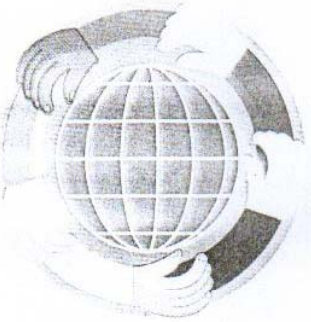
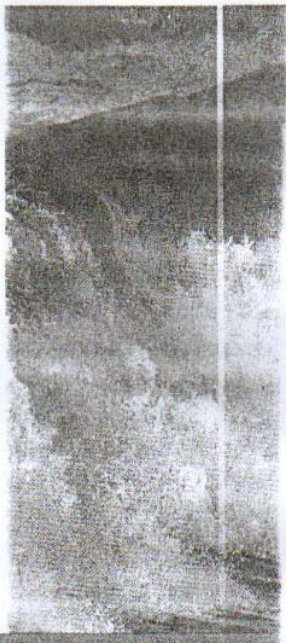
Water, Environment, Energy and Society

Proceedings of WEES-2009





Volume II
Statistical and Systems Analysis Techniques

Proceedings of the International Conference on

Water, Environment, Energy and Society (WEES-2009)

January 12-16, 2009, New Delhi, India

VOLUME–II

Statistical and Systems Analysis Techniques

Organized by

National Institute of Hydrology, Roorkee
Ministry of Water Resources, Government of India
New Delhi–110 001

ALLIED PUBLISHERS PVT. LTD.

New Delhi • Mumbai • Kolkata • Lucknow • Chennai
Nagpur • Bangalore • Hyderabad • Ahmedabad

Contents

VOLUME–II: STATISTICAL AND SYSTEMS ANALYSIS TECHNIQUES

International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

Assessment of the Relation between Rainfall and Flood Frequency Curves through Peak over Threshold Analysis

P. Claps[1], P. Allamano and F. Laio
Department of Hydraulics, Transports and Civil Infrastructures
Polytechnic of Turin - 10129, ITALY
E-mail: [1]claps@polito.it

ABSTRACT: The analysis of the flood production mechanisms from rainfall is crucial for building flood frequency curves in ungauged basins. In this paper, the rainfall-runoff transformation under heterogeneous geomorphoclimatic conditions is investigated by comparing the shape of the flood and rainfall frequency curves in gauged basins. We propose an empirical comparison of the magnitude of equiprobable rain and flood peaks, at the basin scale and for different return periods. In order to obtain reliable estimates of the empirical frequency assigned to each peak we use a revised partial duration series procedure, named Filtered Peaks Over Threshold (FPOT), as an alternative to the customary annual maxima approach. The analysis is carried out on 33 time series of daily discharge and rainfall from basins located in North-Western Italy. Results of the analysis show that for the majority of basins the shape of the rainfall and runoff distributions is statistically indistinguishable. However, a relevant deviation from this rule is represented by basins where the dominating flood production mechanism is snow melting. For these cases, a simple theoretical interpretation of the difference between the shape of the frequency curves of rainfall and runoff is presented.

INTRODUCTION

In the context of the regional flood frequency analysis, good results in terms of transfer of hydrological information to ungauged basins can be achieved if significant information about the characteristics of rainfall is gathered. This is confirmed by the abundant literature on the derived flood distribution approach that follows the seminary work by Eagleson (1972) (see Iacobellis and Fiorentino, 2000 for an updated review on this topic). An important point in this field concerns the necessity to investigate if and why there is a change in shape between rainfall and flood frequency curves for a given basin (e.g. Hashemi *et al.*, 2000). A change in shape is in fact a clear clue that the Rainfall-Runoff (R-R) relation is either non-linear or non-deterministic, and this can address the investigations on appropriate models for the R-R transformation. The scope of this analysis is therefore the assessment of the geomorphoclimatic conditions possibly producing the mentioned changes in shape. In this regard, the present work is akin to that of Hashemi *et al.* (2000) and Franchini *et al.* (2000), who propose a sensitivity analysis for determining climatic and basin factors affecting the features of the flood frequency curve. However the methods employed here for the analysis are completely different, since the mentioned authors follow a numerical simulation approach while we use an empirical, data-based, procedure (see also Claps *et al.*, 2002). The main difficulty of an extended, data-based, analysis of flood and rain frequency curves is in the limited length of the available records, that contrasts with the necessity to have a (data-demanding) estimate of the shape of the frequency curves. For example, if annual maximum time series are used, the fitting of a three-parameter probability distribution often relies on just 25–30 years long time-series. It is easy to verify that in most cases the resulting estimates would have such a large uncertainty to hamper the significance of the whole analysis. In order to reduce this uncertainty, we use daily data of rainfall and discharge extracted by means of a Peak Over Threshold (POT) technique. In this technique the basic idea is to use more than one flood or rainfall peak per year, to increase as much as possible the basic available flood information constituted by annual maxima of discharge.

The revised POT approach (see Claps and Laio, 2003) that is used here is summarized in the next section. The comparison of rain and flood frequency curves is carried out on 33 time series of daily discharge and rainfall from basins located in North-Western Italy. The results of the analysis are shown and discussed in the fourth section followed by a simple theoretical interpretation of the rainfall-runoff mechanism for the basins that present a particular form of the rainfall and discharge frequency curves.

THE FPOT PROCEDURE

In the field of flood frequency modeling, a valid alternative to the usual analysis of annual maxima has been developed under the names of Peak Over Threshold (POT) or Partial Duration Series (PDS) analysis. The characteristics of the method are reviewed by Madsen *et al.* (1997) and by Lang *et al.* (1999). The basic idea is to extract from the daily discharge sequences a sample of peaks containing more than one flood peak per year, in order to increase the available information with respect to the annual maxima analysis. However, in the practical applications ambiguous criteria for peak selection affect the efficiency of the method: for example, the average number of peaks per year, λ, is usually forced to remain in the range between 2 and 3 in order to preserve independence among subsequent flood peaks (e.g., Lang *et al.*, 1999; Madsen *et al.*, 1997), despite the fact that such low values contrast with physical and statistical considerations. This contradiction and the ambiguities in the choice of the appropriate threshold were overcome by Claps and Laio (2003), who proposed a revised procedure for peak selection, named Filtered Peak Over Threshold (FPOT), that is rapidly summarized below: (i) The peak events are identified in correspondence to all the local maxima of the daily discharge time series, and the magnitude of each event is assigned as the absolute ordinate of the maximum. This sequence of peaks presents a noisy component that must be removed (filtered). (ii) A second sequence of peaks is then obtained by subtracting from the magnitude of each event the discharge measured at the first relative minimum preceding the event itself. In this manner we obtain a sequence of Filtered Peaks (FP). (iii) A threshold filter is applied to the FP sequence to retain only the large peaks. The problem of choosing a correct threshold for the analysis is still present in the FPOT procedure; however, in this case one takes advantage by the fact that the noisy component becomes better recognizable when FP events are considered (see next step). (iv) The appropriate threshold is selected by testing the independence of the peaks in the sample (Kendall's τ test), the distribution of occurrences (Cunnane test for the Poisson distribution) and the distribution of the peak magnitudes (Cramer-von Mises goodness of fit test for the Pareto distribution). The threshold is gradually increased until the three tests are jointly met. Note that the three tests are applied by considering the real magnitude of each peak, and not its filtered value: in fact, the FP peaks sequence is used only to preliminarily filter out the noise from the sample. The FPOT approach allows one to increase the number of peaks per year to 5–10, thus reducing the uncertainty of estimate of the flood frequency curve with respect to the classical POT procedures (see Claps and Laio, 2003). The FPOT procedure was formulated for flood frequency analysis, but it can be readily extended to rainfall data. In this case the procedure is even simpler, since the rainfall time series does not have the strong auto correlation that characterizes discharge. It is then not necessary to preliminary filter the series (steps (i) and (ii) mentioned above). When working with rainfall series, it is therefore sufficient to directly apply the threshold filter to the daily data, and increase the threshold until an independent Poisson-Pareto sample is obtained. The only problem with this approach is that rainfall days tend to occur in clusters, which makes the Poisson distribution unsuited for modeling the occurrences (in fact, the Cunnane test is met by the rainfall samples only with very large thresholds, often corresponding to less than one event per year). A more adequate choice for the modeling of occurrences in this case would be the negative binomial distribution (e.g. Onoz and Bayazit, 2001); however, the distribution of occurrences does not affect the shape of the frequency curve, since only the average number of events per year, λ, is relevant to its determination (see for example Lang *et al.*, 1999 and Claps and Laio, 2003). We therefore apply the FPOT procedure to the rainfall data without testing the distribution of occurrences: the threshold is accepted if the independence and goodness-of-fit tests are met. For consistency, the Poisson hypothesis is also disregarded for discharge. The only drawback of this simplification is that the correspondence between the distribution of peak magnitudes and the distribution of annual maxima is lost. In fact, if the occurrences were Poisson distributed, the Pareto distribution of peak magnitudes would correspond to the Generalized Extreme Value distribution of annual maxima. Once the FPOT procedure has been applied, one obtains two samples of independent peaks, one for rainfall and one for discharge. Based on these samples, one can estimate the flood and rainfall frequency curves, or, which is conceptually the same, the relation between the return period T and the corresponding critical rainfall and flood, p_T and q_T. When the Generalized Pareto distribution is used to model the peak magnitudes, this relation assumes the form,

$$q_T = q_0 + \frac{a_Q}{k_Q}\left[1 - \left(\frac{1}{\lambda_Q T}\right)^{k_Q}\right] \qquad \ldots (1)$$

where q_0 and λ_Q are directly obtained from the peak selection procedure, while a_Q and k_Q are estimated from the peak samples (for example, we use maximum likelihood estimators). An equation analogous to (1) applies to the rainfall peaks sequence, with a different set of parameters p_0, λ_P, a_P and k_P.

CASE-STUDY

33 time series of daily data of runoff and spatially averaged rainfall are analysed. The daily runoff time-series are considered in mm/day, in order to have units consistent to those of rainfall. The average rainfall over each basin is found by means of the Thiessen polygons using daily precipitation measurements from a network of 517 rainfall stations covering the entire region of interest. Appropriate corrections are applied to account for sparse missing data in the hystorical records. The drainage basins of interest are mostly located in the North-West of Italy, in the Piemonte and Valle d'Aosta regions (see Figure 1). The Alps influence the majority of basins in the considered group (basins with code number from 1 to 24), but some basins are located in an Apennine (code number 30–33) or in a mixed context (code number 25–29).

The main characteristics of the considered basins are reported in Table 1. The basins in this dataset represent a significant starting point for the variability they show in the geomorphoclimatic features: as reported in Table 1, their area ranges from very small (37 Km^2) to rather large (~8000 Km^2), and their average altitude above the sea level ranges from ~500 m (typical of an hilly region) to ~2600 m (nivo-glacial environment). The rainfall regimes also present high variability, as proved by the total amount of rainfall per year, ranging from 835 mm/year to ~2000 mm/year and by the differences in monthly rainfall seasonality. Also note in Table 1 that the length of the discharge time series is always much shorter than the corresponding rainfall time series, a situation that is often found in the Italian context.

DATA-BASED ANALYSIS OF FLOOD PRODUCTION MECHANISMS

From the daily time series for each of the 33 drainage basins in the dataset, samples of rainfall and flood peaks are selected. These samples contain the critical values to build the empirical frequency curves. If one considers the relation between the T-year event and the return period T as a representation of the frequency curve, one needs to attribute a return period to each measured peak value in order to build the empirical curve.

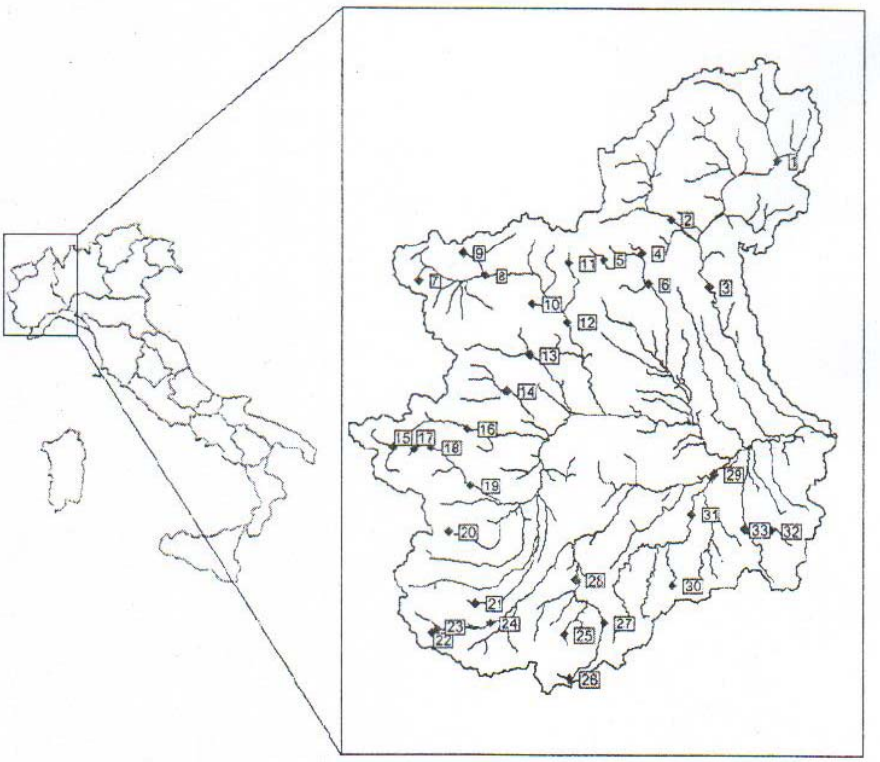


Fig. 1: Location of the considered drainage basins

Table 1: Characteristic features of the drainage basins considered in the application. Along with the drainage area, the average elevation h_m, the annual amount of rainfall P_{tot} and the record length for discharge, t_Q, and rainfall, t_P, are reported for each station

Code	Name	Area (Km^2)	hm (m a.s.l.)	P_{tot} (mm/years)	t_Q (years)	t_p (years)
1.	Ticino at Bellinzona	1515	1615	1560	10	56
2.	Toce at Candoglia	1532	1641	1570	21	56
3.	Ticino at Sesto Calende (Miorina)	6600	1283	1740	26	54
4.	Mastallone at Ponte Folle	149	1350	1995	22	54
5.	Sesia at Campertogno	170	2120	1421	7	54
6.	Sesia at Ponte aranco	695	1480	1696	9	58
7.	Rutor at Promise	50	2616	883	20	34
8.	Dora baltea at Aosta	1840	2270	835	10	50
9.	Artanavaz at St. Oyen	69	2206	1304	16	42
10.	Ayasse at Champorcher	42	2392	1271	22	61
11.	Lys at Gressonay St. Jean	91	2615	1035	7	54
12.	Dora Baltea at Tavagnasco	3313	2080	870	18	48
13.	Orco at Ponte Canavese	617	1930	1254	29	48
14.	Stura Lanzo at Lanzo	582	1751	1268	32	53
15.	Dora Riparia at Oulx (Ulzio)	262	2169	959	10	55
16.	Dora Riparia at S. Antonino di Susa	1048	1613	838	10	59
17.	Chisone at Soucheres	94	2233	849	7	62
18.	Chisone at Fenestrelle	155	2169	856	8	62
19.	Chisone at S. Martino	580	1751	1033	26	51
20.	Po at Crissolo	37	2235	1198	28	69
21.	Grana at Monterosso	102	1540	1224	32	30
22.	Rio Bagni at Bagni Vinadio	63	2124	983	11	62
23.	Stura di Demonte at Pianche	181	2070	964	14	62
24.	Stura di Demonte at Gaiola	562	1817	1071	11	63
25.	Corsaglia at Presa C. Molline	89	1530	1375	18	49
26.	Tanaro at Ponte Nava	148	1623	1182	30	47
27.	Tanaro at nucetto	375	1227	1209	29	46
28.	Tanaro at Farigliano	1522	938	1159	34	46
29.	Tanaro at Montecastello	7985	663	1005	37	52
30.	Erro at Sassello	96	591	1440	16	35
31.	Bormida at Cassine	1483	493	1007	12	44
32.	Borbera at Baracche	202	880	1266	14	43
33.	Scrivia at Serravalle	605	695	1376	14	40

This is done by assigning to the j-th peak in the ordered sample (j is then the rank of the peak) a return period,

$$T_j = \frac{t_Q}{\lambda_Q \cdot t_Q + 0.5 - j} \quad \ldots (2)$$

where t_Q is the record length in years, λ_Q the average number of flood peaks per year, and the 0.5 value comes from assuming the Hazen plotting position. Each peak value is represented by a point in the plane (T, q_T), and the sequence of all points constitutes the empirical frequency curve (see Figure 2). The same procedure can be followed to construct the rainfall frequency curve (obviously, λ_P and t_P should be used in (2)). An analytical approximation of the shape of these frequency curves is given by equation (1), under the FPOT hypotheses of independent and Pareto distributed peaks. By plotting the empirical points and the analytical curves for rainfall and discharge on the

same diagram (Figures 2–4), one can first assess the goodness of the analytical approximation, and then analyse the mutual relations between the two frequency curves.

The example in Figure 2 refers to river Toce at Candoglia and depicts a situation in which the empirical points are well fitted by the Pareto distribution, and the two curves for rainfall and discharge are very close to straight lines on the semi-logarithmic paper. Note that a linear relation between the T-year event and $\log(T)$ is obtained by setting $k_Q = 0$ in equation (1), for the well known simplification of the Pareto to an exponential distribution when $k_Q = 0$ (e.g. *Madsen et al.*, 1997). In any case, for the river Toce at Candoglia one can conclude that the shapes of the rainfall and discharge frequency curves appear very similar to each other. When other drainage basins are considered (see Figures 3 and 4), the main findings are that (i) the Pareto distribution provides a good fit to the empirical data in most of the cases; (ii) the curvature of the rainfall distribution is often weak on the $\{p_T, \log(T)\}$ plane; (iii) the discharge distribution, in contrast, tends in some cases to assume a relevant upward or downward curvature. Note that an upward curvature entails a rapid increase of the T-year floods with the return period (Figure 3), and corresponds to having a large positive skewness of the distribution. In contrast, when the curvature is downward the T-year floods for large return periods tend to a constant value, and the skewness of the distribution is much lower (Figure 4).

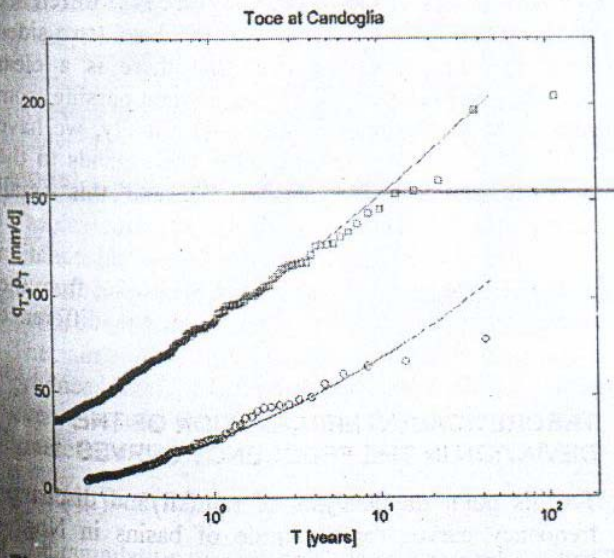


Fig. 2: Rainfall (squares) and discharge (circles) frequency curves for basin *n.* 2, river Toce at Candoglia: the T-year variables are plotted as a function of return period T. The continuous lines correspond to equation (1) with parameters fitted from the data

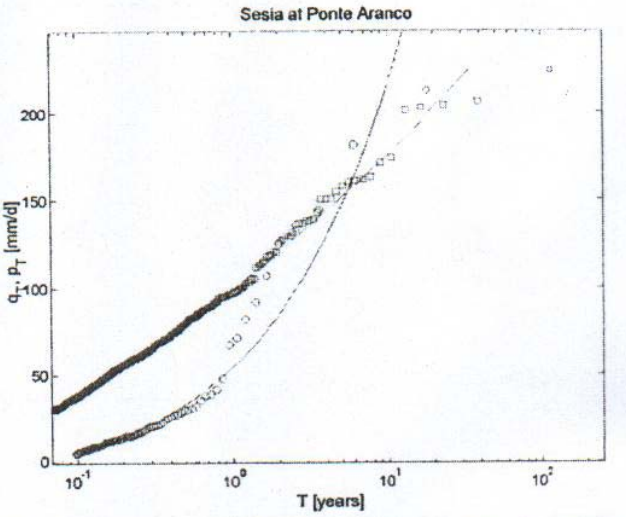


Fig. 3: Rainfall (squares) and discharge (circles) frequency curves for basin *n.* 6, river Sesia at Ponte Aranco: the T-year variables are plotted as a function of return period T. The continuous lines correspond to equation (1) with parameters fitted from the data

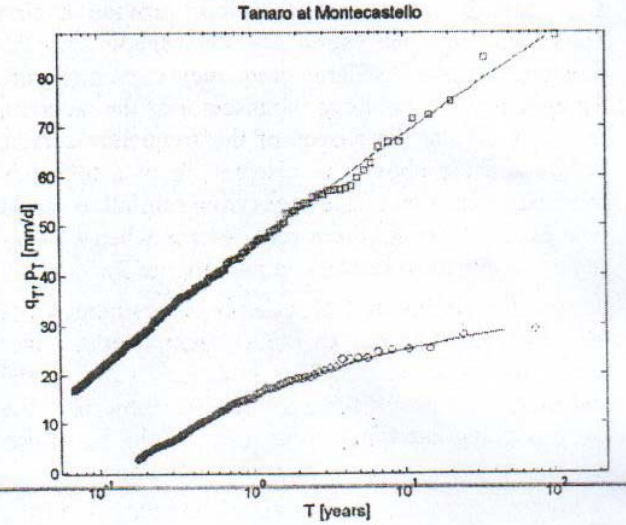


Fig. 4: Rainfall (squares) and discharge (circles) frequency curves for basin *n.* 6, river Tanaro Montecastello: the T-year variables are plotted as a function of return period T. The continuous lines correspond to equation (1) with parameters fitted from the data

A question arises from the observation of Figures 2–4: how many basins behave like river Toce, with rain and runoff frequency curves very similar in shape, and how many like river Sesia or river Tanaro, with a clear change in shape? A possible answer can be given considering the value of the k_P and k_Q coefficients in (1). In fact, k acts as a shape coefficient for the frequency curves: positive k values correspond to a downward curvature (discharge in Figure 4), negative

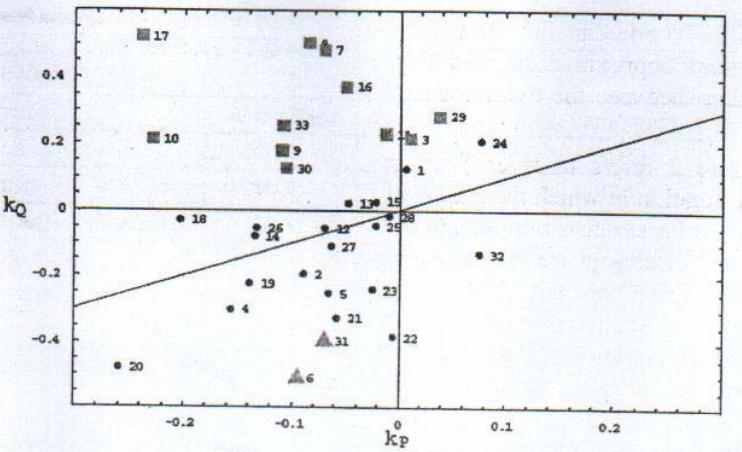


Fig. 5: Scatter plot of the shape parameter of discharge and rainfall frequency curves (k_q and k_p, respectively). The number close to each point refers to the code of the drainage basin, from Table 1. Squares and triangles represent basins where the difference between k_q and k_p is significant at the 10% level (see text for details). The inclined line is the bisector of the diagram

values to an upward curvature (discharge in Figure 3), and when k tends to zero the frequency curves are approximately linear (Figure 2). A scatter plot of the k_Q versus k_P values can therefore provide a first assessment of the similarity in shape of the precipitation and discharge frequency curves (Figure 5): the closer the points to the bisector of the diagram, the more similar the shapes of the frequency curves. Points that are above the bisector show a tendency towards a reduction of skewness from rainfall to runoff (see Figure 4), while the reverse is true when $k_Q < k_P$ (points below the bisector, compare Figure 3).

Also note in Figure 5 that the k_P values range from –0.3 to 0.1, while the k_Q values span over a wider range, from –0.5 to 0.5. This is probably due to the shortness of the discharge records towards the corresponding rainfall records (see Table 1), which greatly increases the variance of the estimate of k.

Figure 5 provides a good visual comparison of the shape coefficients for rainfall and discharge, but a statistically reliable procedure is needed to understand where the differences in shape are significant. This can be done in an approximate but simple manner by considering that the maximum likelihood estimate of k from a Pareto distribution is asymptotically normally distributed with variance:

$$\sigma^2_{k_Q} = \frac{(1-k_Q)^2}{\lambda_Q \cdot t_Q} \qquad \ldots (3)$$

(e.g. *Madsen et al.*, 1997). Since k_Q and k_P are estimated from independent samples, the difference $d = (k_Q - k_P)$ is also asymptotically normally distributed with variance $\sigma^2_{k_Q} + \sigma^2_{k_P}$. It is therefore possible to calculate for each basin the probability that $d \leq 0$, P_D ($d \leq 0$): when this probability is below some significance level r (here, $r = 0.05$), one can conclude that the difference between k_Q and k_P is significantly positive at the r level (squares in Figure 5). Similarly, when P_D ($d \leq 0$) >1–r, ($k_Q - k_P$) is significantly negative at the r level (triangles in Figure 5). In all of the other cases, the difference between the two shape parameters is not statistically relevant (circles in Figure 5). From the analysis of Figure 5 we are now able to conclude that for the majority of basins (20 out of 33) the difference in shape is not significant at the 10% level (two sided test). For 11 basins we find that there is a clear reduction in skewness ($k_Q >> k_P$) when passing from rainfall to discharge (see Figure 4). Finally, we have $k_Q << k_P$ for only 2 basins. This corresponds to the opposite situation (see Figure 3), and this small number can be due to the fact that negative values of k_Q have such a large variance of estimate (see equation 3) to make them statistically indistinguishable from the corresponding k_p values, even when the difference ($k_Q - k_P$) is rather large.

THEORETICAL INTERPRETATION OF THE DEVIATION IN THE FREQUENCY CURVES

To this point the analysis of rainfall and discharge frequency curves for a sample of basins in North-Western Italy has demonstrated that there are several basins for which there is a consistent reduction in skewness from rainfall to runoff. For some of these basins the high average elevation and the geographic

location suggest that this reduction can be due to the influence of snow accumulation and melting in the flood production mechanism. For other cases, e.g. river Tanaro at Montecastello in Figure 4, the downward slope of the discharge frequency curve is probably due to overflow phenomena. Few other basins, however, apparently elude both of the above simple conjectures.

In this section theoretical frequency curves for rainfall and discharge are analyzed to justify changes in the flood production mechanisms that reflect in changes of the return period of the events. To this end, a return period T is set and a corresponding abstraction $\Delta s(T)$ is defined as,

$$\Delta s(T) = s_P(T) - s_Q(T) = p_T - q_T \qquad \ldots (4)$$

where s represents a generic threshold value and p_T and q_T are the corresponding values of $s_P(T)$ and $s_Q(T)$ as indicated in Figures 2–4. A simple theoretical model for the rainfall-runoff transformation is then applied with the aim of providing some insights on the relationship between the current abstraction Δs and the return period T of the events. In the model, rainfall and runoff differ by a simple constant abstraction Φ and a runoff coefficient α,

$$q = \alpha(p - \Phi) \quad \text{if } p > \Phi \qquad \ldots (5)$$
$$q = 0 \quad \text{otherwise.}$$

In Eqn. (5) α has to be taken in the range [0, 1] (see Iacobellis and Fiorentino (2000)). Observe that in this relation q and p do not have the same return period, i.e. $q_T \neq \alpha(p_T - \Phi)$. The following analysis is in fact referred to the event-scale, where the discharge q represents the direct consequence of the rainfall p.

Model 1: Constant Abstraction

If one assumes $\alpha = 1$, the surface runoff is assumed to occur with a saturation from below mechanism, and the soil is supposed to be always completely dry before a rain event and to have a fixed capacity, equal to Φ. The rain events with magnitude lower than Φ do not produce runoff, so that the frequency of the flood events equals the frequency of the rainfall events greater than Φ, i.e.

$$T(q_0) = T(\Phi). \qquad \ldots (6)$$

The cumulative density function of the discharges can be deduced from that of rainfall as a derived distribution truncated to avoid the presence of the atom of probability at $q = 0$. The result is (Benjamin and Cornell, 1970),

$$F_Q(s_Q) = \frac{F_P(s_P + \Phi) - F_P(\Phi)}{1 - F_P(\Phi)} \qquad \ldots (7)$$

A simple explicit relationship exists between the return period of the peak values that exceed a threshold s_X, namely $T_X(s_X)$, and the correspondent cdf's of the exceedences $x > x_0$ being, for time discrete processes, λ_X the average number of independent events in one year,

$$\frac{1}{T_X(s_X)} = \lambda_X\left(1 - F_X(s_X)\right) \qquad \ldots (8)$$

Given (7) and (8) the following relationships hold,

$$\begin{aligned}\frac{1}{T_Q(s_Q)} &= \lambda_Q\left[1 - F_Q(s_Q)\right] \\ &= \lambda_Q\left[1 - \frac{F_P(s_P + \Phi) - F_P(\Phi)}{1 - F_P(\Phi)}\right] \\ &= \lambda_Q \frac{1 - F_P(s_P + \Phi)}{1 - F_P(\Phi)}\end{aligned} \qquad \ldots (9)$$

Equation (8) is also valid for rainfall, i.e.

$$\frac{1}{T_P(s_P)} = \lambda_P\left[1 - F_P(s_P)\right] \qquad \ldots (10)$$

so that, from the equivalence (6), one obtains,

$$\begin{aligned}&\lambda_Q\left[1 - F_Q(q_0)\right] = \lambda_P\left[1 - F_P(\Phi)\right] \\ &\Rightarrow \lambda_Q = \frac{\lambda_P\left[1 - F_P(\Phi)\right]}{\left[1 - F_Q(q_0)\right]}\end{aligned} \qquad \ldots (11)$$

that substituted in (9) and considering that $F_Q(q_0) = 0$ (because $s_Q = q_0$ corresponds to $s_P = 0$ in Eqn. (7)) gives,

$$T_Q(s_Q) = \frac{1}{\lambda_P\left(1 - F_P(s_P + \Phi)\right)} \qquad \ldots (12)$$

Under this hypothesis, the current abstraction $\Delta s(T)$ is constant with varying frequencies, independently of the cdf of the rain events. In fact, for a fixed T one can invert equations (10) and (12) and find,

$$s_P(T) = F_P^{-1}\left(1 - \frac{1}{\lambda_P T}\right) \qquad \ldots (13)$$

and

$$s_Q(T) = F_Q^{-1}\left(1 - \frac{1}{\lambda_P T}\right) - \Phi \qquad \ldots (14)$$

so that the resulting absorption is always,

$$\Delta s(T) = s_P(T) - s_Q(T) = \Phi \qquad \ldots (15)$$

that is consistent with finding the two distributions with the same shape.

Model 2: Linear Relationship

The second model ($\alpha < 1$ in equation (5)) is a slightly complicated version of the rational formula for the rainfall-runoff transformation (see Iacobellis and Fiorentino (2000) for a detailed description of the rationale behind this model). In this case the equivalent of equation (7) is,

$$F_Q(s_Q) = \frac{F_P\left(\frac{s_P}{\alpha} + \Phi\right) - F_P(\Phi)}{1 - F_P(\Phi)} \qquad \ldots (16)$$

and the resulting frequency of the floods events reads,

$$\begin{aligned} \frac{1}{T_Q(s_Q)} &= \lambda_Q\left[1 - F_Q(s_Q)\right] \\ &= \lambda_Q\left[1 - \frac{F_P(s_P/\alpha + \Phi) - F_P(\Phi)}{1 - F_P(\Phi)}\right] \qquad \ldots (17) \\ &= \lambda_P\left(1 - F_P\left(\frac{s_P}{\alpha} + \Phi\right)\right) \end{aligned}$$

The resulting current abstraction is found by inverting equations (10) and (17),

$$\begin{aligned} \Delta s(T) &= s_P(T) - s_Q(T) \\ &= \alpha\Phi + (1-\alpha) F_P^{-1}\left(1 - \frac{1}{\lambda_P T}\right) \qquad \ldots (18) \end{aligned}$$

The shape of the relationship (18) depends on the rainfall cdf, that can be assumed exponential, $F_P(s_P) = 1 - \exp[-s_P/\Theta]$. In this case $F_P^{-1}(s_P) = -\Theta \ln(1 - s_P)$ that gives with equation (18),

$$\begin{aligned} \Delta s(T) &= \alpha\Phi - (1-\alpha)\Theta \ln\left(1 - \left(1 - \frac{1}{\lambda_P T}\right)\right) \qquad \ldots (19) \\ &= \alpha\Phi - (1-\alpha)\Theta \ln\left(\frac{1}{\lambda_P T}\right) \end{aligned}$$

When $\Delta s(T)$ is plotted on a logarithmic axis versus the frequencies, the relation appears as a straight line (Figure 6a). This case is important because an exponential distribution for the magnitudes of the events correspond, in the hypothesis of Poisson distributed occurrences, to a Gumbel distribution of the annual maxima (e.g., Lang *et al.*, 1999). If $F_P(s_P)$ is a Generalized Pareto distribution, that corresponds to GEV distributed annual maxima when the occurrences are Poissonian, one has,

$$F_P(s_P) = 1 - \left[1 - \frac{k_P}{a} s_P\right]^{1/k_P} \qquad \ldots (20)$$

In this case $F_P^{-1}(s_P) = \frac{a}{k_P}\left(1 - (1 - s_P)^{k_P}\right)$, that in equation (18) gives,

$$\Delta s(T) = \alpha\Phi + (1-\alpha)\frac{a}{k_P}\left(1 - \left(\frac{1}{\lambda_P T}\right)^{k_P}\right) \qquad \ldots (21)$$

When the shape parameter k_P is greater than zero (negative skewness), equation (21) appears in the $(\ln(T), \Delta s)$ diagram as a reversed negative exponential, when $k_P < 0$ (positive skewness) one finds a positive exponential (see Figure 6b).

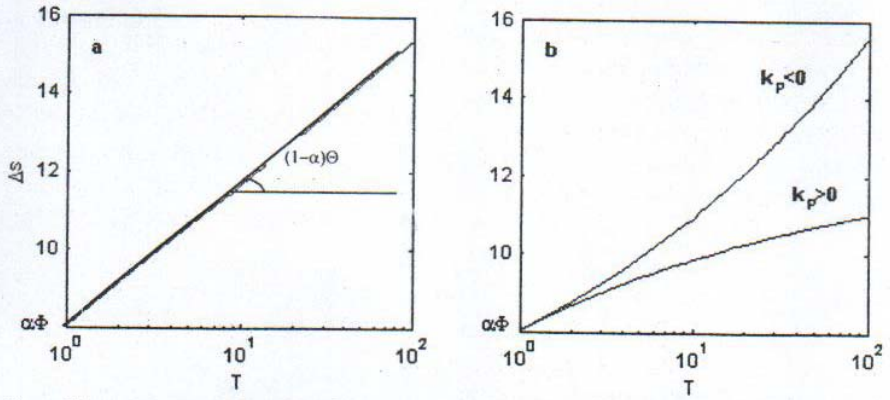


Fig. 6: Variation of the current abstraction Δs versus the return period T. In graph (a), $F_P(S_P)$ is assumed as exponential, in graph (b), $F_P(S_P)$ is a Generalized Pareto Distribution with positive or negative values of the parameter k_P

This implies that a linear model interpreting the rainfall-runoff transformation at the event-scale can capture the dynamics of the frequency curves of the basins that behave like the one in Figure 4. For these cases the distance Δs between the two curves increases with the return period. In Figure 6b, the behavior of Δs is reproduced both with positive and negative values of k_P (in the range observed for the square points in Figure 5). On the contrary, the proposed model does not help to interpret the mechanisms that lay behind cases like those of basins 6 and 31 of our sample.

CONCLUSIONS

In this paper, relations between rain and flood frequency curves is investigated using time series of daily data and a Filtered Peaks Over Threshold technique. The present initial approach suggests that mechanisms of rainfall runoff transformations related to the extremes can be categorized depending on the shape coefficients of the runoff and rainfall frequency curves. Data from 33 basins in North-Western Italy show that for the majority of cases the shapes are statistically indistinguishable. On the other hand, the number of basins for which there is a consistent reduction in skewness from rainfall to runoff (squares in Figure 5) is also significant. The high average elevation and the geographic location of these basins suggest that the reduction in skewness from rainfall to runoff can be due to the influence of snow accumulation and melting in the flood production mechanisms. For these basins a formal mathematical treatment of the rainfall-runoff transformation under simple linear assumptions is presented. The analysis proves able to justify the difference in shape of the frequency curves of rainfall and discharge when they diverge, thus eliminating the problem of matching the single rainfall and discharge peaks at the scale of the event. Few other basins, however, apparently elude the above simple conjectures.

Further investigations can include hydraulic mechanisms, such as discharge attenuation in large floodplains, and consideration of the behaviour of snow volumes, e.g. through rain-on-snow mechanisms.

ACKNOWLEDGEMENTS

This work has been co-funded by Regione Piemonte and Italian Ministry of Education (Prin CUBIST Project).

REFERENCES

Benjamin, J.R. and Cornell, C.A (1970). *Probability, Statistics, and Decision for Civil Engineers*. McGraw-Hill, New York.

Claps, P. and Laio, F. (2003). "Can continuous stremflow data support flood frequency analysis? An alternative to the partial duration series approach". *Water Resour. Res.*, 39(8), 1216, doi: 10.1029/2002WR001868.

Claps, P., Laio, F. and Villani, P. (2002). "Assessment of extreme flood production mechanisms through POT analysis of daily data". *International Conference on Flood Estimation*, CHR report II, 17, pp. 75–85, Fed. Off. For Water and Geol., Berne, Switzerland.

Franchini, M, Hashemi, A.M. and O'Connell, P.E. (2000). "Climatic and basin factors affecting the flood frequency curve: Part II—A full sensitivity analysis based on the continuous simulation approach combined with a factorial experimental design". *Hydrology and Earth System Sciences*, 4(3), 483–498.

Hashemi, A.M., Franchini, M. and O'Connell, P.E. (2000). "Climatic and basin factors affecting the flood frequency curve: Part I—A simple sensitivity analysis based on the continuous simulation approach". *Hydrology and Earth System Sciences*, 4(3), 463–482.

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

Lang, M., Ouarda, T.B.M.J. and Bobee, B. (1999). "Towards operational guidelines for overthreshold modelling". *J. Hydrol.*, 225, 103–117.

Madsen, H., Rasmussen, P.F. and Rosjbierg, D. (1997). "Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. At site modelling". *Water Resour. Res.*, 33(4), 747–757.

Onoz, B. and Bayazit, M. (2001). "Effect of the occurrence process of the peaks over threshold on the flood estimates". *J. Hydrol.*, 244, 86–96.