Research Papers
Issue RP0128
March 2012

Impacts on Soil and Coasts Division

From extreme areal rainfall to flood risk using a derived distribution approach with examples of climate change impacts

By **Renata Vezzoli**
Impacts on Soil and Coasts Division, CMCC
renata.vezzoli@cmcc.it

SUMMARY The aim of the report within GEMINA project is to present a methodology to probabilistically derive peak flood distribution from the statistical representation of extreme rainfall at the fixed durations of 1, 3, 6, 12, 24 hours over an area of interest. The methodology provides an analytical approach to derive the flood peak frequency distribution from the statistical characterisation of rainfall using a simplified description of the catchment response (SCS-CN model) to link rainfall and peak flood. The comparison between empirical and derived peak flood distribution evidences that the model performances have to be improved inserting a scaling factor dependent on the return period. The methodology is applied to the Baganza river basin in Emilia Romagna region (North Italy). Furthermore the methodology is used to investigate the effects of a positive and a negative variation of the average rainfall on peak flood distribution. Future applications will be based on climate and rivers discharge generated within the GEMINA project.

Keywords: Areal rainfall, Climate change, Derived distribution, Extremes, Peak flood

This report is carried on within a MOU agreement between CMCC and ARPA Emilia Romagna SIMC. Renata Vezzoli thanks ARPA Emilia Romagna SIMC personnel for support and collaboration. The research leading to these results has received funding from the Italian Ministry of Education, University and Research and the Italian Ministry of Environment, Land and Sea within the framework of Work Package 6.2.17 under the GEMINA project.

Introduction

The aim of this work is to define and test a procedure to estimate peak flood frequency distribution from the statistical characterisation of precipitation and the main features of a river basin. The methodology proposed would be one of the tools developed within the GEMINA project to evaluate, qualitatively and quantitatively the climate change effects on river discharge. The analysis can be made using a simulation method, see e.g. [1], that generates synthetic time series of flood events from known rainfall and derives the peak flood distribution frequency. In this case, a detailed description of the hydrological response of the basin should be provided. Alternatively, a derived distribution approach can be used. This approach is less demanding in terms of hydrological modelling parametrization and it allows to derive analytically the peak flood distribution. [6, 2] applied the distribution derived approach to watersheds in Liguria (North West Italy). Here a derived distribution approach is followed: rainfall is described through a regional model, hydrological response is modelled through the SCS-CN method, and a lumped model to transform rainfall excess into peak flood. Then the peak flood is scaled through a function of its return period. In the next sections the derived approach is described. Areal rainfall over for fixed area and duration is described by growth curve, intensity-duration curve, and areal reduction factor. The hydrological response is modelled through the SCS-CN method [8] to estimate the effective rainfall and a lumped model to convert effective rainfall into peak flood values. Since rainfall growth curve is described by a GEV distribution the peak flood distribution is also a GEV. Finally the application to Baganza river basin in Emilia Romagna, including the results under hypothetical climate change scenarios, is presented.

Derived peak flood distribution

The maximum rainfall depth, $h_T(d, A)$ in mm, for fixed duration, d in hours, assigned return period, T in years, over an area, A in km^2, is given by:

$$h_T(A, d) = a_1 d^{1-\nu} x_T(d) ARF(A, d) \tag{1}$$

where a_1 is the hourly average depth, ν is the scaling exponent, x_T is the dimensionless quantile related to the return period T and duration d, and ARF is the dimensionless areal reduction factor function of area, A, and duration.

The term

$$\eta_T(d) = a_1 d^{1-\nu} x_T(d) \tag{2}$$

gives an estimate of the maximum point rainfall depth for given duration d and return period T. Under the hypothesis that the rainfall depth is distributed as a GEV, the growth curve factor

$$x_T = \begin{cases} \epsilon_d - \alpha_d \left[\ln\left(-\ln\left(1-\frac{1}{T}\right)\right)\right], & \text{if } k_d = 0 \\ \epsilon_d + \frac{\alpha_d}{k_d}\left[1-\left(-\ln\left(1-\frac{1}{T}\right)\right)\right]^{k_d}, & \text{if } k_d \neq 0 \end{cases} \tag{3}$$

with $\epsilon_d > 0$ as position parameter, $\alpha_d > 0$ as scale parameter, and $k_d \leq 0$ as shape parameter. In this work, the GEV parameters are dimensionless because they are derived from normalised (respect to the average value) time series at the different durations, but their estimates depends on the duration. If the rainfall observations at the different durations are scale invariant a single set of ϵ, α, and k values independent from the duration could be estimated. The case $k = 0$ is a particular case where the GEV distribution reduces to a Gumbel distribution. If the area of interest is homogeneous, data from different raingauges can be pooled together to estimate the parameters over a wider sample. Combining and inverting Eq.(1) and Eq.(3) the cumulative density distribution of point rainfall is obtained. Figures 1 and 2 provide a comparison between the empirical cumulative distribution and the theoretical one at Lagdei and Calestano, respectively.

The areal reduction factor ARF transforms a point rainfall for a given duration and return period into the average areal rainfall, in mm, characterised by the same duration and return period, [3]. The ARF is defined as

$$ARF(A,d) = \left(1 + \omega \left(\frac{A^z}{d}\right)^b\right)^{-\nu/b} \quad (4)$$

where ω is a normalization factor and b a scaling exponent, for the mathematical derivation of ARF see [5, 7]. The estimate of the ARF parameters requires to aggregate at different spatial and temporal scale the observed data and to identify for each spatio-temporal scale the maximum rainfall. Once the matrix of area, time aggregation and maximum rainfall is available, the ARF parameters can be estimated. As result the areal rainfall is given by Eq.(1).
For simplicity, hereafter the duration d is assumed constant and equal to the basin concentration time, t_c, and the notation $_d$ is omitted.
The rainfall depth is transformed into rainfall excess using the SCS-CN method, [8]. According to SCS-CN the rainfall excess, h^*, is a function of rainfall depth, h,

$$h^* = \begin{cases} 0, & \text{if } h \le I_a \\ \frac{(h-I_a)^2}{h-I_a+S}, & \text{if } h > I_a. \end{cases} \quad (5)$$

where $S = 254(100/CN - 1)$, in mm, is the maximum soil potential retention, and $I_a = 0.2S$ is the rainfall lost as initial abstraction. CN is the curve number and depends on the soil type, the land use, and the antecedent moisture condition (AMC). The application is performed assuming the antecedent moisture condition II.

The peak flood, q in m^3/s, associated to a rainfall, h, is described by a simple lumped model

$$q = \begin{cases} 0, & \text{if } h \le I_a \\ \phi\frac{(h-I_a)^2}{h-I_a+S}, & \text{if } h > I_a. \end{cases} \quad (6)$$

where ϕ is a proportionality coefficient. The cumulative distribution of q can be derived by that one of h as

$$F_Q(q) = \begin{cases} \exp\left(-\exp\left(\frac{q+\sqrt{q^2+4\phi S q}+2\epsilon_q)}{2\alpha_q}\right)\right), & \text{if } k_q = 0 \\ \exp\left(-\left(1 - \frac{k\left[q+\sqrt{q^2+4\phi S q}+2\epsilon_q\right]}{2\alpha_q}\right)^{\frac{1}{k_q}}\right), & \text{if } k_q \neq 0 \end{cases} \quad (7)$$

where

$$\begin{cases} \alpha_q = \phi\alpha_p \\ \epsilon_q = \phi(\epsilon_p - I_a) \\ k_q = k \end{cases} \quad (8)$$

and

$$\begin{cases} \alpha_p = \alpha a_1 t_c^{1-\nu} ARF \\ \epsilon_p = \epsilon a_1 t_c^{1-\nu} ARF \end{cases} \quad (9)$$

for q large enough, i.e. if $\sqrt{q^2 + 4q\phi S} \simeq q$, Eq.(7) simplifies as

$$F_Q(q) = \begin{cases} \exp\left(-\exp\left(\frac{q-\epsilon_q}{\alpha_q}\right)\right), & \text{if } k_q = 0 \\ \exp\left(-\left(1 - \frac{k_q(q-\epsilon_q)}{\alpha_q}\right)^{1/k_q}\right), & \text{if } k_q \neq 0. \end{cases} \quad (10)$$

The estimates of $I_a(CN)$ and ϕ are crucial for the application of the methodology presented. Even if the CN parameter can be derived from land use maps and ϕ could be derived as ratio between basin area and concentration time, here they have been derived equating the mean and the variance of the flood peak from the derived distribution to the corresponding values of the sample. Figure 3 shows the comparison between data empirical, theoretical, and derived cumulative distributions: *empirical* is the Weibull plotting position, *theoretical* is the distribution fitted on peak flood observation, *derived* is the distribution with parameters estimates from Eq.(8). The bad fitting of the derived cumulative distribution is due to the hypothesis $k_q = k$, that, for Baganza river is not verified, see Table 1.

Table 1
Values of parameters of theoretical and derived peak flood distributions for Baganza river basin, α_q, ϵ_q are in m^3/s, k_q is dimensionless

	α_q	ϵ_q	k_q
Theoretical	66.82	109.205	-0.25
Derived	69.81	117.66	-0.07

To get ahead of this problem a correction factor β based on the return period, [4], is introduced

$$\beta(T) = \begin{cases} \beta_0 + \beta_1 \ln(10), & \text{if } T \leq 10 \\ \beta_0 + \beta_1 \ln(T), & \text{if } T > 10. \end{cases} \tag{11}$$

The parameters β_0 and β_1 have been estimated from the ratio between theoretical and derived quantiles of the peak flood. The functional form of β is valid also for other basins in Emilia Romagna.

The β function allows to correct the previous estimates of q_T as $q_{T,\beta} = \beta(T)q_T$.

Case study

The Baganza river basin closed at the river section of Ponte Nuovo (Parma) is 177 km^2 and the concentration time is 7.3 hours. Inside the area of interest observations of maximum rainfall at 1, 3, 6, 12, 24 hours are available at Lagdei, period 1994-2010, and Calestano, period 1995-2010. Figures 1 and 2 show rainfall empirical and theoretical distributions. The theoretical distribution is obtained combining and inverting Eq.s (2) and (3) for the different duration, the values of α_d, ϵ_d, and k_d are given in Table 2. The parameter a_1 is 45.1 mm/h$^{(1-\nu)}$ for Lagdei and 29.9 mm/h$^{(1-\nu)}$ at Calestano, ν is equal to 0.53 for both sites.

Table 2
Values of α_d, ϵ_d, and k_d for Baganza river basin, all the parameters are dimensionless

	1 hr	3 hr	6 hr	12 hr	24 hr
α_d	0.30	0.28	0.26	0.25	0.26
ϵ_d	0.82	0.82	0.83	0.83	0.84
k_d	-0.019	-0.077	-0.073	-0.077	-0.031

These values, together with the estimates of ARF parameters ($\omega = 0.0003$, $z = 7.345$, and $b = 0.144$), have been estimated by ARPA Emilia Romagna in a study on rain-intensity-duration-area-frequency relationships from event maxima [7].
Since the concentration time of the Baganza basin is about 7.3 hours, the estimates at 6 hours of the GEV parameters have been used to derive the peak flood distribution parameters from Eq. (8). Peak flood data at Ponte Nuovo are available from 1980 to 2008, the average and the standard deviation of the sample are used to estimate $I_a(CN)$ and ϕ such as the moments of the derived distribution are equal to the sample ones. The parameters values are $I_a = 36.67$ mm for $CN = 58$ and $\phi = 2.59$ m^3/mm/s. The estimates values of β_0 and β_1 are 0.68 and 0.16 respectively. Figure 3 shows the comparison between data empirical, theoretical, and derived cumulative distributions while Figure 4 shows the comparison between theoretical, derived, and β-derived cumulative

distributions.

The derived approach can be used to evaluate the effects of variations in rainfall average on the peak flood distribution. The following scenarios are considered: scenario^0 characterised by the observed average, scenario$^-$ that assumes a reduction of 30% in average rainfall, and scenario$^+$ where average rainfall increases of 30%. Figure 5 compares the $\beta-$derived distributions obtained under the three scenarios, while Table 3 summarized the quantile for 10 and 100 years return periods. The flood peak shows a decrease of 30% in scenario$^-$ for both the return period and in scenario$^+$ an increase of the 24% for $T = 10$ years and of 30% for $T = 100$ years. The return period of the quantile $q^0_{10} = 302$ m^3/s becomes 24 years in scenario$^-$ and 5 years in scenario$^+$, while for the 100 years quantile $q^0_{100} = 702$ m^3/s becomes 334 years in scenario$^-$ and 44 years in scenario$^+$.

Table 3
Quantiles, in m^3/s, of the $\beta-$derived distribution for scenario^0, scenario$^-$, and scenario$^+$ for return periods 10 and 100 years.

	10 yr	100 yr
scenario^0	302	702
scenario$^-$	211	491
scenario$^+$	375	912

Conclusion

Derived distribution approach is a powerful tool to derive peak flood distribution from rainfall and basin features, but its performances depends on the reliability of the hypothesis $k_q = k$ as shown in the case study. At the same time once the parameter are correctly calibrated the approach is valid to evaluate climate change effects on peak flood. The foreseen variation in rainfall statistics over a region can be directly inserted into the model just modifying the values of α_p, ϵ_p, and eventually k such as the moments of the GEV distribution fit the foreseen statistics to obtain with a minimal computational load, the new flood peak quantiles. The climate change scenario used in the case study are just a simple example of application. Applications within GEMINA project will include the analytical derivation of peak flood distribution from rainfall statistics extracted by climate scenarios and the comparison with synthetic flow discharge timeseries simulated using the same scenarios.

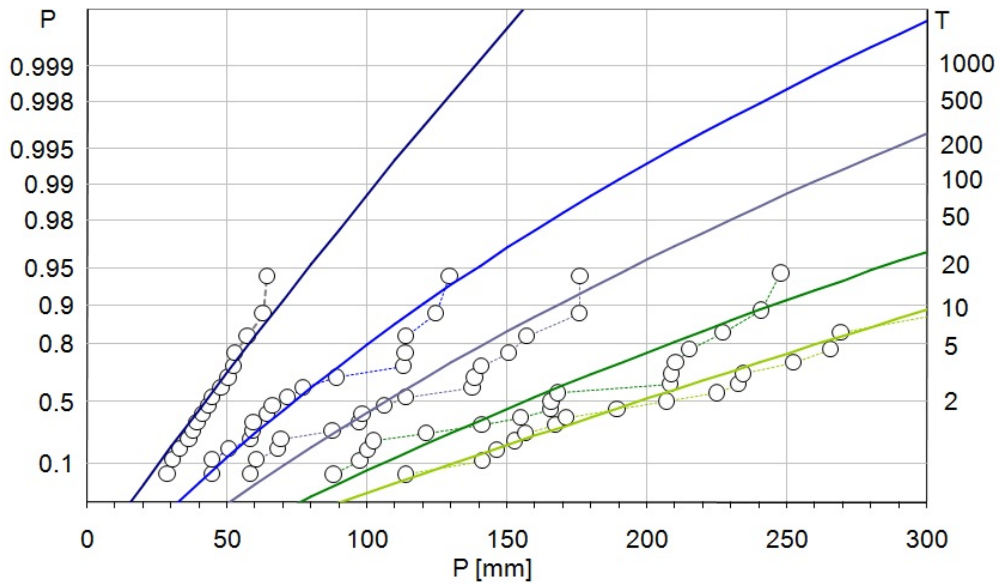

Figure 1:
From left to right comparison between maximum rainfall observed at 1, 3, 6, 12, and 24 hours empirical (o) and theoretical distribution (lines) at Lagdei raingauge station. Return period, T, is in years.

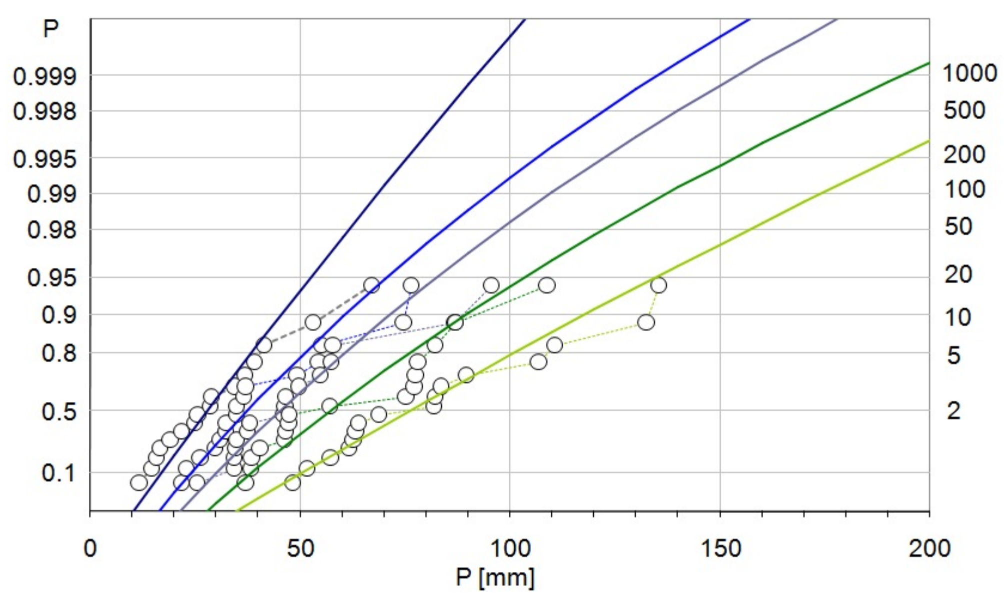

Figure 2:
From left to right comparison between maximum rainfall observed at 1, 3, 6, 12, and 24 hours empirical (o) and theoretical distribution (lines) at Calestano raingauge station. Return period, T, is in years.

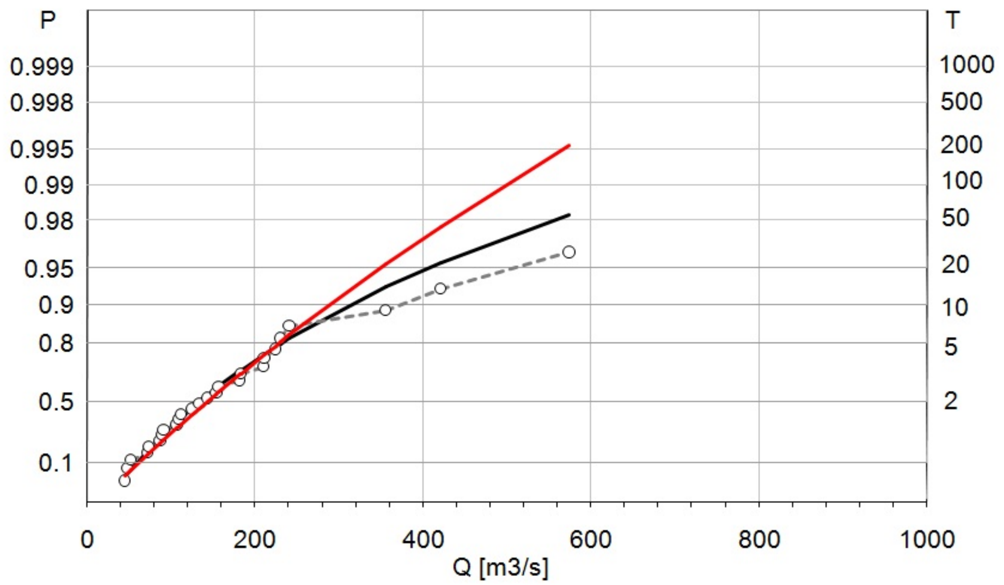

Figure 3:
Comparison between peak flood data plotting position (o), theoretical (black line), and derived (red line) distributions. Return period, T, is in years.

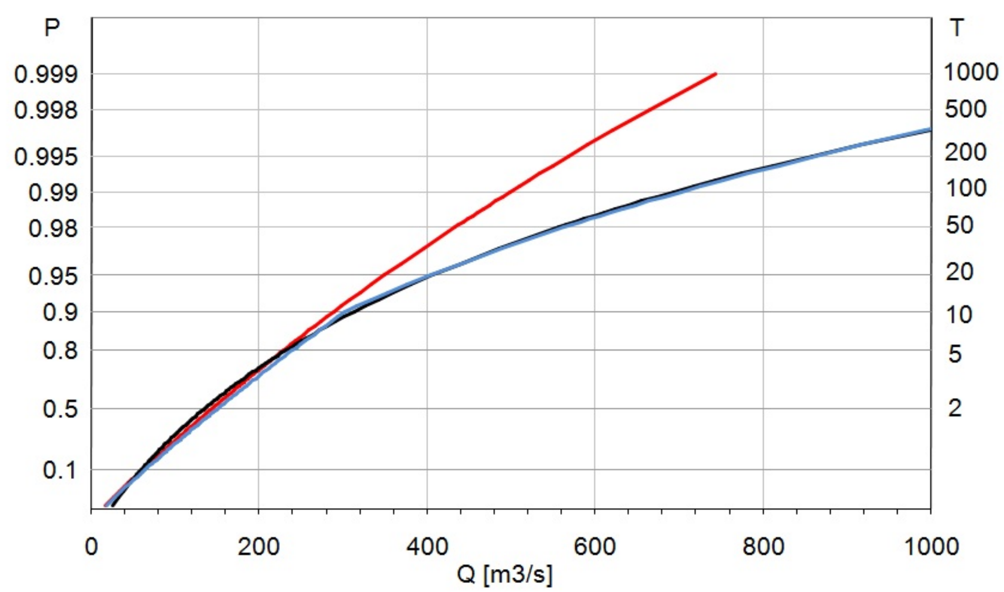

Figure 4:
Comparison among theoretical (black line), derived (red line), and $\beta-$derived (blue line) distributions. Return period, T, is in years.

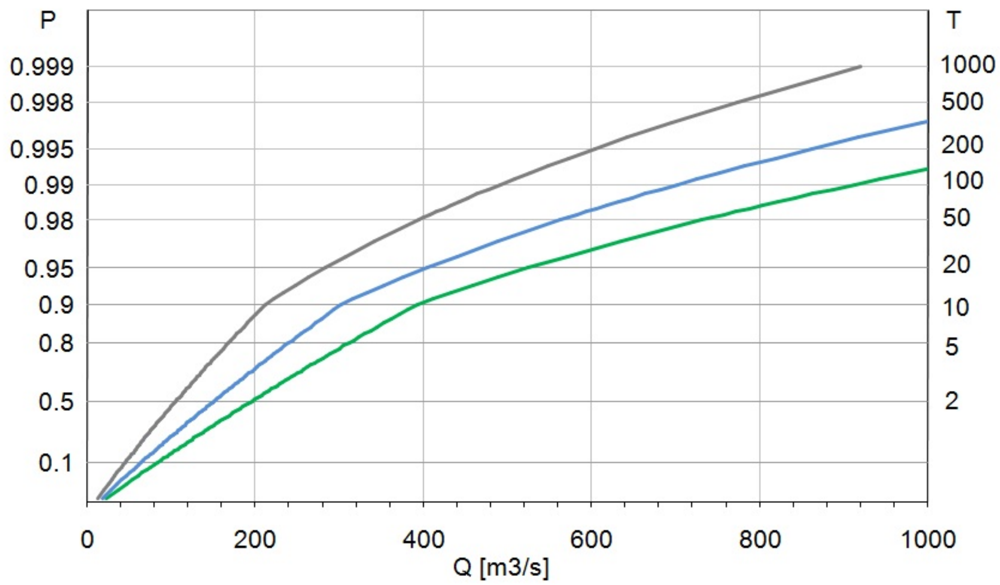

Figure 5:
$\beta-$derived distribution functions under scenario^0 (blue line), scenario$^-$ (grey line), and scenario$^+$ (green line). Return period, T, is in years.

Bibliography

[1] G.T. Aronica and A. Candela. Derivation of flood frequency curves in poorly gauged mediterranean catchments using a simple stochastic hydrological rainfall-runoff model. *Journal of Hydrology*, 347(1-2):132–142, 2007.

[2] D. Bocchiola and R. Rosso. Use of a derived distribution approach for flood prediction in poorly gauged basins: A case study in italy. *Advances in Water Resources*, 32(8):1284–1296, 2009.

[3] Natural Environmental Research Council. *Flood studies report, II, Meteorological studies*. Swindon, England, 1975.

[4] Iowa State University Centre for transportation Research and Education. *Iowa storm water management manual*, 2008.

[5] C. De Michele, N.T. Kottegoda, and R. Rosso. The derivation of areal reduction factor of storm rainfall from its scaling properties. *Water Resources Research*, 37(12):3247–3252, 2001.

[6] C. De Michele and G. Salvadori. On the derived flood frequency distribution: analytical formulation and the influence of antecedent soil moisture condition. *Journal of Hydrology*, 262(1-4):245–258, 2003.

[7] C. De Michele, E. Zenoni, S. Pecora, and R. Rosso. Analytical derivation of rain intensity-duration-are a-frequency relationships from event maxima. *Journal of Hydrology*, 399(3-4):385–393, 2011.

[8] Soil Conservation Service US Department of Agriculture. *National Engineering Handbook - Hydrology*. Washington DC, USA, 1986.

Introduction

The aim of this work is to define and test a procedure to estimate peak flood frequency distribution from the statistical characterisation of precipitation and the main features of a river basin. The methodology proposed would be one of the tools developed within the GEMINA project to evaluate, qualitatively and quantitatively the climate change effects on river discharge. The analysis can be made using a simulation method, see e.g. [1], that generates synthetic time series of flood events from known rainfall and derives the peak flood distribution frequency. In this case, a detailed description of the hydrological response of the basin should be provided. Alternatively, a derived distribution approach can be used. This approach is less demanding in terms of hydrological modelling parametrization and it allows to derive analytically the peak flood distribution. [6, 2] applied the distribution derived approach to watersheds in Liguria (North West Italy). Here a derived distribution approach is followed: rainfall is described through a regional model, hydrological response is modelled through the SCS-CN method, and a lumped model to transform rainfall excess into peak flood. Then the peak flood is scaled through a function of its return period. In the next sections the derived approach is described. Areal rainfall over for fixed area and duration is described by growth curve, intensity-duration curve, and areal reduction factor. The hydrological response is modelled through the SCS-CN method [8] to estimate the effective rainfall and a lumped model to convert effective rainfall into peak flood values. Since rainfall growth curve is described by a GEV distribution the peak flood distribution is also a GEV. Finally the application to Baganza river basin in Emilia Romagna, including the results under hypothetical climate change scenarios, is presented.

Derived peak flood distribution

The maximum rainfall depth, $h_T(d, A)$ in mm, for fixed duration, d in hours, assigned return period, T in years, over an area, A in km^2, is given by:

$$h_T(A, d) = a_1 d^{1-\nu} x_T(d) ARF(A, d) \qquad (1)$$

where a_1 is the hourly average depth, ν is the scaling exponent, x_T is the dimensionless quantile related to the return period T and duration d, and ARF is the dimensionless areal reduction factor function of area, A, and duration.

The term

$$\eta_T(d) = a_1 d^{1-\nu} x_T(d) \qquad (2)$$

gives an estimate of the maximum point rainfall depth for given duration d and return period T. Under the hypothesis that the rainfall depth is distributed as a GEV, the growth curve factor

$$x_T = \begin{cases} \epsilon_d - \alpha_d \left[\ln\left(-\ln\left(1-\frac{1}{T}\right)\right)\right], & \text{if } k_d = 0 \\ \epsilon_d + \frac{\alpha_d}{k_d}\left[1-\left(-\ln\left(1-\frac{1}{T}\right)\right)\right]^{k_d}, & \text{if } k_d \neq 0 \end{cases} \qquad (3)$$

with $\epsilon_d > 0$ as position parameter, $\alpha_d > 0$ as scale parameter, and $k_d \leq 0$ as shape parameter. In this work, the GEV parameters are dimensionless because they are derived from normalised (respect to the average value) time series at the different durations, but their estimates depends on the duration. If the rainfall observations at the different durations are scale invariant a single set of ϵ, α, and k values independent from the duration could be estimated. The case $k = 0$ is a particular case where the GEV distribution reduces to a Gumbel distribution. If the area of interest is homogeneous, data from different raingauges can