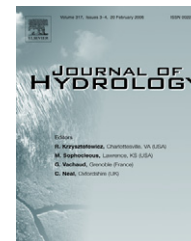


available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/jhydrol

Multivariate probabilistic regional envelopes of extreme floods

Attilio Castellarin ^{a,*}, Richard M. Vogel ^{b,1}, Nicholas C. Matalas ^{c,2}

^a DISTART – Università di Bologna, Viale Risorgimento 2, 40136 Bologna, Italy

^b Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155, United States

^c 709 Glyndon St. S.E., Vienna, VA 22180, United States

Received 20 September 2006; received in revised form 22 December 2006; accepted 9 January 2007

KEYWORDS

Regional flood
frequency analysis;
Flood of record;
Envelope curve;
Intersite correlation;
PUB (predictions in
ungaged basins)

Summary Recently probabilistic regional envelope curves of extreme floods (RECs) were introduced along with an estimator of the exceedance probability of a REC that accounts for the impact of correlation among flood sequences. The dependence of traditional envelope curves on drainage area alone impacts their reliability for estimating the design-flood. We introduce multivariate regional envelopes (MVEs) of extreme floods which are envelope surfaces (or hyper-surfaces) that represent the bound on our flood experience in a region in terms of geomorphologic and climatic basin descriptors. An empirical MVE is derived for a group of 34 unregulated catchments located in north-east-central Italy. A cross-validation procedure that simulates ungaged conditions at every site in the region is used to compare the reliability of design flood estimates resulting from REC, MVE and an index-flood approach. We document that MVEs outperform RECs and provide flood quantile estimates at ungaged sites that are nearly as reliable as index flood quantiles.

© 2007 Elsevier B.V. All rights reserved.

Introduction

The bound on our current experience of extreme floods gained through systematic observation of flood discharges

in a region is defined in terms of the largest observed floods (floods of record) observed at all gaging stations and can be graphically summarized using a regional envelope curve (REC). RECs are curves that provide an upper bound on all observed floods of record plotted versus the basin area. Alternatively, the REC can be represented by a curve defining the upper bound on the unit floods (ratios of floods to drainage areas) versus drainage areas.

The history of RECs is rather long; Myers' envelope curve is one of the first examples (see e.g., Fuller, 1914; Creager et al., 1966) and, since their introduction, RECs

* Corresponding author. Tel.: +39 051 2093354.

E-mail addresses: attilio.castellarin@mail.ing.unibo.it (A. Castellarin), richard.vogel@tufts.edu (R.M. Vogel), nmatalas@aol.com (N.C. Matalas).

¹ Tel.: +1 617 6274260.

² Tel.: +1 703 3194016.

have been and continue to be developed worldwide. For instance, Marchetti (1955) proposed a set of RECs for Italy that is still used as a reference by practitioners; Jarvis (1925) identified the REC for the continental US, that has been updated by Crippen and Bue (1977) and Crippen (1982), Mimikou (1984) constructed RECs for western Greece; Kadoya (1992) for Japan; Costa (1987) compared record flood experience in the US, China and the world. Recent worldwide catalogues of record floods and their graphical envelope summaries can be found in Herschy (2002) and IAHS (2003).

The enduring value and widespread application of RECs can only be explained by their ease of use and understanding due in part to their graphical nature. RECs' long-lasting popularity is increasingly at odds with their traditional deterministic interpretation, which reduces the curves to visual catalogues of maximum observed floods limiting their applicability for design purposes (see e.g., IAHS, 2003). Even though RECs are mainly viewed as deterministic graphical tools, they are amenable to probabilistic statements. According to Meyer (1917), Fuller (1914) was the first to define a regional flood probability in the context of an envelope curve. Only recently Castellarin et al. (2005) proposed a probabilistic interpretation of RECs, and formulated an empirical estimator of the exceedance probability p_{EE} of the expected REC (the subscript EE stands for Expected Envelope). These authors defined the expected REC as the envelope curve that, on average, is expected to bound the flood experience for a given region, where region refers to a set of annual maximum flood sequences that are cross-correlated, concurrent and of equal length. The terms "on average" and "expected REC" highlight the fact that we experience only one realisation of the envelope, and this single realisation will differ from the theoretical expected envelope for a region of given characteristics, to which p_{EE} refers. Subsequently, Castellarin (2006) presented an algorithm for the application of the empirical estimator of p_{EE} to historical annual maximum series of unequal length, and assessed how the selection of a particular cross-correlation formula and plotting position affects the accuracy of $1/p_{EE}$ -year flood quantiles estimated from RECs at ungaged basins. Castellarin (2006) showed that the selection of a particular plotting position significantly affects the quantile accuracy, hence he recommended and developed a quantile-unbiased plotting position. Castellarin (2006) also suggested that the dependence of the envelope on drainage area alone significantly curtails the accuracy of the $1/p_{EE}$ -year flood estimates.

The primary objectives of this study are then: (1) to drop the restriction that a single factor, the drainage area, controls the flood envelope by introducing multivariate regional envelopes of extreme floods (MVEs); (2) to show that MVEs yield a more effective definition of the envelope than envelopes which depend only on drainage area or only on some other single factor; (3) to demonstrate, under the same assumptions adopted by Castellarin et al. (2005), that the exceedance probability of the expected multivariate envelope coincides with p_{EE} , and therefore can be estimated using the same algorithm proposed for RECs; and (4) to show that probabilistic MVEs are significantly more reliable than RECs for estimating the $1/p_{EE}$ -year flood at ungaged basins.

Probabilistic interpretation of regional envelope curves

Regional flood assumptions

Several studies (e.g., Jarvis, 1925; Marchetti, 1955; Castellarin et al., 2005) define a REC using,

$$\ln\left(\frac{Q}{A}\right) = a + b \ln(A), \quad (1)$$

where Q is the envelope flood for a given basin, A is drainage area (i.e., Q/A is the unit flood of record), a and b are two regional coefficients (see the example depicted in Fig. 1). Castellarin et al. (2005) adopted two fundamental assumptions: (i) the region (i.e., pooling-group of sites (e.g., Burn, 1990; Castellarin et al., 2001)) is homogeneous in the sense of the index-flood hypothesis (see e.g., Dalrymple, 1960); and (ii) the relationship between the index-flood μ_X (e.g., mean annual flood) and A is of the form,

$$\mu_X = CA^{b+1}, \quad (2)$$

where b and C are constants and b is the same as in (1). Under these assumptions the authors developed an estimator of the exceedance probability p_{EE} of the expected REC. Castellarin et al. (2005) identified the expected REC through a series of Monte Carlo simulation experiments by repeatedly generating sets of synthetic cross-correlated Gumbel sequences to form regions. Each region consisted of M concurrent synthetic annual maximum flood sequences (AMS) of floods, each of length n .

Castellarin et al. (2005) showed that under the adopted hypotheses the problem of estimating p_{EE} reduces to estimating the exceedance probability of the largest value in a regional sample of standardized annual maximum peak flows [i.e., observed peak flows divided by the mean annual flood]. The primary challenge of their work involved estimation of the regional information content of cross-correlated

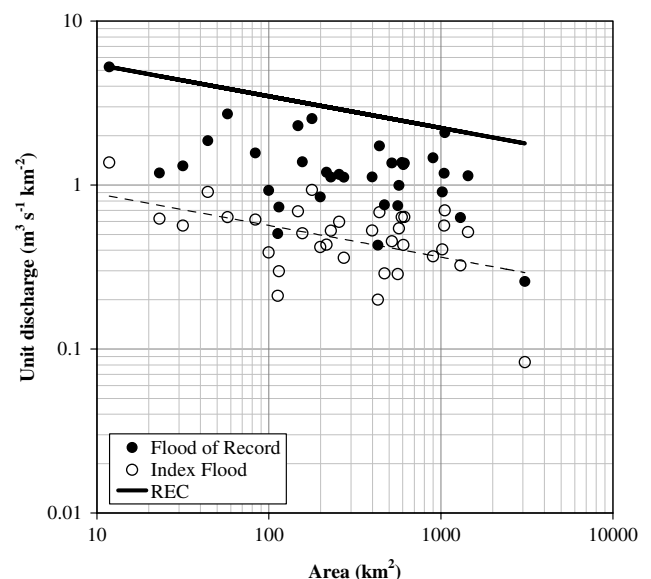


Figure 1 Representation of empirical index-flood (μ_X) and flood of record (Q) values, simple regression between μ_X and A (dashed line), regional envelope curve, REC (bold line).

flood series. [Castellarin et al. \(2005\)](#) used results introduced by [Matalas and Langbein \(1962\)](#) and [Stedinger \(1983\)](#) to quantify the regional information content using the concept of the equivalent number of independent annual maxima. [Castellarin et al. \(2005\)](#) expressed the equivalent number of independent observations, or number of effective observations n_{eff} , as n times the equivalent number of independent sequences M_{EC} , which can be estimated from,

$$\hat{M}_{\text{EC}} = \frac{M}{1 + \overline{\rho^\beta}(M-1)}, \quad \text{with } \beta: 1.4 \frac{(nM)^{0.176}}{(1-\rho)^{0.376}}, \quad (3)$$

where $\overline{\rho^\beta}$ and $(1-\rho)^{0.376}$ are average values of the corresponding functions of the correlation coefficients (i.e., $\overline{\rho^\beta}$ is the average of the $M(M-1)/2$ values of $\rho_{k,j}^\beta$, where $\rho_{k,j}$ is the correlation coefficient between annual maximum floods at sites k and j , with $1 \leq k < j \leq M$).

Exceedance probability of the expected REC

[Castellarin \(2006\)](#) presented an algorithm that relaxes the need for concurrent series, enabling the estimation of n_{eff} for real-world datasets. For a regional dataset consisting of M individual AMS that globally span n years, the actual distribution of the flood series in time (e.g., missing data, different installation years for different gages, etc.) can be taken into account as follows. First, one identifies the number of years, n_1 , for which the original dataset includes only one observation of the annual maximum discharge, that is $M-1$ observations are missing (for example, some gages may not be operational, or may not be installed yet). These n_1 observations are effective by definition. Second, the dataset containing the $n-n_1$ remaining years is subdivided into $N_{\text{sub}} \leq (n-n_1)$ subsets; each one of them (say subset s) is selected in such a way that all its $L_s \leq M$ sequences are concurrent and of equal length L_s and therefore suitable for the application of the estimator proposed by [Castellarin et al. \(2005\)](#). Using this splitting criterion, the effective number of observations n_{eff} can be calculated as the summation of the effective sample years of data estimated for all N_{sub} subsets,

$$\hat{n}_{\text{eff}} = n_1 + \sum_{s=1}^{N_{\text{sub}}} \hat{n}_{\text{eff},s} = n_1 + \sum_{s=1}^{N_{\text{sub}}} \frac{L_s L_s}{1 + \left[\overline{\rho^\beta} \right]_{L_s} (L_s - 1)}, \quad \text{with} \quad (4)$$

$$\beta := 1.4 \frac{(L_s L_s)^{0.176}}{\left[(1-\rho)^{0.376} \right]_{L_s}}.$$

As described previously, n_1 represents the number of times annual floods were observed at one site only (and possibly single observations or indirect measurements at miscellaneous sites), that is, the total number of years in which $L_s = 1$. The notation $[\cdot]_{L_s}$ in (4) indicates that the average terms $\overline{\rho^\beta}$ and $(1-\rho)^{0.376}$, which have the same meaning as in (3), are to be computed with respect to the $L_s > 1$ annual flood sequences which form subset s .

The p_{EE} value can be estimated by representing the inter-site correlation from a suitable model of cross-correlation versus distance between sites (see e.g., [Tasker and Stedinger, 1989](#); [Troutman and Karlinger, 2003](#)) and by using an appropriate plotting position with the overall sample-years of data set equal to \hat{n}_{eff} . [Castellarin \(2006\)](#) showed that the selection of the cross-correlation model has limited impact

on the reliability of estimated p_{EE} values. [Castellarin \(2006\)](#) used a model introduced by [Tasker and Stedinger \(1989\)](#) to approximate the true annual peak cross-correlation function $\rho_{i,j}$ as a function of the distance $d_{i,j}$ among sites i and j ,

$$\rho_{i,j} = \exp \left(- \frac{\lambda_1 d_{i,j}}{1 + \lambda_2 d_{i,j}} \right), \quad (5)$$

where $\lambda_1 > 0$ and $\lambda_2 \geq 0$ are the regional parameters which may be estimated by either ordinary or weighted least squares procedures (see e.g., [Tasker and Stedinger, 1989](#)). Eq. (5) describes the tendency of peak flow correlations to decrease as a function of distance due to both network effects and regional storm effects, and should be used in the absence of a strong control of the channel network on the spatial structure of the intersite correlation model (see e.g., [Troutman and Karlinger, 2003](#)).

[Castellarin \(2006\)](#) addressed the problem of selecting a suitable plotting position for estimating p_{EE} . [Cunnane \(1978\)](#) introduced the general plotting position

$$\hat{p}_{\text{EE}} = 1 - \frac{\hat{n}_{\text{eff}} - \eta}{\hat{n}_{\text{eff}} + 1 - 2\eta}, \quad (6)$$

where η is the plotting position parameter and \hat{n}_{eff} is the empirical estimate of n_{eff} given in (4). Each plotting position is characterized by a particular η value (see e.g., [Cunnane, 1978](#); [Stedinger et al., 1993](#) for selection criteria). The results reported in [Castellarin \(2006\)](#) indicate that, among several possible options, a quantile unbiased-plotting position should be used. [Castellarin \(2006\)](#) derived a quantile-unbiased plotting position for use with the Generalized Extreme Value (GEV) distribution ([Jenkinson, 1955](#)). The GEV distribution has been shown to satisfactorily reproduce the sample frequency distribution of hydrological extremes around the world (see e.g., [Stedinger et al., 1993](#); [Vogel and Wilson, 1996](#); [Robson and Reed, 1999](#); [Castellarin et al., 2001](#) and others). The proposed plotting position is a very compact and easy to apply asymptotic formula for the estimation of the exceedance probability of the largest value in a GEV sample, in which the parameter η of (6) depends on the shape parameter k of the fitted GEV distribution,

$$\eta(k) = \frac{\exp(\gamma) - 1}{\exp(\gamma)} - \frac{\pi^2}{12 \exp(\gamma)} k; \quad \eta(k) = 0.439 - 0.462k, \quad (7)$$

where $\gamma = 0.5772$ is the Euler's constant. Eq. (7) should only be applied when $n_{\text{eff}} \geq 10$ and $-0.5 < k < 0.5$ (see [Castellarin, 2006](#)).

Empirical RECs and REC flood quantiles

The construction of empirical RECs and estimation of p_{EE} involves the following steps: (i) a homogeneous region (or a pooling-group of sites) is identified (see e.g., the procedure for identification of homogeneous regions in [Hosking and Wallis, 1997](#); Chapter 4); (ii) estimation of the REC slope, \hat{b} , in (1) is obtained by regressing the empirical values of the index-flood (i.e., at-site estimates of mean annual flood) against the drainage areas of the corresponding basins; (iii) the value of the intercept a in (1) is computed as,

$$a = \max_{j=1, \dots, M} \left\{ \ln \left(\frac{Q_j}{A_j} \right) - \hat{b} \ln(A_j) \right\}, \quad (8)$$

where Q_j denotes the maximum flood observed at site j , where $j = 1, 2, \dots, M$ and M is the number of sites in the region, and A_j is the drainage area of site j ; (iv) a suitable regional cross-correlation function is identified [e.g., estimation of coefficients λ_1 and λ_2 of (5)]; (v) \hat{n}_{eff} is computed using (4); (vi) a suitable regional parent distribution is chosen (see e.g., the procedure for the regional parent selection in Hosking and Wallis, 1997; Chapter 5) and (vii) a quantile-unbiased plotting position suitable for this distribution is utilized for estimating p_{EE} as a function of \hat{n}_{eff} as in (6).

Following these steps an empirical REC can be constructed for any group of AMS of peak flows and the regional p_{EE} value can then be estimated. From the empirical REC one may then easily (graphically) estimate the $1/p_{\text{EE}}$ -year flood, that hereafter is referred to as REC flood quantile. A REC flood quantile may be estimated for any ungaged basin in the region of interest using its drainage area.

Through a comprehensive cross-validation for a wide geographical region in northern central Italy, Castellarin (2006) showed that the reliability of REC flood quantiles for ungaged sites is comparable with the reliability of regional estimates of flood quantiles generated by the index-flood approach. More importantly, Castellarin (2006) found that the most limiting factor of the REC approach is the dependence of flood quantiles on drainage area alone; thus this restriction is removed in the following section.

Probabilistic multivariate envelopes of extreme floods

Multivariate envelopes

The theoretical framework of probabilistic RECs can be generalized to take into account basin area and other factors that have a bearing on flood magnitude. We consider a multivariate envelope (MVE) of record floods in a region using

$$\ln(Q) = a + \alpha_1 \ln(\omega_1) + \alpha_2 \ln(\omega_2) + \dots + \alpha_v \ln(\omega_v), \quad (9)$$

where, Q is the envelope flood and ω_i , $i = 1, 2, \dots, v$, are geomorphoclimatic basin descriptors, while a and α_i , for $i = 1, 2, \dots, v$, are regional coefficients. Analogous to the univariate case (i.e., REC), once estimates of the coefficients $\hat{\alpha}_i$, with $i = 1, 2, \dots, v$, are obtained through multivariate regression techniques (see next section), the intercept a in (9) may be computed from,

$$a = \max_{j=1, \dots, M} \left\{ \ln(Q_j) - \sum_{i=1}^v \hat{\alpha}_i \ln(\omega_{ij}) \right\}, \quad (10)$$

where Q_j denotes the flood of record at site $j = 1, 2, \dots, M$ and M is the number of sites in the region, while ω_{ij} is the i th geomorphoclimatic parameter of basin j .

Probabilistic multivariate envelopes

The index-flood μ_X (i.e., mean annual flood) can be assumed to depend upon several geomorphologic and climatic basin descriptors through a power-law relationship,

$$\mu_X = \alpha_0 \omega_1^{\alpha_1} \omega_2^{\alpha_2} \dots \omega_v^{\alpha_v}. \quad (11)$$

Multivariate power-law relationships such as (11) are commonly adopted in practice to express the index-flood as a function of several geomorphologic and climatic parameters (see e.g., Brath et al., 2001). The identification of a multivariate regression model consists of selecting candidate explanatory variables (geomorphoclimatic descriptors) using, for instance, a *stepwise regression analysis* (e.g., Wiesberg, 1985; Brath et al., 2001). Instead of *stepwise regression analysis*, alternative multivariate procedures can also be adopted for identifying multivariate estimators of μ_X , such as artificial neural network, principal component or canonical correlation analysis (see e.g., Shu and Burn, 2004; Ouara et al., 2001; Chokmani and Ouara, 2004).

If, as assumed in the studies by Castellarin et al. (2005) and Castellarin (2006), the study region, or pooling-group of sites, is homogeneous in the sense of the index-flood hypothesis (see e.g., Dalrymple, 1960), then the probability distribution of standardized annual maximum peak flows is the same for all sites. The standardized annual maximum peak flow, X' , is defined for a given site as the annual maximum peak flow, X , divided by the site-dependent scale factor, μ_X . Under this assumption, the flood quantile with exceedance probability p , x_p , is,

$$x_p = \mu_X X'_p, \quad (12)$$

where X'_p is the regional dimensionless flood-quantile with exceedance probability p .

Combining (11) and (12) leads to an expression of $\ln(x_p/A)$ that defines the multivariate envelope (in terms of unit floods) of exceedance probability p ,

$$\begin{aligned} \ln\left(\frac{x_p}{A}\right) &= \ln\left(\frac{\mu_X X'_p}{A}\right) \\ &= \ln(\alpha_0 X'_p) - \ln(A) + \alpha_1 \ln(\omega_1) + \alpha_2 \ln(\omega_2) + \dots \\ &\quad + \alpha_v \ln(\omega_v). \end{aligned} \quad (13)$$

The structure of (13) is analogous to (1), the difference being the number of basin descriptors which are employed to describe the envelope.

Assume that an empirical MVE is identified by a bounding surface (or hyper surface) on all record floods to the present, say up to the year n . Let X_j^i denote the annual maximum flood in year $i = 1, 2, \dots, n$ at site $j = 1, 2, \dots, M$, where M is the number of sites in the region. Let $X_j^{(i)}$ denote the flood flow of rank (i) at site j , where ranking is from smallest (1) to largest (n) (i.e., $X_j^{(n)} = Q_j$). The theoretical MVE intercept a up to the year n can then be expressed as,

$$\begin{aligned} a &= \max_{j=1, \dots, M} \left\{ \ln(X_j^{(n)}) - \sum_{i=1}^v \alpha_i \ln(\omega_{ij}) \right\} \\ &= \max_{j=1, \dots, M} \left\{ \ln(\mu_{X,j} X_j^{(n)}) - \sum_{i=1}^v \alpha_i \ln(\omega_{ij}) \right\}, \end{aligned} \quad (14)$$

where $X_j^{(n)}$ is the dimensionless record flood at site j , defined as $X_j^{(n)}/\mu_{X,j}$, with $\mu_{X,j}$ index-flood of site $j = 1, 2, \dots, M$. Combining (14) and (11) leads to,

$$\begin{aligned} a &= \max_{j=1, \dots, M} \{ \ln(\alpha_0) + \ln(X_j^{(n)}) \} \\ &= \ln(\alpha_0) + \ln\left(\max_{j=1, \dots, M} \{X_j^{(n)}\}\right) \end{aligned} \quad (15)$$

in which $\max_{j=1,\dots,M}\{X_j^{(n)}\}$ is the standardized maximum flood (referred to as the regional record flood by [Castellarin et al. \(2005\)](#)).

Analogous to the probabilistic interpretation of a REC introduced by [Castellarin et al. \(2005\)](#) a probabilistic statement can also be associated with the multivariate envelope through the intercept a of (10) and (15). As illustrated in (15), a is controlled by the regional record flood. Therefore, the exceedance probability of the probabilistic MVE coincides with the exceedance probability (or p -value) of the regional record flood.

The problem of estimating the p -value of the MVE reduces to estimating the p -value of the regional record-flood, the maximum of $n \cdot M$ standardized annual maximum peak flows, and therefore is completely equivalent to the problem of estimating the p -value of a REC. Hence the generalization of a REC to a MVE does not alter the theoretical framework of probabilistic envelope curves. We can still define an expected MVE for a cross-correlated region of given characteristics. We can also recognize that the p -value of the expected MVE reduces to the exceedance probability of the largest value in a regional sample of standardized annual maximum peak flows (i.e., observed peak flows divided by the mean annual flood, see [Castellarin et al., 2005](#)).

Empirical MVEs and MVE flood quantiles

We have shown above that the same empirical estimator of p_{EE} used for expected RECs (see steps iv–vi of Section “Empirical RECs and REC flood quantiles”) can be applied for estimating the exceedance probability of the expected MVE. Thus an estimate of the $1/p_{EE}$ -year flood quantile, which is hereafter referred to as the MVE flood quantile, can be computed for any ungaged basin in the region of interest.

The steps required to construct the envelope in the multivariate case are analogous to the steps listed in Section “Empirical RECs and REC flood quantiles” for constructing empirical RECs. First, the homogeneous pooling-group of sites needs to be identified (identical to step i for the uni-

variate case, see Section “Empirical RECs and REC flood quantiles”). Second, a suitable multiregression model for estimating the index-flood is identified, and its coefficients β_i , with $i = 1, 2, \dots, v$, estimated (analogous to step ii for the univariate case, see Section “Empirical RECs and REC flood quantiles”). This task can be accomplished by regressing the empirical values of the index-flood (i.e., at-site estimates of mean annual flood) against a set of basin descriptors through, for instance, a *stepwise regression* analysis. Thirdly, the value of the intercept a of (9) is computed as in (10) (analogous to step iii for the univariate case, see Section “Empirical RECs and REC flood quantiles”). Finally, p_{EE} is estimated analogous to the univariate case (see Section “Empirical RECs and REC flood quantiles”, steps from iv to vi).

Graphical representation of MVEs

Traditional envelope curves have been used widely in part because they are graphical, and thus very easy to use and understand. Thus, it is important to extend this graphical feature to multivariate envelopes of extreme floods. Graphical representation of MVEs is straightforward if the multivariate power-law relationship in (11) consists of only two independent basin descriptors in which case one may construct 3-dimensional scatter plots or contour plots. In this instance, MVEs are represented by a plane in log space or, more generally, a surface in real space. A graphical representation is even possible for multiregression models with more than two explanatory variables. Suppose the first k explanatory variables of the model in (9) are geomorphologic parameters (i.e., ω_i for $i = 1, 2, \dots, k$), while the remaining $n-k$ variables are climatic parameters (i.e., ω_i for $i = k+1, k+2, \dots, v$). We can define a geomorphologic factor (GF) and a climatic factor CF as follows:

$$GF = \omega_1^{\alpha_1} \omega_2^{\alpha_2} \dots \omega_k^{\alpha_k}; \quad CF = \omega_{k+1}^{\alpha_{k+1}} \omega_{k+2}^{\alpha_{k+2}} \dots \omega_v^{\alpha_v}; \quad (16)$$

and use GF and CF to represent the MVE using contour plots or three-dimensional scatter plots. Alternatively, one may

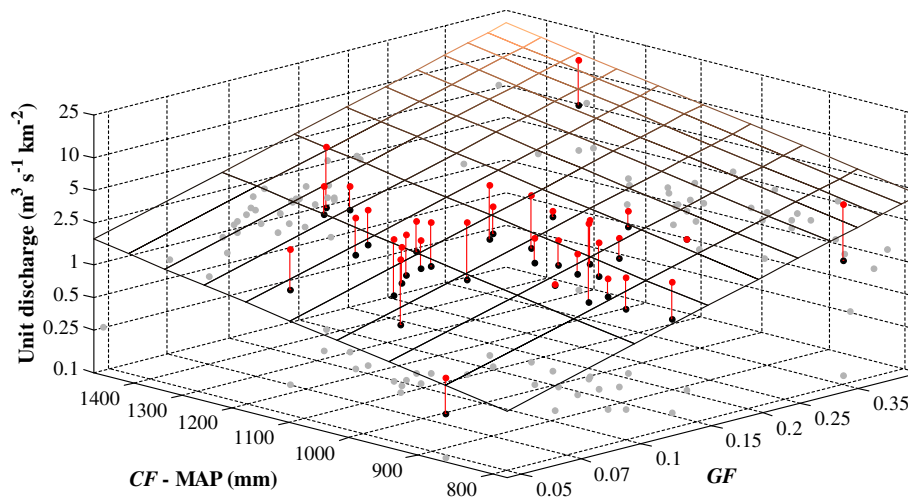


Figure 2a A three-dimensional scatterplot representation of a multivariate envelope of extreme floods (MVE), empirical flood of record values are reported as black dots, their projections onto the x - y , x - z and y - z planes as grey dots and onto the envelope as red dots, the red segments are the distances between the empirical values and their projections.

represent the MVE by reporting the unit discharges against a geomorphoclimatic factor defined as $GF \cdot CF$. This latter representation is analogous to the usual representation of a REC (Fig. 1), and is suitable for comparing the univariate and multivariate envelopes of extreme floods for a given region. Three examples of a graphical representation of an empirical MVE are provided in Fig. 2.

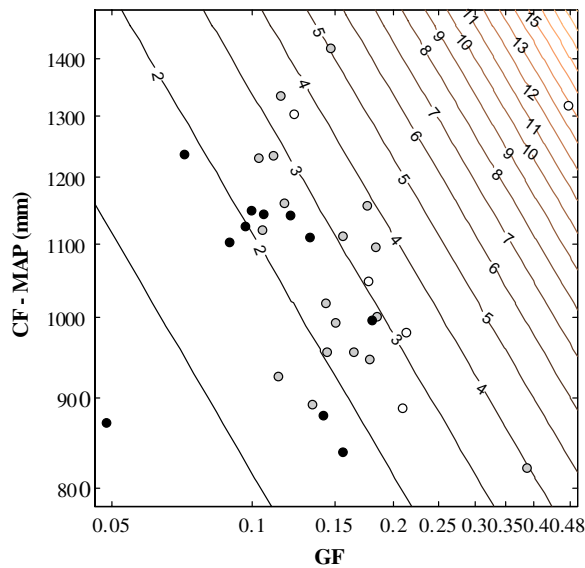


Figure 2b A two-dimensional contour-plot representation of the same MVE, the contour-lines indicate the specific discharge in $\text{m}^3 \text{s}^{-1} \text{km}^{-2}$; empirical flood of record values are reported as black dots (specific discharge $Q/A < 1 \text{ m}^3 \text{s}^{-1} \text{km}^{-2}$), grey dots ($1 \text{ m}^3 \text{s}^{-1} \text{km}^{-2} \leq Q/A \leq 2 \text{ m}^3 \text{s}^{-1} \text{km}^{-2}$) and white dots ($Q/A > 2 \text{ m}^3 \text{s}^{-1} \text{km}^{-2}$).

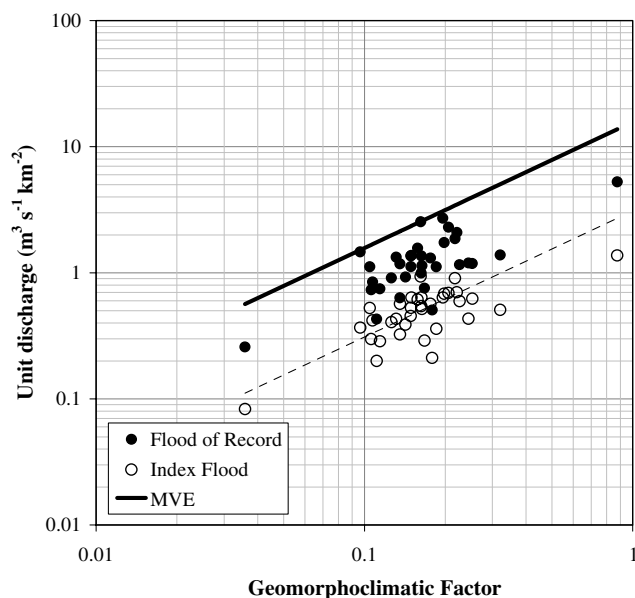


Figure 2c Representation of empirical index-flood (μ_X) and flood of record (Q) values, multivariate regression of μ_X values (dashed line), multivariate envelope of extreme floods, MVE (bold line).

Application

Study area

Fig. 3 shows the study region along with outlines of the 34 unregulated river basins considered here. The AMS of peak discharges were obtained for all basins from the stream-gauges belonging to the former National Hydrographic Service of Italy (SIMN) over the period 1920–1997. Table 1 reports the record lengths, the empirical values of the index-flood, μ_X (average of annual maximum peak flows) and the flood of record (Q) for the group of 34 catchments. Table 1 also provides relevant geomorphologic and climatic parameters of the basins, such as the basin area A [km^2], impervious area A_{imp} [km^2], maximum, H_{max} , mean, H_{mean} , and minimum, H_{min} , elevations [m above the sea level, m asl], main channel length L [km], and mean annual precipitation at basin scale MAP [mm]. We also considered several derived measures, such as the mean slope of the main channel, and, for each basin, factors of shapes and elongation and estimates of the time of concentration. Finally, Table 1 lists the unit flood of record (Q/A) over the study region.

The study region consists of 34 small to large Apenninic basins, which are characterized by significantly different percentages of impervious area (from roughly 20% to 100%). Mean annual precipitation varies significantly over the study area. In view of these differences among the catchments one may expect the regional flood frequency regime to be strongly heterogeneous.

We analysed the homogeneity of this group of sites in terms of the frequency regime of annual maximum floods. The analysis used the heterogeneity measure proposed by Hosking and Wallis (1997), which is briefly illustrated in Appendix A. Considering the most selective heterogeneity measure proposed by Hosking and Wallis (i.e., H_1 , see Appendix A), the results of the test indicate that the region should be regarded as “definitely heterogeneous”, given $H_1 = 4.46$. Nevertheless, the study region can be regarded as possibly homogeneous with respect to the other two heterogeneity measures proposed by Hosking and Wallis (1997) ($H_2 = 0.10$ and $H_3 = -1.04$; see Appendix A for the definition of H_2 and H_3). Hosking and Wallis (1997) recommend H_1 as the most reliable heterogeneity measure. Despite the results obtained in terms of H_1 , we accepted the region as a possibly homogeneous region, which is a necessary assumption for the probabilistic REC and MVE theoretical framework, in order to illustrate the incorporation of factors in addition to basin area in developing and utilising envelopes. We are aware that this approximation may impact the reliability of REC and MVE flood quantiles for the study region. The robustness of our results when the assumption of regional homogeneity is violated needs further investigation, as discussed in the results and discussions section.

Empirical REC and MVE

Following steps (ii) and (iii) of the procedure summarized in Section “Empirical RECs and REC flood quantiles” we identified the REC for the study area. The REC is illustrated in Fig. 1 (black thick line). The estimated REC’s slope \hat{b} is equal to -0.193 , and REC’s intercept a , computed from (8), is equal to $2.135 [\ln(\text{m}^3 \text{s}^{-1} \text{km}^{-2})]$.

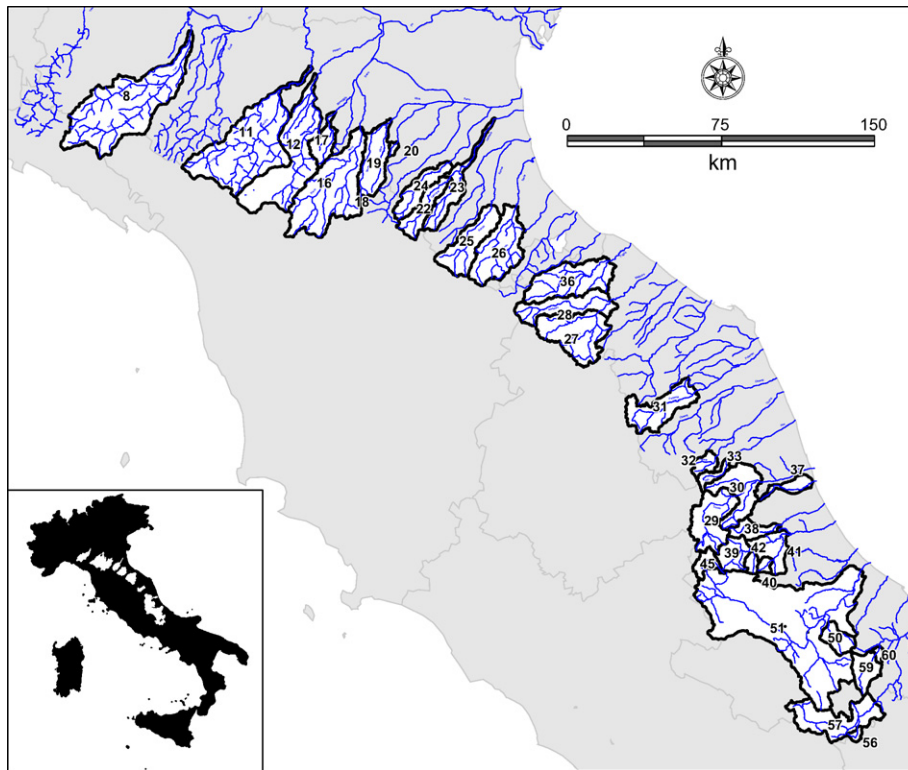


Figure 3 Study area: locations of the 34 study basins.

The regression of the empirical index-flood values versus drainage area alone is illustrated in Fig. 4 (left panel, Regional Estimates). The regression of $\ln(\mu_X)$ versus $\ln(A)$ (termed regression model μ_X-A) produced an overall Nash and Sutcliffe (1970) efficiency measure $E = 0.498$, where $E \in (-\infty, 1]$: $E = 1$ for a perfect fit; $E < 0$ for a model that performs worse than using a single regional mean value for site.

We assessed the robustness of the regression through a (leave-one-out) jackknife cross-validation procedure (see e.g., Shao and Tu, 1995; Brath et al., 2001; Castellarin et al., 2004). The procedure simulates ungaged conditions at each and every site of the study region and can be summarized as follows: (1) all $N_{\text{site}} = 34$ streamgages are considered; (2) one of these gaging stations, say station s , is removed from the set; (3) regression model μ_X-A is re-estimated using only the values of μ_X and A at the remaining $N_{\text{site}} - 1$ gaged sites; (4) using the regression model estimated in step (3), the index-flood is estimated for station s (Jackknife Estimate, left panel of Fig. 4); (5) steps (2)–(4) are repeated $N_{\text{site}} - 1$ times, considering in turn each of the remaining streamgages. The cross-validation procedure provides useful insights into the robustness and reliability of the regression model for the entire study region. The cross-validation procedure resulted in an overall Nash and Sutcliffe (1970) efficiency measure $E_{jk} = 0.150$, which is significantly smaller than the E value obtained from the μ_X-A regression model ($E = 0.498$). One's initial reaction is to conclude that the robustness of the regression model is limited. A closer look at the left panel of Fig. 4 shows instead that the regression model μ_X-A is rather robust be-

cause the differences between the regional and jackknife estimates are quite small. The low overall goodness of fit (represented by E values with and without cross-validation) results mainly from the few discordant sites (four sites). As shown in Fig. 4 (left panel), a basic regression model that assumes a simple scaling law between μ_X and A is incapable of correctly reproducing the empirical values of μ_X for all of these basins.

The construction of the empirical MVE requires the identification of a regional multivariate predictive model for μ_X . The empirical index-flood values for the 34 sites were regressed against the available geomorphoclimatic basin descriptors (see Section "Study area" and Table 1) using multivariate stepwise regression (see e.g., Wiesberg, 1985). Models of the form,

$$\ln(\hat{\mu}_X) = C_0 + C_1 \ln(\omega_1) + C_2 \ln(\omega_2) + \dots + C_v \ln(\omega_v) + \vartheta \quad (17)$$

were considered, where $\hat{\mu}_X$ indicates the empirical at-site index-flood estimate for a particular basin, ω_i , for $i = 1, 2, \dots, v$, are the explanatory variables of the model, C_i , for $i = 0, 1, \dots, v$, are parameters and ϑ is the residual of the model. The optimal subset of explanatory variables and the estimates of C_i , with $i = 0, 1, \dots, v$ were identified using stepwise regression, based on a weighted least-squares (WLS) algorithm. The WLS algorithm weights each squared residual proportionally to the length of the AMS of flood flows at each site. The stepwise procedure begins with a simple model that estimates the dependent variable $\hat{\mu}_X$ as

Table 1 Annual flood sequences: variability of the number of observations, empirical index-flood, μ_X , flood of record, Q ; and relevant geomorphoclimatic basin descriptors

Site	No. Obs.	μ_X (m ³ /s)	Q (m ³ /s)	A (km ²)	A_{imp} (km ²)	H_{max} (m asl)	H_{mean} (m asl)	H_{min} (m asl)	L (km)	MAP (mm)	Q/A (m ³ s ⁻¹ km ⁻²)
8	23	743.8	1640.0	1439	1281	1803.0	660.0	26.3	108.9	1231.8	1.14
11	55	421.3	823.0	1301	1053	2120.0	606.0	21.5	98.5	1149.7	0.63
12	60	413.3	925.0	1019	713	2165.0	662.0	18.4	106.0	1126.1	0.91
16	74	739.0	2200.0	1055	1055	1945.0	639.0	58.0	84.2	1303.2	2.09
17	33	166.5	452.0	178	178	902.0	375.0	44.0	37.7	889.5	2.54
18	20	16.2	62.0	12	12	1276.0	1005.0	730.0	6.0	1317.1	5.26
19	23	210.3	446.0	398	398	1303.0	430.0	29.1	57.7	955.7	1.12
20	26	14.4	27.4	23	23	608.0	285.0	91.6	10.2	822.8	1.18
22	27	152.9	297.0	256	249	1242.0	511.0	49.7	53.7	1096.2	1.16
23	15	235.4	706.0	518	511	1242.0	439.0	2.0	103.0	993.8	1.36
24	33	98.0	304.0	272	272	1187.0	438.0	32.7	45.7	1001.1	1.12
25	41	300.0	762.0	440	426	1657.0	569.0	57.0	57.5	1112.3	1.73
26	42	378.4	814.0	593	580	1408.0	525.0	41.5	66.5	1018.3	1.37
27	29	391.8	834.0	614	454	1702.0	600.0	182.6	56.3	1160.1	1.36
28	29	591.6	1230.0	1044	835	1702.0	560.0	109.7	75.7	1120.0	1.18
29	25	136.1	356.0	470	371	2478.0	1135.0	268.2	55.8	1109.3	0.76
30	41	331.1	1320.0	901	756	2478.0	959.0	90.5	77.2	926.7	1.47
31	41	86.1	185.0	431	185	1570.0	616.0	168.0	58.4	1103.7	0.43
32	41	38.7	92.5	100	29	2334.0	1170.0	424.8	19.4	1145.0	0.93
33	22	51.4	131.0	84	48	2478.0	1100.0	437.0	26.0	946.9	1.57
36	43	261.0	805.0	606	605	1415.0	384.0	38.8	80.0	955.7	1.33
37	9	34.1	84.0	114	62	1815.0	214.0	4.2	36.0	839.9	0.73
38	31	102.3	340.0	148	140	2435.0	930.0	221.8	35.1	981.1	2.30
39	20	93.6	259.0	216	184	2914.0	1530.0	505.1	24.3	1157.3	1.20
40	40	36.6	156.0	58	28	2914.0	1950.0	406.8	11.9	1047.9	2.71
41	15	40.0	82.3	44	10	2914.0	1200.0	400.0	10.0	1334.0	1.86
42	9	310.2	567.0	570	428	2914.0	1065.0	136.0	51.8	1142.8	0.99
45	30	23.8	57.0	113	74	1616.0	1026.0	760.0	20.8	996.4	0.51
50	14	79.6	217.0	157	89	2795.0	1077.0	198.4	18.8	1420.5	1.39
51	39	256.2	796.0	3082	1294	2795.0	940.0	4.5	159.9	872.2	0.26
57	32	17.9	41.4	32	6	1260.0	1080.0	816.3	10.6	1233.8	1.31
58	38	160.9	420.0	562	197	2795.0	1320.0	720.0	61.2	1236.1	0.75
60	6	83.2	168.0	199	119	2795.0	1191.0	270.9	24.6	880.1	0.85
61	9	120.6	256.0	229	135	2795.0	1090.0	225.0	30.1	892.5	1.12
Min	6	14.4	27.4	12	6	608.0	214.0	2.0	6.0	822.8	0.26
Mean	30	209.9	525.2	508	377	1993.3	831.8	223.2	52.3	1074.1	1.37
Max	74	743.8	2200.0	3082	1294	2914.0	1950.0	816.3	159.9	1420.5	5.26

a constant value. Each step results in an additional explanatory variable in the model, choosing the variable that maximizes the adjusted efficiency of the model (see e.g., Wiesberg, 1985). At each step with p explanatory variables, the procedure then tests the performance of all the p models including $p - 1$ explanatory variables that can be obtained from the p -variable model by dropping one variable at a time. If none of the simpler models performs better than the p -variable model, the procedure searches for the best multivariate model with $p + 1$ explanatory variables. The stepwise procedure ends when no further increases in the efficiency of the model are obtained.

The identified model reads,

$$\ln(\hat{\mu}_X) = 1.130 + 0.699 \ln(A_{\text{imp}}) + 2.214 \ln\left(\frac{\text{MAP}}{1000}\right) + \psi', \quad (18)$$

where $\hat{\mu}_X$ is expressed in m³/s, A_{imp} in km² and MAP in mm. The stepwise regression analysis indicated a strong connection exists between μ_X (average value of the maximum annual flood) and MAP (mean value of long-term precipitation). This result is not surprising, as the strong relationship between MAP and the second- and third-order statistics of rainfall extremes is well documented (Schaefer, 1990; Brath et al., 2003), and rainfall dominates the hydrologic response of abasin.

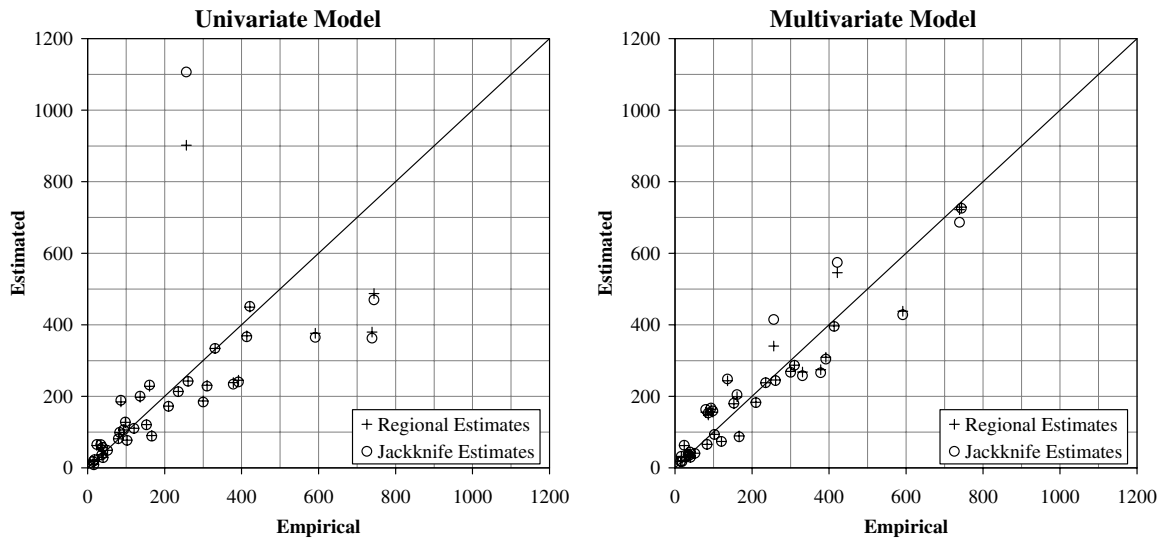


Figure 4 Simple regression between μ_X and A (left panel) and multivariate model (right panel): the empirical index-flood values are reported vs. the corresponding Regional and Jackknife Estimates.

The overall efficiency measure of the regional model (19) is $E = 0.913$. We assessed the robustness of this model using a jackknife cross-validation as described earlier. The cross-validation resulted in an overall efficiency measure equal to $E_{jk} = 0.881$. The rather high E and E_{jk} values and the marked agreement between regional and jackknife estimates that is visible in the scatter plots of Fig. 4 (right panel) indicate that the application of the jackknife procedure does not significantly influence the performance of the models, implying a strong robustness of the models (see e.g., Brath et al., 2001).

Once the model in (19) is identified, the intercept of the empirical MVE can be computed from (10) where the $\hat{\alpha}_i$ values for $i = 1, 2, \dots, v$, coincide with the estimates of C_i coefficients of (19), with $i = 1, 2, \dots, v$. The resulting MVE can be expressed in terms of unit flood of record, Q/A , as follows:

$$\ln\left(\frac{Q}{A}\right) = 2.756 + 0.699 \ln(A_{\text{imp}}) + 2.214 \ln\left(\frac{\text{MAP}}{1000}\right) - \ln(A) = 2.756 + \text{GCF}, \quad (19)$$

where GCF is the geomorphoclimatic factor used for illustrating the MVE in Fig. 2c. Graphical visualizations of the MVE in (20) are also provided in Figs. 2a and 2b, where the geomorphologic factor GF is defined as follows:

$$\text{GF} = A^{-0.301} \left(\frac{A_{\text{imp}}}{A}\right)^{0.699}, \quad (20)$$

with usual notation and units, and $\text{CF} = \text{MAP}$ (mm).

Exceedance probability of the expected REC and MVE

MVEs are generalizations of RECs that assume a multivariate relationship between index-flood and catchment characteristics. As documented in Section "Probabilistic multivariate envelopes", under the adopted hypotheses, the same prob-

abilistic statement can be assigned to both types of envelopes, hence the expected REC and the expected MVE for a given region share the same exceedance probability p_{EE} . p_{EE} can be expressed as a function of the effective number of observations n_{eff} , which, in turn, can be estimated as detailed in Section "Exceedance probability of the expected REC".

In the cross-correlation formula (5), estimates the parameters $\lambda_1 = 4.69 \cdot 10^{-05} \text{ (m}^{-1}\text{)}$ and $\lambda_2 = 2.42 \cdot 10^{-05} \text{ (m}^{-1}\text{)}$ (see Fig. 6) were obtained using WLS with weights equal to the corresponding number of concurrent annual floods. The λ_1 and λ_2 estimates enabled us to apply the algorithm in (4) to the regional dataset, which counted a total of 1035 sample-years of data, $n_1 = 10$ single observations, and a modeled

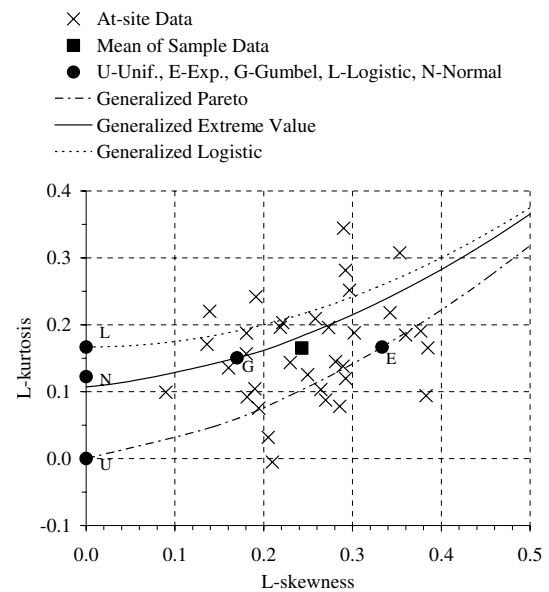


Figure 5 L moment ratios diagram for the study area (see Hosking and Wallis, 1997).

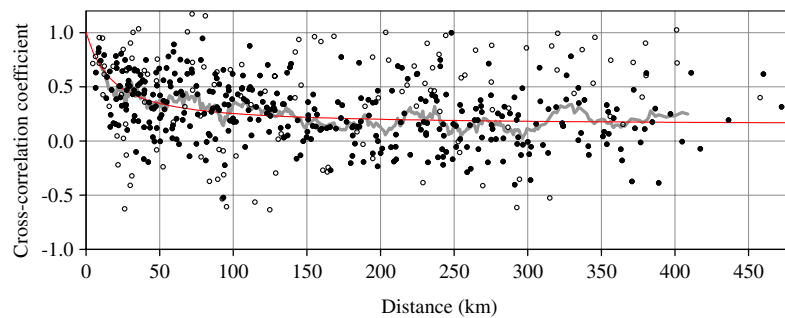


Figure 6 Empirical cross-correlation coefficients for flood sequences with length < 10 (open circles) and ≥ 10 (black dots); moving weighted average curves (thick line); cross-correlation formula proposed by Tasker and Stedinger (1989) fitted onto the empirical data (thin line).

average cross-correlation among the series equal to 0.272. The application of the algorithm produced an estimated number of effective observations equal to $\hat{n}_{\text{eff}} = 593$ sample-years of data. Fig. 6 illustrates significant scatter associated with the estimated cross-correlation coefficients, which is in part due to sampling variability. The weak connection between correlation and distance conveys little information. Fortunately, Castellarin (2006) observed that the cross-correlation has a marginal impact on the reliability of estimated p_{EE} values.

The exceedance probability p_{EE} in (6) requires the selection of suitable regional parent distribution to enable selection of a suitable parameter η . The GEV distribution is a suitable probabilistic model for representing the annual maximum series of flood flows in the study area, as shown by the L moment ratios diagram reported in Fig. 5 (see Hosking and Wallis, 1997; Chapter 5). The shape parameter $k = -0.111$ of the GEV distribution was estimated using the L moments (see e.g., Hosking and Wallis, 1997, Chapter 6). The quantile-unbiased GEV plotting-position proposed by Castellarin (2006), whose $\eta(k)$ parameter results in this case equal to 0.491, produces an estimate of the recurrence interval for the expected envelope $T_{EE} = 1/p_{EE}$ equal to 1165 years. This estimate of the recurrence interval applies both to expected REC and expected MVE for the considered group of 34 AMS of flood flows. Importantly, this estimate of T_{EE} takes into account the cross-correlation among the series and the actual distribution of data in time (e.g., missing data, different installation years for different gages, etc.).

The REC illustrated in Fig. 1 and the MVE reported in Fig. 2 can be applied for any ungaged basin in the study area to produce an estimate of the T_{EE} -year flood quantile, where $T_{EE} = 1165$ years. The REC flood quantile can be readily estimated on the basis of the drainage area of the basin, whereas the estimation of MVE flood quantile requires knowledge of the three geomorphoclimatic descriptors: A , A_{imp} , and MAP.

Evidently, the uncertainty and reliability of REC and MVE flood quantiles can be rather different. The previous section documented that the multiregression model outperforms the simple scaling relationship in (2). The improved performance of the multiregression model must result in a higher reliability of MVE flood quantiles with respect to the REC flood quantiles. In the following section we quantify the reliability of REC and MVE flood quantiles for ungaged basins.

Accuracy of REC and MVE flood quantiles

Reference regional design flood estimates

The proposed probabilistic interpretation of univariate and multivariate envelopes of extreme floods relies on the assumption of regional homogeneity in the sense of the index-flood hypothesis (Dalrymple, 1960). Hence, the index-flood method is a natural choice as a reference method for generating regional estimates of the design flood for comparison with REC and MVE quantiles (see also Castellarin, 2006).

Two different reference regional estimates of the T_{EE} -year flood are used for comparison with REC and MVE flood quantiles. Both regional estimates are computed assuming that the index-flood hypothesis holds (i.e., same probability distribution of standardized annual maximum peak flows for all sites belonging to the homogeneous region). We termed the first reference estimate the “true flood quantile”, $\hat{Q}(i, T_{EE})$. $\hat{Q}(i, T_{EE})$ refers to the ideal situation in which an observed AMS of flood peaks is available at a gaged site i . The second reference, $\hat{Q}(i, T_{EE})_{jk}$, represents the best regional estimate that one can obtain via of the index-flood procedure if site i were an ungaged site. $\hat{Q}(i, T_{EE})_{jk}$, as indicated by subscript jk , is evaluated through a jackknife resampling approach by neglecting the observations collected at site i . Appendix B details how $\hat{Q}(i, T_{EE})$ and $\hat{Q}(i, T_{EE})_{jk}$ were computed.

Uncertainty affects both $\hat{Q}(i, T_{EE})$ and $\hat{Q}(i, T_{EE})_{jk}$, and is obviously higher for $\hat{Q}(i, T_{EE})_{jk}$. Nevertheless, we termed the estimate $\hat{Q}(i, T_{EE})$ the “true flood quantile” to characterize it as the best possible estimate within the index-flood framework. The second index-flood quantile, $\hat{Q}(i, T_{EE})_{jk}$, is utilized in order to quantify the additional uncertainty that affects the index-flood approach when it is applied to ungaged sites.

Cross-validation procedure

The comparison of the different approaches (probabilistic univariate and multivariate envelopes of extreme floods) focuses on the estimation of the design-flood at ungaged sites. Consequently, we assessed the reliability of REC and MVE flood quantiles through a cross-validation procedure that simulates the ungaged conditions in turn at each and every

site. The structure of the cross-validation can be summarized as follows:

1. one gaging station (site i) is removed from the set of 34 streamgages;
2. resampled (or jackknife) empirical REC and MVE, indicated as REC_i and MVE_i respectively, are constructed by ignoring data belonging to site i , and their $T_{EE}(i) = T$ is then estimated (see Sections "Exceedance probability of the expected REC" and "Exceedance probability of the expected REC and MVE");
3. flood quantiles $\hat{Q}(i, T)_{REC}$ and $\hat{Q}(i, T)_{MVE}$ are estimated from REC_i and MVE_i respectively;
4. using T evaluated at step 2, the true and jackknife flood quantiles, $\hat{Q}(i, T)$ and $\hat{Q}(i, T)_{jk}$ respectively, are estimated for site i as described in Appendix B;
5. the following relative errors are then computed:

$$\varepsilon_{REC} = [\hat{Q}(i, T)_{REC} - \hat{Q}(i, T)] / \hat{Q}(i, T), \quad (21a)$$

$$\varepsilon_{MVE} = [\hat{Q}(i, T)_{MVE} - \hat{Q}(i, T)] / \hat{Q}(i, T), \quad \text{and} \quad (21b)$$

$$\varepsilon_{R-LMOM} = [\hat{Q}(i, T)_{jk} - \hat{Q}(i, T)] / \hat{Q}(i, T); \quad (21c)$$

where R-LMOM stands for regional L moments implementation of the index-flood approach;

6. steps 1–5 are repeated $M - 1$ times, with $M = 34$, for each one of the remaining streamgages.

It is important to note that $T_{EE}(i) = T$ refers to the effective regional sample-years of data that characterizes REC_i and MVE_i , and therefore varies for the 34 sites because the regional sample-years of data vary for each site during the cross-validation. Accordingly, $\hat{Q}(i, T)$ and $\hat{Q}(i, T)_{jk}$ are evaluated for the T value obtained for site i .

A comprehensive evaluation of the reliability of REC flood quantiles has been already performed with a very similar procedure by Castellarin (2006). Hence, we focus here on an evaluation of the improvement in the reliability of flood quantiles, if any, that can be acquired through the generalization of probabilistic envelope curves and the introduction of probabilistic multivariate envelopes of extreme floods. Our comparisons of the errors ε_{REC} and ε_{MVE} enables us to understand whether or not the reliability of flood quantiles benefits from the multivariate generalization of the envelope. In addition, the comparison of these errors with ε_{R-LMOM} shows how far the reliability of envelope flood quantiles is from the target reliability associated with the index-flood approach, a standard and proven design approach.

Results and discussion

Figs. 7 and 8 illustrate the results of the cross-validation experiments. The box-plots in Fig. 7 depict the distributions of relative errors obtained for the probabilistic univariate (REC) and multivariate (MVE) envelopes of flood extremes and for the index-flood approach (R-LMOM), for ungaged conditions. The bar-diagram in Fig. 8 compares the relative error values separately at all 34 sites.

The comparison between the ε_{R-LMOM} , ε_{REC} and ε_{MVE} values reported in Figs 7 and 8 shows that the reliability of the

three considered approaches is similar. The comparison between ε_{R-LMOM} and ε_{REC} confirms the results obtained by Castellarin (2006), who showed that, REC flood quantiles exhibit a reliability and accuracy which is comparable to index-flood quantiles. Fig. 7 also shows that ε_{REC} values are positively biased and present the largest dispersion around the expected error (largest interquartile distance). Fig. 7 documents the advantages associated with a multivariate generalization of the regional envelope of food extremes. The distribution of ε_{MVE} values shows a low interquartile distance (low dispersion around the expected error) and a small positive bias, when compared to ε_{REC} values.

The bias associated with the MVE flood quantiles (see Fig. 7) is in part due to the heterogeneity of the study region (see Section "Study area"). As noted by Castellarin et al. (2005, see Fig. 8 on p. 10), in a heterogeneous pooling-group of sites, a few discordant sites characterized by a distribution with a heavier right tail can exert a strong control on the envelope, resulting in a positive bias of the quantile estimates. On the contrary, classical regional flood frequency analysis (RFFA) studies indicate that limited degrees of regional heterogeneity do not significantly affect the reliability of flood quantile estimates resulting from the application of the index-flood hypothesis (see e.g., Lettenmaier et al., 1987; Stedinger and Lu, 1995). For this reason the comparison of REC and MVE quantiles with the "true flood quantiles" exhibits a positive bias. For the REC flood quantiles the bias is even greater (see Fig. 7), due to the fact that the regression model $\mu_X - A$ does not incorporate any information on the impervious portion of the catchment area nor on the climatic regime of each catchment (summarized by MAP).

Although the advantages of a multivariate approach over the index-flood approach are not evident (see Fig. 4), the improvements resulting from a multivariate interpretation

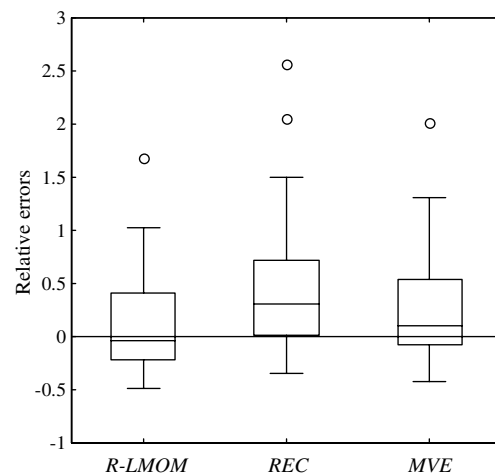


Figure 7 Results of the cross-validation: distributions of relative errors of the T -year flood quantiles estimated for the 34 considered sites with the index-flood approach (R-LMOM) and retrieved from the probabilistic regional envelope curve (REC) and multivariate envelope of extreme floods (MVE); each box-plots reports the minimum and maximum values, the 25th, 50th and 75th percentiles, highlighting the presence of outliers (circles).

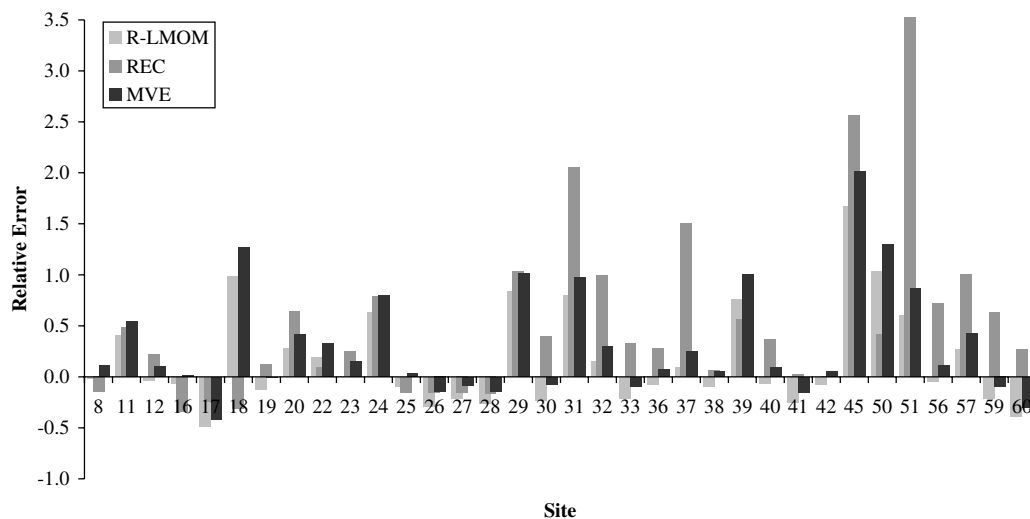


Figure 8 Results of the cross-validation: relative errors of the T_{EE} -year flood quantiles estimated for the 34 considered sites with the index-flood approach (R-LMOM) and retrieved from the probabilistic regional envelope curve (REC) and multivariate envelope of extreme floods (MVE).

of probabilistic regional envelopes is rather striking in terms of improved reliability of estimated flood quantiles. This outcome was expected because the multivariate model of index-flow enables us to integrate information on climate and permeability of the catchment. This information is relevant to μ_X and is crucial for characterizing the physical processes that generate extreme floods, and hence for a better identification of the envelope.

The results in Fig. 7 refer to the region as a whole whereas the bar diagram in Fig. 8 enables us to analyze the performance of each methodology across sites. In particular, it is interesting to observe that the sign of the three errors is not always consistent as in some cases one approach overestimates the flood quantiles and the two remaining approaches underestimate it, and vice versa. We also observe that the best performing approach depends on the site being considered. This suggests that RECs and, particularly, MVEs can provide valuable complements to design-flood estimates derived from traditional regionalization approaches such as the index flood method.

RECs and MVEs were not developed as design flood methods, hence it is important to highlight that with respect to traditional RFFA techniques, probabilistic envelopes of extreme floods suffer from the disadvantage of being associated with a single (arbitrary) recurrence interval, which is the recurrence interval of the expected envelope T_{EE} . Therefore, RECs and MVEs can only be utilized for the estimation of T_{EE} -year floods. Nevertheless, Figs. 7 and 8 show that, under ungaged conditions, the reliability of REC based flood quantiles is similar to the reliability of flood quantiles estimated through traditional RFFA approaches and MVE based flood quantiles are practically as reliable as RFFA flood quantiles. More importantly, with respect to traditional RFFA techniques, RECs and MVEs present the noteworthy feature of providing a comprehensive and readily interpretable visualization of the bound on our current experience of extreme flood flows in a region (see Figs. 1 and 2).

Note that we decided to construct the empirical envelopes for the univariate (REC) and multivariate (MVE) case by fitting the empirical annual floods (i.e., index-floods) and subsequently shifting the univariate and multivariate regression relations upward by modifying their intercept. This approach is consistent with the adopted assumption of regional homogeneity in the sense of the index-flood hypothesis. Nevertheless, alternative multivariate estimation procedures can be applied, such as fitting the envelope by constraining all the residuals to be positive while minimizing distance from envelope to observations. Future analysis will address this issue, investigating whether or not the use of different estimation methods would produce a better envelope and thus a higher reliability of the resulting design floods.

As a final remark, it cannot be denied that the index-flood approach (R-LMOM) outperforms the two considered envelope-based approaches (REC and MVE) for the whole study region. In fact, our goal here is not to present a replacement for RFFA approaches. Nevertheless, our results confirm the value of envelope-based approaches to design-flood estimation at ungaged sites and indicate that MVEs are very promising evolutions of RECs.

Conclusions

Our study presents a multivariate generalization of probabilistic regional envelope curves (RECs), recently proposed by Castellarin et al. (2005) and Castellarin (2006) for estimating the design flood at ungaged sites. Our main objectives were to:

- provide a multivariate generalization of previous univariate probabilistic RECs;
- demonstrate that the exceedance probability of the generalized envelope, p_{EE} , can be estimated using the same algorithm proposed for RECs;

- quantify the reliability of $1/p_{EE}$ -year flood quantiles estimated at ungaged sites from generalized multivariate envelopes of extreme floods (MVEs) and compare it with the reliability of flood quantiles based on (1) univariate envelope curves (RECs) and (2) traditional regional flood frequency analysis (RFFA).

We document a multivariate extension of RECs which we term probabilistic multivariate regional envelopes of extreme floods (MVEs). We also propose a two- or three-dimensional representation of MVEs that provides a graphical visualization of the current bound on our experience of extreme floods in a multivariate framework.

We prove that expected RECs and expected MVEs (i.e., the univariate and multivariate envelopes that, on average, are expected to bound the extreme flood experience for a group of cross-correlated flood sequences) are associated with the same exceedance probability, p_{EE} . Therefore, the algorithm developed by [Castellarin \(2006\)](#) for estimating p_{EE} from real world dataset can be applied for both expected RECs and MVEs.

A cross-validation for a region in north-central Italy enables us to assess the reliability of flood quantiles estimated for ungaged sites using a traditional RFFA procedure, and using an empirical REC and MVE. The results of the cross-validation demonstrate the superiority of MVE over REC and also show that the MVE indices of performance are practically equivalent to the RFFA indices of performance.

In conclusion, our study documents that a multivariate generalization of RECs, which we term MVEs, represent practical and easy-to-use tools to (1) graphically and quantitatively summarize the extreme flooding experience in a region in a multivariate framework; (2) determine plausible extreme-flood values at ungaged sites and (3) provide a realistic estimate of the recurrence intervals associated with such extreme floods. The general applicability of MVEs, and their real potential for predicting design floods at ungaged sites needs further experimentation in different geographical contexts. More importantly, MVEs, along with RECs, should be seen as valuable and useful complements to traditional RFFA techniques, rather than substitutes.

Extensions to the multivariate regional envelope methodology introduced here will likely benefit from: (1) application of generalized least squares (GLS) regression methods (see [Kroll and Stedinger, 1998](#)) for both identification and estimation of the multivariate envelope surface and for estimation of the relationship between cross-correlation and distance between sites; (2) an improved understanding of the theoretical properties of record floods and their covariance structure; and (3) an improved understanding of the geomorphic and climatologic factors which control and/or dictate the upper bound on flood discharges.

Acknowledgements

The research was partially supported by the Italian MIUR (Ministry of Education, University and Research) through the research grant titled "Characterisation of average and extreme flows in Ungaged Basins by Integrated use of data-based methods and hydrological modelling – CUBIST".

Appendix A. Homogeneity testing

The homogeneity test proposed by [Hosking and Wallis \(1997, Chapter 4\)](#) assesses the homogeneity of a group of sequences at three different levels by focusing on three measures of dispersion for different orders of the sample L moment ratios (see [Hosking and Wallis \(1997\)](#) for an explanation of L moments),

1. A measure of dispersion for the L coefficient of variation, $L-Cv$,

$$V_1 = \sum_{i=1}^R n_i (t_{2(i)} - \bar{t}_2)^2 / \sum_{i=1}^R n_i. \quad (A1)$$

2. A measure of dispersion for both the $L-Cv$ and the L -skewness coefficients in the $L-Cv$ – L -skewness space,

$$V_2 = \sum_{i=1}^R n_i \left[(t_{2(i)} - \bar{t}_2)^2 + (t_{3(i)} - \bar{t}_3)^2 \right]^{1/2} / \sum_{i=1}^R n_i. \quad (A2)$$

3. A measure of dispersion for both the L -skewness and the L -kurtosis coefficients in the L -skewness– L -kurtosis space,

$$V_3 = \sum_{i=1}^R n_i \left[(t_{3(i)} - \bar{t}_3)^2 + (t_{4(i)} - \bar{t}_4)^2 \right]^{1/2} / \sum_{i=1}^R n_i, \quad (A3)$$

where \bar{t}_2 , \bar{t}_3 , and \bar{t}_4 are the group mean of $L-Cv$, L -skewness, and L -kurtosis, respectively; $t_{2(i)}$, $t_{3(i)}$, $t_{4(i)}$, and n_i are the values of $L-Cv$, L -skewness, L -kurtosis and the sample size for site i ; and R is the number of sequences.

The underlying concept of the test is to measure the sample variability of the L moment ratios and compare it to the variation that would be expected in a homogeneous group. The expected mean value and standard deviation of these dispersion measures for a homogeneous group, μ_{V_k} and σ_{V_k} respectively, are assessed through repeated simulations, by generating homogeneous groups of basins having the same record lengths as those of the observed data following the methodology proposed by [Hosking and Wallis \(1997, Chapter 4\)](#). The heterogeneity measures are then evaluated using the following expression:

$$H_k = \frac{V_k - \mu_{V_k}}{\sigma_{V_k}}; \quad \text{for } k = 1, 2, 3. \quad (A4)$$

Hosking and Wallis suggest that a group of sites may be regarded as "acceptably homogeneous" if $H_k < 1$, "possibly heterogeneous" if $1 \leq H_k < 2$, and "definitely heterogeneous" if $H_k \geq 2$. The authors also point out that H_2 and H_3 lack power to discriminate between homogeneous and heterogeneous regions, whereas H_1 , has much better discriminatory power.

Appendix B. Computation of reference regional design flood estimates

Under the index-flood assumption, the T -year flood for site i , $Q(i, T)$, is,

$$Q(i, T) = \mu_X(i) \cdot x'(T), \quad (B1)$$

where $\mu_X(i)$ is the index-flood for site i (assumed in this study to be equal to the mean annual flood), whereas $x'(T)$ is the regional dimensionless flood-quantile with exceedance probability $1/T$ for the homogeneous region that contains site i .

The first reference regional estimate, $\hat{Q}(i, T)$, represents the “true flood quantile” and refers to an optimal scenario, for which an observed AMS of flood peaks is available at site i (gaged conditions). $\hat{Q}(i, T)$ is computed as the average of the observed annual maxima (sample estimate of $\mu_X(i)$) times a regional estimate of $x'(T)$ resulting from a GEV regional parent distribution, whose parameters are identified by using the L moments method (see e.g., Hosking and Wallis, 1997, Chapter 6). The estimation of the regional parent is performed considering all AMS observed in the study region.

The second reference regional estimate, $\hat{Q}(i, T)_{jk}$, represents the best regional estimate that one can obtain through the application of the index-flood procedure if site i were an ungaged site (ungaged conditions). $\hat{Q}(i, T)_{jk}$ is evaluated using a jackknife resampling approach by neglecting the observations collected at site i , and is the product of an indirect estimate of $\mu_X(i)$, $\hat{\mu}_X(i)_{jk}$, and a resampled regional quantile, $\hat{x}'(i, T)_{jk}$:

- $\hat{\mu}_X(i)_{jk}$ is estimated using a multiregression model that is identified by discarding the flood data observed at site i (see e.g., Brath et al., 2001; Castellarin et al., 2001). In particular, we the estimates $\hat{\mu}_X(i)_{jk}$ are based on 34 estimates resulting from the cross-validation of the multiregression model (19) presented in Section “Empirical REC and MVE”; these estimates are illustrated in the right panel of Fig. 4 (Jackknife Estimates);
- $\hat{x}'(i, T)_{jk}$ is estimated from a GEV regional parent, whose parameters are estimated by applying the method of L moments illustrated by Hosking and Wallis (1997, Chapter 6) to all sequences except for the one observed at site i .

References

- Brath, A., Castellarin, A., Franchini, M., Galeati, G., 2001. Estimating the index flood using indirect methods. *Hydrol. Sci. J.* 46 (3), 399–418.
- Brath, A., Castellarin, A., Montanari, A., 2003. Assessing the reliability of regional depth-duration-frequency equations for gaged and ungaged sites. *Water Resour. Res.* 39 (12), 1367. doi:10.1029/2003WR002399.
- Burn, D.H., 1990. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resour. Res.* 26 (10), 2257–2265.
- Castellarin, A., 2006. Application of probabilistic envelope curves for design-flood estimation at ungaged sites, *Water Resour. Res.*, accepted for publication.
- Castellarin, A., Burn, D.H., Brath, A., 2001. Assessing the effectiveness of hydrological similarity measures for regional flood frequency analysis. *J. Hydrol.* 241 (3–4), 270–285.
- Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A., Brath, A., 2004. Regional flow-duration curves: reliability for ungaged basins. *Adv. Water Resour.* 27, 953–965.
- Castellarin, A., Vogel, R.M., Matalas, N.C., 2005. Probabilistic behavior of a regional envelope curve. *Water Resour. Res.* 41, w06018. doi:10.1029/2004wr003042.
- Chokmani, K., Ouarda, T.B.M.J., 2004. Physiographical space-based kriging for regional flood frequency estimation at ungaged sites. *Water Resour. Res.* 40, W12514. doi:10.1029/2003WR002983.
- Costa, J.E., 1987. A comparison of the largest rainfall-runoff floods in the united States with those of the people’s Republic of China and the World. *J. Hydrol.* 96, 101–115.
- Creager, W.P., Justin, J.D., Hinds, J., 1966. *Engineering for Dams*, vol. I. General Design, 9th ed. John Wiley and Sons, New York, pp. 99–140.
- Crippen, J.R., 1982. Envelope curves for extreme flood events. *J. Hydraul. Eng., ASCE* 108 (HY10), 1208–1212.
- Crippen, J.R. Bue, C.D., 1977. Maximum flood flows in the conterminous United States, Geological Survey Water-Supply Paper 1887, United States Printing Office, Washington DC, 1977.
- Cunnane, C., 1978. Unbiased plotting positions – a review. *J. Hydrol.* 37, 205–222.
- Dalrymple, T., 1960. Flood frequency analyses, US Geol. Survey Wat. Supply Paper 1543-A, Reston, VA.
- Fuller, W.E., 1914. Flood flows. *Trans. Am. Soc. Civ. Eng., Paper No.* 1293, 77, pp. 564–617.
- Hersch, R.W., 2002. The world’s maximum observed floods. *Flow Meas. Instrum.* 13, 231–235.
- Hosking, J.R.M., Wallis, J.R., 1997. *Regional frequency analysis – an approach based on L-moments*. Cambridge University Press, New York, p. 224.
- International Association of Hydrological Sciences (IAHS), 2003. *World Catalogue of Maximum Observed Floods*, IAHS Publ. 284, Compiled by R. Hersch, IAHS Press, Wallingford, UK, p. 285.
- Jarvis, C.S., 1925. Flood flow characteristics. *Trans. ASCE* 88, 985–1032.
- Jenkinson, A.F., 1955. The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quart. J. Roy. Met. Soc.* 81, 158–171.
- Kadoya, M., 1992. Study on record flood peaks in Japan. *Proc. Jpn. Acad., Ser. B* 68, 133–138.
- Kroll, C.N., Stedinger, J.R., 1998. Regional hydrologic analysis: ordinary and generalized least squares revisited. *Water Resour. Res.* 34 (1), 121–128.
- Lettenmaier, D.P., Wallis, J.R., Wood, E.F., 1987. Effect of regional heterogeneity on flood frequency estimation. *Water Resour. Res.* 23 (2), 313–323.
- Marchetti, G., 1955. Sulle massime portate di piena osservate nei corsi d’acqua italiani a tutto il 1953 (in Italian). *Giorn. Genio Civ.* 93, 3–4.
- Matalas, N.C., Langbein, W.B., 1962. Information content of the mean. *J. Geophys. Res.* 67 (9), 3441–3448.
- Meyer, A.F., 1917. *Elements of Hydrology*. John Wiley and Sons, New York, p. 487.
- Mimikou, M., 1984. Envelope curves for extreme flood events in north-western and western Greece. *J. Hydrol.* 67, 55–66.
- Nash, J.E., Sutcliffe, J.E., 1970. River flow forecasting through conceptual models, Part 1 – A discussion of principles. *J. Hydrol.* 10 (3), 282–290.
- Ouarda, T.B.M.J., Girard, C., Cavadias, G.S., Bernard, B., 2001. Regional flood frequency estimation with canonical correlation analysis. *J. Hydrol.* 254, 157–173.
- Robson, A.J., Reed, D.W., 1999. 3: Statistical procedures for flood frequency estimation. *Flood Estimation Handbook*, vol. 3. Institute of Wallingford, Wallingford, p. 338.
- Schaefer, M.G., 1990. Regional analysis of precipitation annual maxima in Washington State. *Water Resour. Res.* 26 (1), 119–131.
- Shao, J., Tu, D., 1995. *The Jackknife and Bootstrap*. Springer-Verlag, New York.
- Shu, C., Burn, D.H., 2004. Artificial neural network ensembles and their application in pooled flood frequency analysis. *Water Resour. Res.* 40 (9), W09301.
- Stedinger, J.R., 1983. Estimating a regional flood frequency distribution. *Water Resour. Res.* 19 (2), 503–510.

- Stedinger, J.R., Lu, L., 1995. Appraisal of regional and index flood quantile estimators. *Stochastic Hydrol. Hydraul.* 9 (1), 49–75.
- Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events. In: Maidment, D.A. (Ed.), *Handbook of Hydrology*. McGraw-Hill, New York (Chapter 18).
- Tasker, G.D., Stedinger, J.R., 1989. An operational GLS model for hydrologic regression. *J. Hydrol.* 111, 361–375.
- Troutman, B.M., Karlinger, M.R., 2003. Regional flood probabilities. *Water Resour. Res.* 39 (4), 1095. doi:10.1029/2001WR001140.
- Vogel, R.M., Wilson, I., 1996. Probability distribution of annual maximum, mean, and minimum streamflows in the United States. *J. Hydrol. Eng.* 1 (2), 69–76.
- Wiesberg, S., 1985. *Applied Linear Regression*, 2nd ed. John Wiley, New York.