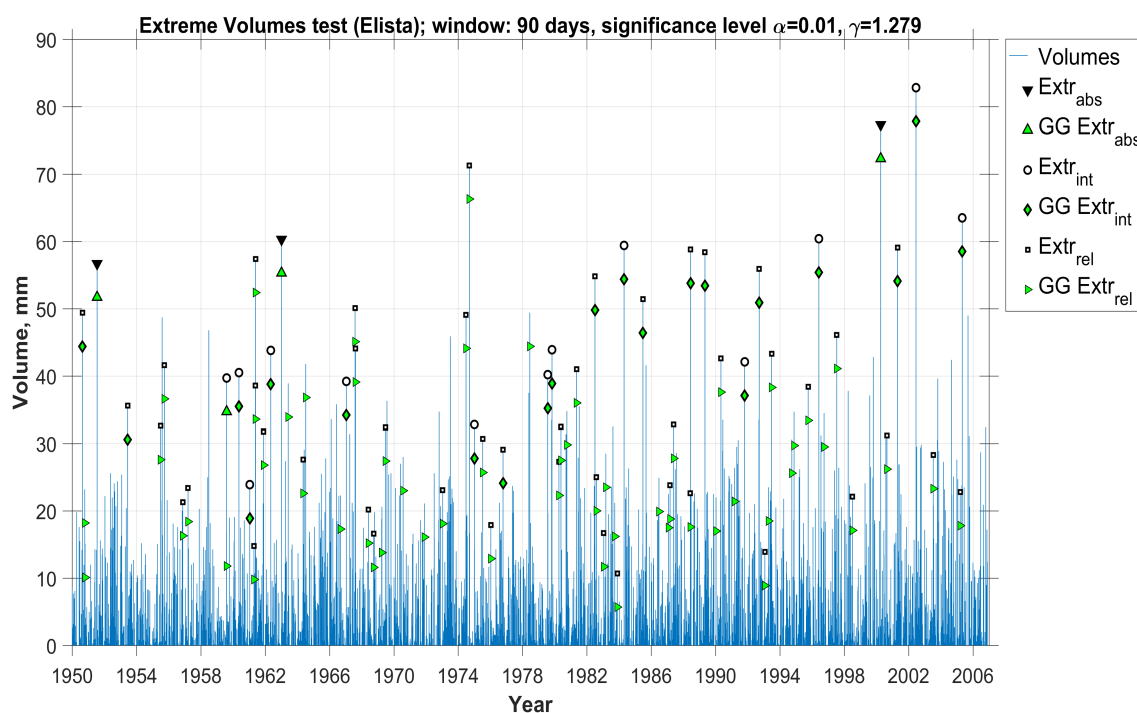


Figure 9. Abnormal precipitation volumes (Elista, 90 days).

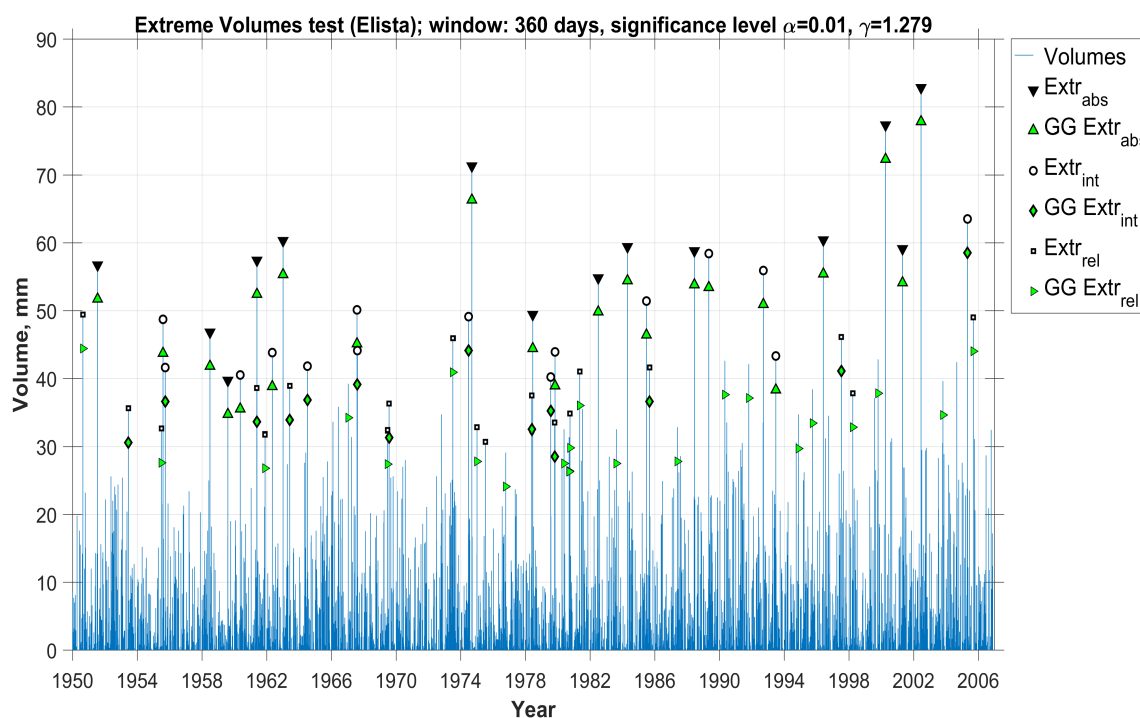


Figure 10. Abnormal precipitation volumes (Elista, 360 days).

8. Comparison of GG-Based Statistical Test and Peaks over Threshold Methodology for Extreme Precipitation Intensities

One important precipitation indicator is the precipitation intensity that is defined as the ratio of the total precipitation volume over a wet period to the duration of this wet period measured in days. The extreme precipitation volumes and intensities are relevant to various problems of climatology and hydrometeorology (see, for example, [44–47]). Traditionally, these phenomena are investigated for

different geographical regions or countries [48,49]. In particular, the issue of determining threshold values, the excess of which leads to the extreme events, for example, in daily rainfalls or their intensities, is the key point of the study. Precipitation intensities are important not only for forecasting floods but also for solving problems such as runoff and soil erosion [50,51]. It can be explained by the contemporary climate change scenarios that predict a significant increase in the frequency of high intensity rainfall events, primarily in the dry areas. Moreover, precipitation can induce shallow landslides [52,53] and debris flows [54].

Statistical analysis of real data shows that the probability distribution of the precipitation intensity can be approximated by the gamma distribution with very high accuracy. In [55], some theoretical arguments were presented to justify the gamma model for the distribution of precipitation intensities. So, the statistical approach described in Section 6 and in [2] can be also used for identification of abnormally large intensities. For the analysis, the precipitation intensities in Potsdam and Elista (verified samples without missing values) are used as the initial data. This section presents a comparison of a non-parametric approach based on the extreme value theory as well as modified Peaks over Threshold (PoT) methodology [56] with the parametric approach that significantly involves testing parametric statistical hypotheses to determine extreme intensities of wet periods (see Section 6). The classical version of PoT [57] is quite popular for solving a wide range of climatic problems. In particular, the following results can be mentioned: the inverse Weibull distribution as an extreme wind speed model [58], a time-dependent versions of the PoT model for severe storm waves [59] and daily temperatures [60], probability model for rainfalls of high magnitude [61], analysis of precipitation extremes in a changing climate [62]. Most applications of the extreme value theory assume stationarity, but it is well-known that real events are not stationary. So, the generalized results analogous to Theorem 7 are required. All the numerical methods are implemented as a MATLAB program. Algorithm A4 demonstrates the method based on the PoT and GG-test (see Appendix A).

Figures 11 and 12 present the results obtained by the modified PoT algorithm in which the Weibull distribution is considered as the distribution of time between extreme events. Starting from the maximum threshold value that coincides with the maximum of the analyzed data, the hypothesis that the time intervals between the moments of excess of a certain threshold have the Weibull distribution is tested. The corresponding P -value is saved, and the threshold is shifted down by a certain (small) step (in this case, 0.01). It is worth noting that a similar procedure was suggested in [56] for precipitation volumes under the assumption that the time intervals between excesses have the exponential distribution. For a given significance level (in both cases, α is chosen as 0.01), the corresponding hypothesis is not rejected for all thresholds for which the P -values are located to the right of the red vertical line in the upper graphs in Figures 11 and 12. The lower graphs show the parameters of the fitted Weibull distribution.

On Figures 13 and 14 the results of the test (see Section 6) for both the GG and usual gamma distribution are compared with those of the PoT method based on the exponential and Weibull distributions for the intensities in Potsdam and Elista. The following notation is used:

- The thresholds with the indices low correspond to the minimum levels at which the hypothesis of exponential or Weibull distribution is not rejected (the lowest point to the right of the red line on the upper graphs in Figures 11 and 12);
- The thresholds with the indices maxval correspond to the maximum P -value (the rightmost point in the upper graphs);
- The thresholds with the indices high correspond to the upper level, when the corresponding hypothesis is not rejected (the highest point to the right of the red line on the graphs).

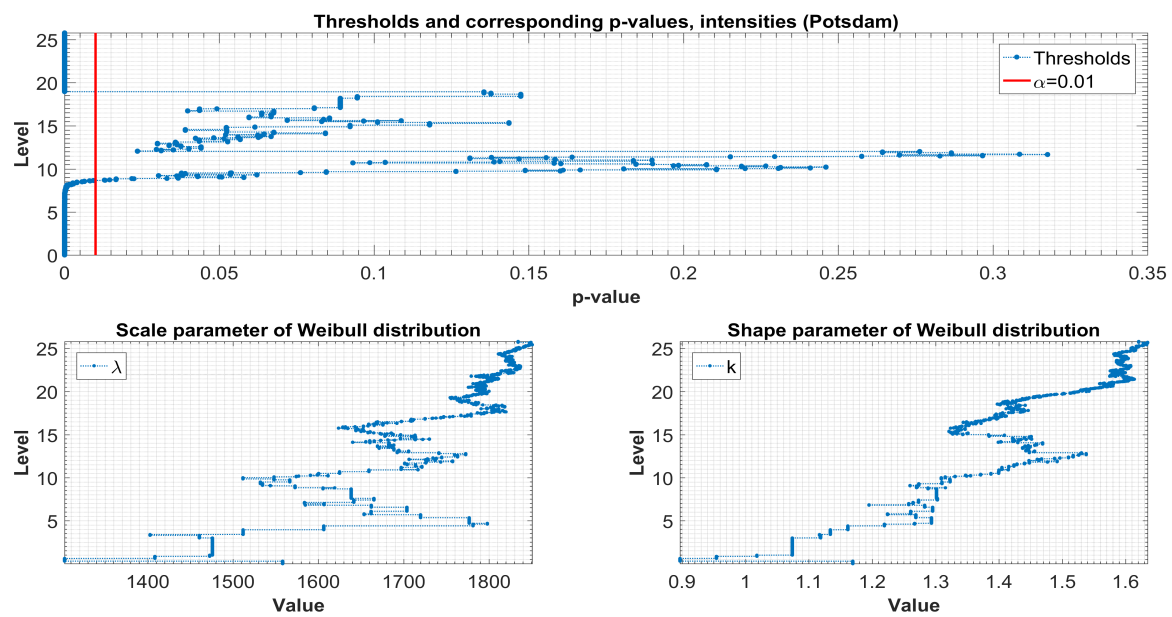


Figure 11. P-values that correspond to the various thresholds (Potsdam).

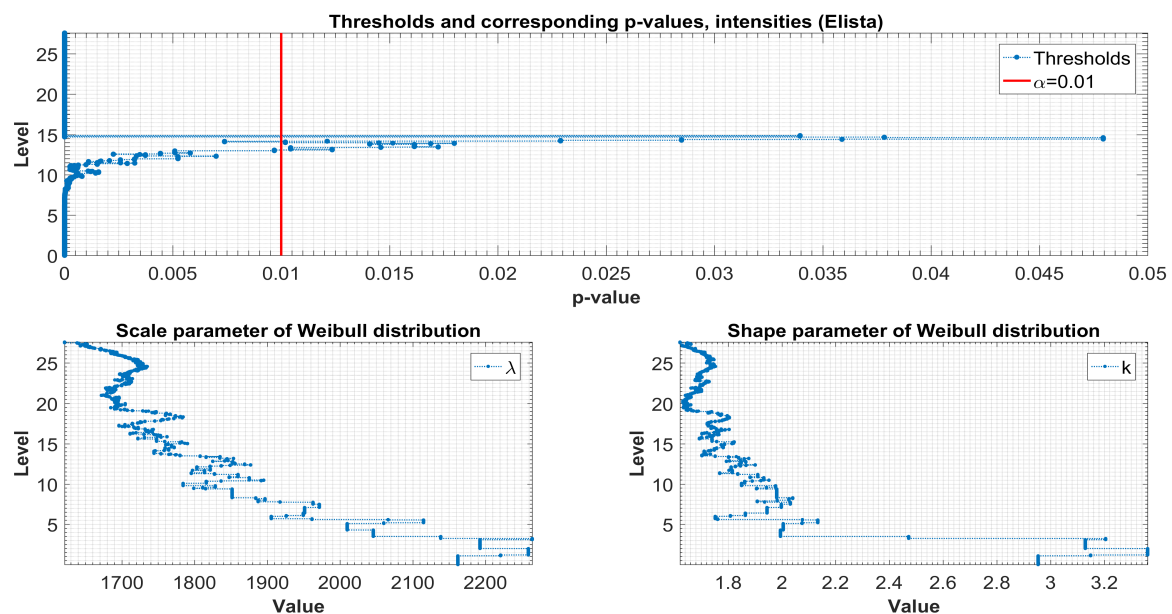


Figure 12. P-values that correspond to the various thresholds (Elista).

The green-filled downward-pointing triangles mark the intensities, which are classified as absolutely abnormal based on the GG test (see Section 6). The black upward-pointing triangles correspond to the decision based on the classical gamma distribution test (that is, $\gamma = 1$, see (31)). The circles denote intermediate extreme observations, and the squares mark relatively extreme ones. This classification is described in Section 7. It is worth noting that for the GG test in Potsdam the value of γ is 1.0775 and for Elista $\gamma = 1.1257$.

For Potsdam, the results of gamma and GG tests are good and close. In addition, the PoT method is also effective in the case where the threshold is chosen with the maximum P -value. However, for Elista, with less rainfalls with lower intensities, the results are quite different. Indeed, the decisions of the PoT method are close for the exponential and Weibull cases (the thresholds differ by only 0.29).

However, a statistical test based on the gamma distribution identifies only four intensities as absolutely extreme, while the GG test identifies more absolute extremes, including those below the thresholds mentioned above.

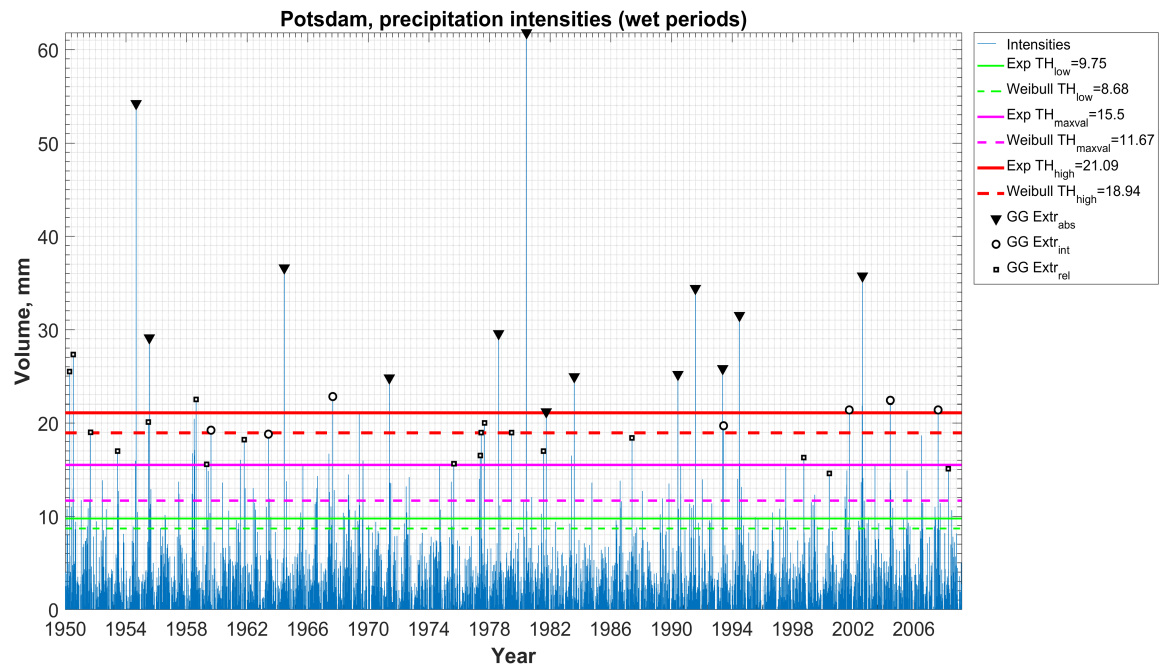


Figure 13. Comparison of statistical tests and peaks over threshold methodology for extreme precipitation intensities (Potsdam).

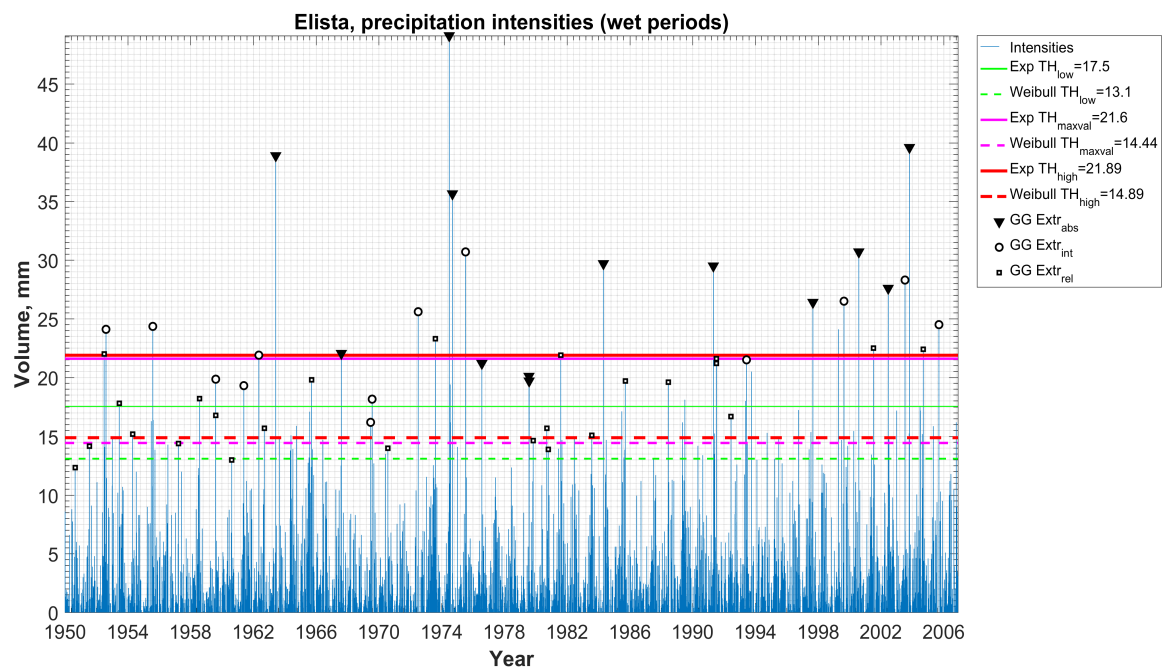


Figure 14. Comparison of statistical tests and peaks over threshold methodology for extreme precipitation intensities (Elista).

9. Conclusions and Discussion

In the paper, asymptotic models for some precipitation characteristics based on GNB distributions were considered. Also, a statistical test based on the GG distribution was proposed for the determination of the type of precipitation extremes. The GG and GNB distributions are not quite widespread, so the methods for the estimation of their parameters are, as a rule, not implemented in standard statistical packages. Therefore, the implementation of appropriate procedures requires the creation of specialized software solutions, for example, based on the functional approach, as it was done in the study described in this paper using the MATLAB programming language. However, as was demonstrated in the paper, the results of fitting such distributions to real data turned out to be better as compared to conventional models. Therefore, for processing spatial meteorological data from a large number of stations, the proposed methods and models can be effectively implemented as high-performance computing services.

Author Contributions: Conceptualization, V.K., A.G.; formal analysis, V.K., A.G.; funding acquisition, A.G.; investigation, A.G., V.K.; methodology, V.K., A.G.; project administration, V.K., A.G.; resources, A.G.; software, A.G.; supervision, V.K.; validation, A.G.; visualization, A.G.; writing—original draft, V.K., A.G.; writing—review and editing, A.G., V.K. All authors have read and agreed to the published version of the manuscript.

Funding: The results presented in Section 8 were obtained by Andrey Gorshenin to meet the research goals of grant No 18-71-00156 of the Russian Science Foundation.

Acknowledgments: Authors thank the reviewers for their valuable comments that helped to improve the presentation of the material.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Algorithms

This section presents the above-described algorithms in the form pseudo-code. All of these algorithms have been implemented using Matlab programming language without any tools specific to this development environment. They can be successfully implemented, for example, with the Python programming language taking into account some minor changes.

Algorithm A1. GNB approximations

```

1: LOAD(Data);                                     // Loading initial data (Potsdam, Elista)
2:  $\alpha=0.05$ ;                                     // Significance level
3: WP=WETPERIODS(Data);                             // Finding wet periods in Data
4:  $r=\text{NBFIT}(\text{WP}-1, \alpha)$ ;                       // Finding parameter  $r$ 
5:  $[\gamma_{l1}, \mu_{l1}, \text{err}_{l1}]=\text{GNBAPPROX}(l_1, r)$ ; // GNB approximation based on  $\ell_1$ -distance minimization
6:  $[\gamma_{l2}, \mu_{l2}, \text{err}_{l2}]=\text{GNBAPPROX}(l_2, r)$ ; // GNB approximation based on  $\ell_2$ -distance minimization
7:  $[\gamma_{l\infty}, \mu_{l\infty}, \text{err}_{l\infty}]=\text{GNBAPPROX}(l_{\infty}, r)$ ; // GNB approximation based on  $\ell_{\infty}$ -distance minimization
   // Plotting initial histograms and GNB approximations for all cases
8: HISTOGRAMS( $\gamma_{l1}, \mu_{l1}, \text{err}_{l1}, \gamma_{l2}, \mu_{l2}, \text{err}_{l2}, \gamma_{l\infty}, \mu_{l\infty}, \text{err}_{l\infty}$ );

```

Algorithm A2. Stabilization of averages

```

1: LOAD(Potsdam, Elista);                           // Loading initial data
2:  $m=3000$ ;                                           // The value can be chosen empirically
3: DATAPREPROCESSING(Potsdam, Elista);
4:  $[\beta_P, a_P]=\text{StabParams}(\text{Potsdam})$ ; // Search for stabilization parameters using Formulas (24) and (25)
5:  $[\beta_E, a_E]=\text{StabParams}(\text{Elista})$ ;
6: PLOTAVERAGES(Potsdam, Elista,  $\beta_P, a_P, \beta_E, a_E$ ); // Drawing results

```

Algorithm A3. Statistical test for extreme volumes

```

1:  $\alpha=0.01$ ; // Significance level
2: LOAD(Data); // Loading initial data (Potsdam, Elista)
// Correspondence between astronomical time and sample elements
3: window=DAYS TO OBSERVATIONS(Days);
4: Vols=VOLUMES(Data); // Volumes obtained from raw data
5:  $\gamma$ =GAMMAFIT(Vols,  $\alpha$ ); // Finding parameter  $\gamma$ 
6: [SR, SRGG]=GGSTATISTICS(Vols,  $\gamma$ ); // Finding values of statistics based on Formula (31)
7: [Extrabs, GG Extrabs, Extrint, GG Extrint, Extrrel, GG Extrrel]=GGTEST(SR, SRGG);
// Plotting initial data and decisions based on statistics SR and SRGG
8: PLOT EXTREMES(Extrabs, GG Extrabs, Extrint, GG Extrint, Extrrel, GG Extrrel);

```

Algorithm A4. PoT and GG-based test for intensities

```

1:  $\alpha=0.01$ ; // Significance level
2: LOAD(Data); // Loading initial data (Potsdam, Elista)
3: Ints=INTENSITIES(Data); // Intensities obtained from raw data
4: [GG Extrabs, GG Extrint, GG Extrrel]=GGTEST(Ints, SRGG);
// Modified PoT
5: level=MAX(Ints); // Initial PoT level (threshold) equal to maximum data value
6: L=LENGTH(Ints);
7: k=0;
// Determining the dependence of the level on the p-value
8: while level ≥ 0 do
9:   I= FIND(Ints > level);
// Minimum sufficient number of elements in the sample exceeding the threshold
10:  if LENGTH(I) < MinNum then
11:    level=level-step;
12:    CONTINUE;
13:  end if
14:  [ExpParam(k), ExpPval(k)] = FITEXP(I,  $\alpha$ );
15:  [WeibullParams(k), WeibullPval(k)] = FITWEIBULL(I,  $\alpha$ );
16:  k++;
17:  level=level-step;
18: end while
19: PLOT EXTREMES(GG Extrabs, GG Extrint, GG Extrrel); // Plotting intensity extremes
20: PLOT THRESHOLDS( ); // Plotting thresholds

```

References

1. Korolev, V.; Gorshenin, A.; Gulev, S.; Belyaev, K.; Grusho, A. Statistical Analysis of Precipitation Events. *AIP Conf. Proc.* **2017**, *1863*, 090011. [[CrossRef](#)]
2. Korolev, V.; Gorshenin, A.; Belyaev, K. Statistical tests for extreme precipitation volumes. *Mathematics* **2019**, *7*, 648. [[CrossRef](#)]
3. Korolev, V.; Zeifman, A. Generalized negative binomial distributions as mixed geometric laws and related limit theorems. *Lith. Math. J.* **2019**, *59*, 1461–1466. [[CrossRef](#)]
4. Zolina, O.; Simmer, C.; Belyaev, K.; Kapala, A.; Gulev, S.; Koltermann, P. Changes in the duration of European wet and dry spells during the last 60 years. *J. Clim.* **2013**, *26*, 2022–2047. [[CrossRef](#)]
5. Bening, V.; Korolev, V. *Generalized Poisson Models and Their Applications in Insurance and Finance*; VSP: Utrecht, The Netherlands, 2002.

6. Gnedenko, B.; Korolev, V. *Random Summation: Limit Theorems and Applications*; CRC Press: Boca Raton, FL, USA, 1996.
7. Grandell, J. *Doubly Stochastic Poisson Processes*; Lecture Notes Mathematics; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 1976; Volume 529.
8. Grandell, J. *Mixed Poisson Processes*; Chapman and Hall: London, UK, 1997.
9. Korolev, V.; Chertok, A.; Korchagin, A.; Zeifman, A. Modeling high-frequency order flow imbalance by functional limit theorems for two-sided risk processes. *Appl. Math. Comput.* **2015**, *253*, 224–241. [[CrossRef](#)]
10. Korolev, V.; Skvortsova, N. *Stochastic Models of Structural Plasma Turbulence*; VSP: Utrecht, The Netherlands, 2006.
11. Stacy, E. A generalization of the gamma distribution. *Ann. Math. Stat.* **1962**, *38*, 1187–1192. [[CrossRef](#)]
12. Greenwood, M.; Yule, G. An inquiry into the nature of frequency-distributions of multiple happenings, etc. *J. R. Stat. Soc.* **1920**, *83*, 255–279. [[CrossRef](#)]
13. Holla, M. On a Poisson-inverse Gaussian distribution. *Metrika* **1967**, *11*, 115–121. [[CrossRef](#)]
14. Sichel, H. On a family of discrete distributions particular suited to represent long tailed frequency data. In Proceedings of the 3rd Symposium on Mathematical Statistics, Pretoria, South Africa, 19–22 July 1971; pp. 51–97.
15. Korolev, V.; Korchagin, A.; Zeifman, A. Poisson theorem for the scheme of Bernoulli trials with random probability of success and a discrete analog of the Weibull distribution. *Informat. Appl.* **2016**, *10*, 11–20. [[CrossRef](#)]
16. Steutel, F.; van Harn, K. *Infinite Divisibility of Probability Distributions on the Real Line*; Marcel Dekker: New York, NY, USA, 2004.
17. Korolev, V. Limit distributions for doubly stochastically rarefied renewal processes and their properties. *Theory Probab. Appl.* **2016**, *61*, 753–773. [[CrossRef](#)]
18. Korolev, V.; Korchagin, A.; Zeifman, A. On doubly stochastic rarefaction of renewal processes. *AIP Conf. Proc.* **2017**, *1863*, 090010. [[CrossRef](#)]
19. Zaks, L.; Korolev, V. Generalized variance gamma distributions as limit laws for random sums. *Informat. Appl.* **2013**, *7*, 105–115. [[CrossRef](#)]
20. Korolev, V.; Bening, V.; Shorgin, S. *Mathematical Foundations of Risk Theory*, 2nd ed.; FIZMATLIT: Moscow, Russia, 2011.
21. Zolotarev, V. *One-Dimensional Stable Distributions*; American Mathematical Society: Providence, RI, USA, 1986.
22. Schneider, W. *Stable Distributions: Fox Function Representation and Generalization*; Springer: Berlin, Germany, 1986; pp. 497–511.
23. Uchaikin, V.; Zolotarev, V. *Infinite Divisibility of Probability Distributions on the Real Line*; VSP: Utrecht, The Netherlands, 1999.
24. Shanbhag, D.; Sreehari, M. On certain self-decomposable distributions. *Z. Für Wahrscheinlichkeitstheorie Und Verwandte Geb.* **1977**, *38*, 217–222. [[CrossRef](#)]
25. Korolev, V. Product representations for random variables with the Weibull distributions and their applications. *J. Math. Sci.* **2016**, *218*, 298–313. [[CrossRef](#)]
26. Gleser, L. The gamma distribution as a mixture of exponential distributions. *Am. Stat.* **1989**, *43*, 115–117. [[CrossRef](#)]
27. Korolev, V. On convergence of distributions of compound Cox processes to stable laws. *Theory Probab. Appl.* **1999**, *43*, 644–650. [[CrossRef](#)]
28. Korolev, V. Convergence of random sequences with independent random indexes. I. *Theory Probab. Appl.* **1994**, *39*, 313–333. [[CrossRef](#)]
29. Korolev, V. Convergence of random sequences with independent random indexes. II. *Theory Probab. Appl.* **1995**, *40*, 770–772. [[CrossRef](#)]
30. Korolev, V.; Gorshenin, A.; Sokolov, I. Max-compound Cox processes. I. *J. Math. Sci.* **2019**, *237*, 789–803. [[CrossRef](#)]
31. Korolev, V.; Gorshenin, A. The probability distribution of extreme precipitation. *Dokl. Earth Sci.* **2017**, *477*, 1461–1466. [[CrossRef](#)]
32. Gorshenin, A.; Korolev, V. Scale mixtures of Frechet distributions as asymptotic approximations of extreme precipitation. *J. Math. Sci.* **2018**, *234*, 886–903. [[CrossRef](#)]

33. Johnson, N.; Kot, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Wiley: New York, NY, USA, 1995; Volume 2.
34. Goldie, C. A class of infinitely divisible distributions. *Math. Proc. Camb. Philos. Soc.* **1967**, *63*, 1141–1143. [[CrossRef](#)]
35. Korolev, V.; Gorshenin, A.; Sokolov, I. Max-compound Cox processes. III. *arXiv* **2020**, arXiv:1912.02237v2.
36. Galambos, J. *The Asymptotic Theory of Extreme Order Statistics*; Wiley: New York, NY, USA, 1978.
37. Rényi, A. A Poisson-folyamat egy jellemzése. *Magy. Tud. Acad. Mat. Kut. Int. Közl.* **1956**, *1*, 519–527. [[CrossRef](#)]
38. Rényi, A. On an extremal property of the Poisson process. *Ann. Inst. Stat. Math.* **1964**, *16*, 129–133. [[CrossRef](#)]
39. Kalashnikov, V. *Geometric Sums: Bounds for Rare Events with Applications*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1997.
40. Zolotarev, V. Approximation of Distributions of Sums of Independent Random Variables with Values in Infinite-Dimensional Spaces. *Theory Probab. Appl.* **1976**, *21*, 721–737. [[CrossRef](#)]
41. Zolotarev, V. Ideal Metrics in the Problem of Approximating Distributions of Sums of Independent Random Variables. *Theory Probab. Appl.* **1977**, *22*, 433–449. [[CrossRef](#)]
42. Zolotarev, V. *Modern Theory of Summation of Random Variables*; VSP: Utrecht, The Netherlands, 1997.
43. Korolev, V.; Zeifman, A. Bounds for convergence rate in laws of large numbers for mixed Poisson random sums. *arXiv* **2020**, arXiv:2003.12495.
44. Cheng, L.; AghaKouchak, A. Nonstationary Precipitation Intensity-Duration-Frequency Curves for Infrastructure Design in a Changing Climate. *Sci. Rep.* **2014**, *4*, 7093. [[CrossRef](#)]
45. Wasko, C.; Sharma, A. Steeper temporal distribution of rain intensity at higher temperatures within Australian storms. *Nat. Geosci.* **2015**, *8*, 527–U166. [[CrossRef](#)]
46. Mo, C.; Ruan, Y.; He, J.; Jin, J.; Liu, P.; Sun, G. Frequency analysis of precipitation extremes under climate change. *Int. J. Climatol.* **2019**, *39*, 1373–1387. [[CrossRef](#)]
47. Donat, M.; Angelil, O.; Ukkola, A. Intensification of precipitation extremes in the world’s humid and water-limited regions. *Environ. Res. Lett.* **2019**, *14*. [[CrossRef](#)]
48. Groisman, P.; Knight, R.; Karl, T. Changes in Intense Precipitation over the Central United States. *J. Hydrometeorol.* **2012**, *13*, 47–66. [[CrossRef](#)]
49. Xu, C.; Qiao, Y.; Jian, M. Interdecadal Change in the Intensity of Interannual Variation of Spring Precipitation over Southern China and Possible Reasons. *J. Clim.* **2013**, *32*, 5865–5881. [[CrossRef](#)]
50. Ziadat, F.; Taimeh, A. Effect of Rainfall Intensity, Slope, Land Use and Antecedent Soil Moisture on Soil Erosion in an Arid Environment. *Land Degrad. Dev.* **2013**, *24*, 582–590. [[CrossRef](#)]
51. Jomaa, S.; Barry, D.; Rode, M.; Sander, G.; Parlange, J. Linear scaling of precipitation-driven soil erosion in laboratory flumes. *Catena* **2017**, *152*, 285–291. [[CrossRef](#)]
52. Bezak, N.; Auflic, M.J.; Mikos, M. Application of hydrological modelling for temporal prediction of rainfall-induced shallow landslides. *Landslides* **2019**, *16*, 1273–1283. [[CrossRef](#)]
53. Bliznak, V.; Kaspar, M.; Muller, M.; Zacharov, P. Sub-daily temporal reconstruction of extreme precipitation events using NWP model simulations. *Atmos. Res.* **2019**, *224*, 65–80. [[CrossRef](#)]
54. Huang, W.; Nychka, D.; Zhang, H. Estimating precipitation extremes using the log-histospline. *Environmetrics* **2019**, *30*. [[CrossRef](#)]
55. Martinez-Villalobos, C.; Neelin, J. Why Do Precipitation Intensities Tend to Follow Gamma Distributions? *J. Atmos. Sci.* **2019**, *76*, 3611–3631. [[CrossRef](#)]
56. Gorshenin, A.; Korolev, V. Determining the extremes of precipitation volumes based on a modified “Peaks over Threshold”. *Inform. I Ee Primen.* **2018**, *12*, 16–24. [[CrossRef](#)]
57. Leadbetter, M. On a basis for “Peaks over Threshold” modeling. *Stat. Probab. Lett.* **1991**, *12*, 357–362. [[CrossRef](#)]
58. Simiu, E.; Heckert, N. Extreme wind distribution tails: A “peaks over threshold” approach. *J. Struct. Eng.-ASCE* **1996**, *122*, 539–547. (539). [[CrossRef](#)]
59. Mendez, F.; Menendez, M.; Luceno, A.; Losada, I. Estimation of the long-term variability of extreme significant wave height using a time-dependent Peak Over Threshold (POT) model. *J. Geophys. Res.-Ocean.* **2006**, *111*, C07024. [[CrossRef](#)]

60. Kyselý, J.; Píček, J.; Beranova, R. Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Glob. Planet. Chang.* **2010**, *72*, 55–68. [[CrossRef](#)]
61. Begueria, S.; Angulo-Martinez, M.; Vicente-Serrano, S.; Lopez-Moreno, I.; El-Kenawy, A. Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: A case study in northeast Spain from 1930 to 2006. *Int. J. Climatol.* **2011**, *31*, 2102–2114. [[CrossRef](#)]
62. Roth, M.; Buishand, T.; Jongbloed, G.; Tank, A.; van Zanten, J. A regional peaks-over-threshold model in a nonstationary climate. *Water Resour. Res.* **2012**, *48*, W11533. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).