

Theoretical derivation of the index flood

A. Gioia¹, M. Fiorentino², V. Iacobellis¹, and M. R. Margiotta²

¹Dipartimento di Ingegneria delle Acque e di Chimica, Politecnico of Bari, Italia

²Dipartimento di Ingegneria e Fisica dell'Ambiente, University of Basilicata, Italia

Received: 17 November 2004 – Revised: 27 May 2005 – Accepted: 6 June 2005 – Published: 20 July 2005

Abstract. The expected value of the annual flood is always needed for attaining reliable estimates of the return period of floods. In particular, the accuracy of its estimate is crucial for ungauged basins, where regional analysis is to be invoked.

The main goal of this research was to search for an index flood formula suitable to explicate the influence of basin characteristics such as geomorphology, vegetation, soil and climate. This formula, due to unavoidable overparameterization, more than being suggested for technical use, allows sensitivity analyses of simpler formulae to basin features.

This study was mainly based on the theoretical distribution of floods developed by Iacobellis and Fiorentino (2000). In particular, the relationship that they derived for the first order moment was analyzed and simplified by assuming the Gumbel distribution as a first order approximation for the distributions of floods and rainfall intensities. The results were validated with regard to data recorded in some basins located in Southern Italy, in a climatically and geologically heterogeneous territory.

1 Introduction

In regional methods for flood prediction, the estimation of the index flood is a crucial point in order to attain reliable predictions in ungauged basins.

Nevertheless, in spite of a huge amount of research work devoted to develop methods for achieving accurate and unbiased estimations of high order distribution moments, much less attention has been paid to reduce uncertainty in the estimation of the expected flood. If this is not a relevant problem when even a few annual maximum flood observed values are available, it remains the weakest factor in regionalization procedures. Moreover, as shown by Fiorentino et al. (1987b), the lower the order of the distribution moment the higher is its spatial variability. Then, the more is the physical con-

trol on the moment variability and the less is the information content of regional hydrometric data.

The great majority of index flood formulae proposed in the literature are more or less empirical, and include coefficients (often estimated by regression analyses) thought to account for complex phenomena such as routing effects and soil losses.

Empirical models are often used to relate the index flood to physical characteristics of rivers basins such as climate, geomorphology, and land use. In such cases, simple or multiple linear regression models are generally exploited to estimate the parameters of the selected relationships (e.g. Versace et al., 1989; Franchini and Galeati, 1999). A comparison of some of those methods has been carried out by Grover et al. (2002), who analyzed the use of Ordinary Least Squares, weighted Least Squares, Generalized Least Squares methods in linear regression models. They compared also classic geographical regionalization methods to a “region of influence” approach (Burn, 1990), founding this last method less biased.

Amongst the other index flood estimation procedures, the rational formula, based on the assumption that rainfall is constant in space and time is probably the most used. Rossi and Villani (1994), Claps et al. (1994), Versace et al. (1989) proposed rational formulae with coefficients of physical soundness and Brath et al. (1992) developed a model based on the SCS-CN method.

In this work an attempt is made to analyze the control that physical and hydrological basin features play on such coefficients. In particular, a theoretical analysis is presented, following the approach of Iacobellis and Fiorentino (2000), which allows to identify the physical meaning of both coefficient and exponent of a power law relationship widely used to relate index flood to the basin area. The original probabilistic model, was simplified by assuming the Gumbel distribution as a first order approximation for the annual maxima of both flood and rainfall intensity (in a duration equal to the basin lag time) probability distributions.

2 Index flood estimation and scaling

The theoretical probability distribution of floods proposed by Iacobellis and Fiorentino (2000), that will be hereafter indicated as I&F model, is based on the hypothesis that the peak of direct streamflow Q is the product of two random variables: the runoff per unit contributing area u_a and the portion a of the basin area A that contributes to direct runoff. Indeed, these quantities are controlled by intensity, duration, and size of storm, basin routing factors and hydrological losses. In particular, the peak discharge per unit contributing area u_a , is linearly related (by means of a constant routing factor ξ) to the excess rainfall fallen on the contributing part a of the basin. To that end, a fixed rainfall duration, equal to a characteristic lag time τ_a , is used and this lag time is related to the contributing area a by a power law. Thus, the runoff formula is:

$$u_a = \xi(i_{a,\tau} - f_a) \quad (1)$$

where $i_{a,\tau}$ is the rainfall intensity on the area a during the duration τ and the parameter f_a represents the average loss rate within the duration τ_a ; it takes into account both initial absorption and infiltration and it is considered independent from rainfall intensity.

Therefore, the index flood formula derived in I&F model is:

$$Q_I = \xi \left[i_1 \frac{\gamma (1 + 1/k, \log(\Lambda_p/\Lambda_q))}{\Gamma(1 + 1/k)} - f_1 \right] \Lambda_q S_{\Lambda_q} r^{1-\varepsilon} A^{1-\varepsilon} + q_o \quad (2)$$

where i_1 and ε are the coefficients of a power law relationship between rainfall intensity ($i_{a,\tau}$) and contributing area a , f_1 is the coefficient of a power law relationship between average loss rate (f_a) and contributing area a , Λ_p/Λ_q is the ratio between mean annual number of independent rainfall events and independent flood events, r is the ratio of the expected contributing area to the basin area, q_o is base flow, k is the shape parameter of the weibull distribution assumed to describe rainfall intensity and $\Lambda_q S_{\Lambda_q}$ is a factor that allows to elapse from the mean of the base process to the mean of maxima.

The simplification adopted in this work consists of assuming the probability distribution of $i_{a,\tau}$ conditional on a to be an exponential law. Then, the average peak flow can be expressed as:

$$\mu_q = E[Q] = E[u_a a] = E[u_{E[a]}] E[a] \quad (3)$$

and the expected peak unit runoff becomes:

$$\begin{aligned} E[u_a] &= E[\xi(i_{a,\tau} - f_a) | i_{a,\tau} > f_a] \\ &= \xi \int_{f_a}^{\infty} i_{a,\tau} g_{i_{a,\tau} > f_a}(i_{a,\tau}) di_{a,\tau} \end{aligned} \quad (4)$$

where, as mentioned before, the exponential function is the probability density function of $i_{a,\tau}$ conditional on rainfall

values greater than f_a . As for the compound Poisson processes the expectation of exceedances is independent of the threshold, describing the expected value of contributing area as a fraction of total area by means of a factor r , $E[a] = rA$, and taking into account the routing factor ξ , the Eq. (3) becomes:

$$\mu_q = \xi E[i_{rA}] r A \quad (5)$$

In addition, for the properties of Poisson compound processes based on an exponential distribution of the base process, the expectation Q_I of the annual maximum flood is a simple function of the expected peak flow μ_q , and is given by:

$$Q_I = \mu_q (\ln \Lambda_q + 0.577) \quad (6)$$

in which Λ_q is the mean annual number of independent flood events correspondent to the process of exceedances with a null threshold.

From Eqs. (5) and (6), and taking into account the base flow q_o , we obtain:

$$Q_I = \xi (\ln \Lambda_q + 0.577) E[i_{rA}] r A + q_o \quad (7)$$

Accordingly, under similar assumptions, the expectation of the rainfall intensity in a duration equal to the basin lag-time τ_A can be related to the expected value of its annual maximum, I_{rA} , by:

$$E[i_{rA}] = \frac{I_{rA}}{\ln \Lambda_p + 0.577} \quad (8)$$

where I_{rA} can be calculated through the relationship $I_{rA} = p_1 \tau_{rA}^{n-1}$, in which p_1 and n are the IDF Intensity Duration Frequency parameters for the expected annual maximum rainfall. The same relationship holds for any contributing area $a \leq A$ by assuming that the lag time τ_a of the contributing area a is given by $\tau_a = \tau_1 a^v$ with parameters τ_1 and v defined as in Fiorentino and Iacobellis (2001).

From (7) and (8) and taking into account the rainfall intensity areal reduction factor κ (e.g. U.S. Weather Bureau, 1958), it follows:

$$Q_I = \xi \kappa \frac{I_{rA}}{\ln \Lambda_p + 0.577} (\ln \Lambda_q + 0.577) r A + q_o \quad (9)$$

in which Λ_p is the mean annual number of independent rainfall events. For values of $\Lambda_q/\Lambda_p > 0.023$, Eq. (9) can be approximated as:

$$Q_I = \xi \kappa \left(1 + \frac{4}{15} \ln \left(\frac{\Lambda_q}{\Lambda_p} \right) \right) I_{rA} r A + q_o \quad (10)$$

The above relationship depends on climatic and geomorphologic river characteristics through the routing factor ξ , the parameters Λ_p and Λ_q , the expected annual maximum rainfall referred to the expected value of contributing area rA , the ratio r and basin area A .

Briefly, Eq. (10) can be expressed as:

$$Q_I = c A^{1-\theta} + q_o \quad (11)$$

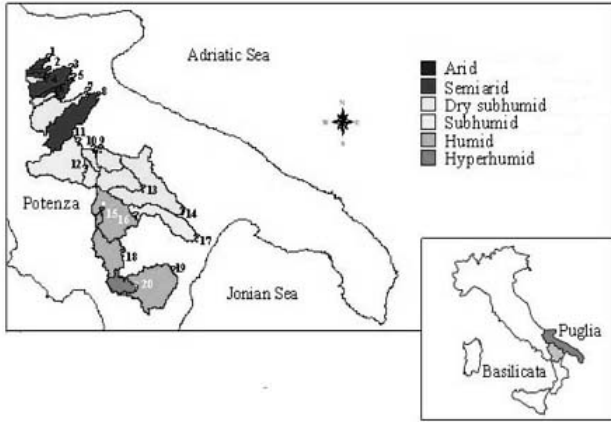


Fig. 1. Investigated basins in Puglia and Basilicata.

with

$$c = \psi \kappa \xi p_1 \tau_1^{n-1} r^{1-\theta} \left(1 + \frac{4}{15} \ln \frac{\Lambda_q}{\Lambda_p} \right) \text{ and } \theta = -\nu(n-1) \quad (12)$$

Thus, a link between Eq. (11) and the classical scaling relationship relating the expected flood to the basin area is provided. In addition, the influence of climatic characteristics (Λ_p , p_1 , n), geomorphologic characteristics (A , ν , τ_1), absorption infiltration and routing basin process (Λ_q/Λ_p , r , ξ) is also identified by (12). It is noteworthy that typical values of ν and n , ranging around $0.45 \div 0.6$ and $0.25 \div 0.4$ respectively, lead θ , and then the exponent of the flood peak-area relationship, to the most encountered values in the real world.

Finally, it is to remark that in equation (12) a coefficient ψ has been included to account for the possible bias due to simplifications made in this paper. In facts, one should acknowledge that the assumption of a Gumbel variate for the rainfall annual maxima could lead to overestimate the mean rainfall intensity given in Eq. (8), whenever the skewness of the probability distribution of the annual maximum rainfall is so high that cannot be accounted by a Gumbel distribution. On the other hand the expectation of the base process of peak flow, under the hypothesis that the annual maximum flood distribution follows the Gumbel distribution, is likely to generate, in Eq. (6), an underestimation of the index flood which depends on the ratio between the skewness of rainfall and flood probability distributions.

3 Application

The proposed model was applied to 20 basins located in Southern Italy (Puglia and Basilicata regions), shown in Fig. 1, whose climatic and geomorphologic characteristics are reported in Table 1. These basins have been also investigated in a number of previous works. In particular, for each of them an estimate of $r = E[a]/A$ was achieved by Fiorentino and Iacobellis (2001) by equating Q_I as in (2) to the respective observed mean annual flood.

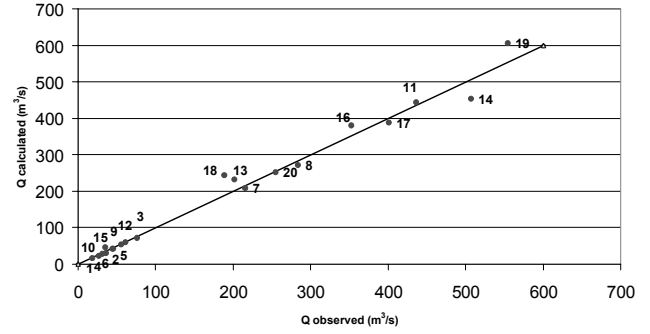


Fig. 2. Comparison between calculated (Eq. 10) and observed mean annual floods.

Although it is not relevant to the aim of this paper, it is noteworthy that the estimated values of r showed strongly different patterns in semiarid and humid climates. In particular, in the former case this pattern was rather controlled by soil characteristics, while in the latter a very interesting correlation was found between r estimates and the variation coefficient of the topographic index (Fiorentino et al., 2002).

In Fig. 2 the comparison between observed and estimated (Eq. 11 and 12) values of Q_I is shown. The reported good fit was attained by taking ψ in (12) equal to 1.2.

It is remarkable that a unique ψ value in the entire region was able to account for the bias due to approximations. This result reinforces the idea that bias in (10) is mainly due to the inability of the Gumbel law to completely account for the skewness of observed rainfall and flood data. In facts, amongst the probability distribution features depending on the first three order moments, the skewness is that showing the lowest spatial variability at the regional scale.

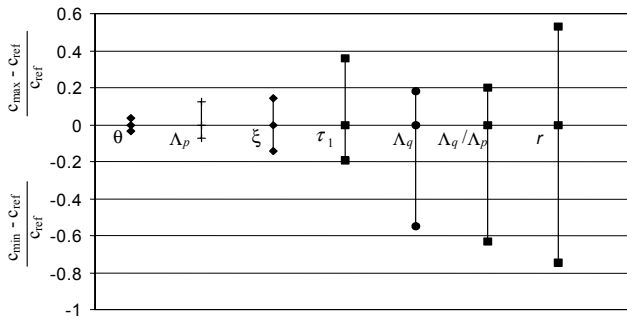
4 Sensitivity analysis

In this chapter a sensitivity analysis on the quantities that control the constant c in Eq. (12) is presented. The range of variability of each quantity was chosen with regard to typical values found in Mediterranean climates. In particular, the following rationale was followed.

The mean annual number of flood events depends generally on the mean annual number of rainfall events exceeding the given runoff threshold. Within humid areas, where the probability of finding wet antecedent moisture condition is high, the ratio Λ_q/Λ_p tends to be accordingly high, approaching unity in hyper-humid basins. Instead the ratio Λ_q/Λ_p tends to become very low in arid climates, where the probable dryness of soils at the time of the storm allows high absorption capacity, and generates a significantly lower number of floods for a given number of rainfall events. In Italy, with particular regard to the South, typical values of Λ_q vary from 2 to 30 [events/year], while Λ_p is found to vary between 20 to 50 [events/year].

Table 1. Climatic and geomorphologic characteristics of the river basins investigated.

Basin	N	A(Km ²)	p_1 (mm/h)	n	Λ_p	Λ_q	τ_1 (hours)	θ	q_o (m ³ /s)	r
Santa Maria at Ponte Lucera Torremaggiore	1	58	24.27	0.29	44.6	2.6	0.39	0.34	0.2	0.15
Triolo at Ponte Lucera Torremaggiore	2	56	34.5	0.3	44.6	3.1	0.39	0.34	0.3	0.2
Salsola at Ponte Foggia San Severo	3	455	23.29	0.27	44.6	5	0.39	0.35	2.8	0.12
Casanova at Ponte Lucera Motta	4	57	24.57	0.3	44.6	3.7	0.39	0.34	0.5	0.18
Celone at Ponte Foggia San Severo	5	233	23.33	0.27	44.6	6.6	0.39	0.35	2.2	0.08
Celone at San Vincenzo	6	92	24.08	0.29	44.6	6.1	0.39	0.34	1.2	0.11
Cervaro at Incoronata	7	539	23.86	0.28	44.6	5.2	0.39	0.35	8	0.54
Carapelle at Carapelle	8	715	24.13	0.28	44.6	8.5	0.39	0.35	7	0.45
Venosa at Ponte Sant' Angelo	9	263	24.13	0.26	44.6	4.2	0.39	0.36	1.4	0.14
Arcidiaconata at Ponte Rapolla Lavello	10	124	24.13	0.26	44.6	4.1	0.39	0.36	1.2	0.2
Ofanto at Rocchetta Sant' Antonio	11	1111	24.13	0.26	21	4.7	0.39	0.36	26	0.7
Atella at Ponte sotto Atella	12	176	24.13	0.26	21	6.3	0.39	0.36	3	0.13
Bradano at Ponte Colonna	13	462	22.2	0.28	21	4	0.19	0.35	5	0.35
Bradano at San Giuliano	14	1657	23.52	0.29	21	2.9	0.19	0.34	10	0.36
Basento at Pignola	15	42	21	0.31	21	19.6	0.39	0.33	1.5	0.25
Basento at Gallipoli	16	853	20.41	0.31	21	8.5	0.19	0.33	25	0.26
Basento at Menzena	17	1382	21.48	0.31	21	6.6	0.19	0.33	25	0.18
Agri at Tarangelo	18	511	21.56	0.36	21	16.8	0.39	0.31	10	0.26
Sinni at Valsinni	19	1140	23.13	0.4	21	19.1	0.19	0.29	45	0.19
Sinni at Pizzutello	20	232	21.56	0.36	32	31	0.19	0.31	15	0.27

**Fig. 3.** Sensitivity of parameter c (Eq. 12) to geomorphoclimatic quantities.

The routing factor ξ can generally fluctuate from about 0.6 to about 0.8, depending on the storage capacity of the basin (Fiorentino et al., 1987).

The τ_1 parameter is the lag time per unit contributing area and, in the considered region varies around an average of 0.2 h. The ν parameter tends to be very close to 0.5, which is, for instance, the estimated value for the Basilicata region, while a slightly lower value (0.48) was estimated for the basins located in Puglia (Ermini and Fiorentino, 1994). In the considered region, the IDF parameter n tends to get values as low as 0.2 in semi-arid climates and considerably higher, around 0.4, in more humid areas. Thus, following (12), θ may be assumed to be quite stable around 0.3.

Therefore, starting from the following arbitrary set of parameters: $p_1=22.2$ mm/h; $n=0.28$; $\xi=0.7$; $\tau_1=0.29$ h; $r=0.375$; $\theta=0.32$; $\Lambda_q=16.8$; $\Lambda_p=32.8$, the sensitivity anal-

ysis was carried out changing them one at a time within the following ranges: $\xi=0.6-0.8$; $\tau_1=0.19-0.39$; $r=0.05-0.7$; $\theta=0.29-0.36$; $\Lambda_q=2.6-31$; $\Lambda_p=21-44.6$; $\Lambda_q/\Lambda_p=0.058-0.969$. The IDF coefficient p_1 was not varied as it is quite stable in large regions. The monthly regional average value observed in January was ascribed to base flow q_o .

Thus for each parameter, three values of c were calculated by (12): c_{\min} is the lowest value in the range, c_{\max} is the highest value and c_{ref} is obtained with the initial set of parameters. The results are shown in Fig. 3, which demonstrates that c is highly sensitive to r , Λ_q/Λ_p , and, although a little less, to the dynamic parameter τ_1 .

The new index flood expression is, therefore, a remarkable contribution to understand how some physical and hydrological basin features control such coefficient, but it keeps a strong dependence, yet, from the r parameter not directly measurable but certainly connected to characteristics of runoff generation.

5 Final remarks

The need to reduce the prediction uncertainty in ungauged catchments, drives the research toward a better understanding of the physical phenomena that influence flood generation mechanisms.

In this paper the empirical power law relationship relating the mean annual flood to the basin area, $Q_I=c A^{1-\theta}+q_o$, was theoretically derived. Thus, a link between geomorphoclimatic characteristics of the basin and parameters c and θ was identified.

In particular, θ was shown to be dependent on both the exponents $n-1$ of the IDF and ν of the lag time–area relationship. It is noteworthy that commonly estimated values of n and ν bring $1-\theta$ into the range more frequently shown by real world data.

As regards the coefficient c , it was shown to depend on certain basin characteristics and to be particularly controlled by the expected partial area rA contributing to the flood peak and by the mean annual number of independent floods Λ_q (or the ratio Λ_q/Λ_p). About the parameter r one could note that in the hydrologic practice it is usually mimicked by the runoff coefficient, then, results achieved in this paper could have been somehow expected.

Not the same can be said about the result regarding Λ_q , indeed, its control on the mean annual flood sounds quite new. Nevertheless, this result seems to be strongly consistent if one notes, as at the previous point, that in arid climates low values of Λ_q/Λ_p leads to a significant number of no-floods because of a large amount of soil storage during the storm. As most of the times this storage affects the flood peak also, it must be expected, for a given r , a lower peak discharge.

Acknowledgements. This work was supported by PRIN2003 (2003084552) “Analysis and prediction of hydrologic extremes in mediterranean semi-arid basins” research project funded by the Italian Ministry of University and Scientific Research (MIUR).

Edited by: L. Ferraris

Reviewed by: V. Kveton and another referee

References

- Brath, A., Bacchi, B., and Rosso, R.: La derivazione geomorfoclimatica della distribuzione di probabilità delle portate di piena, *Idrotecnica* N.4, Luglio-Agosto 1992.
- Burn, D. H.: An appraisal of the ‘Region of Influence’ approach to flood frequency analysis, *Hydrol. Sci. J.*, 35, 149–165, 1990.
- Claps, P., Copertino, V. A., and Fiorentino, M.: Analisi regionale dei massimi annuali delle portate al colmo di piena, in: *Valutazione delle piene in Puglia*, edited by: Copertino, V. A. and Fiorentino, M., DIFA-University of Basilicata and CNR-GNDCI, 1994.
- Ermini, R. and Fiorentino, M.: I tempi di ritardo caratteristici dei bacini idrografici, in: *Valutazione delle piene in Puglia*, edited by: Copertino, V. A. and Fiorentino, M., DIFA- University of Basilicata and CNR-GNDCI, 1994.
- Fiorentino, M., Rossi, F., and Villani, P.: Effect of the basin geomorphoclimatic characteristics on the mean annual flood reduction curve, *proc. of the 18th Annual Conference on Modeling and Simulation, Pittsburgh, Part 5, 1777–1784, 1987.*
- Fiorentino, M., Gabriele, S., Rossi, F., and Versace, P., Hierarchical approach for regional flood frequency analysis, in: *Regional flood frequency analysis*, edited by: Singh, V. P., D. Reidel, Norwell, Mass, 35–49, 1987b.
- Fiorentino, M. and Iacobellis, V.: New insights about the climatic and geologic control on the probability distribution of floods, *Water Resour. Res.*, 37, 3, 721–730, 2001.
- Fiorentino, M., Carriero, D., Laguardia, G., Manfreda, S., Margiotta, M., Rossano, R., Sole, A., and Iacobellis, V.: Una proposta metodologica per la mappatura della variabilità spaziale delle perdite idrologiche durante i fenomeni di piena, *Proc. of the “La Difesa della Montagna”*, Assisi, 11–12 December 2002, pp. 147–154, 2002.
- Franchini, M. and Galeati, G.: Sintesi del rapporto regionale per i compartimenti di Bologna, Pisa, Roma, e zona emiliana del bacino del Po, in: *GNDCI Linea 1. Rapporto di sintesi sulla valutazione delle piene in Italia*, 5.1–5.10, Roma 1999.
- Grover, P. L., Burn, D. H., and Cunderlik, J. M.: A comparison of index flood estimation procedures for ungauged catchments, *Can. J. Civil Eng.*, 29(5), 734–741, 2002.
- Iacobellis, V. and Fiorentino, M.: Derived distribution of floods based on the concept of partial area coverage with a climatic appeal, *Water Resour. Res.*, 36, 2, 469–482, 2000.
- Rossi, F. and Villani, P.: *Valutazione delle Piene in Campania*, CNR-GNDCI, Pubbl. N. 1472 Grafica Matelliana & C., Cava de’ Tirreni (SA), 1994.
- U.S. Weather Bureau: *Rainfall-intensity-frequency regime, Part 2*, Southwestern United States Technical Report no. 29, Washington, 1958.
- Versace, P., Ferrari, E., Gabriel, S., and Rossi, F.: *Valutazione delle Piene in Calabria*, Bozza Finale, CNR-IRPI e GNDCI, Geodata, Cosenza 1989.