Analysis of peaks over threshold within daily data for flood frequency curve assessment

P. CLAPS[a] & P. VILLANI[b]

[a] *DIFA – Università della Basilicata – Potenza (Italy) [claps@unibas.it]*
[b] *DIC – Università di Parma (Italy) [villani@parma1.eng.unipr.it]*

ABSTRACT

We report an attempt to relate basic properties of stochastic processes of basin intense rainfalls and floods, in order to improve the prediction of flood risk by moving information from rainfalls to floods. This approach is based on a stochastic model of the extreme precipitation in which a wet and a dry season can be distinguished. During the wet period, the process of rainfall occurrences can be assumed as Poissonian and the distribution of marks can be derived trough the analysis of the peak over the threshold (POT) process. The basin operates on the rainfall process as a stochastic filter, constituted by a stochastic threshold *S* on precipitation, that reduces the number of runoff events. Theoretical analysis of the filter mechanism provides the relationships between stochastic parameters of maximum rainfall and flood. An application to data from a Southern Italy basin showed that magnitude of S is higher at low probability level of X and Y than at high probabilities, and the decay is rather steep. This finding seems to suggest a peculiar two-state behaviour for the basin antecedent moisture conditions: a *wet* state, in correspondence of annual maximum events, and a *dry* state for base process, both during the *wet* season (October-May).

1 INTRODUCTION

Peaks Over Threshold (POT) models have been frequently employed for the study of flood frequency analysis, even though with much less emphasis than the popular Annual Maximum Flood (AMF) sampling. Based on the applications available in literature, the POT approach seems to present under-employed potential and its reconsideration looks promising for the information content it can transfer into the flood analysis (see *e.g.* Lang et al., 1999; Sirangelo and Iiritano, 1999). In particular, one of the less underemployed features of the POT analysis is its capability to give information on the peak distribution and on the stochastic process of peak occurrence from the characteristics of processes derived from the cut operated by different thresholds.

The aim of this work is to analyse peak occurrence in daily data of rainfall and runoff and compare parameters of both POT processes in order to shed light on the probabilistic aspects of the transformation operated by the basin. In this analysis we consider a stochastic model simple yet consistent with the possibility of analysing the basic aspects of the transformation mechanism under a probabilistic viewpoint. In particular, we show that a Poisson model of occurrences and a quasi-homogeneous process structure over seasons and over return periods (no separation probabilistic structure) seem to meet this requirement.

The problem under study admits a feasible theoretical approach, based on the theory of derived distributions. According to this approach, it is possible to derive the peak discharge probability distribution from the analysis of the joint probability distribution of the rainfall peaks and of the variables describing basin transformation. Usually, the latter are considered as deterministic parameters (e.g., Bierkiens and Puente, 1990). However, basin antecedent soil moisture conditions are decisive in determining the magnitude of the peak discharge (see *e.g.* Todini, 1996).

In this paper, we will take explicitly into account the stochastic nature of some basin parameters and we will use a non-parametric technique to derive the stochastic structure of a basin parameter related to soil moisture from daily rainfall and discharge data. In particular, our attention is focused on the threshold over which rainfall peaks produce discharge peaks. We will show, as a result, the presence of a two-state behaviour, for low and high probability of rainfalls and floods. This finding, if confirmed on regional basis, could help explaining the relative role played by the basin antecedent moisture condition and the peak rainfall in determining very high flood peaks.

2 STOCHASTIC MODEL OF PEAK OCCURRENCES

2.1 – Mechanism of peak occurrences

From the analysis of peak occurrences in basins of the Mediterranean area it results evident that this stochastic process is not homogeneous in time. A first basic distinction can be made observing the process patterns in two different climatic seasons: a *dry season* during the summer period, and a *wet season* otherwise (Figure 1). This approximation looks reasonable and parsimonious as a starting point (see also Sirangelo, 1994).

As said above, we assume a Poissonian scheme for modelling time occurrences of flood peaks during the wet season, based on the observation that the exponential probability distribution of the inter-arrival time of the peaks can be valid as a first approximation. If necessary, one could also distinguish a finer time structure during the wet season.

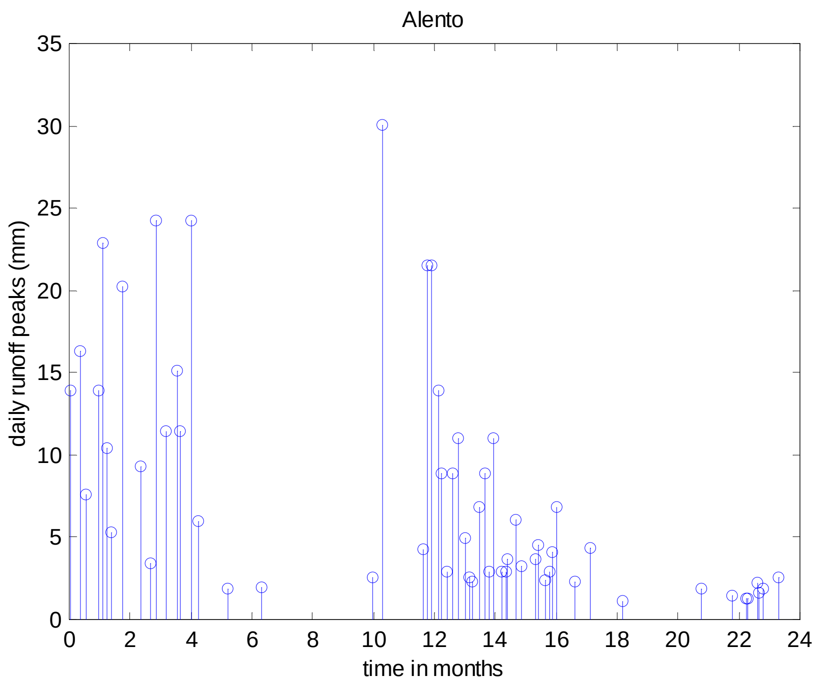

Figure 1. Process of occurrence of peaks of mean daily discharge.

2.2 – Probability distribution of peaks

As a consequence of the Poissonian hypothesis, if $\Lambda(x)$ is the mean annual number of peak occurrences over a threshold of magnitude x and $\Lambda(x_o)$ is the analogue for a value x_o low enough to have a cumulative probability $F_X(x_o) \approx 0$, but high enough to allow the Poissonian hypothesis to be valid, then the following relationship holds (see *e.g.* Çinlar, 1975, p. 94]

$$\Lambda(x)/\Lambda(x_o) = 1 - F_X(x) \qquad (x \geq x_o) \qquad (1)$$

where $F_X(x)$ is the cumulative distribution function (CDF) of the random variate X. In practice, the importance of the choice of x_o decreases as much as X becomes high enough to consider valid the Poissonian hypothesis on occurrences.

2.3 Statistical relation between rainfall and discharge POT processes

Let us consider:

X = total rainfall depth in 1 day, over the basin (in *mm*);

Y = total discharge volume in 1 day, from the basin outlet (in *mm*).

We assume Y to be the result of a deterministic/stochastic filter operated by the basin over the stochastic process of the daily rainfall. The transformation we consider is a simple non-linear, threshold model as

$$\begin{cases} Y = 0 & \text{if} \quad X \leq S \\ Y = g(X-S;\, \alpha) & \text{otherwise} \end{cases} \qquad (2)$$

where $g(\bullet;\alpha)$ is a general transformation function, that depends on the parameter vector α, and contains information on the basin characteristics. Filter parameters S and α can be both deterministic and stochastic. As a first approximation, we consider the basin characteristics α as deterministic, while the filter threshold S can take into account the stochastic state of the system preceding the intense rainfall generating the flood peaks.

Based on these assumptions, discharge peaks will be considered equivalent to over-threshold rainfalls ($Y=X-S$), producing a relationship between the annual rate of flood peaks Λ_Y and rainfall peaks Λ_X of the type of eq. (1), that gives here the probability of the (external) filter threshold S:

$$\Lambda_Y = \Lambda_X\,[1 - F_X(S)] \qquad (3)$$

Eq. (3) indicates that, for any transformation function $g(\bullet;\alpha)$, the mean annual number of flood peak occurrences can be derived through the knowledge of the probability distribution of rainfall peaks $F_X(\bullet)$ and of the mean annual number of rainfall peak occurrences, given a state of the system, i.e. a value of S. Both Λ_Y and Λ_X have a probabilistic structure that depends on the threshold level, as shown in (1). The control operated by the threshold filter *S* on the transformation processes will be then inferred by the ratio of the empirical occurrence parameters Λ_Y and Λ_X computed through POT analysis. In particular, using historical records of areal daily rainfall and of mean daily discharges we will show how it is possible to characterise the probability distribution of *S* and the relative role that its stochastic variability plays in producing high flood peaks.

3 POT ANALYSIS AND CASE STUDY

3.1 Criteria for identification of peaks

On daily discharge records, consecutive and different peak occurrences (in the statistical sense of independence) must be suitably identified. To this end, a standard procedure adapted from the Matlab© library, namely the `findpeaks` function [The Mathworks] was applied to runoff records, leading to the selection of *isolated* peaks (see Figure 2). Analysis of interarrival times during the wet season (Figure 3) showed that additional filtering was not necessary, at least at the present stage of analysis.

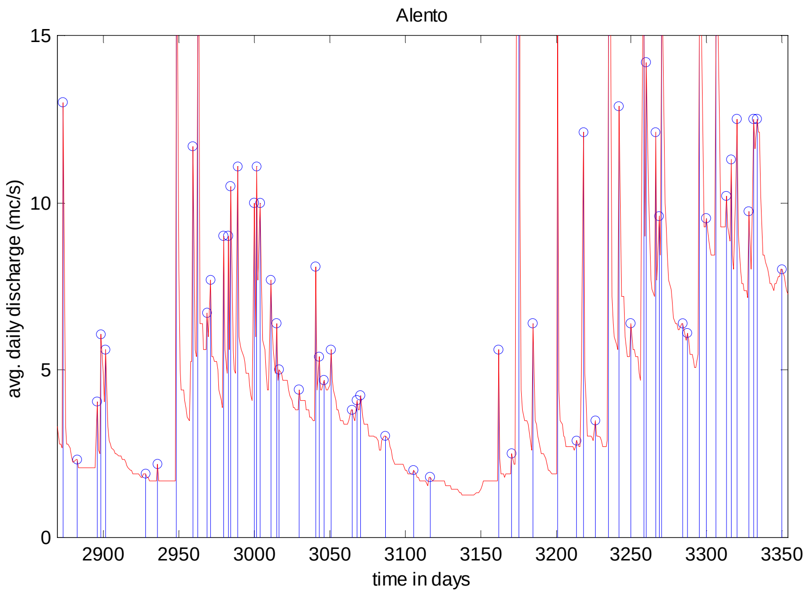

Figure 2. temporal process of peak discharge occurrence

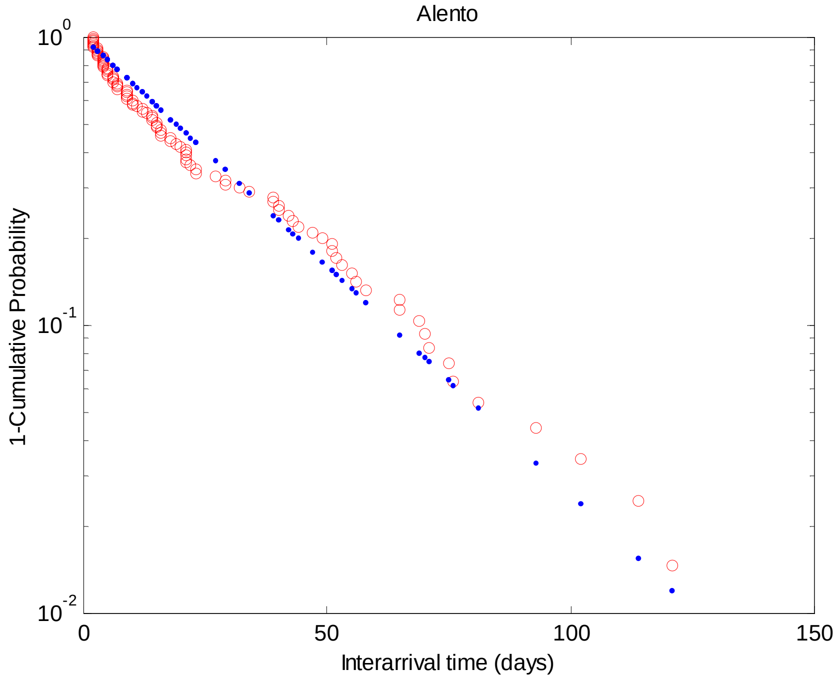

Figure 3. Probability distribution of the inter-arrival time of discharge peak occurrences during the wet season: estimates (circles) and fitted exponential distribution (dots).

3.2 Relation between filter threshold *S* and probability of the events

To identify the statistical relationship between stochastic filter threshold *S* and probability of the peak event we used the following procedure (graphically represented in Figures 4a and 4b):

(i) Given a probability $1 - F_X(X)$ for rainfall peaks, $\Lambda_o \cdot \Lambda_X / \Lambda_o$ gives an occurrence level Λ_X and a corresponding threshold X (Fig. 4a).

(ii) Since (3) is valid, we search a value of the discharge peak Y such that $\Lambda_X = \Lambda_Y = \Lambda$ (Fig. 4b).

(iii) The difference between X and Y values taken at the same Λ represents the estimate of S_X corresponding to the probability $1 - F_X(X)$.

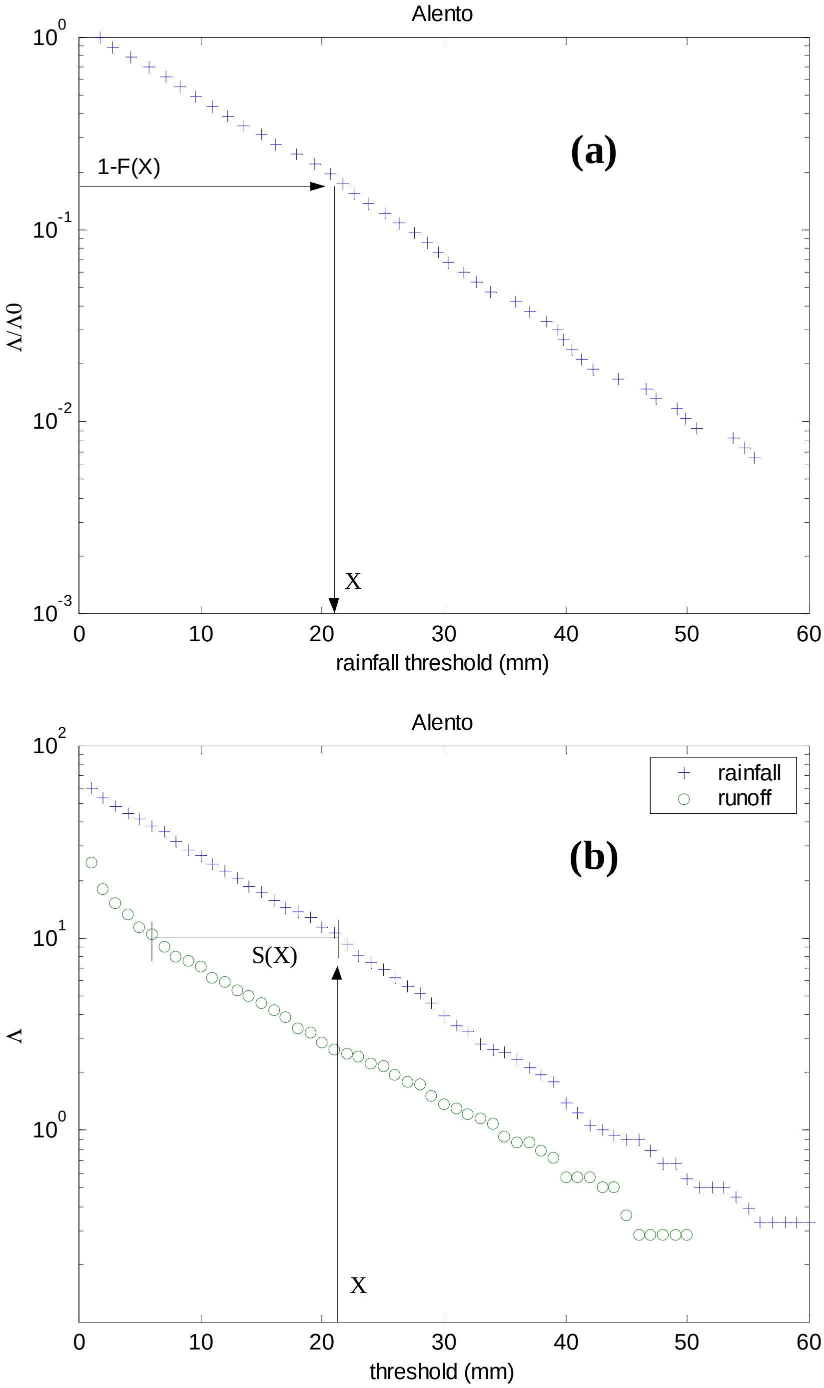

Figure 4. Sketch of the procedure of deriving the statistical relationship between the threshold *S* and the probability of the rainfall peak.

3.2 Basin description and application of the procedure

Areal rainfall and discharge data from a basin located in southern Italy have been analysed. The basin belongs to the Alento river, in Campania, has an area of 285 Km^2 and presents only a very small fraction of carbonate formations, so to allow us to classify it as almost completely impermeable. Based on the basin area, order of magnitude of discharges can be determined from the figures below considering that 1 *mm* of equivalent discharge measures about 3.3 m^3/s.

On the basin under study 13 years of daily average discharge and 20 years of daily rainfall were available. All of the figures presented above are referred to these hydrological records.

The procedure discussed in the above paragraph was applied to the curves of $\Lambda(X)$ (figure 4) obtained for different thresholds on rainfalls and discharges.

Applying the proposed procedure for probability levels ranging from 0.1 to over 0.99 we obtained the curve represented in Figure 5 , that shows a peculiar relationship between the probability level F of the rainfall and the stochastic threshold *S* of the basin. Note that using the representation in terms of 1-F, in reverse scale, provides a correct viewpoint in terms of recurrence interval increasing toward the right part of the scale.

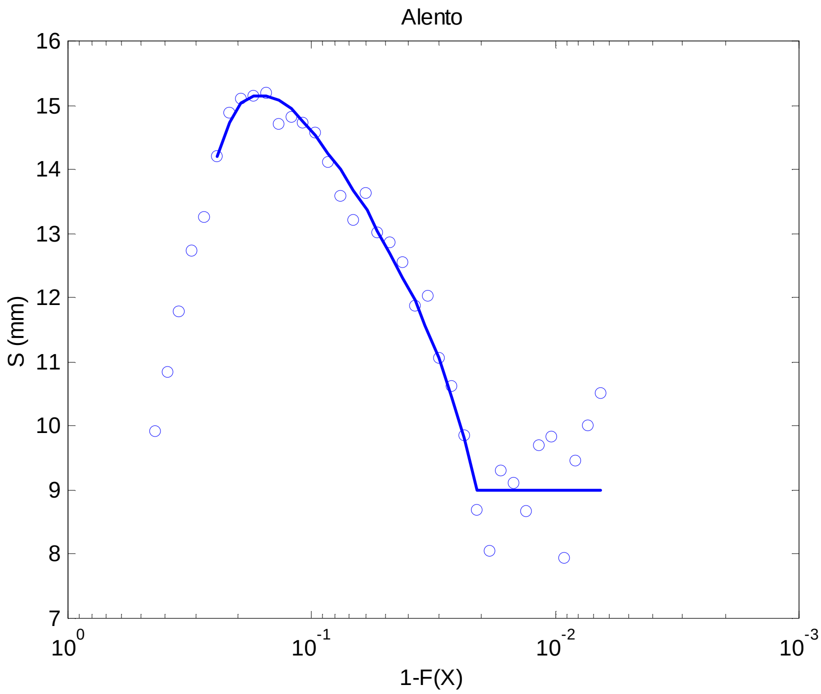

Figure 5. Statistical relationship between the threshold *S* and the probability of the rainfall peak for the Alento river

In the figure we can distinguish different zones:

i) very low value of F (*non-Poissonian zone*): we disregarded values obtained for $F<0.5$, that are of no interest for the goals of this application, and for which the Poissonian hypothesis in eq. (1) cannot be considered valid, even in first approximation. Moreover, F is the cumulative probability of the whole ensemble of rainfall and runoff peaks (not annual maximum), while annual maximum rainstorms and floods cover only the higher part of the CDF;
ii) low value of F (*dry state*): for low return period of rainfall peaks ($F\sim0.85$) the basin threshold S is relatively constant and high, being 14 mm, that is 36% of mean annual daily flood;
iii) high value of F (*wet state*): for high return period of rainfall peaks ($F>0.99$) the basin threshold S is relatively constant and low, being 8 mm, that is less than 60% of the value assumed in the dry state;
iv) intermediate value of F (*transition phase*): during this phase the threshold S is not a constant and must be considered a stochastic variable.

Based on this result, it seems possible to identify a two-state pattern that characterises the system transformation in dependence of the probability level of rainfalls and floods.

4. FINAL DISCUSSION

In order to identify the mechanism leading to flood peak distribution Y starting from rainfall peak distribution X, we considered a general framework in which the basin is a deterministic filter that operates through a stochastic threshold parameter *S*.

We did not address the theoretical derivation of the distribution of Y that considers the joint probability distribution of X and *S*. This distribution is difficult to derive from usual hydrological information, because the analytical approach has to face the dependence of *S* on antecedent moisture condition.

Our alternative approach is based on peak over threshold (POT) analysis of daily rainfall and discharge: under the simple, but realistic assumption of Poissonian occurrence of peaks, we showed how it is possible to derive a relationship between the magnitude of S and the probability level of X and Y.

Previous analyses considered only the annual maximum series (AMS) of daily rainfall and flood: Ferrari et al. [1993] considered the ratio Λ_Y/Λ_X, that can be considered a surrogate of S, only as a deterministic parameter of the basin response, constant over a climatic region. Recently, an attempt was made to explain the observed spatial variability of that ratio taking into account the permeability characteristic of the basin and a surrogate of the antecedent soil moisture condition considered constant and related to the Thorntwaite annual climatic index [Iacobellis et al., 1998]: no estimation of the probability level associated with this ratio can be made, because of the use of the AMS. The presented procedure, taking the advantage of using the POT techniques of daily rainfall and discharge, is the first attempt to explicitly estimate the relative role played by deterministic and stochastic parameter of the basin, in transforming heavy rain into flood peak discharge.

An application to a case study in a Southern Italy basin has shown a peculiar probabilistic structure of S: the magnitude of S is higher at low probability level of X and Y than at high probabilities, and the decay is quite steep. This findings seems to suggest a peculiar two-state behaviour for the basin antecedent moisture conditions: a *wet* state, in correspondence of annual maximum events, and a *dry* state for base process. It is important to consider that both states were found during the *wet* season (October-May). Other authors found a two-state basin behaviour, but in two different seasons (e.g. Grayson et al., 1997).

Such a mechanism could explain some extraordinary flood events experienced in the past in the same region of the case study and not justified by extraordinary rainfall peaks: inundations could have been produced by a combination of high rainfall peak and extraordinary wet condition of the basin, represented by a very low value of S. Further investigations are needed to evaluate the possibility that POT give information useful for the separation process.

Identification of the probability structure of S can be made in terms of: (i) magnitude characterising *wet* and *dry* state; (ii) magnitude of difference between the two states (or *jump*), and (iii) location, in probability, of the *break-points* between the two states. The ensemble of information resulting on S have suggested the possibility of a regionalization of the proposed procedure, whose goal is the consideration of this stochastic parameter as a function of basin hydro-geological characteristics: to this aim, further investigation are needed, in order to extend the different environments in which the procedure can be applied and to test and, eventually, relax, the base hypotheses of Poissonian processes.

6. REFERENCES

Bierkens, M.F. P. and C. E. Puente, Analytically Derived Runoff Models Based on Rainfall Point Processes, *Water Resour. Res.*, 1990, vol. 26, no. 11, pp. 2653-2665.

Çinlar, E., *Introduction to Stochastic Processes*, Prentice-Hall, Inc., Englewood Cliffs, N.J. (USA), 1975, pp. 402.

Iacobellis, V., Claps, P., & Fiorentino, M. (1998) Sulla dipendenza dal clima dei parametri della distribuzione di probabilità delle piene, *Proc. XIV Conf di*

Idraulica e Costruzioni Idrauliche, vol II, 213-224, CUECM, Italy, (in Italian).

Ferrari, E., S. Gabriele and P. Villani, Combined regional frequency analysis of extreme rainfalls and floods, *6th IAHS/IAMAP Symp.*, Yokohama (Japan), 1993.

Grayson R.B., A.W. Western and F.H.S. Chiew, Preferred states in spatial soil moisture patterns: local and non-local controls, *Water Resour. Res.*, 1997, vol. 33, no. 12, pp. 2897-2908.

Lang, M., T.B.M.J. Ouarda and B. Bobée, Toward operational guidelines for over-threshold modeling, *J. Hydrol.*, 225, 1999, pp. 103-117.

Sirangelo, B., Un modello Poissoniano non omogeneo per l'analisi delle precipitazioni cumulate su lunghi intervalli temporali, *Idrotecnica*, 3, 1994, pp. 125-158.

Sirangelo, B. and G. Iiritano, I processi puntuali nello studio delle occorrenza delle precipitazioni intense, *L'Acqua*, 6, 1999, pp. 35-60.

The MathWorks, Natick, MA (http://www.mathworks.com).

Todini, E., The ARNO rainfall-runoff model, *J. Hydrol.*, 175, 1996, pp. 339-382.