Journal of Hydrology 262 (2002) 245–258
Journal of Hydrology
www.elsevier.com/locate/jhydrol

On the derived flood frequency distribution: analytical formulation and the influence of antecedent soil moisture condition

C. De Michele[a,*], G. Salvadori[b,1]

[a]Dipartimento di Ingegneria Idraulica, Ambientale e del Rilevamento, Sezione Idraulica, Politecnico di Milano, 32, Piazza Leonardo da Vinci, I-20133 Milano, Italy
[b]Dipartimento di Matematica "Ennio De Giorgi", Università di Lecce, Provinciale Lecce-Arnesano, P.O. Box 193, I-73100 Lecce, Italy

Received 4 September 2001; revised 16 January 2002; accepted 25 February 2002

Abstract

In this paper we present an analytical formulation of the *derived distribution* of *peak flood* and *maximum annual peak flood*, starting from a simplified description of rainfall and surface runoff processes, and we show how such a distribution is useful in practical applications. The assumptions on rainfall dynamics include the hypotheses that the maximum storm depth has a *Generalized Pareto* distribution, and that the temporal variability of rainfall depth in a storm can be described via power–law relationships. The *SCS-CN* model is used to describe the soil response, and a lumped model is adopted to transform the rainfall excess into peak flood; in particular, we analyse the influence of *antecedent soil moisture condition* on the flood frequency distribution. We then calculate the analytical expressions of the derived distributions of peak flood and maximum annual peak flood. Finally, practical case studies are presented and discussed. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Antecedent soil moisture condition; Derived distribution; Extreme event; Generalized Pareto; Maximum annual peak flood; Peak flood

1. Introduction

The dynamics of flood frequency is historically analysed through the determination of the derived flood frequency distribution via a simplified representation of rainfall and runoff processes. This approach follows the route proposed by Eagleson (1972), who first developed the idea of deriving flood statistics from a simplified schematization of storm and basin characterization. Indeed, the mechanism of *derived distribution* is well established in probability theory, where a variable $Y = Y(\mathbf{X})$ is functionally related to a *random vector* $\mathbf{X}$, whose components are random variables with joint density function $f_{\mathbf{X}}$ and joint distribution function $F_{\mathbf{X}}$. Due to the randomness of $\mathbf{X}$, also Y is expected to be a random variable, with distribution function F_Y given by, for $y \in \mathbb{R}$:

$$F_Y(y) = P\{Y(\mathbf{X}) \le y\} = \int_{\{x:Y(x)\le y\}} f_X(x)\mathrm{d}x$$

For instance, Y may represent the peak flow rate, and the components of $\mathbf{X}$ may include, e.g. soil and vegetation characteristics parametrized via both deterministic and random variables.

This route can be followed by either analytical derivation of the distribution function (Eagleson, 1972, 1978; Wood, 1976; Klemeš, 1978; Cordova and

* Corresponding author. Tel.: +39-02-2399-6233; fax: +39-02-2399-6207.
E-mail addresses: carlo.demichele@polimi.it (C. De Michele), gianfausto.salvadori@unile.it (G. Salvadori).
[1] Tel.: +39-0832-320-584; fax: +39-0832-320-410.

PII: S0022-1694(02)00025-2

Rodriguez-Iturbe, 1983; Diaz-Granados et al., 1984; Wood and Hebson, 1986; Raines and Valdes, 1993, among others), or via the statistical moments using the *second-order second-moment approximation* (SOSM) of extreme flood flows (Robinson and Sivapalan, 1997a). The first approach provides complex analytical formulations which may require numerical methods for the computations. The second one gives approximate estimates of the statistical moments of maximum annual flood, useful for calculating the parameters of the distributions of interest; its practical applicability is paid in terms of the requirement of the existence of these moments. Alternatively, *Monte Carlo* methods can be run to estimate either flood quantiles or the moments (Sivapalan et al., 1990; Robinson and Sivapalan, 1997b). Accordingly, the derived distribution approach is an attempt to provide a physically based description of flood processes with an acceptable computational effort for practical applications: often, a simplified description of the physical processes is a necessity in order to obtain mathematically tractable models. Indeed, the derived flood frequency distribution approach provides an attractive and alternative solution for ungauged catchments; in particular, given the explicit use of physically meaningful basin parameters, the impact of various land use changes on flood magnitude and frequency can be directly investigated.

The targets of this paper are: (i) derive the analytical expression of the distributions of peak flood and maximum annual peak flood using simple, but physical, assumptions on rainfall dynamics and catchment response, in order to provide a formulation useful in practical applications; (ii) analyse the influence of *antecedent soil moisture condition* (AMC) on flood frequency distribution, and calculate the derived distribution of peak flood conditioned by the AMC statistical law.

In the next sections the dynamics of rainfall is described. It is represented by maximum storm depth for a fixed duration, and the corresponding random variable is given a *Generalized Pareto* (GP) distribution; we also show how proper power–law relationships for the *position* and the *scale* parameters of the GP law may yield a scaling behaviour. Introducing the *SCS-CN* method as a model for direct *rainfall excess*, we derive the distribution of the effective rainfall. Then, using the distribution of the rainfall excess, and a lumped model to transform the rainfall excess into peak flood, we obtain the distribution of peak flood, and consequently the distribution of maximum annual peak flood. The derived (asymptotic) behaviour shows that a *Generalized Extreme Values* (GEV) law may account for the dynamics of the process for large values of the variable. Ad hoc techniques for estimating specific parameters are also shown in Appendix A. Following Wood (1976), and using the SCS-CN method to transform the rainfall in excess rainfall, we incorporate the probability distribution of AMC into the flood frequency distribution, and we illustrate the importance of AMC knowledge for the determination of flood quantiles. Finally, a model application is given.

2. Derived flood frequency distribution

2.1. The rainfall model

The rainfall storm is represented here as the *maximum rainfall depth* observed in a given time period within the considered storm. This is different from the canonical *Poisson Rectangular Pulses* (PRP) model, introduced by Eagleson (1972) and widely used in literature, where the storm event is described by an *average intensity* and an *average duration*, both exponentially distributed; however, such a model is unable to describe the observed scaling properties of temporal rainfall for extreme events (Rodriguez-Iturbe, 1986; Burlando and Rosso, 1993). On the one hand, the present model simplifies the representation of the phenomenon and, on the other hand, it captures the fundamental components of the precipitation useful to represent the flood frequency distribution.

Let Δt be a given *time duration* (e.g. $\Delta t = 1$ h), and denoted by $P_{\Delta t}$ the maximum rainfall depth observed in a generic period of length Δt within the considered storm. We assume that $P_{\Delta t}$ has the following GP distribution, for $x > b_{\Delta t}$:

$$F_{P_{\Delta t}}(x) = 1 - \left(1 - \frac{k_{\Delta t}}{c_{\Delta t}}(x - b_{\Delta t})\right)^{1/k_{\Delta t}} \quad (1)$$

where $b_{\Delta t} \geq 0$ is a position parameter, $c_{\Delta t} > 0$ a scale parameter, and $k_{\Delta t} < 0$ is a *shape* parameter. Note that $P_{\Delta t}$ is *non-negative* and *upper-unbounded*.

The use of a GP distribution with negative shape parameter (instead of the classical choice of an exponential law, widely adopted and recommended in literature) may provide both a valuable tool for modelling extreme events, due to the presence of an algebraic upper-tail, and a mathematical framework suitable for describing possible scaling features of the phenomenon analysed. Indeed, the GP law considered in the present paper has been widely used in extreme values analysis, and especially in hydrology (Davison, 1984; van Montfort and Witter, 1985, 1986; Hosking and Wallis, 1987; Castillo, 1994).

Let $\Delta t' = r\Delta t$ denotes a generic time duration, where $r > 0$ represents the *scale ratio*, and let $P_{\Delta t'}$ be the rainfall depth associated with the time scale $\Delta t'$. Then, if the following relations hold:

$$\begin{cases} b_{\Delta t'} = b_{\Delta t} r^{\delta} \\ c_{\Delta t'} = c_{\Delta t} r^{\delta} \\ k_{\Delta t'} = k_{\Delta t} = k \end{cases} \tag{2}$$

where $\delta \in \mathbb{R}$ is a *scaling exponent*, it is easy to show (Salvadori and De Michele, 2001) that P is *strict sense simple scaling*, and consequently also *wide sense simple scaling*, in the temporal domain (Gupta and Waymire, 1990). It is important to note that, once the parameters δ and $\{b_{\Delta t}, c_{\Delta t}, k_{\Delta t}\}$ are known for a given duration Δt, in principle it would be possible to calculate the distribution of P for any given duration $\Delta t'$ (within the limits of physical validity of the scaling régime). Clearly, such a scale invariance property may provide a characterization of the temporal variability of rainfall within a storm; thus, a scaling GP law might be used to model the stochastic behaviour of rainfall rate in storms of any duration.

Fixing a *reference time period* $\Delta T > \Delta t$ (in the present case, $\Delta T = 1$ year), and assuming that the sequence of storms has a *Poissonian chronology*, the random number N_P of storms in ΔT is a Poisson r.v., with distribution given by:

$$\mathrm{P}\{N_P = n\} = \mathrm{e}^{-\Lambda_P} \frac{\Lambda_P^n}{n!}, \qquad n \in \mathbb{N} \tag{3}$$

where $\Lambda_P > 0$ represents the *annual storm rate*.

Let $P^*_{\Delta t} = \max\{P_{\Delta t}\}$ in ΔT, i.e. the *maximum annual* (*within-storm*) *rainfall depth* observed in a generic period of length Δt. Conditioning upon N_P, the distribution of $P^*_{\Delta t}$ is given by:

$$F^*_{P_{\Delta t}}(x) = \mathrm{e}^{-\Lambda_P(1 - F_{P_{\Delta t}}(x))} \tag{4}$$

for all suitable values of x. If $P_{\Delta t}$ is distributed as in Eq. (1), then:

$$F_{P^*_{\Delta t}}(x) = \exp\left(-\left(1 - \frac{\kappa_{\Delta t}}{\alpha_{\Delta t}}(x - \xi_{\Delta t})\right)^{1/\kappa_{\Delta t}}\right) \tag{5a}$$

for $x > \xi_{\Delta t} + \alpha_{\Delta t}/\kappa_{\Delta t}$, where:

$$\begin{cases} \xi_{\Delta t} = b_{\Delta t} + \dfrac{c_{\Delta t}}{k_{\Delta t}}\left(1 - \Lambda_P^{-k_{\Delta t}}\right) \\ \alpha_{\Delta t} = c_{\Delta t}\Lambda_P^{-k_{\Delta t}} \\ \kappa_{\Delta t} = k_{\Delta t} \end{cases} \tag{5b}$$

Therefore, $P^*_{\Delta t}$ is a GEV *upper-unbounded* r.v., where $\xi_{\Delta t}$ is a position parameter, $\alpha_{\Delta t} > 0$ is a scale parameter, and $\kappa_{\Delta t} < 0$ is a shape parameter. In Salvadori and De Michele (2001) it is shown how to derive, for the GEV process of the maxima modelled by Eqs. (5a) and (5b), the same scaling features found for the parent GP process modelled by Eq. (1), and the relations linking the parameters of the GP–GEV laws involved. In particular, if the following relations hold (where the notation is the same as in Eq. (2)):

$$\begin{cases} \xi_{\Delta t'} = \xi_{\Delta t} r^{\delta} \\ \alpha_{\Delta t'} = \alpha_{\Delta t} r^{\delta} \\ \kappa_{\Delta t'} = \kappa_{\Delta t} = \kappa \end{cases} \tag{6}$$

then also P^* is *strict/wide sense simple scaling* in the temporal domain.

Note that we considered only negative values of the shape parameters because in our regions (North-Western Italy) the statistics of rainfall and flood is characterized by steep frequency curves (De Michele and Rosso, 2000): they estimated the GEV parameters of maximum annual rainfall depth for the duration of 1, 3, 6, 12, and 24 h collected by rain gauge stations located in Thyrrenian Liguria. For 58 rainfall stations with a sample dimension larger than 30 years they found that 86% of the stations presents a negative shape parameter, while for the remaining ones (14%) it is not rejectable the hypothesis of a shape parameter equal to zero. Furthermore, investigating the regionalization of maximum annual peak flood in North-Western Italy, they estimated the shape parameter of the normalized GEV distribution of the five homogeneous regions, and found values of the shape parameters in the range (−0.32, −0.09). In addition, Meigh et al. (1997) analysed the variability of the regional flood frequency distribution with climate, using a GEV model. They considered 22 regions throughout the world and showed how in tropical regions the frequency curve is not very steep (and the shape parameter may also assume positive values), whereas in subtropical regions (like the ones considered in our manuscript) the variability of peak flood is high, and rare floods can be extremely large. These results support the use of a negative shape parameter in our region.

Inverting Eqs. (5a) and (5b) we may obtain the *depth duration frequency* curves (DDF). As noted before, the rainfall model considered is different from the canonical PRP model, but it is consistent with the DDF curves. We observe that, assuming $b_{\Delta t} = 0$ and using Eq. (5b), in principle it would be possible to calculate the parameters Λ_P and $\{c_{\Delta t}, k_{\Delta t}\}$ simply through the estimate of the parameters $\{\xi_{\Delta t}, \alpha_{\Delta t}, \kappa_{\Delta t}\}$ (for an alternative approach, see Appendix A). Therefore, using the series of maximum annual rainfall data collected at different durations Δts (generally 1, 3, 6, 12 and 24 h), it would be possible to calculate the distribution of $P^*_{\Delta t}$, and then make inference on the law of $P_{\Delta t}$.

For simplicity, in the present model we consider the rainfall duration as constant, and equal to the time of equilibrium of the basin. Such an hypothesis, physically acceptable for small basins, will allow to calculate the analytical expression of the distribution of peak flood and maximum annual peak flood (note that the same assumption is also adopted by Temez (1991)). Accordingly, we shall abandon the Δt subscript, thus simplifying the mathematical notation.

2.2. The distribution of rainfall excess

We use here the SCS-CN method adopted by USDA Soil Conservation Service (1986) to transform the rainfall depth in rainfall excess. Accordingly, the total volume of rainfall excess P_e can be expressed in terms of the rainfall depth P as:

$$P_e = P_e(P) = \begin{cases} \dfrac{(P - I_A)^2}{P + S - I_A}, & P > I_A \\ 0, & P \le I_A \end{cases} \tag{7}$$

where I_A is the rainfall lost as *initial abstraction*, and $S \ge 0$ is the *maximum soil potential retention*. Here S is expressed in mm, and is given by $S = 254((100/\mathrm{CN}) - 1)$, where CN is the *curve number*; also note that I_A is generally estimated as $I_A \approx 0.2S$. The parameter CN depends upon the soil type, the land use and the *antecedent moisture conditions*. The Soil Conservation Service (1986) provides tables to estimate the CN knowing the soil type, the land use, and the AMC. In these tables, the CN values are referred to an average antecedent moisture condition of the basin. The SCS-CN method considers three AMC classes (I, II, and III) dependent on the total 5-days antecedent rainfall and the season category (*dormant* or *growing*). Condition I involves a dry catchment, and it is characterized by a total 5-days antecedent rainfall less than 13 mm in the dormant season, and less than 36 mm in the growing season. Condition II is characterized by a total 5-days antecedent rainfall ranging from 13 to 28 mm in the dormant season, and from 36 to 53 mm in the growing season. Condition III occurs when the soil is almost saturated, with a total 5-days antecedent rainfall greater than 28 mm in the dormant season, and greater than 53 mm in the growing season. In Ponce (1989)—and references therein—can be found formulas to calculate the CN for AMC I and AMC III from the values of CN corresponding to AMC II. As we shall see later, the parameters tuning the dynamics of the basin may depend upon the AMC, and thus the probability distributions involved may be conditioned by the actual AMC condition.

Note that if $P \le I_A$ then $P_e = 0$. Therefore, since P has a GP law, we obtain:

$$\mathrm{P}\{P_e = 0\} = P\{P \le I_A\} = 1 - \left(1 - \frac{k}{c}(I_A - b)\right)^{1/k} \tag{8}$$

Thus, the distribution of P_e is characterized by an *atom* (mass point) at zero. Using Eq. (7) we may derive the conditional distribution of P_e given that $P > I_A$:

$$\mathrm{P}\{P_e \le x | P > I_A\} = P\{P \le I_A + \frac{x + \sqrt{x^2 + 4xS}}{2} | P > I_A\} = 1 - \frac{\left(1 - \frac{k}{c}\left(I_A + \frac{x + \sqrt{x^2 + 4xS}}{2} - b\right)\right)^{1/k}}{\left(1 - \frac{k}{c}(I_A - b)\right)^{1/k}} \tag{9}$$

for $x > 0$. Then, from Eqs. (8) and (9) we may obtain the derived distribution of rainfall excess:

$$F_{P_e}(x) = 1 - \left(1 - \frac{k}{c}\left(I_A + \frac{x + \sqrt{x^2 + 4xS}}{2} - b\right)\right)^{1/k} \tag{10}$$

for $x \geq 0$, which is *right-continuous* at zero. Note that $\sqrt{x^2 + 4xS} \approx x$ for x large enough; hence, for $x \gg 1$:

$$F_{P_e}(x) \approx 1 - \left(1 - \frac{k_e}{c_e}(x - b_e)\right)^{1/k_e} \tag{11}$$

where $b_e = b - I_A$ is a position parameter, $c_e = c$ is a scale parameter, and $k_e = k$ is a shape parameter. Thus, the limit distribution of P_e is again a GP law with the given parameters. Indeed, such a result could also be derived recalling that the GP distribution is *stable* with respect to *excess-over-threshold* operations (Castillo and Hadi, 1997), and noting that Eq. (7) is asymptotically linear for $P \gg 1$.

Note that the distribution of rainfall excess is calculated assuming the AMC of the basin as constant. This may not be consistent with the actual situation in many catchments, as pointed out by Wood (1976) and later by Sivapalan et al. (1990): in fact, the former author found that initial moisture condition of the catchment could have a substantial effect on the estimation of the flood frequency distribution. Thus, later we shall relax this hypothesis, and calculate the flood distribution considering also the AMC as a random variable.

2.3. The flood frequency distribution (AMC constant)

Let Q denote the peak flood produced by a precipitation P according to the transformation:

$$Q = Q(P) = \begin{cases} \phi \dfrac{(P - I_A)^2}{P + S - I_A}, & P > I_A \\ 0, & P \leq I_A \end{cases} \tag{12}$$

with $\phi = A/t_c$, where A is the *area* of the basin and t_c is the *time of concentration* of the basin. The transform function is *non-linear* in P (but *linear* in P_e, since $Q = \phi P_e$), and invertible for $P > I_A$. Thus, by using Eq. (10), we may calculate the distribution of Q as:

$$F_Q(q) = 1 - \left(1 - \frac{k}{c}\left(I_A + \frac{q + \sqrt{q^2 + 4q\phi S}}{2\phi} - b\right)\right)^{1/k} \tag{13}$$

for $q \geq 0$. Note that $\sqrt{q^2 + 4q\phi S} \approx q$ for q large enough; hence, for $q \gg 1$:

$$F_Q(q) \approx 1 - \left(1 - \frac{k_Q}{c_Q}(q - b_Q)\right)^{1/k_Q} \tag{14}$$

where $b_Q = \phi(b - I_A)$ is a position parameter, $c_Q = \phi c$ a scale parameter, and $k_Q = k$ is a shape parameter. Thus, the limit distribution of the peak flood Q is again a GP law with the given parameters.

Clearly, only a fraction of the N_P annual storm events produces an effective rainfall $P_e > 0$, and hence a peak flood $Q > 0$. This corresponds to a random *Bernoulli selection* over the Poissonian chronology of the actual storms; thus, the random sequence of flood events has again a Poissonian chronology, with annual rate parameter Λ_Q given by:

$$\Lambda_Q = \Lambda_{P_e} = \Lambda_P \mathrm{P}\{P > I_A\} = \Lambda_P\left(1 - \frac{k}{c}(I_A - b)\right)^{1/k} \tag{15}$$

which, in turn, specifies the distribution of the random number N_Q of annual peak floods. Clearly, if $I_A \leq b$ (i.e. if the minimum rainfall b is already larger than the initial abstraction I_A), then $\Lambda_Q = \Lambda_{P_e} = \Lambda_P$, since the storm automatically generates a flood event. Also, note that Eq. (15) if properly modified, may provide the average number of peaks over a given threshold.

Note that the initial abstraction I_A is a function of both the *soil properties*, the maximum soil potential retention, and the AMC, through the simple empirical relation $I_A = 0.2S_{\mathrm{AMC}}$; thus the soil characteristics *do* influence the expected number of annual flood events. Indeed, rewriting Eq. (15) as:

$$\rho_{Q,P} = \frac{\Lambda_Q}{\Lambda_P} = \left(1 - \frac{k}{c}(0.2S - b)\right)^{1/k} \tag{16}$$

it is easy to study the behaviour of the ratio $\rho_{Q,P}$ as a function of the *absorptive characteristics* of the basin. As an illustration, and using the values of the

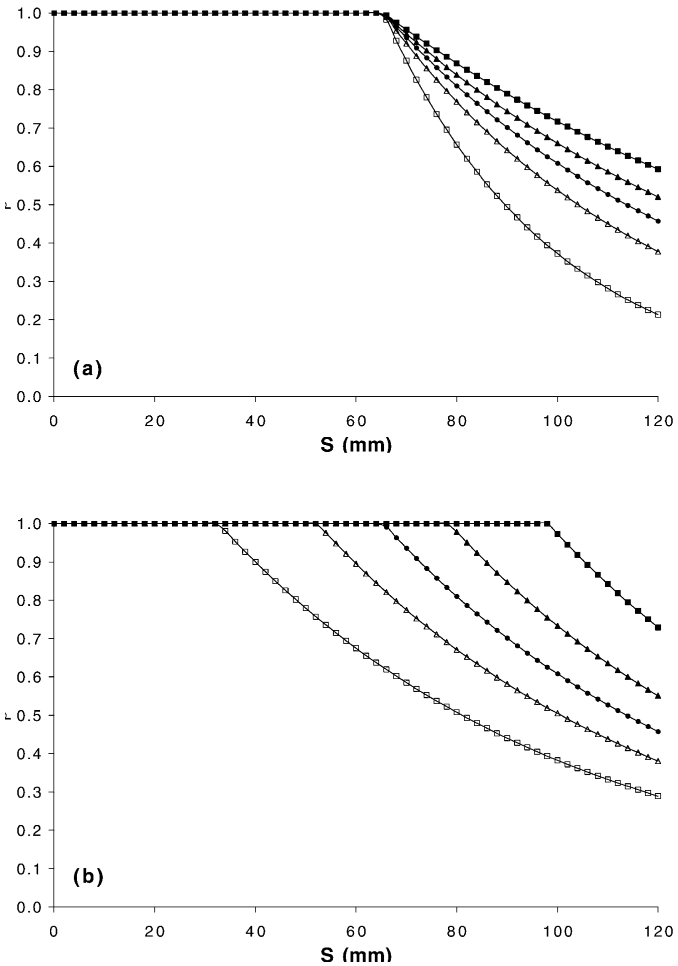

Fig. 1. Behaviour of the ratio $\rho_{Q,P}$as a function of the *absorptive* characteristics of the basin S (see text). (a) Here $b = 13.08$ mm while $c_1 = 6.90$ mm (empty squares), $c_2 = 11.04$ mm (empty triangles), $c_3 = 13.80$ mm (circles), $c_4 = 16.56$ mm (full triangles), and $c_5 = 20.70$ mm (full squares). (b) Here $c = 13.80$ mm while $b_1 = 6.54$ mm (empty squares), $b_2 = 10.46$ mm (empty triangles), $b_3 = 13.08$ mm (circles), $b_4 = 15.70$ mm (full triangles), and $b_5 = 19.62$ mm (full squares).

parameters calculated in Section 4 investigating the Bisagno basin (see Table 3), in Fig. 1 we plot $\rho_{Q,P}$ as a function of S. More particularly, for the sake of completeness, we proceed as follows: in Fig. 1a we fix $b = 13.08$ mm and select five different values $c_1, \ldots, c_5$ of c equal to, respectively, 50, 80, 100, 120, and 150% of $c = 13.80$ mm; in Fig. 1b we fix $c = 13.80$ mm and select five different values $b_1, \ldots, b_5$ of b equal to, respectively, 50, 80, 100, 120, and 150% of $b = 13.08$ mm. In both cases, it is evident how the ratio $\rho_{Q,P}$ becomes smaller than one, depending upon the values of the rainfall parameters and the absorptive features of the basin. The influence of the remaining parameter k is negligible, and it is not shown. It is interesting to note that the results shown are similar to (and supported by) those obtained by Iacobellis and Fiorentino (2000).

Furthermore, conditioning upon N_Q, the distribution of the maximum annual peak flood Q is given by:

$$F_{Q^*}(q) = \mathrm{e}^{-\Lambda_Q(1-F_Q(q))} \tag{17}$$

for all suitable values of q. If Q is asymptotically distributed as in Eq. (14), then

$$F_{Q^*}(q) \approx \exp\left(-\left(1 - \frac{\kappa_Q}{\alpha_Q}(q - \xi_Q)\right)^{1/\kappa_Q}\right) \quad (18a)$$

for $q \gg 1$, where:

$$\begin{cases} \xi_Q = b_Q + \dfrac{c_Q}{k_Q}\left(1 - \Lambda_Q^{-k_Q}\right) \\ \alpha_Q = c_Q \Lambda_Q^{-k_Q} \\ \kappa_Q = k_Q \end{cases} \quad (18b)$$

Therefore, Q^* is (asymptotically) a GEV *upper-unbounded* r.v., where ξ_Q is a position parameter, $\alpha_Q > 0$ is a scale parameter, and $\kappa_Q < 0$ is a shape parameter.

Thus, the shape parameter of the flood distribution is the same as that of the rainfall distribution; in other terms, asymptotically, the curve of maximum annual flood quantiles becomes parallel to the curve of maximum annual rainfall quantiles. A similar result is given by the 'Gradex method' (Guillot and Duband, 1967): there, using a Gumbel distribution for the maximum annual rainfall depth, and assuming that during the extreme flood event the basin saturation is approached, the probability law of the specific flood volume is calculated; such a derived distribution is again a Gumbel law, with the location parameter depending on the initial conditions of the basin, and the scale parameter (*gradex*) equal to that of the rainfall distribution.

2.4. The flood frequency distribution (AMC variable)

Wood (1976) pointed out how the AMC of the basin could be the most important factor influencing the estimation of flood frequency distribution. In order to take into account the variability of initial soil moisture condition prior to the storm, we now consider the AMC as a random variable, and associate to it a discrete probability distribution:

$$\begin{cases} P\{\text{AMC} = \text{I}\} = \pi_{\text{I}} \geq 0 \\ P\{\text{AMC} = \text{II}\} = \pi_{\text{II}} \geq 0 \\ P\{\text{AMC} = \text{III}\} = \pi_{\text{III}} \geq 0 \\ \pi_{\text{I}} + \pi_{\text{II}} + \pi_{\text{III}} = 1 \end{cases} \quad (19)$$

where $\{\pi_{\text{I}}, \pi_{\text{II}}, \pi_{\text{III}}\}$ are the probabilities of occurrence of the three different moisture conditions of the basin. It is clear that these probabilities are dependent on climatic conditions. For example, Gray et al. (1982) analysed 17 stations in Kentucky and Tennessee to determine the distribution of the AMC: they found a predominance of AMC I (85%), whereas AMC II (7%) and AMC III (8%) were much less probable. More recently, a similar study was carried out by Silveira et al. (2000), who calculated the AMC distribution considering rainfall events collected between 1992 and 1995 in a river basin in Uruguay, with return period equal or greater than one.

The distribution of the peak flood conditioned by the AMC distribution is given by:

$$F_Q(q) = \sum_{i=\text{I,II,III}} \pi_i \left(1 - \left(1 - \frac{k}{c} \times \left(I_A + \frac{q + \sqrt{q^2 + 4q\phi S}}{2\phi} - b\right)\right)^{1/k}\right)_{\text{AMC}=i} \quad (20)$$

for $q \geq 0$, where F_Q turns out to be the weighted sum of three terms; as indicated, the expression in parentheses depends upon the AMC conditions via the basin parameters tuning the dynamics of the phenomenon. Combining Eqs. (17) and (20) it is possible to calculate the distribution F_Q^* of the maximum annual peak flood Q^* conditioned by the AMC:

$$F_{Q^*}(q) = \exp\left(-\Lambda_Q\left(1 - \sum_{i=\text{I,II,III}} \pi_i \left(1 - \left(1 - \frac{k}{c} \times \left(I_A + \frac{q + \sqrt{q^2 + 4q\phi S}}{2\phi} - b\right)\right)^{1/k}\right)_{\text{AMC}=i}\right)\right) \quad (21)$$

for all suitable values of q. Note that, if AMC is constant, then all the probabilities π_is except one are zero: thus, Eq. (21) is also able to model the *deterministic* case described in Section 2.3.

As an illustration, Fig. 2 shows the function F_Q^* for five different AMC distributions. Overall, the strong influence of the AMC on the flood frequency distribution is evident: for example, passing from AMC I to AMC III, the 100-years flood quantile changes from 119 to 398 m^3/s. Thus, it is essential to stress the fundamental importance of the AMC distribution in order to properly identify the distribution of peak

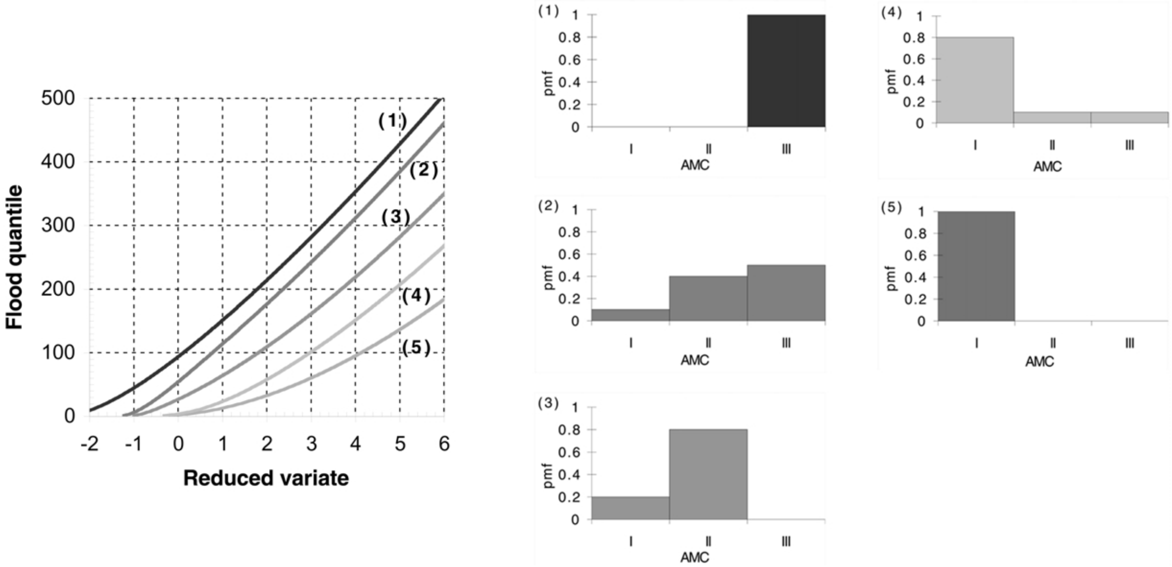

Fig. 2. Probability distribution of maximum annual peak flood for five different AMC probability distributions. Here the parameters are: $k = -0.031$, $c = 13.39$ mm, $b = 12.69$ mm, $\Lambda_Q = 9$ storm/year, $A = 34$ km^2, $t_c = 1.97$ h, and $S_{II} = 76$ mm.

flood and maximum annual peak flood. Note that the shape parameter of the flood distribution is the same as that of the rainfall distribution; in other terms, asymptotically the curve of maximum annual flood quantile becomes parallel to the curve of maximum annual rainfall quantile. Thus, the variability of initial moisture condition significantly affects the scale parameter of the flood frequency curve but not the shape parameter, and does not influence the asymptotic behaviour of the flood distribution, as is also evident in Fig. 2. Such a result is not new to hydrologic literature: see, e.g. the Gradex Method (Guillot and Duband, 1967), where a Gumbel law is used for the rainfall and the flood distribution, i.e. both laws have the shape parameter equal to zero. As a consequence, the estimation of the low return period quantiles (up to 100 years) is reliable, since the error on the shape parameter produces smaller consequences on the estimation of the low quantiles with respect to a similar error on the scale parameter (which might have unacceptable consequences in the same range of the return period values). This is also evident from the *sensitivity analysis* shown in Section 3.

3. Sensitivity analysis

An important theoretical and practical tool in hydrologic modelling is represented by sensitivity analysis. Such a tool provides a systematic means to examine the response of a hydrologic model in a way that is free of the 'error variation' that exists when dealing with measured data. This freedom makes it possible to assess more easily the rationality of the model, as well as examining the effect of the error in the input.

From an operative point of view, the *flood quantile function* $q(T)$, for any given return period T, represents the object of maximum interest. Inverting the fundamental expression of the law of Q, as given by Eq. (13), we obtain:

$$q(T) = \phi \frac{\left(\frac{k}{c}(1 - T^{-k}) - (0.2S - b) \right)^2}{S + k/c(1 - T^k) - (0.2S - b)} \tag{22}$$

Here we perform a *linearized sensitivity* analysis (McCuen and Snyder, 1986), which provides an analytical derivation of the sensible dependence of the derived quantiles upon the (rainfall and soil) parameters involved in Eq. (22), for several standard

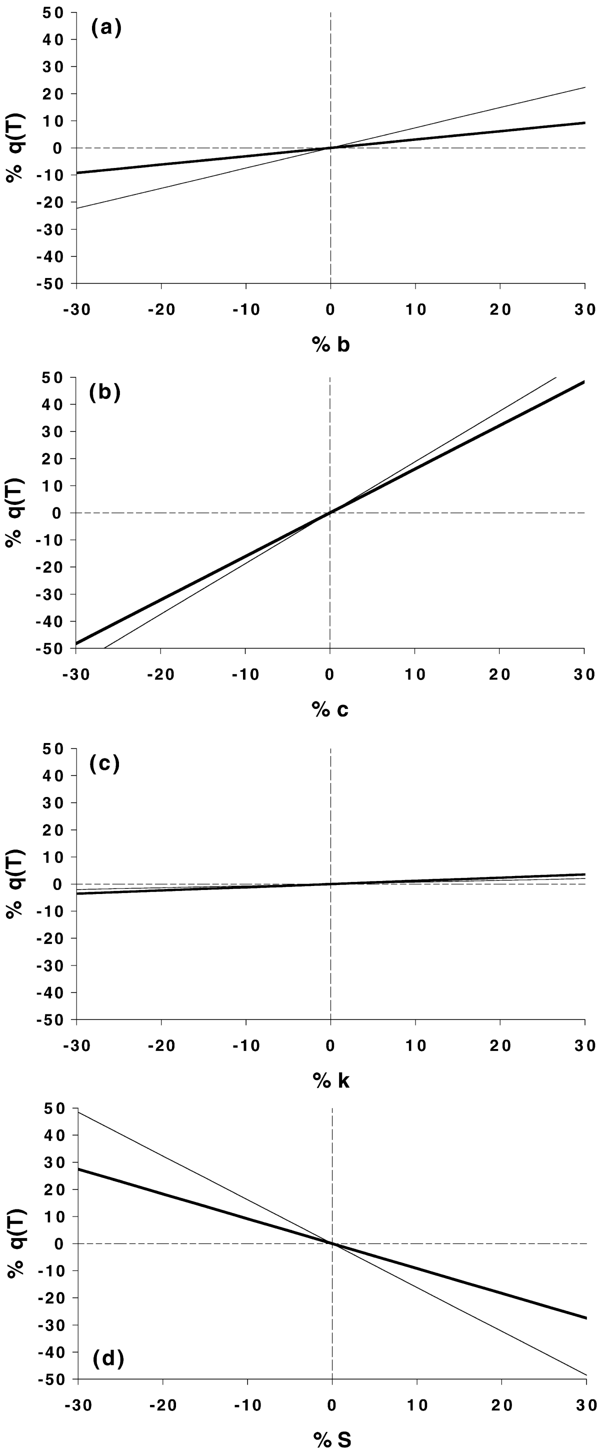

Fig. 3. Sensitivity analysis for the flood quantile function $q(T)$ given in Eq. (22), using two return periods: $T = 10$ years (thin line) and $T = 100$years (thick line). (a) % dependence on b; (b) % dependence on c; (c) % dependence on k; (d) % dependence on S.

Table 1
Characteristics of the basins investigated

	Bisagno	Vara	Arroscia
Area (km^2)	34.2	206	201
Relief (m)	1063	1640	2141
Rainfall (cm/year)	167	187	116
Soil	Limestone 62%; clay, clayey 31%	Sandy, Marly 22%; clay, clayey 56%	Sandy, Marly 36%; Calcar.-Marly 58%
Land use	Trans. wood/shrub 60%; agroforest 18%	Trans. wood/shrub 59%; sown field in well-water 18%; mixed forest 10%	Trans. wood/shrub 56%; agroforest 11%; mixed forest 10%
Rain G., # years	Scoffera, 35; Viganego, 39	P. Cento C., 20; Tavarone, 44; Varese L., 43	Pieve di Teco, 25
Flood G, # years	La Presa, 48	Naseto, 38	Pogli, 55

return periods. More particularly, we proceed as follows: we select, one at a time, one of the four parameters of interest (here ϕ is naturally discarded, being a simple multiplying factor, and thus introducing only a perfect linear dependence), leaving the remaining three fixed, and fix a return period T; then, we calculate the relative percentual variation of $q(T)$ as a function of the percentual variation of the selected parameter.

The results are shown in Fig. 3; the vertical scale is the same for all plots, to make the comparisons easier. Evidently, the dependence of $q(T)$ on k (plot (c) is weak, independently of T. In all other cases, the larger the T the smaller the dependence of $q(T)$ on the parameters involved; in particular, only the parameters c and S seem to affect in a significant way the behaviour of the flood quantile function, and the sensitivity analysis also shows how an error on the parameters yields a larger error on the flood quantiles corresponding to lower return periods (considering a fixed time of equilibrium of the basin t_c or, equivalently, a fixed ϕ). Indeed, the analysis shows how c and S are the most sensible parameters: a 30% error on c corresponds to a 50% error on the 100-year flood quantile, and a 30% error on S corresponds to a 30% error on the 100-year flood quantile. The sensitivity of the model with respect to b and k is rather modest: a 30% error on b corresponds to a 10% error on the 100-year flood quantile, and a 30% error on k corresponds to a 3% error on the 100-year flood quantile. Incidentally, this last result supports our use of a shape parameter for flood distribution equal to the one adopted for rainfall.

4. Model applications

The derived distribution approach outlined in the

Table 2
Estimates of the parameters $\alpha_{\Delta t}$ and $\xi_{\Delta t}$ (in mm), for different durations Δt. Also reported are the estimates of the corresponding scaling exponents δ_α and δ_ξ, together with their standard errors (s.e.). The bottom row shows the estimates of the parameter κ

Δt	Bisagno		Vara		Arroscia	
	α	ξ	α	ξ	α	ξ
1 h	12.10	35.31	9.86	28.29	9.26	24.49
3 h	17.09	56.27	16.54	46.28	15.40	37.64
6 h	20.17	75.22	20.69	61.71	19.99	51.71
12 h	25.95	103.39	27.02	79.15	30.24	76.34
24 h	31.40	134.04	31.50	101.90	40.90	102.97
δ	0.300	0.423	0.368	0.403	0.470	0.460
s.e.	0.008	0.006	0.026	0.013	0.016	0.017
κ	− 0.031		− 0.183		− 0.057	

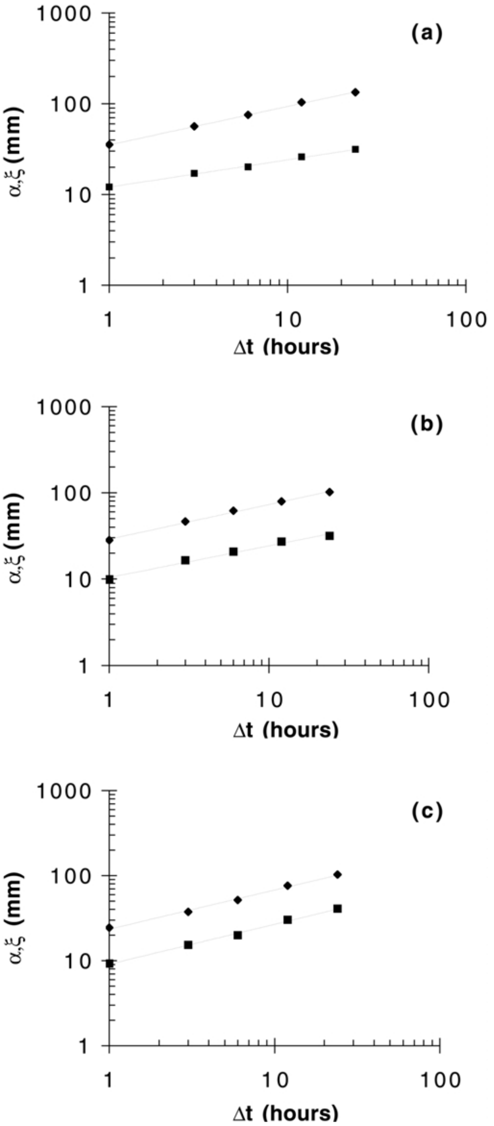

Fig. 4. Temporal scaling (in a log–log plane) of the parameters $\alpha_{\Delta t}$ (squares) and $\xi_{\Delta t}$ (diamonds) and the corresponding fits (lines) of the scaling exponents δ_α and δ_ξ for three basins: (a) Bisagno, (b) Vara, and (c) Arrosica.

previous sections is applied to three river basins: Bisagno at La Presa, Vara at Naseto, and Arroscia at Pogli, all located in Thyrrhenian Liguria, North-Western Italy. In Table 1 some basin characteristics are illustrated: area, relief, mean annual rainfall, soil type, land use, and information on rainfall and streamflow gauges, including the number of records. The

Table 3
Parameters of the derived distributions of peak flood (Eq. (20)) and maximum annual peak flood (Eq. (21))

	Unit	Bisagno	Vara	Arroscia
$b_{\Delta t=t_c}$	mm	13.08	17.10	20.00
$c_{\Delta t=t_c}$	mm	13.80	14.40	28.53
k		− 0.031	− 0.183	− 0.057
A	km^2	34.2	206	202
t_c	h	1.97	8.00	15.60
S_{II}	mm	80	89	99
π_I		0.25	0.26	0.27
π_{II}		0.12	0.16	0.08
π_{III}		0.63	0.58	0.65
Λ_Q	st./year	9	13	9

maximum annual rainfall depth is collected in each raingauge for the durations of 1, 3, 6, 12, and 24 h. Making the hypothesis that rainfall features are homogeneous in basins of small dimensions, the historical information concerning the hanging raingauges in a basin are pooled together. The parameters of the GEV distribution are calculated (at the basin scale) using the *L-moments* technique (Hosking, 1990): the shape parameter κ is constant (i.e. independent of Δt) in each basin, and is estimated using all the rainfall information at different durations, normalizing each time series with respect to the sample mean value; the estimated values are shown in Table 2. Then, using Eq. (6), it is possible to evaluate the scaling exponents δ_α and δ_ξ (see Table 2): these are statistically the same (within 99% confidence intervals) for both the Vara and the Arroscia basins; for the Bisagno basin, the difference between the two is statistically significant, and thus the corresponding distribution is scaling but not simple scaling. In Fig. 4 the temporal scaling of $\alpha_{\Delta t}$ and $\xi_{\Delta t}$ is shown for all the three basins considered, together with the fits of δ_α and δ_ξ : the scaling behaviour is well evident in all cases.

Using Eq. (5b), and the procedure described in Appendix A, it is possible to calculate the parameters Λ_P and $\{b_{\Delta t}, c_{\Delta t}, k_{\Delta t}\}$ simply through the estimate of the corresponding parameters $\{\xi_{\Delta t}, \alpha_{\Delta t}, \kappa_{\Delta t}\}$; then, using Eq. (15), the parameter Λ_Q can be estimated. In Table 3 the parameters of the derived distributions of peak flood (Eq. (20)) and maximum annual peak flood (Eq. (21)) are given. The maximum soil potential retention S_{II} (for AMC II), at the basin scale, is obtained from spatial integration over digital maps at resolution 225 m $\times$ 225 m (the estimates are also

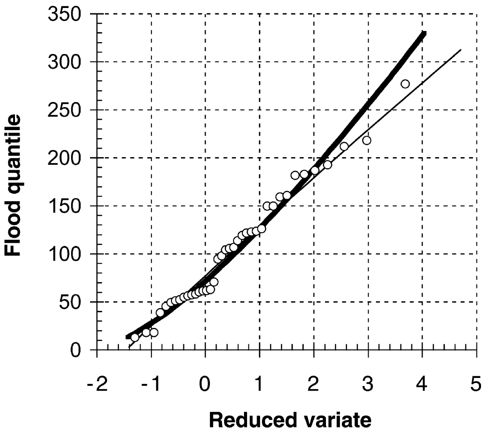

Fig. 5. Flood frequency curves (in m³/s) for the Bisagno river basin at La Presa: observations (circles), GEV fitted on data (thin solid line), derived distribution (thick solid line).

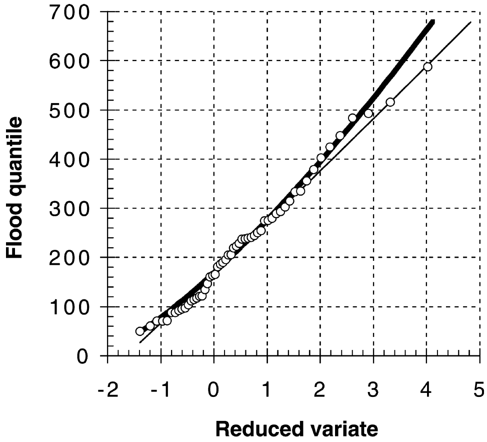

Fig. 7. Flood frequency curves (in m³/s) for the Arroscia river basin at Pogli: observations (circles), GEV fitted on data (thin solid line), derived distribution (thick solid line).

cross-checked with the ones obtained from the analysis of some observed rainfall-runoff events); in the present analysis we collected coupled observations and information about hyetograph and hydrograph for some flood events, and we used them to check and calibrate the mean value of S_{II} (at basin scale). Such a procedure is recommended by several authors (Gray et al., 1982; Hawkins et al., 1985; Silveira et al., 2000). Then, S_I and S_{III} are calculated as described by Ponce (1989). Finally, the estimated distributions of the AMC conditions for the three basins

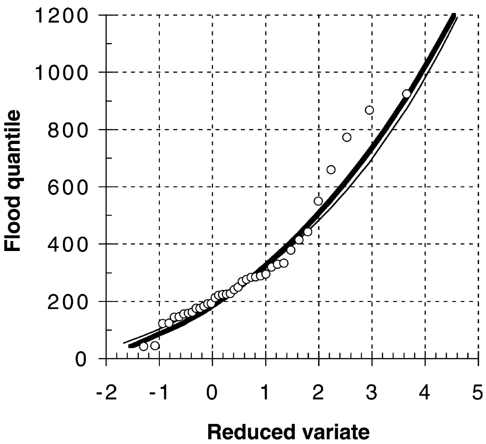

Fig. 5. Flood frequency curves (in m³/s) for the Vara river basin at Naseto: observations (circles), GEV fitted on data (thin solid line), derived distribution (thick solid line).

considered are shown. These are obtained by calculating the total 5-days antecedent rainfall for the series of maximum annual flood event in the Bisagno and Arroscia basins; instead, for the Vara basin we consider the 31 main flood events between 1993 and 1997. It is important to note that the distributions obtained are practically identical, thus indicating homogeneous climatic conditions for the area considered.

In Figs. 5–7 the T-year flood quantile obtained via the derived distribution is compared to the observed data of maximum annual peak flood for, respectively, Bisagno, Vara, and Arroscia catchments. Also shown is the GEV distribution fitted on the observed data. It is important to note that, for small return periods (say, $T \leq 100$ years), the derived distribution almost matches the GEV distribution calculated from the observed data, and both show a good agreement with the observations; in particular, the agreement is very good for Vara basin. From the analysis of Figs. 5–7 it is evident that the statistical information extracted from the series of maximum annual peak floods (and represented by the distributions fitted on the data) is also available considering the rainfall data (for small return periods) using the derived distribution approach proposed here.

Overall, the derived distribution is able to represent, up to a practically significant degree of approximation, the flood quantile curves, and thus it performs fairly well when considering small return periods.

5. Conclusions

In the present paper, we put the emphasis on several important facets of the problems concerning the determination of flood frequency distribution, starting from the statistical properties of rainfall and the hydrologic modelling of extreme floods. In particular, starting from simplified but physically based considerations, we provide the analytic expressions of the derived distribution of both peak flood and maximum annual peak flood. The following points must be stressed.

- *The introduction of the GP distribution for modelling the maximum storm depth, and the (temporal) simple scaling of the rainfall depth to represent the rainfall storm*. Apparently, the GP law provides a valuable model of the phenomena considered. In addition, the scaling feature may considerably simplify the mathematical tractability and description of the dynamics investigated: on the one hand, it offers a flexible tool for making inference at different temporal scales without changing the model adopted; on the other hand, it also provides a synthesis of the mechanisms underlying (extreme) storm events and (peak) flows, and provides a general conceptual framework for data analysis and modelling. Indeed, the simple rainfall model presented describes the fundamental rainfall properties necessary for calculating the flood frequency distribution for small return periods (say, $T \leq 100$ years); this model is also consistent with the DDF curves.
- *The importance of the antecedent soil moisture condition in the determination of flood frequency distribution*. In particular, using the SCS-CN method to transform the rainfall in excess, it is easy to incorporate the probability distribution of the AMC into the flood frequency distribution, and to show the importance of the knowledge of AMC for the estimation of the flood quantiles.
- *The comparison between the derived distribution and the available flood data*. It is evident that the statistical information extracted from the series of maximum annual peak floods (and represented by the distribution fitted on the data) is practically equivalent to the one obtained using the proposed derived distribution approach for small return periods.

As a conclusion, the derived distributions of peak flood and maximum annual peak flood proposed in this paper show a fairly good agreement with both the historical data and the distributions fitted to the data. These results support their application to river basin with no flood information and small return periods (say, $T \leq 100$ years).

Acknowledgements

The research was partially supported as a part of the project 'FRAMEWORK' (Flash-flood Risk Assessment under the impact of land use changes and river Engineering works) granted by the European Commission, D.G. XII, Contract no. ENV4-CT97-0529. The support of CNR-GNDCI via the RIVERS-project and funds by 'Progetto Giovani Ricercatori' are also acknowledged.

Appendix A

Considering Eq. (5b), suppose that some estimates of the parameters $\{\xi_{\Delta t}, \alpha_{\Delta t}, \kappa_{\Delta t}\}$ are available, and that the remaining parameters Λ_P and $\{b_{\Delta t}, c_{\Delta t}, k_{\Delta t}\}$ need to be estimated; clearly, there would be only three equations to estimate four parameters. However, we may suggest the following procedure to solve the problem.

Taking twice the logarithm of Eq. (4) we obtain:

$$\ln\Big(-\ln F_{P^*_{\Delta t}}(x)\Big) = \ln \Lambda_P + \frac{1}{k_{\Delta t}} \ln\left(1 - \frac{k_{\Delta t}}{c_{\Delta t}}(x - b_{\Delta t})\right)$$

Let $x^*_{(1)} < \cdots < x^*_{(n)}$ be the sample *order statistics* corresponding to n available observations of $P^*_{\Delta t}$. Since for (small) $x \approx b_{\Delta t}$ the above formula can be approximated by:

$$\ln\Big(-\ln F_{P^*_{\Delta t}}(x)\Big) \approx \ln \Lambda_P - \frac{1}{c_{\Delta t}}(x - b_{\Delta t})$$

we see that (assuming, as empirically reasonable, that the smallest order statistics of $P^*_{\Delta t}$ are close to $b_{\Delta t}$):

$$\varphi^*_i \approx \ln\Big(-\ln F_{P^*_{\Delta t}}(x^*_{(i)})\Big) \approx \ln \Lambda_P - \frac{1}{c_{\Delta t}}(x^*_{(i)} - b_{\Delta t})$$

where $\varphi^*_i = \ln(-\ln(i/(n+1)))$ and the index i is small. Thus, we may provide an estimate $\hat{c}_i$ of $c_{\Delta t}$ which involves none of the remaining parameters as follows:

$$\hat{c}_i \approx \frac{x^*_{(i+1)} - x^*_{(i)}}{\varphi^*_i - \varphi^*_{i+1}}$$

which, in turn, can be combined with the last two formulas of Eq. (5b) to obtain an independent estimate of Λ_P. Practically it turns out that the average of several estimates $\hat{c}_i$s (for small indices is) provides reliable estimates of the parameter c.

References

Burlando, P., Rosso, R., 1993. Stochastic models of temporal rainfall: reproducibility, estimation and prediction of extreme events. In: Salas, J.D., Harboe, R., Marco-Segura, J. (Eds.). Stochastic Hydrology and its Use in Water Resources Systems, Simulation and Optimization. Kluwer Academic Publishers, Dordrecht, pp. 137–173.

Castillo, E., 1994. Extremes in engineering applications. In: Galambos, J. (Ed.). Proceedings of the Conference on Extreme Value Theory and its Applications. Kluwer Academic Publishers, Gaithersburg, MD, pp. 15–42.

Castillo, E., Hadi, A.S., 1997. Estimating the parameters of the Generalized Pareto law. J. Amer. Stat. Ass. 92 (440), 1609–1620.

Cordova, J.R., Rodriguez-Iturbe, I., 1983. Geomorphoclimatic estimation of extreme flow probabilties. J. Hydrol. 65, 159–173.

Davison, A.C., 1984. Modelling excesses over high thresholds, with an application. In: Tiago de Oliveira, J. (Ed.). Statistical Extremes and Applications. NATO ASI Series. Reidel, Dordrecht, pp. 461–482.

De Michele, C., Rosso, R., 2000. Valutazione delle piene nell'Italia nord-occidentale: bacino padano e Liguria tirrenica. La Valutazione delle Piene in Italia. CNR-GNDCI, Roma pp. 4.1–4.26, in Italian.

Diaz-Granados, M.A., Valdes, J.B., Bras, L.R., 1984. A physically based flood frequency distribution. Water Resour. Res. 20 (7), 995–1002.

Eagleson, P.S., 1972. Dynamics of flood frequency. Water Resour. Res. 8 (4), 878–898.

Eagleson, P.S., 1978. Climate, soil and vegetation 5, a derived distribution of storm surface runoff. Water Resour. Res. 14 (5), 741–748.

Gray, D.D., Katz, P.G., deMonsabert, S.M., Cogo, N.P., 1982. Antecedent moisture condition probabilities. J. Irrig. Drain. Div., ASCE 108 (2), 107–114.

Guillot, P., Duband, D., 1967. La méthode du Gradex pour le calcul de la probabilité des crues á partir des pluies. IAHS, 560–569 Publ no. 84.

Gupta, V.K., Waymire, E., 1990. Multiscaling properties of spatial rainfall and river flow distributions. J. Geophys. Res. 95 (D3), 1999–2009.

Hawkins, R.H., Hjelmfelt Jr, A.T., Zevenbergen, A.W., 1985. Runoff probability, storm depth, and curve numbers. J. Irrig. Drain. Div., ASCE 111 (4), 330–340.

Hosking, M., 1990. *L*-moments: analysis and estimation of distribution using linear combinations of order statistics. J. Royal Stat. Soc. B (52), 105–124.

Hosking, J.R.M., Wallis, J.R., 1987. Parameter and quantile estimation for the Generalized Pareto distribution. Technometrics 29 (3), 339–349.

Iacobellis, V., Fiorentino, M., 2000. Derived distribution of floods based on the concept of partial area coverage with climatic appeal. Water Resour. Res. 36 (2), 469–482.

Klemeš, V., 1978. Physically based stochastic hydrologic analysis. In: Chow, V.T. (Ed.). Advances in Hydrosciences, 11. Academic Press, New York, pp. 285–352.

McCuen, R.H., Snyder, W.M., 1986. Hydrologic Modeling: Statistical Methods and Applications. Prentice-Hall, New Jersey.

Meigh, J.R., Farquharson, F.A.K., Sutcliffe, J.V., 1997. A worldwide comparison of regional flood estimation methods and climate. Hydrol. Sci. J., Ser. B 42 (2), 225–244.

van Montfort, M.A.J., Witter, J.V., 1985. Testing exponentiality against Generalized Pareto distribution. J. Hydrol. 78, 305–315.

van Montfort, M.A.J., Witter, J.V., 1986. The Generalized Pareto distribution applied to rainfall depths. Hydrol. Sci. J., Ser. B 52, 392–442.

Ponce, V.M., 1989. Engineering Hydrology. Principles and Practices. Prentice Hall, New Jersey.

Raines, T.H., Valdes, J.B., 1993. Estimation of flood frequencies for ungaged catchments. J. Hydraul. Engng 119 (10), 1138–1155.

Robinson, J.S., Sivapalan, M., 1997a. An investigation into the physical causes of scaling and heterogeneity of regional flood frequency. Water Resour. Res. 33 (5), 1045–1059.

Robinson, J.S., Sivapalan, M., 1997b. Scaling of flood frequency: temporal scales and hydrologic regimes. Water Resour. Res. 33 (12), 2981–2999.

Rodriguez-Iturbe, I., 1986. Scale of fluctuation of rainfall models. Water Resour. Res. 22 (9), 15S–37S.

Salvadori, G., De Michele, C., 2001. From Generalized Pareto to extreme values laws: scaling properties and derived features. J. Geophys. Res. 106 (D20), 24063–24070.

Silveira, L., Charbonnier, F., Genta, J.L., 2000. The antecedent soil moisture condition of the curve number procedure. Hydrol. Sci. J. 45 (1), 3–12.

Sivapalan, M., Wood, E.F., Beven, K.J., 1990. On hydrologic similarity 3, a dimensionless flood frequency model using a generalized geomorphic unit hydrograph and partial area ruonff generation. Water Resour. Res. 26 (1), 43–58.

Temez, J.R., 1991. Extended and improved rational method. Proceedings of XXIV Congress IAHR, Madrid, Spain, A pp. 33–40.

US Department of Agriculture, Soil Conservation Service, 1986. National Engineering Handbook, Hydrology, Sec. 4 Washington DC.

Wood, E.F., 1976. An analysis of the effects of parameter uncertainty in deterministic hydrologic models. Water Resour. Res. 12 (5), 925–932.

Wood, E.F., Hebson, C., 1986. On hydrologic similarity 1, derivation of the dimensionless flood frequency curve. Water Resour. Res. 22 (11), 1549–1554.