Regional analysis of runoff thresholds behaviour in Southern Italy based on theoretically derived distributions

M. Fiorentino¹, A. Gioia², V. Iacobellis², and S. Manfreda¹

¹Dipartimento di Ingegneria e Fisica dell’Ambiente, Università degli Studi della Basilicata, Italy
²Dipartimento di Ingegneria delle Acque e di Chimica, Politecnico di Bari, Italy

Received: 8 March 2010 – Revised: 24 April 2010 – Accepted: 30 August 2010 – Published: 1 February 2011

Abstract. The analysis of runoff thresholds and, more in general, the identification of main mechanisms of runoff generation controlling the flood frequency distribution is investigated, by means of theoretically derived flood frequency distributions, in the framework of regional analysis. Two nested theoretically-derived distributions are fitted to annual maximum flood series recorded in several basins of Southern Italy. Results are exploited in order to investigate heterogeneities and homogeneities and to obtain useful information for improving the available methods for regional analysis of flood frequency.

1 Introduction

Inferring the flood frequency distribution in ungauged basins is one of the most important goals in the hydrological applied research. In many areas of the world, as well as in some parts of Southern Italy, the absence of reliable discharge measurements or the limited length of observed flood time series still imposes to determine the design floods through the statistical analysis of rainfall distribution coupled with deterministic rainfall-runoff models. This choice is often facilitated by the availability of more dense networks and long lasting records of rainfall measurements. Besides this, since the years '80s and '90s in Southern Italy regional analysis for flood prediction and flood frequency analysis has been applied, using methods based on the index flood approach and regional growth curves on homogeneous areas (e.g. Rossi and Villani, 1992). Today, this kind of approach still represents the most robust methodology for flood prediction in ungauged basins but it shows two main strong limitations. The first one, which heavily affects the accuracy of prediction, is related to the intrinsic difficulty to find reliable and robust relationships for the index flood estimation based on available physical measures of basin features. The second one is related to the use of homogeneous areas whereas the physical spatial variability affecting the flood frequency distribution is neglected within homogeneous areas. The physical spatial variability of parameters, usually, is hard to detect by means of the available datasets of annual maximum series of flood peaks (AMFS), being masked by high sample variability. In this research we try to tackle this problem by analyzing, at the regional scale, the behavior of some key parameters of theoretically derived distributions. We use two different theoretically derived distributions, the IF (Iacobellis and Fiorentino, 2000) distribution and the TCIF (Two Component IF) distribution (Gioia et al., 2008). The two distributions are “nested” because the IF is a particular case of the TCIF distribution. In particular the TCIF is able to represent the occurrence of ordinary and extraordinary floods: when the second of these components is not present the IF is applied. The IF distribution has been already applied to basins of Puglia and Basilicata (Fiorentino and Iacobellis, 2001; Fiorentino et al., 2007), respectively South-Eastern and South-Central regions of Italy, while the TCIF distribution has been tested (so far) only in ten highly-skewed basins of Puglia, Basilicata and Calabria (South-Western region). In this paper we report results of the application of these distributions to thirty-three basins of the three regions and provide useful insights for the regional estimation of their parameters. The used AMFS cover a wide range of natural basins, different for climate, geomorphology, vegetation coverage, soil type and permeability.

The overall procedure is composed of three steps. First, we report values of all those parameters of the IF and TCIF distributions which are estimated without using AMFS. Among these are all (local and regional) parameters related to precipitation. These parameters are obtained by regional studies performed on annual maxima rainfall series. In the second step, the remaining unknown parameters are calibrated using...
AMFS, and the best distribution is selected between IF and TCIF. Being both distributions physically based, this procedure allows one to identify how many and which are the main components of the rainfall-runoff process playing as sub-models of the theoretical distribution. Finally, heterogeneities and homogeneities are analyzed in terms of spatial variability of the calibrated parameters.

2 Case studies and application

We applied the IF and TCIF distributions to thirty-three gauged basins in Southern Italy shown in Fig. 1 where the regions of Puglia, Basilicata and Calabria are highlighted. This sample of basins is extremely interesting for the variety of attributes that it displays (see Table 1). Basins have area ranging from 15 to 1657 km² and the series of annual maximum floods are characterized by highly variable skewness coefficient (0.08 < Cs < 3.18). The mean annual rainfall ranges from 600 mm in Puglia to more than 1800 mm in some parts of Basilicata and Calabria. The basin average Thornthwaite climatic index I (e.g. Thornthwaite, 1948), varies from −0.28 in Puglia, corresponding to semi-arid climate, to 1.66, corresponding to hyper-humid climate, in Calabria.

3 Model description and parameter estimation

3.1 The IF and TCIF distributions

The IF distribution (Iacobellis and Fiorentino, 2000) assumes that the exceedance probability function of the peak direct streamflow \( Q \), \( G_Q(q) \), is derived integrating, over the appropriate domain, the joint probability density function (pdf) of two stochastic variables (e.g. Eagleson, 1972): the source area contributing to runoff peak \( a \) (whose probability density function is assumed Gamma distributed with position parameter depending on the ratio, \( r \), of the expected contributing area to the basin area) and the runoff peak per unit of contributing area, \( u_a \). Both random variables are controlled by rainfall intensity, duration and areal extension, runoff concentration and hydrological losses.

The runoff peak per unit area, \( u_a \), is linearly dependent on the areal rainfall intensity (which is assumed Weibull distributed with shape parameter \( k \)) in a time interval equal to lag-time (\( r_A \)) with a constant routing factor equal to 0.7. The IF distribution assumes that both average rainfall intensity (\( E[i_{a,1}] \)) and average hydrologc loss (\( f_a \)) scale with contributing area \( a \) according to the following power law relationships:

\[
E[i_{a,1}] = E[i_{A,1}](a/A)^{-\varepsilon} \quad f_a = f_A(a/A)^{-\varepsilon'}
\]

where \( E[i_{A,1}] \) and \( f_A \) are the average rainfall intensity and the average hydrologic loss referred to the entire basin area \( A \), while \( \varepsilon \) and \( \varepsilon' \) are scaling exponents. Thus, under the hypothesis of Poissonian occurrence of independent annual maximum floods, the cumulative distribution function \( F_{Q_p}(q_p) \) of the annual maximum flood peak \( q_p = Q + q_o \), is derived, with \( q_o \) the base flow. Moreover the following relationship between the mean annual number of rainfall (\( \Lambda_p \)) and flood events (\( \Lambda_q \)) holds:

\[
\Lambda_q = \Lambda_p \exp \left( -f_A^k / E[i_{A,1}^k] \right).
\]  

Starting from the consideration that different mechanisms may arise, in any basin, with different frequency and magnitude (e.g., Sivapalan et al., 1990), Gioia et al. (2008) generalized the IF theoretical probability distribution introducing a two-component derived distribution called “Two Component IF” distribution (TCIF). In this new framework two different response types linked to different runoff thresholds-driven processes are identified:

- “L-type” (frequent) response, occurring when a lower threshold \( f_{a,L} = f_{A,L}(a_L/A)^{-\varepsilon_L} \) is exceeded, and responsible of ordinary floods likely produced by a relatively small portion of the basin \( a_L \).

- “H-type” (rare) response, occurring when a higher threshold \( f_{a,H} = f_{A,H}(a_H/A)^{-\varepsilon_H} \) is exceeded, and providing extraordinary floods mostly characterized by larger contributing areas \( a_H \).

The flood-peak contributing areas \( a_L \) and \( a_H \) are assumed, in analogy with the IF distribution, as Gamma distributed with expected contributing area respectively equal to \( r_L A \) and \( r_H A \).

In both mechanisms, the runoff threshold \( f_a \) is described by a simple power law with exponent \( \varepsilon' \) as in Eq. (1). Fiorentino and Iacobellis (2001) showed that such exponent provides an important signature of the expected behavior of the flood.
soil infiltration. In humid climates typically one finds $s' = 0$ that implies a constant infiltration rate, while $s' = 0.5$ is expected in dry climate where the threshold $f_a$ shows a storage behavior. These two different mechanisms were used by Gioia et al. (2008), working on a dataset that included two dry cases out of ten basins. They found $s_L = 0$ and $s_H = 0.5$ thus associating a constant infiltration behavior for the L-type response which develops in areas closer to the river network, and a storage behavior for the H-type response arising in areas far from the river network and, then, characterized by lower soil moisture in conditions antecedent the rainfall events.

Assuming that L-type and H-type events are independent and that both rates of occurrence are Poisson distributed, the overall process of exceedances is also a Poisson process and the cdf of the annual maximum floods is

$$F_Q(q_p) = \exp\{-\Lambda_L [G_{Q,L}(q_p)] - \Lambda_H [G_{Q,H}(q_p)]\}, \quad (3)$$

where $G_{Q,L}$ and $G_{Q,H}$ are the peak flow distributions corresponding to components of L-type and H-type events; $\Lambda_L$ and $\Lambda_H$ are respectively the mean annual number of independent flood events for L-type and H-type processes and are related to the runoff thresholds by means of the following relationships:

Table 1. Main features of the investigated basins.

<table>
<thead>
<tr>
<th>n.</th>
<th>Sample</th>
<th>Size (km²)</th>
<th>A</th>
<th>Cs</th>
<th>I</th>
<th>r</th>
<th>r_L</th>
<th>r_H</th>
<th>f_A (mm/h)</th>
<th>f_AL (mm/h)</th>
<th>f_AH (mm/h)</th>
<th>$\Lambda_q$</th>
<th>$\Lambda_L$</th>
<th>$\Lambda_H$</th>
<th>Best Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15</td>
<td>58</td>
<td>0.99</td>
<td>-0.28</td>
<td>-</td>
<td>-</td>
<td>0.03</td>
<td>0.59</td>
<td>-</td>
<td>6.04</td>
<td>10.37</td>
<td>-</td>
<td>2.38</td>
<td>0.74</td>
<td>TCIF</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>56</td>
<td>0.45</td>
<td>-0.25</td>
<td>0.19</td>
<td>-</td>
<td>-</td>
<td>7.78</td>
<td>7.76</td>
<td>27.08</td>
<td>3.90</td>
<td>3.86</td>
<td>0.06</td>
<td>IF</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>455</td>
<td>0.28</td>
<td>-0.27</td>
<td>0.12</td>
<td>-</td>
<td>-</td>
<td>1.69</td>
<td>1.69</td>
<td>4.82</td>
<td>6.10</td>
<td>5.65</td>
<td>0.45</td>
<td>IF</td>
<td></td>
</tr>
</tbody>
</table>

(Continued...)
\[ \Lambda_q = \Lambda_L + \Lambda_H = \Lambda_p \exp \left( -\frac{f_{A,L}^k}{E[i_{A,t}]} \right) \]  
(4a)

\[ \Lambda_H = \Lambda_p \exp \left( -\frac{f_{A,H}^k}{E[i_{A,t}]} \right) \]  
(4b)

The IF and the TCIF distributions are characterized respectively by twelve and fifteen parameters (see Gioia et al., 2008); nevertheless, for their estimation, based on current knowledge, only four parameters of the TCIF, \( r_L, \gamma_H, \Lambda_H, \Lambda_L \), and two parameters of the IF, \( r \) and \( \Lambda_q \), require to be calibrated by using the recorded AMFS. In facts, all the remaining parameters can be retrieved from information other than AMFS and principally from rainfall regional datasets.

### 3.2 Parameter estimation and results

The parameters estimated without using AMFS are common to the IF and TCIF distributions and were all available from previous studies (see Gioia et al., 2008; Fiorentino and Iacobellis, 2001; Iacobellis and Fiorentino, 2000; Claps et al., 2000).

The loss threshold scaling factor of the IF distribution \( \varepsilon' \) was assumed, following Fiorentino and Iacobellis (2001), \( \varepsilon' = 0 \) for humid basins \( (I > 0) \) and \( \varepsilon' = 0.5 \) for basins characterized by dry climate \( (I < 0) \). The two loss threshold scaling factors \( \varepsilon_L \) and \( \varepsilon_H \) of the TCIF were estimated following the regional procedure performed by Gioia et al. (2008). This procedure is based on the structural similarity of the TCIF and TCEV distribution (Rossi et al., 1984). Both models, in fact, arise as distributions of annual maxima of a Poisson compound process characterized by two independent base processes. Then, the TCEV includes two parameters \( \Lambda_1 \) and \( \Lambda_2 \) which represent the mean annual number of events of the ordinary and extraordinary components, respectively. Assuming, for one moment, that the L-type and H-type events correspond to the TCEV ordinary and extraordinary components, \( \Lambda_1 \) and \( \Lambda_2 \) can replace \( \Lambda_L \) and \( \Lambda_H \) in Eq. (4). Thus, from the available AMFS, we evaluated the at-site TCEV-ML (Two Component Extreme Value- Maximum Likelihood) estimates of \( \Lambda_1 \) and \( \Lambda_2 \) and assuming \( \Lambda_L = \Lambda_1 \) and \( \Lambda_H = \Lambda_2 \), we obtained some approximated estimates of \( f_{A,L} \) and \( f_{A,H} \), using Eq. (4) re-written as:

\[ f_{A,L} = \left[ E[i_{A,t}] \right] \log \left( \frac{\Lambda_p}{\Lambda_L + \Lambda_H} \right) \]  
(5a)

\[ f_{A,H} = \left[ E[i_{A,t}] \right] \log \left( \frac{\Lambda_p}{\Lambda_H} \right) \]  
(5b)

where, \( k, \Lambda_p, E[i_{A,t}] \) were estimated from the annual maximum rainfall series.

Observing the regional behavior of such estimates we found that for all basins, independently from climate, the higher threshold \( f_{A,H} \) scales with exponent \( \gamma_H = 0.5 \). The lower threshold, on the other hand, showed a strong dependence on climate being \( \varepsilon_L = 0 \) in humid basins \( (I > 0) \) and \( \varepsilon_L = 0.5 \) in dry basins \( (I < 0) \).

Then we used the AMFS for the estimation of parameters \( \Lambda_q \) and \( r \) of the IF distribution and \( \Lambda_L, \Lambda_H, r_L, r_H \) of the TCIF distribution adopting the at-site evaluation procedure which is described in Iacobellis et al. (2010), where for each river basin, the best parameter-dataset is chosen as the one maximizing the log-likelihood function of the observed sample of annual maximum floods. The selected parameter-datasets are reported in Table 1 for all basins while in Fig. 2 we display the TCIF-cdf, the IF-cdf and the Weibull plotting positions of the AMFS of the thirty-three basins in a Gumbel probability plot. It is worth noting that for twenty-one out of thirty-three basins the IF distribution was selected as the best-fitting while the others are TCIF distributed. Iacobellis et al. (2010) tested and discussed the suitability of different selection criteria based on the maximization of the log-likelihood function and accounting for model parsimony, by using a cross-validation technique for benchmark. They concluded, accordingly with Busemeyer and Wang (2000), that the use of a log-likelihood criterion, without any penalty factor accounting for model parsimony, is suggested when dealing with the IF and TCIF nested distributions and small sample size. It is also important to mention that, being the IF and TCIF nested distributions, the IF is selected whenever the parameter estimation procedure of the TCIF provides values of \( \Lambda_H \) which tend to 0, meaning that the TCIF curve collapses into the IF curve. From the analysis of data in Table 1 it is possible to observe that many of the TCIF-distributed series are characterized by a high skewness coefficient, but this is not a general rule. These results confirm that the presence of two different mechanisms of runoff generation may lead to highly-skewed distribution, nevertheless it is also possible to observe basins with two components and not particularly high skewness coefficient and it is also possible to have a high skewness coefficient in basins characterized by a single runoff component.

In Fig. 3 we show the scaling relationship of the runoff thresholds for all river basins: the red markers represent the values of \( f_{A,H} \) for dry basins (stars) and humid basins (squares); in blue we report the \( f_{A,L} \) values (stars) and the \( f_A \) values (circles) observed in dry basins; in green are shown the \( f_{A,L} \) values (squares) and the \( f_A \) values (circles) observed in humid basins. Four different patterns can be distinguished related to the considered components and the climatic conditions. These patterns are highlighted by the continuous lines representing power laws with exponent 0.5 (blue, red and magenta) and 0 (green). These scaling relationships are consistent with the behaviors above discussed for the scaling factors \( \varepsilon', \varepsilon_L \) and \( \varepsilon_H \). More important, they demonstrate that when comparing dry and humid basins the behavior of
the lower threshold is still different, although in both cases it is quite homogeneous and independent on the presence of a second higher threshold. The scaling behavior of the second higher threshold, with exponent 0.5 denotes a capacitive behavior related to water storage in the soil-vegetation package in both humid and dry basins. Nevertheless the soil storage capacity is higher for humid basins than for the dry ones. Also, for humid basins the first runoff threshold is identified as a constant infiltration rate (which may be related to the shallow groundwater drainage) while in dry basins it corresponds to a water storage capacity whose nature will be discussed in conclusions.

4 Conclusions

The performances of two nested theoretically derived distributions, IF and TCIF, were tested on several heterogeneous river basins of Southern Italy. Results show that, despite the statistical homogeneity which was claimed over the entire region using the flood index method (e.g. Fiorentino et al., 1987), a significant spatial variability can be detected by means of theoretically derived distributions. Almost half of the basins, and precisely twelve out of thirty-three, show evidence of two different mechanisms of runoff generation.
Results show that there is not a direct relationship between the presence of two components and a high skewness coefficient. Dynamics of different processes are summarized in Fig. 3 where the average hydrological losses during a flood event are plotted as a function of the basin area. This graph depicts the characteristics of different mechanisms of runoff generation. The behavior of humid basins was already assessed in other papers and in particular Gioia et al. (2008) recognized a saturation excess mechanism as responsible of the lower threshold and an infiltration excess mechanism for the higher threshold. On the other hand different interpretations and different real situations are feasible in dry basins. Results indicate that two different capacitive thresholds are responsible, in dry basins, for the lower and higher thresholds. These may be due to different patterns of vegetation, soil, geology and geomorphological factors. Also the presence of a saturation excess mechanism can be included among possible causes, although this one should be distinguished from the one occurring in humid climate. In humid climates the saturation excess mechanism occurs as a consequence of the watertable rise which is seasonally recharged. In dry basins this kind of mechanism may still arise when a shallow groundwater, originated by an impermeable bedrock layer underlying the river network, is filled during rain events. This conceptual interpretation is also consistent with the conceptual model developed by McGrath et al. (2007) which use a threshold storage for the saturation excess process. Further light about the physical interpretation of the observed spatial variability of flood frequency distribution may arise from the analysis of the estimates of the average ratios of contributing areas $r_L$ and $r_H$ and of the residual, local variability of $f_{A,L}$ and $f_{A,H}$ which is still object of research.

Acknowledgements. The authors gratefully acknowledge the support of MIUR (Italian Ministry of Instruction, University and Research) under the grant PRIN – Cubist-CoFin2007. The authors also thank R. Deidda (editor), and the anonymous referee for their constructive suggestions, that greatly helped improving the quality of the manuscript.

Edited by: R. Deidda
Reviewed by: one anonymous referee

References


