Harry Pavlopoulos; Jan Picek; Jana Jurečková

Heavy tailed durations of regional rainfall

Applications of Mathematics, Vol. 53 (2008), No. 3, 249--265

Persistent URL: `http://dml.cz/dmlcz/140319`

53 (2008) APPLICATIONS OF MATHEMATICS No. 3, 249–265

HEAVY TAILED DURATIONS OF REGIONAL RAINFALL*

Harry Pavlopoulos, Athens, Jan Picek, Liberec,
Jana Jurečková, Praha

(Invited)

Abstract. Durations of rain events and drought events over a given region provide important information about the water resources of the region. Of particular interest is the shape of upper tails of the probability distributions of such durations. Recent research suggests that the underlying probability distributions of such durations have heavy tails of hyperbolic type, across a wide range of spatial scales from 2 km to 120 km. These findings are based on radar measurements of spatially averaged rain rate (SARR) over a tropical oceanic region. The present work performs a nonparametric inference on the Pareto tail-index of wet and dry durations at each of those spatial scales, based on the same data, and compares it with conclusions based on the classical Hill estimator. The results are compared and discussed.

Keywords: wet and dry durations of regional rainfall, quantile multiscaling, heavy tails, Pareto tail-index, semi-parametric statistical inference

MSC 2010: 62G32, 62P12

1. Introduction

Rainfall is a physical phenomenon characterized by intermittency between dry and wet states across a wide range of scales of observations, both spatial and temporal. This type of intermittency combined with the "wild" variability of rainfall intensity where and when it rains, from mild to moderate or higher and occasionally cataclysmic levels, have led to the recognition of rain rate fields as multifractal structures (Lovejoy and Mandelbrot [23], Lovejoy and Schertzer [24], [25], Gupta and Waymire [11], [12], Tessier et al. [40], Over and Gupta [30], [31], Marsan et al. [29], Foufoula-Georgiou [8]). Rainfall intensity is measured by *rain rate* in mm/hr units,

* The authors express sincere thanks to the Mathematisches Forschungsinstitut Oberwolfach (MFO) for facilitating their collaboration under a "*Research in Pairs*" project hosted at MFO during March 5–25, 2006. The research of the second and third authors was supported by the project LC06024.

representing the flux of water volume carried by rain droplets passing through (or landing on) an elementary surface of unitary area per unit of time. The dry state of rainfall is represented by zero rain rate and its wet state is represented by positive values of rain rate. If a field of rain rates is mapped over a (two-dimensional) geographic region at a given time instant, then the support of wet states constitutes a multifractal subset of the mapped region.

Alternatively, if the observed rain rate field at a fixed time is spatially averaged over the region, or if the small region can be approximately considered as a geometric point, then the support of wet states becomes a multifractal subset of the time interval during which rainfall is recorded. Dry epochs are defined as maximal time segments with zero rain rate (everywhere in the region), and wet epochs are defined as maximal time segments with positive rain rate (somewhere in the region). The lengths of such segments are referred to as dry and wet durations, respectively.

The shape of probability tails of durations of wet and dry epochs is of importance for practice (e.g., for the management of storage and consumption of water resources). Some authors used hydrological models for regionalization of hydrologic extremes in large basins of river networks (Gupta and Waymire [13], Gupta [10]). Empirical conclusions, pointing to hyperbolic probability tails of wet and dry epoch durations, can be justified by the perception of rain fields as intermittent multifractal structures (Mandelbrot [28]). The Pareto tail-index $(\Delta + 1)$ of dry durations is directly related with the (capacity) fractal dimension (Δ) of the (temporal) support of wet states (Lowen and Teich [26], Schmitt et al. [37], Pavlopoulos and Gupta [34]). Moreover, the probability of wet states (i.e. the probability of raining) in aggregated intermittent records or maps of rain rate, behaves as a power-law of the scale of aggregation, with the exponent determined by the fractal dimension of the initial (i.e. before aggregation) record or map (Kedem and Chiu [21], Over and Gupta [30], Kundu and Bell [22]). There is also emerging evidence that heavier tail probabilities of wet and dry epoch durations have an impact on the global memory properties of the underlying rainfall process (De Michele and Pavlopoulos [5]). Heavy tailed distributions of OFF durations are often causally associated with the long range dependence (LRD) (Taqqu and Levy [39], Willinger et al. [41], Heath et al. [15], Adler et al. [1], Doukhan et al. [6], Lowen and Teich [27]).

Pavlopoulos and Gupta [34] studied the effect of the scale of the considered region on the higher orders quantiles of durations of wet and dry epochs of spatially averaged rain rate (SARR). Their main result is that durations of regionally wet and dry epochs have probability tails of hyperbolic type, in all the probed spatial scales, with specific estimated values of the associated Pareto tail index. The data used in that study and in the present work are briefly presented in Section 2, while the scaling properties of their sample quantiles are summarized in Section 3. The main goal of the present paper is to test whether the predicted estimates of Pareto tail index, obtained by multiscaling analysis of wet and dry durations of SARR, are acceptable as trustworthy indicators of tail heaviness. Our main tool is a class of non-parametric tests of hypotheses, confidence intervals and associated point estimators, recently developed by the present as well as other authors. After a brief description of these methods in Section 4, they are applied to real data in Section 5.

2. Description of working data and related issues

The raw data used by Pavlopoulos and Gupta [34] is a time series of digital maps of radar reflectivity measurements. These maps were obtained during the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean-Atmosphere Response Experiment (COARE) by a shipboard Doppler precipitation radar (MIT). Each map corresponds to a single radar scan, probing a fixed oceanic region of reference S, with an area $240 \times 240\,\mathrm{km}^2$, in the tropical sector of South Pacific Ocean (China Sea: 2°S, 156°E). The temporal resolution between successive scans is (approximately) 20 minutes. Reflectivity measurements Z from each scan, binned over square pixels of area $2 \times 2\,\mathrm{km}^2$, have been converted to instantaneous rain rate R by the Z-R relationship $R = (Z/230)^{0.8}$, rendering a series of retrieved rain rate digital maps. The entire series corresponds to the full period of Cruise 1 (November 10, 1992 through December 9, 1992), consisting of 1992 scans, and to the early part of Cruise 2 (December 21, 1992 through December 29, 1992), consisting of 617 scans. A good source of detailed information about TOGA-COARE and its objectives is Short et al. [38].

The experience showed that the multiscaling analysis of wet and dry durations can be restricted to spatial scales ranging from 120 km down to 2 km, following the rule of half (approximately). This amounts to a total of seven scales, 120, 60, 30, 16, 8, 4, 2 km, of which the largest is referred to as *scale of reference* and the smallest as *pixel scale*. It is convenient to refer to spatial scales in terms of a unit-free *scale index* $\lambda \in (0, 1]$, formally defined as the ratio of diameters of two geometrically similar subregions, say A and A_λ, such that $A_\lambda \subset A$. For the set of 7 scales, 25 different nestings of square (i.e. geometrically similar) subregions were sampled according to a certain symmetric design of spatial sampling. A time series of spatially averaged rain rate was obtained separately for Cruise 1 and Cruise 2 on each sampled subregions. Spells of zeros and spells of positive values were identified as dry and wet epochs, respectively. The integer-valued lengths of these spells, multiplied by 1/3, provide "*quantized*" working data of dry and wet durations in units of hours (hr).

To suppress some bias of the data and to reduce the effect of the skewness, working data from both cruises were pooled temporally and spatially. The final product of the overall pooling amounts to 14 sets (7 wet and 7 dry) of *spatio-temporally pooled* working data of durations. The spatio-temporal pooling strategy is justifiable only under conditions of “Temporal Homogeneity” and “Spatial Homogeneity” of probability distributions of durations, which was nonparametrically verified in Pavlopoulos and Gupta [34, Section 4].

3. Multiscaling of tail quantiles

The spatio-temporally pooled working data of wet and dry durations can be considered as samples from the populations of random variables W_λ and D_λ, respectively, with the corresponding quantile functions $Q_\lambda^{(w)}(p)$ and $Q_\lambda^{(d)}(p)$. Pavlopoulos and Gupta [34] formulated the following parametric models for *tail quantile functions* of wet and dry epoch durations:

$$Q_\lambda^{(w)}(p) = \mathrm{e}^{\alpha \ln \lambda + \beta} \cdot (1-p)^{\gamma \ln \lambda + \delta}, \tag{3.1}$$

$$Q_\lambda^{(d)}(p) = \mathrm{e}^{\alpha^* \lambda + \beta^*} \cdot (1-p)^{\gamma^* \lambda + \delta^*}, \tag{3.2}$$

where $\alpha = 0.3652$, $\beta = 0.8746$, $\gamma = 0.0285$, $\delta = -0.5006$, $\alpha^* = -0.5117$, $\beta^* = 0.2327$, $\gamma^* = 0.379$, $\delta^* = -0.57$ are estimates of parameter values obtained through logarithmic regression in the range of $0.8 \leqslant p \leqslant 0.995$.

This implies that the scaling of wet tail-quantiles is of power-law type,

$$Q_\lambda^{(w)}(p) = \lambda^{\alpha + \gamma \ln(1-p)} Q_1^{(w)}(p), \tag{3.3}$$

while the scaling of dry tail-quantiles is of exponential type,

$$Q_\lambda^{(d)}(p) = \mathrm{e}^{[\alpha^* + \gamma^* \ln(1-p)](\lambda - 1)} Q_1^{(d)}(p), \tag{3.4}$$

with respect to λ for a given tail probability level p near to 1. Figs. 1–2 depict QQ-plots between sample tail quantiles and predicted tail quantiles according to the multiscaling models (3.1) and (3.2), respectively. Both models constitute quite significant improvements when compared against power-law *simple scaling* models (i.e. $\gamma = 0$ in (3.3)). Simple scaling for both wet and dry duration quantiles has been investigated as a potentially valid theory for small scales on the basis of TOGA-COARE releases (Gritsis [9], Pavlopoulos and Gritsis [32], Pavlopoulos and Gupta [33]).

Inverting formulae (3.1) and (3.2), one easily obtains (upper) tail probabilities of wet durations

$$P(W_\lambda > u) = \mathrm{e}^{-(\alpha \ln \lambda + \beta)/(\gamma \ln \lambda + \delta)} \cdot u^{1/(\gamma \ln \lambda + \delta)} \tag{3.5}$$

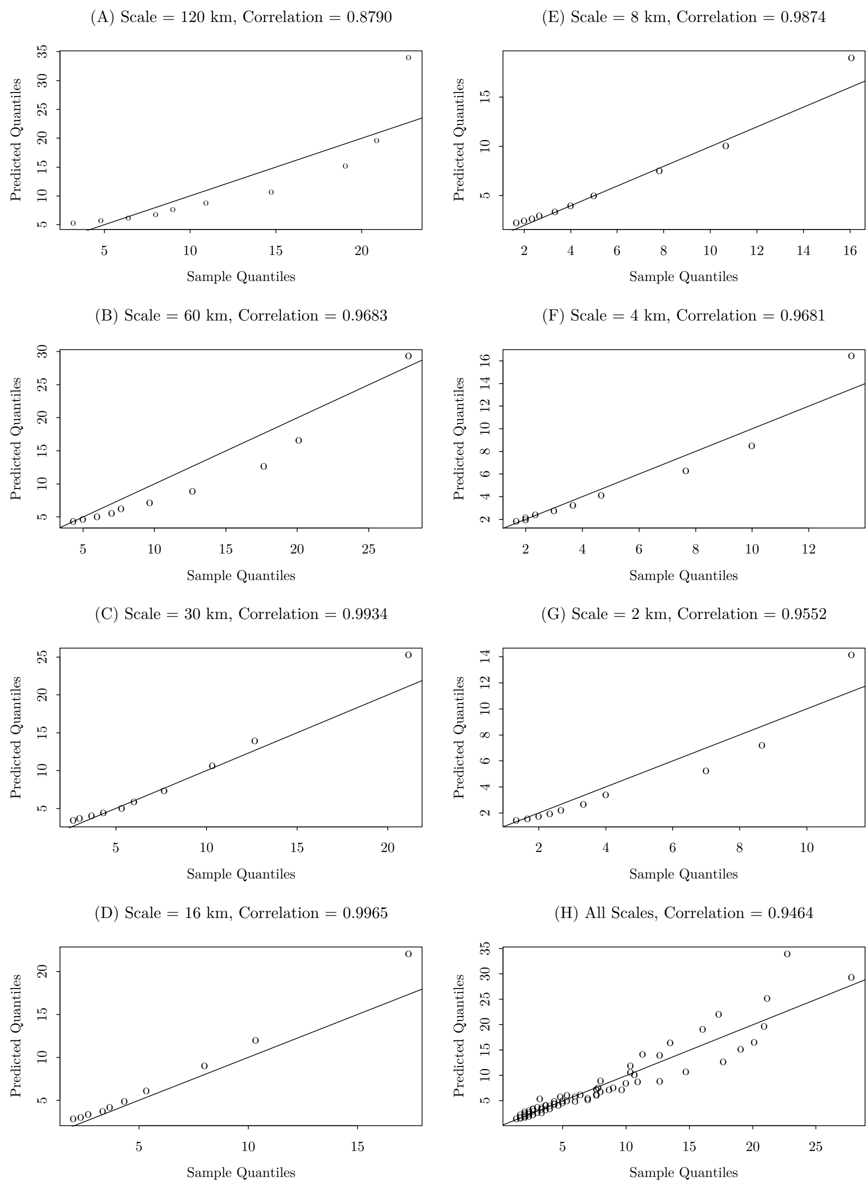

Figure 1. Q-Q-plots of wet duration quantiles predicted by the *power-law* multiscaling model (3.1), versus sample estimates of wet duration quantiles, at probability levels 0.8, 0.825, 0.85, 0.875, 0.9, 0.925, 0.95, 0.975, 0.985, 0.995, for each of the seven scales considered (A-G plots) and collectively across all scales combined (H plot). The correlation coefficient reported on each plot was obtained from simple linear regression of predicted tail-quantiles against their sample estimates. The line drawn in each plot is the diagonal through the origin.

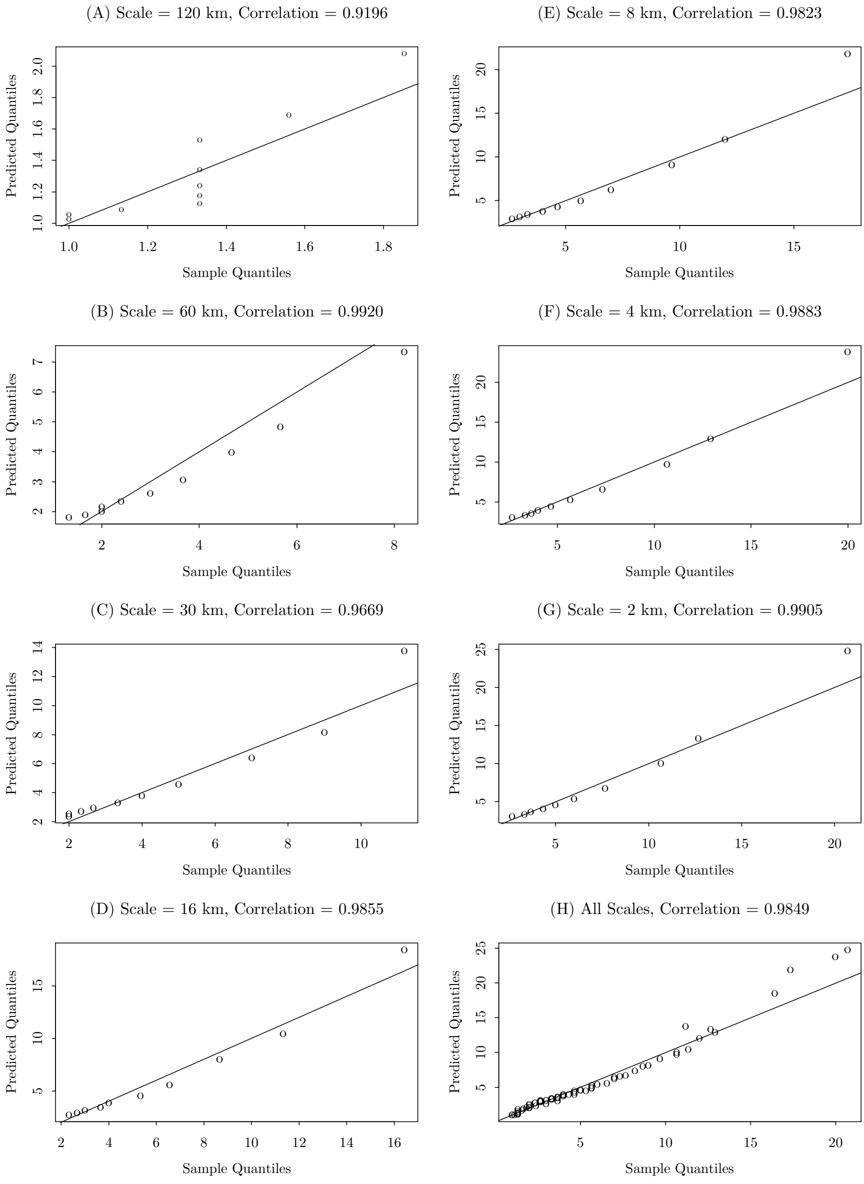

Figure 2. Q-Q-plots of dry duration quantiles predicted by the *exponential* multiscaling model (3.2), versus sample estimates of dry duration quantiles, at probability levels 0.8, 0.825, 0.85, 0.875, 0.9, 0.925, 0.95, 0.975, 0.985, 0.995, for each of the seven scales considered (A-G plots) and collectively across all scales combined (H plot). The correlation coefficient reported on each plot was obtained from simple linear regression of predicted tail-quantiles against their sample estimates. The line drawn in each plot is the diagonal through the origin.

for $u \geqslant Q_\lambda^{(w)}(0.8) = \mathrm{e}^{\alpha \ln \lambda + \beta} \cdot 0.2^{\gamma \ln \lambda + \delta}$, and (upper) tail probabilities of dry durations

$$P(D_\lambda > u) = \mathrm{e}^{-(\alpha^* \lambda + \beta^*)/(\gamma^* \lambda + \delta^*)} \cdot u^{1/(\gamma^* \lambda + \delta^*)} \tag{3.6}$$

for $u \geqslant Q_\lambda^{(d)}(0.8) = \mathrm{e}^{\alpha^* \lambda + \beta^*} \cdot 0.2^{\gamma^* \lambda + \delta^*}$. Formulae (3.5) and (3.6) not only reveal that durations of wet and dry epochs have hyperbolic tails in all the spatial scales, but also provide specific estimates of the Pareto tail index corresponding to each scale, namely

$$\begin{aligned} m_0(\lambda) &= -(\gamma \ln \lambda + \delta)^{-1} \quad \text{for wet tails,} \\ m_0^*(\lambda) &= -(\gamma^* \lambda + \delta^*)^{-1} \quad \text{for dry tails,} \end{aligned}$$

respectively (see also Tabs. 1–4).

These estimates reveal that the tail indices of both wet and dry durations increase in the scale index λ and that the tails might be heavier in smaller regions than in larger ones. Also, the tails of the wet duration appear to be potentially heavier than the tails of the dry duration at each given scale of observations.

4. Semiparametric inference for Pareto tail index

4.1. Inference based on the Hill estimator

We work with the *heavy tailed* probability distribution whose cumulative distribution function F satisfies

$$\lim_{u \to \infty} \frac{-\log(1 - F(u))}{m \ln u} = 1 \tag{4.1}$$

for some $0 < m < \infty$. The number m in (4.1) is referred to as the *Pareto tail index* of F and (4.1) implies that the tail probability has the form

$$1 - F(u) = u^{-m} L(u), \quad u \in \mathbb{R}, \tag{4.2}$$

with L being a positive function slowly varying at infinity, i.e. $\lim_{u \to \infty} L(au)/L(u) = 1$ for all $a > 0$ (see Embrechts et al. [7, Theorem 3.3.7]).

Our interest is focused partly on obtaining confidence intervals and partly on testing hypotheses on the Pareto tail index.

The standard approach to obtain confidence intervals for the tail index is by exploiting asymptotic normal distribution of the Hill estimator

$$H(k) = \frac{1}{k} \sum_{i=1}^{k} \log X_{(l-i+1:l)} - \log X_{(l-k:l)},$$

where $X_{(i:l)}$ denotes the ith order statistic of the sample $X_1, \ldots, X_l$. In this spirit, Cheng and Pan [2] considered a one-term Edgeworth expansion of the distribution function of the Hill estimator in cases where the asymptotic bias is zero, while Cheng and Peng [3] provided an algorithm for computing a *plug-in* value of the theoretically optimal sample fraction k^*, in the sense of minimizing the *absolute coverage error* of the confidence interval. The optimal sample fraction k^* has been derived under the instrumental condition that the scalar function L is of a specific form, such that the underlying heavy tailed distribution corresponds to a special case of second-order regular variation (see de Haan and Stadtmüller [4]).

The two-sided confidence interval for the Pareto tail index $m = 1/\xi$ obtained by this approach, at level $0 < \alpha < 1$, is given by

$$I_2(\alpha, k) = \left(\frac{\sqrt{k}}{\sqrt{k}H(k) + \Phi^{-1}(1-\alpha/2)H(k)}, \frac{\sqrt{k}}{\sqrt{k}H(k) - \Phi^{-1}(1-\alpha/2)H(k)} \right),$$

while the one-sided right and left intervals are respectively given by

$$I_1(\alpha, k) = \left(\frac{\sqrt{k}}{\sqrt{k}H(k) + \Phi^{-1}(1-\alpha)H(k)}, \infty \right) \& \left(0, \frac{\sqrt{k}}{\sqrt{k}H(k) - \Phi^{-1}(1-\alpha)H(k)} \right),$$

where Φ denotes the CDF of the standard normal law. These are also referred to as *Wald* intervals, among several other types of intervals reviewed in a recent paper by Haeusler and Segers [14], elaborating some new developments on Hill-based confidence intervals for $\xi = 1/m$ from Edgeworth expansions under certain asymptotic conditions on the bias.

4.2. Inference based on sub-sample statistics

To verify the reliability of estimates $m_0(\lambda)$ and $m_0^*(\lambda)$ of the Pareto tail index associated with hyperbolic tails of wet and dry durations of SARR, we can test the hypotheses of the form $\mathbf{H}_{m_0}\colon\ m \leqslant m_0$ (or $\mathbf{H}^*_{m_0}\colon\ m \geqslant m_0$) against one-sided alternatives $\mathbf{K}_{m_0}\colon\ m > m_0$, (or $\mathbf{K}^*_{m_0}\colon\ m < m_0$, respectively), for some motivated values $m_0 > 0$ of $m_0(\lambda)$ and $m_0^*(\lambda)$.

Several semiparametric tests of $\mathbf{H}_{m_0}$ and $\mathbf{H}^*_{m_0}$ have been recently developed by Jurečková [16], [17], Jurečková and Picek [19], Picek and Jurečková [36] and Jurečková et al. [18]. The idea in this approach is to split the sample into a set of N independent subsamples $\mathbf{X}_j = (X_{j1}, \ldots, X_{jn})'$, for $j = 1, \ldots, N$, and to represent each one by a suitable statistic, say S_j, $j = 1, \ldots, N$. The consistent and asymptotically normal testing procedures are based on the empirical distribution functions of $S_1, \ldots, S_N$. Specific choices of S_j, $j = 1, \ldots, N$ are as follows:

sub-sample maxima: $X^{(j)}_{(n)} = \max\{X_{j1}, \ldots, X_{jn}\}$, $j = 1, \ldots, N$,

sub-sample means: $\overline{X}_n^{(j)} = \dfrac{1}{n}\sum_{i=1}^{n} X_{ji}$, $j = 1, \dots, N$,

sub-sample averaged block maxima: $\hat{\theta}_n^{(j)} = (X_j^{(1)} + X_j^{(2)})/4$, where $X_j^{(1)} = X_{(\nu)}^{(j)} = \max\{X_{j1}, \dots, X_{j\nu}\}$, $X_j^{(2)} = \max\{X_{j(\nu+1)}, \dots, X_{jn}\}$ for $j = 1, \dots, N$ and some fixed $1 \leqslant \nu \leqslant n-1$.

In particular, the test based on sub-sample maxima rejects $\mathbf{H}_{m_0}\colon\ m \leqslant m_0$ in favor of $\mathbf{K}_{m_0}\colon\ m > m_0$ at the asymptotic significance level $0 < \alpha < 1$, when

either $1 - \hat{F}_N^*(u_{N,m_0}) = 0$,
or $1 - \hat{F}_N^*(u_{N,m_0}) > 0$ and
$N^{\delta/2}[-\log(1 - \hat{F}_N^*(u_{N,m_0})) - (1-\delta)\log N] \geqslant \Phi^{-1}(1-\alpha)$,

where $u_{N,m} := (nN^{1-\delta})^{1/m}$ for a chosen constant $0 < \delta < \frac{1}{2}$.

On the other hand, the test rejects $\mathbf{H}^*_{m_0}\colon\ m \geqslant m_0$ in favor of $\mathbf{K}^*_{m_0}\colon\ m < m_0$, when

either $\hat{F}_N^*(u_{N,m_0}) = 0$,
or $\hat{F}_N^*(u_{N,m_0}) > 0$ and
$N^{\delta/2}[-\log(1 - \hat{F}_N^*(u_{N,m_0})) - (1-\delta)\log N] \leqslant \Phi^{-1}(\alpha)$.

The tests lead to one-sided confidence intervals (Picek [35]). The right-sided interval is:

$$J_1(\alpha,\delta) = \left(\frac{\log(nN^{1-\delta})}{\log\bigl(\hat{F}_N^{*-1}(1 - \exp\{-N^{-\delta/2}\Phi^{-1}(1-\alpha) - (1-\delta)\log N\})\bigr)}\,,\infty\right)$$

and the left-sided interval is

$$J_1(\alpha,\delta) = \left(0\,, \frac{\log(nN^{1-\delta})}{\log\bigl(\hat{F}_N^{*-1}(1 - \exp\{N^{-\delta/2}\Phi^{-1}(1-\alpha) - (1-\delta)\log N\})\bigr)}\right).$$

Moreover, inverting the tests in the Hodges-Lehmann manner, one obtains strongly consistent point estimators of m (Jurečková and Picek [20]), given by

$$M^*(\delta) = \frac{1}{2}(M_+^* + M_-^*),$$
$$M_+^* := \sup\{m\colon\ 1 - \hat{F}_N^*(u_{N,m}) < N^{-(1-\delta)}\},$$
$$M_-^* := \inf\{m\colon\ 1 - \hat{F}_N^*(u_{N,m}) > N^{-(1-\delta)}\}.$$

Extensive simulation studies show that these tests distinguish very well the tail behavior among different types of distributions, even for moderate sample sizes. Yet, the performance of these procedures is affected by the chosen value of the constant δ.

5. Results and discussion

The procedures described in Section 4 were applied to each of the 14 sets of spatio-temporally pooled working data of wet and dry durations of SARR across the considered seven spatial scales of averaging. Our focus of interest is mainly on confidence intervals and on testing one-sided hypotheses about the Pareto tail index m of the underlying probability distributions of wet and dry durations at each spatial scale. The assumption of heavy-tailed distributions, as well as other assumptions, are adopted in light of the evidence established by the multiscaling analysis of sample tail quantiles of the same data. The numerical results are based on the use of sub-sample maxima, as described in Section 4. Our conclusions are based on the following Tabs. 1–4.

Tab. 1 tabulates P-values for testing one-sided hypotheses $\mathbf{H}_{m_0}$ and $\mathbf{H}^*_{m_0}$, with m_0 equal to the predicted benchmark values $m_0(\lambda)$ and $m^*_0(\lambda)$. The P-values marked by $*$ indicate the conclusion that Pareto tail index does not exceed the benchmark values. Then we conclude that tails are at least as heavy as predicted by the multiscaling analysis of sample tail quantiles. Pairs of P-values marked in boldface indicate the admissibility of the corresponding predicted benchmark value m_0. For example, wet durations at the 4 km scale for $\delta = 0.25$ give P-value 0.982 to $\mathbf{H}_{1.67}\colon\ m \leqslant 1.67$ and P-value 0.018 to $\mathbf{H}^*_{1.67}\colon\ m \geqslant 1.67$, indicating that the formal intersection $\mathbf{H}_{1.67} \cap \mathbf{H}^*_{1.67}\colon\ m = 1.67$ may be considered non-rejectable at the 0.01 level of significance. In this sense, the predicted benchmark value 1.67 is considered *admissible*.

Tabs. 2 and 3 tabulate the right and left one-sided confidence intervals for the Pareto tail index of wet and dry epoch durations, at 95% confidence level. Tab. 2 gives intervals coming from the tests of $\mathbf{H}_{m_0}\colon\ m \leqslant m_0$, while Tab. 3 gives the intervals based on testing $\mathbf{H}^*_{m_0}\colon\ m \geqslant m_0$, both based on empirical CDF's of sub-sample maxima. The $I_1(0.05)$ intervals are based on the Hill estimator for three different options of choosing the sample fraction k. Intervals containing the corresponding benchmark values of Pareto tail index m_0 are marked with $*$. The unmarked intervals indicate that the values m_0 might underestimate (Tab. 2) or overestimate (Tab. 3) the true index of heavy tailed durations, with probability not greater than 5% for each of these two events.

Tab. 4 tabulates values of the Jurečková-Picek point estimator $M^*(\delta)$ and the Hill estimator $H(k)$ for the Pareto tail index of wet and dry epoch durations, along with two-sided 95% confidence intervals $I_2(0.05)$ based on the Hill estimator. Intervals containing the estimated values of Pareto tail index m_0 by multiscaling are again $*$-marked. The value being "closest" to m_0 predicted by multiscaling is marked in boldface.

Scale of spatial averaging	2 km	4 km	8 km	16 km	30 km	60 km	120 km
Wet durations sample size (l)	2824	2897	2983	3179	3038	1421	25
Tail-index $m_0 = m_0(\lambda)$	1.62	1.67	1.73	1.79	1.85	1.92	1.99
$\mathbf{H}_{m_0}$ P-values, $\delta = 0.05$	0.000	0.000	* **0.666**	* **0.893**	* 0.998	* 1.000	* **0.852**
$\mathbf{H}_{m_0}$ P-values, $\delta = 0.25$	***0.110**	* **0.982**	* 1.000	* 1.000	* 1.000	* 1.000	* **0.865**
$\mathbf{H}_{m_0}$ P-values, $\delta = 0.45$	*1.000	* 1.000	* 1.000	* 1.000	* 1.000	* 1.000	* **0.743**
$\mathbf{H}^*_{m_0}$ P-values, $\delta = 0.05$	1.000	1.000	**0.334**	**0.107**	* 0.002	* 0.000	**0.148**
$\mathbf{H}^*_{m_0}$ P-values, $\delta = 0.25$	**0.890**	**0.018**	0.000	* 0.000	* 0.000	* 0.000	**0.135**
$\mathbf{H}^*_{m_0}$ P-values, $\delta = 0.45$	*0.000	*0.000	* 0.000	* 0.000	* 0.000	* 0.000	**0.257**
Dry durations sample size (l)	2530	2618	2789	3087	3145	1679	45
Tail-index $m_0 = m^*_0(\lambda)$	1.77	1.79	1.83	1.92	2.10	2.62	5.23
$\mathbf{H}_{m_0}$ P-values, $\delta = 0.05$	*0.993	*0.990	***0.958**	*0.990	***0.986**	*1.000	***0.986**
$\mathbf{H}_{m_0}$ P-values, $\delta = 0.25$	*1.000	*1.000	*1.000	*1.000	*1.000	*1.000	*0.992
$\mathbf{H}_{m_0}$ P-values, $\delta = 0.45$	*1.000	*1.000	*1.000	*1.000	*1.000	*1.000	***0.987**
$\mathbf{H}^*_{m_0}$ P-values, $\delta = 0.05$	*0.007	*0.010	**0.042**	*0.010	**0.014**	*0.000	**0.014**
$\mathbf{H}^*_{m_0}$ P-values, $\delta = 0.25$	*0.000	*0.000	*0.000	*0.000	*0.000	*0.000	*0.008
$\mathbf{H}^*_{m_0}$ P-values, $\delta = 0.45$	*0.000	*0.000	*0.000	*0.000	*0.000	*0.000	**0.013**

Table 1. P-values for testing one-sided hypotheses $\mathbf{H}_{m_0}:\ m \leqslant m_0$ (vs. $\mathbf{K}_{m_0}:\ m > m_0$) and $\mathbf{H}^*_{m_0}:\ m \geqslant m_0$ (vs. $\mathbf{K}^*_{m_0}:\ m < m_0$) for the Pareto tail index of wet and dry epoch duration.

Scale of spatial averaging	2 km	4 km	8 km	16 km	30 km	60 km	120 km
Wet durations sample size (l)	2824	2897	2983	3179	3038	1421	25
Tail-index $m_0 = m_0(\lambda)$	1.62	1.67	1.73	1.79	1.85	1.92	1.99
$J_1(0.05)$, $\delta = 0.05$	$(1.750, \infty)$	$(1.753, \infty)$	*$(1.705, \infty)$	*$(1.590, \infty)$	*$(1.691, \infty)$	*$(1.451, \infty)$	*$(0.760, \infty)$
$J_1(0.05)$, $\delta = 0.25$	*$(1.569, \infty)$	*$(1.468, \infty)$	*$(1.517, \infty)$	*$(1.436, \infty)$	*$(1.425, \infty)$	*$(1.255, \infty)$	*$(0.669, \infty)$
$J_1(0.05)$, $\delta = 0.45$	*$(1.406, \infty)$	*$(1.337, \infty)$	*$(1.281, \infty)$	*$(1.280, \infty)$	*$(1.211, \infty)$	*$(1.063, \infty)$	*$(0.587, \infty)$
$I_1(0.05)$, $k =$ plug-in k^*	*$(1.231,\infty)$	*$(1.449, \infty)$	$(1.859, \infty)$	*$(1.599, \infty)$	$(2.563, \infty)$	$(2.441, \infty)$	*$(0.713, \infty)$
$I_1(0.05)$, $k = l/10$	*$(1.538,\infty)$	*$(1.545, \infty)$	*$(1.629, \infty)$	*$(1.652, \infty)$	$(1.964, \infty)$	*$(1.579, \infty)$	*$(0.408, \infty)$
$I_1(0.05)$, $k = 2\sqrt{l}$	$(1.731,\infty)$	*$(1.666, \infty)$	$(1.803, \infty)$	*$(1.776, \infty)$	$(2.305, \infty)$	$(2.224, \infty)$	*$(0.390, \infty)$
Dry durations sample size (l)	2530	2618	2789	3087	3145	1679	45
Tail-index $m_0 = m_0^*(\lambda)$	1.77	1.79	1.83	1.92	2.10	2.62	5.23
$J_1(0.05)$, $\delta = 0.05$	*$(1.660, \infty)$	*$(1.668, \infty)$	*$(1.693, \infty)$	*$(1.770, \infty)$	*$(1.820, \infty)$	*$(1.909, \infty)$	*$(2.185, \infty)$
$J_1(0.05)$, $\delta = 0.25$	*$(1.425, \infty)$	*$(1.432, \infty)$	*$(1.484, \infty)$	*$(1.494, \infty)$	*$(1.594, \infty)$	*$(1.740, \infty)$	*$(2.025, \infty)$
$J_1(0.05)$, $\delta = 0.45$	*$(1.191, \infty)$	*$(1.202, \infty)$	*$(1.263, \infty)$	*$(1.283, \infty)$	*$(1.421, \infty)$	*$(1.482, \infty)$	*$(1.986, \infty)$
$I_1(0.05)$, $k =$ plug-in k^*	$(2.365, \infty)$	$(2.046, \infty)$	$(2.342, \infty)$	$(2.778, \infty)$	$(2.844, \infty)$	*$(2.163, \infty)$	*$(2.719, \infty)$
$I_1(0.05)$, $k = l/10$	*$(1.720, \infty)$	*$(1.601, \infty)$	*$(1.784, \infty)$	*$(1.573, \infty)$	*$(1.791, \infty)$	*$(2.294, \infty)$	*$(2.403, \infty)$
$I_1(0.05)$, $k = 2\sqrt{l}$	$(1.981, \infty)$	$(1.860, \infty)$	$(2.215, \infty)$	$(2.103, \infty)$	$(2.360, \infty)$	*$(2.285, \infty)$	*$(1.225, \infty)$

Table 2. Right-sided 95% confidence intervals for the Pareto tail index of wet and dry epoch durations.

Scale of spatial averaging	2 km	4 km	8 km	16 km	30 km	60 km	120 km
Wet durations sample size (l)	2824	2897	2983	3179	3038	1421	25
Tail-index $m_0 = m_0(\lambda)$	1.62	1.67	1.73	1.79	1.85	1.92	1.99
$J_1(0.05)$, $\delta = 0.05$	*(0, 2.071)	*(0, 1.809)	*(0, 1.899)	*(0, 1.853)	(0, 1.786)	(0, 1.547)	*(0, 4.596)
$J_1(0.05)$, $\delta = 0.25$	*(0, 2.193)	*(0, 2.103)	*(0, 2.064)	*(0, 1.982)	*(0, 2.006)	(0, 1.624)	*(0, 4.918)
$J_1(0.05)$, $\delta = 0.45$	*(0, 2.948)	*(0, 2.639)	*(0, 2.776)	*(0, 2.608)	*(0, 2.558)	*(0, 2.101)	*(0, 75.391)
$I_1(0.05)$, $k =$ plug-in k^*	(0, 1.310)	*(0, 1.834)	*(0, 2.822)	*(0, 10.497)	*(0, 4.023)	*(0, 3.668)	(0, 1.434)
$I_1(0.05)$, $k = l/10$	*(0, 1.873)	*(0, 1.876)	*(0, 1.972)	*(0, 1.988)	*(0, 2.374)	*(0, 2.084)	*(0, 2.681)
$I_1(0.05)$, $k = 2\sqrt{l}$	*(0, 2.390)	*(0, 2.297)	*(0, 2.477)	*(0, 2.430)	*(0, 3.162)	*(0, 3.267)	(0, 1.235)
Dry durations sample size (l)	2530	2618	2789	3087	3145	1679	45
Tail-index $m_0 = m_0^*(\lambda)$	1.77	1.79	1.83	1.92	2.10	2.62	5.23
$J_1(0.05)$, $\delta = 0.05$	(0, 1.735)	(0, 1.752)	(0, 1.828)	(0, 1.874)	(0, 2.041)	(0, 2.187)	(0, 3.283)
$J_1(0.05)$, $\delta = 0.25$	(0, 1.469)	(0, 1.494)	(0, 1.571)	(0, 1.590)	(0, 1.769)	(0, 1.862)	(0, 2.958)
$J_1(0.05)$, $\delta = 0.45$	(0, 1.322)	(0, 1.309)	(0, 1.359)	(0, 1.377)	(0, 1.500)	(0, 1.631)	(0, 2.506)
$I_1(0.05)$, $k =$ plug-in k^*	*(0, 3.512)	*(0, 2.883)	*(0, 3.157)	*(0, 4.961)	*(0, 5.286)	*(0, 3.161)	*(0, 8.614)
$I_1(0.05)$, $k = l/10$	*(0, 2.117)	*(0, 1.964)	*(0, 2.174)	*(0, 1.899)	*(0, 2.158)	*(0, 2.963)	*(0, 8.236)
$I_1(0.05)$, $k = 2\sqrt{l}$	*(0, 2.761)	*(0, 2.584)	*(0, 3.062)	*(0, 2.881)	*(0, 3.229)	*(0, 3.307)	(0, 3.282)

Table 3. Left-sided 95% confidence intervals for the Pareto tail index of wet and dry epoch durations.

Scale of spatial averaging	2 km	4 km	8 km	16 km	30 km	60 km	120 km
Wet durations sample size (l)	2824	2897	2983	3179	3038	1421	25
Tail-index $m_0 = m_0(\lambda)$	1.62	1.67	1.73	1.79	1.85	1.92	1.99
JP-estimate, $\delta = 0.05$	1.78	**1.77**	**1.74**	**1.68**	**1.70**	**1.50**	**0.82**
JP-estimate, $\delta = 0.25$	**1.69**	1.48	1.57	1.53	1.47	1.27	0.76
JP-estimate, $\delta = 0.45$	1.43	1.39	1.32	1.34	1.26	1.09	0.78
Hill-estimate, $k =$ plug-in k^*	1.27	1.62	2.24	2.78	3.13	2.93	**0.95**
Hill-estimate, $k = l/10$	**1.69**	**1.70**	**1.78**	**1.80**	**2.15**	**1.80**	0.71
Hill-estimate, $k = 2\sqrt{l}$	2.01	1.93	2.09	2.05	2.67	2.65	0.59
$I_2(0.05)$, $k =$ plug-in k^*	$(1.22, 1.32)$	*$(1.42, 1.88)$	$(1.80, 2.97)$	*$(1.48, 2.48)$	$(2.48, 4.26)$	$(2.36, 3.85)$	*$(0.68, 1.59)$
$I_2(0.05)$, $k = l/10$	*$(1.51, 1.91)$	*$(1.52, 1.92)$	*$(1.60, 2.01)$	*$(1.62, 2.03)$	$(1.93, 2.42)$	*$(1.54, 2.15)$	*$(0.38, 5.74)$
$I_2(0.05)$, $k = 2\sqrt{l}$	$(1.69, 2.48)$	*$(1.62, 2.38)$	$(1.76, 2.57)$	*$(1.73, 2.52)$	$(2.25, 3.28)$	$(2.16, 3.42)$	$(0.37, 1.56)$
Dry durations sample size (l)	2530	2618	2789	3087	3145	1679	45
Tail-index $m_0 = m_0^*(\lambda)$	1.77	1.79	1.83	1.92	2.10	2.62	5.23
JP-estimate, $\delta = 0.05$	**1.71**	**1.72**	**1.76**	**1.80**	**1.86**	**2.06**	**2.70**
JP-estimate, $\delta = 0.25$	1.44	1.44	1.50	1.55	1.67	1.79	2.34
JP-estimate, $\delta = 0.45$	1.24	1.25	1.32	1.34	1.46	1.57	1.99
Hill-estimate, $k =$ plug-in k^*	2.826	2.394	2.689	3.561	3.698	2.569	**4.134**
Hill-estimate, $k = l/10$	**1.898**	**1.764**	**1.960**	**1.721**	**1.958**	**2.586**	3.720
Hill-estimate, $k = 2\sqrt{l}$	2.307	2.163	2.571	2.431	2.727	2.702	1.785
$I_2(0.05)$, plug-in k^*	$(2.29, 3.68)$	$(1.99, 3.00)$	$(2.28, 3.27)$	$(2.66, 5.36)$	$(2.72, 5.76)$	*$(2.10, 3.31)$	*$(2.55, 10.9)$
$I_2(0.05)$, $k = l/10$	*$(1.69, 2.16)$	*$(1.57, 2.01)$	*$(1.75, 2.22)$	*$(1.55, 1.94)$	*$(1.76, 2.20)$	*$(2.25, 3.05)$	*$(2.25, 10.7)$
$I_2(0.05)$, $k = 2\sqrt{l}$	$(1.93, 2.87)$	$(1.81, 2.68)$	$(2.16, 3.18)$	$(2.05, 2.99)$	$(2.30, 3.35)$	*$(2.22, 3.46)$	$(1.16, 3.91)$

Table 4. Point estimates $M^*(\delta)$ and $H(k)$ for the Pareto tail index of wet and dry epoch durations, along with two-sided 95% confidence intervals based on the Hill estimator.

Overall, we conclude that there is good agreement between the semiparametric inference procedures and the predicted values of the Pareto tail index. This is quite reassuring, because the performance of the semiparametric procedures apparently have not yet been applied to any real data, but exclusively to simulated data.

The data used for this study represent the climate of a tropical oceanic region during the monsoon period, which is very rich in rainfall represented by long wet epochs, during which extremely high rain rates are locally experienced due to predominantly convective clouds. The tail heaviness of *both* wet and dry epochs might be challenged in other types of climate.

References

[1] *R. J. Adler, R. E. Feldman, M. S. Taqqu* (eds.): A Practical Guide to Heavy Tails. Statistical Techniques and Applications. Birkhäuser, Boston, 1998.

[2] *S. Cheng, J. Pan*: Asymptotic expansions of estimators for the tail index with applications. Scand. J. Stat. *25* (1998), 717–728.

[3] *S. Cheng, L. Peng*: Confidence intervals for the tail index. Bernoulli *7* (2001), 751–760.

[4] *L. de Haan, U. Stadtmüller*: Generalized regular variation of second order. J. Austr. Math. Soc., Ser. A *61* (1996), 381–395.

[5] *C. De Michele, H. Pavlopoulos*: The variance-scale plot of intermittent time series and the ranges of scales. Preprint of working paper. 2007.

[6] *P. Doukhan, G. Oppenheim, M. S. Taqqu* (eds.): Theory and Applications of Long-Range Dependence. Birkhäuser, Boston, 2003.

[7] *P. Embrechts, C. Klüppelberg, T. Mikosch*: Modelling Extremal Events for Insurance and Finance. Springer, Berlin, 1997.

[8] *E. Foufoula-Georgiou*: On scaling theories of space-time rainfall: Some recent results and open problems. Stochastic Methods in Hydrology: Rainfall, Landforms and Floods. Advanced Series on Statistical Sciences and Applied Probability, Vol. 7 (O. E. Barndorff-Nielsen, V. K. Gupta, V. Perez-Abreu, E. C. Waymire, eds.). World Scientific, Singapore, 1998, pp. 129–171.

[9] *J. Gritsis*: On the probability distribution of the duration of dry and wet spells in processes of spatially averaged rain rate. Master thesis. Department of Statistics, Athens University of Economics and Business, Athens, 1997.

[10] *V. K. Gupta*: Emergence of statistical scaling in floods on channel networks from complex runoff dynamics. Chaos Solitons Fractals *19* (2004), 357–365.

[11] *V. K. Gupta, E. C. Waymire*: Multiscaling properties of spatial rainfall and river flow distributions. J. Geophys. Res. *95* (1990), 1999–2009.

[12] *V. K. Gupta, E. C. Waymire*: A statistical analysis of mesoscale rainfall as a random cascade. J. Appl. Meteorology *32* (1993), 251–267.

[13] *V. K. Gupta, E. C. Waymire*: Spatial variability and scale invariance in hydrologic regionalization. In: Scale Dependence and Scale Invariance in Hydrology (G. Sposito, ed.). Cambridge University Press, Cambridge, 1998, pp. 88–135.

[14] *E. Haeusler, J. Segers*: Assessing confidence intervals for the tail index by Edgeworth expansions for the Hill estimator. Bernoulli *13* (2007), 175–194.

[15] *D. Heath, S. Resnick, G. Samorodnitsky*: Heavy tails and long range dependence in on/off processes and associated fluid models. Math. Oper. Res. *23* (1998), 145–165.

[16] *J. Jurečková*: Tests of tails based on extreme regression quantiles. Stat. Probab. Lett. *49* (2000), 53–61.

[17] *J. Jurečková*: Statistical tests on tail index of a probability distribution. METRON—International Journal of Statistics *LXI* (2003), 151–190.
[18] *J. Jurečková, H. L. Koul, J. Picek*: Testing the tail index in autoregressive models. Annals of the Institute of Statistical Mathematics. To appear. Published online: 10.1007/s10463-007-0155-z.
[19] *J. Jurečková, J. Picek*: A class of tests on the tail index. Extremes *4* (2001), 165–183.
[20] *J. Jurečková, J. Picek*: Estimates of the tail index based on nonparametric tests. Theory and Applications of Recent Robust Methods. Series: Statistics for Industry and Technology (M. Hubert, G. Pison, A. Struyf, S. Van Aelst, eds.). Birkhäuser, Basel, 2004, pp. 141–152.
[21] *B. Kedem, L. S. Chiu*: Are rain rate processes self similar? Water Resources Research *23* (1987), 1816–1818.
[22] *P. K. Kundu, T. L. Bell*: A stochastic model of space-time variability of mesoscale rainfall: Statistics of spatial averages. Water Resources Research *39* (2003), 1328, doi:10.1029/2002WR001802.
[23] *S. Lovejoy, B. B. Mandelbrot*: Fractal properties of rain and a fractal model. Tellus *37A* (1985), 209–232.
[24] *S. Lovejoy, D. Schertzer*: Generalized scale invariance in the atmosphere and fractal models of rain. Water Resources Research *21* (1985), 1233–1250.
[25] *S. Lovejoy, D. Schertzer*: Multifractals and rain. New Uncertainty Concepts in Hydrology and Water Resources (Z. W. Kunzewicz, ed.). Cambridge University Press, Cambridge, 1995.
[26] *S. B. Lowen, M. C. Teich*: Fractal renewal processes generate $1/f$ noise. Phys. Rev. E *47* (1993), 992–1001.
[27] *S. B. Lowen, M. C. Teich*: Fractal-Based Point Processes. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, 2005.
[28] *B. B. Mandelbrot*: The Fractal Geometry of Nature. W. H. Freeman, San Francisco, 1982.
[29] *D. Marsan, S. Lovejoy, D. Schertzer*: Causal space-time multifractal processes: Predictability and forecasting of rain fields. J. Geophys. Res. *101 (D21)* (1996), 26333–26346.
[30] *T. Over, V. K. Gupta*: Statistical analysis of meso-scale rainfall: Dependence of a random cascade generator on large-scale forcing. J. Appl. Meteor. *33* (1994), 1526–1542.
[31] *T. Over, V. K. Gupta*: A space-time theory of mesoscale rainfall using random cascades. J. Geophys. Res. *101* (1996), 26319–26331.
[32] *H. Pavlopoulos, J. Gritsis*: Wet and dry epoch durations of spatially averaged rain rate, their probability distributions and scaling properties. Environ. Ecol. Stat. *6* (1999), 351–380.
[33] *H. Pavlopoulos, V. K. Gupta*: On the intermittence of rainfields: A space-time approach. Technical Report No. 143. Department of Statistics, Athens University of Economics, Athens, 2001.
[34] *H. Pavlopoulos, V. K. Gupta*: Scale invariance of regional wet and dry durations of rain fields: A diagnostic study. J. Geophys. Res. *108 (D8)* (2003), 8387, doi:10.1029/2002JD002763.
[35] *J. Picek*: Confidence intervals of the tail index. In: Proceedings in Computational Statistics (A. Rizzi, M. Vichi, eds.). Physica, Heidelberg, 2006, pp. 1301–1308.
[36] *J. Picek, J. Jurečková*: A class of tests on the tail index using the modified extreme regression quantiles. ROBUST'2000 (J. Antoch, G. Dohnal, eds.). Union of Czech Mathematicians and Physicists, Prague, 2001, pp. 217–226.
[37] *F. Schmitt, S. Svannistsem, A. Barbosa*: Modeling of rainfall time series using two-state renewal processes and multifractals. J. Geophys. Res. *103 (D18)* (1998), 181–194.

[38] *D. A. Short, P. A. Kucera, B. S. Ferrier, J. C. Gerlach, S. A. Rutledge, O. W. Thiele*: Shipboard radar rainfall patterns within the TOGA/COARE IFA. Bull. Am. Meteor. Soc. *78* (1997), 2817–2836.

[39] *M. S. Taqqu, J. B. Levy*: Using renewal processes to generate long-range dependence and high variability. Dependence in Probability and Statistics (E. Eberlein, M. S. Taqqu, eds.). Birkhäuser, 1986, pp. 73–89.

[40] *Y. Tessier, S. Lovejoy, D. Schertzer*: Universal multifractals: Theory and observations of rain and clouds. J. Appl. Meteor. *32* (1993), 223–250.

[41] *W. Willinger, M. S. Taqqu, R. Sherman, D. V. Wilson*: Self-similarity through high variability: Statistical analysis of Ethernet LAN traffic at the source level (Extended Version). IEEE/ACM Transactions on Networking *5* (1997), 71–86.

Authors' addresses: *H. Pavlopoulos*, Athens University of Economics and Business, Department of Statistics, 76 Patission Str., GR-10434 Athens, Greece, e-mail: `hgp@aueb.gr`; *J. Picek*, Technical University in Liberec, Department of Applied Mathematics, Studentská 2, 461 17 Liberec, Czech Republic, e-mail: `jan.picek@tul.cz`; *J. Jurečková*, Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8, Czech Republic, e-mail: `jurecko@karlin.mff.cuni.cz`.