Advances in Water Resources 79 (2015) 121–126

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

# A metastatistical approach to rainfall extremes

Marco Marani [a,b,*], Massimiliano Ignaccolo [a,c]

[a] Division of Earth & Ocean Sciences, Nicholas School of the Environment, Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC, USA
[b] Department of Civil, Environmental, and Architectural Engineering, University of Padova, Padova, Italy
[c] DIIAR, Politecnico di Milano, Milano, Italy

ARTICLE INFO

*Article history:*
Received 28 August 2013
Received in revised form 1 March 2015
Accepted 3 March 2015
Available online 13 March 2015

*Keywords:*
Extreme events
Generalized Extreme Value distribution
Metastatistics

ABSTRACT

The traditional statistical theory of extreme events assumes an asymptotic regime in which the number of events per year is large enough for a limiting Generalized Extreme Value distribution to apply. This has been shown not to be applicable to many practical cases. We introduce here a Metastatistical Extreme Value (MEV) approach which is defined in terms of the distribution of the statistical parameters describing "ordinary" daily rainfall occurrence and intensity. The method does not require an asymptotic assumption, and naturally accounts for the influence of the bulk of the distribution of ordinary events on the distribution of annual maximum daily rainfall. Building on existing observations showing the distribution of daily rainfall to be Weibull right-tail equivalent, the MEV approach is then specialized to yield a compact and easily applicable formulation. We apply this formulation to Monte Carlo experiments based on Weibull statistics derived from the 3-century long rainfall time series observed in Padova (Italy). We find an excellent agreement between MEV estimates and the 'observed' frequency of occurrence of extreme events in the synthetic time series generated. GEV and Gumbel estimates, on the contrary, exhibit systematic errors. Tests with different rates of occurrence of rainfall events show slight improvements of the GEV and Gumbel estimation bias when the number of events/year is increased. However, a constant bias in GEV and Gumbel estimates is seen for (synthetic) climates where the number of events and the distribution of intensities is varied stochastically. The estimation root mean square error is also larger for the GEV and Gumbel distributions than for the MEV approach. Hence, GEV and Gumbel quantile estimates are more likely to be further away from the actual value than MEV estimates. Finally, the application of the new MEV approach to subsets of the long Padova time series identifies marked variabilities in rainfall extremes at the centennial time scale.



## 1. Introduction

The definition and estimation of extreme rainfall events is of central importance in the analysis of past and projected rainfall regimes, as well as in the design of any water resources management and flood control infrastructure. For a given event duration of interest (here we will focus on the important case of daily duration), extreme value analysis usually studies the distribution of yearly maxima, *y*, either directly or by considering the distribution of rainfall values over a high threshold [1]. Under the assumptions that (i) rainfall intensity may be assumed independent and identically distributed (i.i.d.) and (ii) the number of events per year tends to infinity, the classical Extreme Value Theory (EVT) identifies a Generalized Extreme Value (GEV) distribution of yearly maxima [2–6], which has been and still is widely applied [[7–12], e.g.]. It is important to emphasize here that the GEV is not an exact distribution of yearly maxima, and that the actual extreme value distribution may converge to a GEV distribution only as the number of events/year is "large enough", a potentially problematic concept as the number of events/year (wet days in the present case) is necessarily limited. However, little work has addressed the conditions under which the actual distribution may be considered to be close to the limiting GEV form [13,14, e.g.] or how the possible variability of the rainfall depth distribution (i.e. a violation of the i.i.d. hypothesis e.g. due to seasonality), can affect the resulting extreme value distribution [15]. These analyses show that indeed the actual extreme value distribution of rainfall may in practice be quite far from the asymptotic GEV form.

We propose here a non-asymptotic approach to the definition and evaluation of an extreme value distribution based on a metastatistical approach (also referred to as superstatistics [16], compound distributions [17], or doubly stochastic processes [18] in

* Corresponding author at: Division of Earth & Ocean Sciences, Nicholas School of the Environment, Department of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC, USA.

http://dx.doi.org/10.1016/j.advwatres.2015.03.001
0309-1708/

different contexts). Our approach avoids the need of assuming an infinite collection of events, i.e. it avoids the asymptotic assumption, and allows for interannual variability to be accounted for.

The manuscript is organized as follows. In Section 2 we describe the data used in our analyses. In Section 3 we briefly summarize the classical extreme event theory and introduce the new Metastatistical Extreme Value formulation (MEV). A Results section compares GEV and MEV performances, and a Discussion and Conclusion section closes the paper.

## 2. Data

We analyze extremes in the daily rainfall time series observed in Padova (Italy) over a span of almost three centuries, as well as on synthetic data derived from its statistical properties. The Padova dataset, comprised of 275 complete years of daily observations is described in detail elsewhere [19,20], and provides an exceptionally long record, particularly suitable to test estimates of high return period extremes.

## 3. Methods

We first briefly summarize the EVT, as typically used in hydrology, and then introduce a metastatical approach to the definition of extreme value distributions.

### 3.1. Extreme value theorem

We use the random variable $X > 0$ to indicate daily rainfall depth, $f(x)$ being its probability density function, $F(x) = P(X \leqslant x)$ its cumulative distribution function, and $\Psi(x) = 1 - F(x)$ the exceedance probability. Notice that having considered $X > 0$, no probability atom at $X = 0$ need be considered to represent the finite probability of zero rainfall. The maximum of $n$ realizations of the stochastic variable $X$, $Y_n = \max(x_1, x_2, \ldots, x_n)$, is also a random variable, often termed an n-maximum or a maximum with *cardinality* $n$ of the "parent" stochastic variable. In hydrologic practice $n$ will be the number of wet days in a given year, itself a discrete random variable. If the events generating the $n$ realizations of $X$ are independent, the cumulative distribution, $H_n(y)$, of $Y_n$ may be expressed as

$$H_n(y) = [F(y)]^n \tag{1}$$

Upon definition of a renormalized variable $S_n = (Y_n - b_n)/a_n$ ($a_n > 0$ and $b_n$ being constants), the EVT establishes that [2–4]

$$\lim_{n\to\infty} P(S_n < s) = \lim_{n\to\infty} H_n(s) = \lim_{n\to\infty} [F(a_n \cdot s + b_n)]^n = H(s) \tag{2}$$

The limiting distribution $H(s)$ in Eq. (2), depending on the tail behaviour of $\Psi(y)$, can only be one of three distributions: (i) the Gumbel distribution (Extreme Value 1 – EV1, or double exponential), when the tail of $\Psi(y)$ decreases faster than a power law; (ii) the Frechét distribution (EV2), when the tail of $\Psi(y)$ behaves as a power law for large values of $x$; and (iii) the Weibull distribution (EV3), when $x$ has a finite upper limit [2–4].

In terms of the non renormalized variable $y$, the three asymptotic types, EV1–EV3, can be thought of as special cases of a single Generalized Extreme Value distribution [6]:

$$H_{\text{GEV}}(y) = \exp\left\{-\left(1 + k\frac{y-\mu}{\sigma}\right)_+^{-1/k}\right\} \tag{3}$$

where $(\cdot)+ = \max(\cdot, 0)$, $\mu$ is the location parameter, $\sigma > 0$ is the scale parameter, and $k$ is a shape parameter. The limit $k = 0$ corresponds to the EV1 distribution, $k > 0$ to the EV2 distribution, and $k < 0$ to the EV3 distribution.

As noted, the distribution describing the $n$-sample maximum will strictly be a GEV distribution, independent of the specific value $n$, only for 'large enough' values of the number of wet days. When the number of wet days $n$ is not large enough for the asymptotic regime of the EVT to apply (e.g. this has been shown to be the case in practice for Weibull variates [13,21]) one must use Eq. (1). However, a useful approximation of $H_n(y)$ that does not require $n \to \infty$ can be obtained by considering $U_n$, the expected largest value of the variable $X$ in $n$ realizations. Because $U_n$ is on average exceeded once every $n$ realizations of $X$ [22,23]:

$$\Psi(U_n) = \frac{1}{n} \tag{4}$$

(note that a Weibull plotting position estimate, $\Psi(U_n) = 1/(n+1)$, could also be used with no consequence of substance). Using this result we can rewrite the cumulative probability for the $n$-sample maximum $Y_n$ as

$$H_n(y) = [F(y)]^n = [1 - \Psi(y)]^n = \left[1 - \frac{\Psi(y)}{n\Psi(U_n)}\right]^n \tag{5}$$

For $y > U_n$ (i.e. for an extreme value larger than the average maximum value in the observations) the term $\Psi(y)/\Psi(U_n) < 1$. Therefore, for large values of $y$, i.e. for extremes, we can use the Cauchy approximation: $(1-z)^n \cong 1 - n \cdot z \cong \exp(-n \cdot z)$, valid for $z \ll 1$. Hence, Eq. (5) can be approximated as:

$$H_n(y) = \exp\left(-\frac{\Psi(y)}{\Psi(U_n)}\right) \tag{6}$$

Eq. (6) is sometimes referred to as the "penultimate" approximation [24,22], the "ultimate" approximation being Eq. (2), only valid when $n$ is very large. The error associated with the penultimate approximation can be quantified through the relative error $\varepsilon(y) = \{[\exp(-\Psi(y)/\Psi(U_n)] - [1 - \Psi(y)/(n\Psi(U_n))]^n\}/[1 - \Psi(y)/\,(n\Psi(U_n))]^n$. For $y = U_n$ [24]: $\varepsilon(U_n) = (\exp(-1) - [1 - 1/n]^n)/[1 - 1/n]^n$. For example, for $n = 50$ the relative error is $\varepsilon(U_{50}) = 0.01$. Note that for values $y > U_n$, of greatest applicative interest, the relative error is smaller than $\varepsilon(U_n)$, as $\Psi(y) < \Psi(U_n)$. The penultimate approximation has been used in the evaluation of extreme values in some geophysical contexts, such as in modelling wind power [24,23] or of drought severity [25], but very rarely has it been applied to rainfall extremes [26,15].

### 3.2. The case of Weibull variates

Daily rainfall as been shown to be accurately modelled as a Weibull variate [27]. Hence we consider here the important case of $\Psi(x) = \exp(-(x/C)^w)$. Under these assumptions, the yearly maximum daily rainfall depth, i.e. the maximum depth over the $n$ wet days occurred in a generic year, is distributed as:

$$H_n(y) = \left[1 - \exp\left(-\left(\frac{y}{C}\right)^w\right)\right]^n \tag{7}$$

Hence, the penultimate approximation takes the following forms:

$$H_n(y) \cong \left[1 - n \cdot \exp\left(-\frac{y^w}{C^w}\right)\right] \cong \exp\left[-\exp\left(-\frac{y^w}{C^w} + \ln n\right)\right] \tag{8}$$

Note that these expressions (and later results in this Section) are valid also for distributions that are only right-tail equivalent to a Weibull distribution [24,27] (two distributions $F_1$ and $F_2$ are right-tail equivalent if $(1 - F_1(x))/(1 - F_2(x)) \to 1$ when $x \to +\infty$).

### 3.3. A metastatistical approach

The cumulative probability, $H_n(y)$, of the n-maximum $Y_n$ depends on the number of wet days, $n$, and on the parameters,