Bivariate Statistical Approach to Check Adequacy of Dam Spillway

C. De Michele[1]; G. Salvadori[2]; M. Canossi[3]; A. Petaccia[4]; and R. Rosso[5]

Abstract: The problem of selecting the appropriate design flood is a constant concern to dam engineering and, in general, in the hydrological practice. Overtopping represents more than 40% of dam failures in the world. The determination of the design flood is based in some cases on the *T*-year quantile of flood peak, and in other cases considering also the *T*-year quantile of flood volume. However, flood peak and flood volume have a positive (strong or weak) dependence. To model properly this aspect a bivariate probability distribution is considered using the concept of 2-Copulas, and a bivariate extreme value distribution with generalized extreme value marginals is proposed. The peak–volume pair can then be transformed into the correspondent flood hydrograph, representing the river basin response, through a simple linear model. The hydrological safety of dams is considered checking adequacy of dam spillway. The reservoir behavior is tested using a long synthetic series of flood hydrographs. An application to an existing dam is given.

DOI: 10.1061/(ASCE)1084-0699(2005)10:1(50)

CE Database subject headings: Dam safety; Hydrographs; Probability distribution; Spillways; Statistics.

Introduction

A large number of existing dams in the world were built during the 20th century under engineering, social, economic, and climate conditions different from those to be faced in this century. Data availability, process knowledge, and modeling techniques were at that time less sophisticated than today. During the last century, a large amount of historical and proxy information, i.e., hourly reservoir levels and the rules to the operating outflows, were collected by the dam regulators. This information can improve the knowledge of the dam inflows and consequently the dam safety issues.

Dam failures have been significantly reduced in the last few decades (Berga 1998); the percentage of failures before 1950 was 2.3%, while for dams constructed from 1951 to 1982 it reduced to 0.2%, and since 1982 is only 0.09%. This reduction indicates that progress has been achieved in dam safety. Recent regulations, codes, and guidelines emphasize the importance of the spillway design flood as a key factor to dam safety (De Almeida and Viseu 1997). It is interesting to note that overtopping represents more than 40% of dam failures in the world and has been the cause of many other accidents (Committee on Failures and Accidents to Large Dams of the United States Committee on Large Dams 1975). In the United States, over 2,000 dams (3% of the 75,000 United States dams) have been identified as potential hazards to lives in upstream or downstream areas, due to problems of inadequate spillway capacity (ASCE 2000).

The main purpose of the paper is to outline a very general model describing the possible bivariate behavior of the random variables flood peak and flood volume, which are of primary interest in hydrological practice. In particular, here the attention is focused on testing the adequacy of dam spillway. A methodology for evaluating flood hydrographs is provided: it is based on a bivariate analysis of the maximum annual values of flood peak and flood volume. A bivariate extreme value distribution is considered using the mathematical concept of 2-Copulas (see details below). A hydrograph is obtained using the flood-peak and flood-volume pair with the river basin response represented through a lumped model. Successively, a synthetic series of flood-peak and flood-volume pairs is generated using Monte Carlo simulation. From this, a series of flood hydrographs is then obtained. Operating the reservoir routing, it is possible to test adequacy of the dam spillway. An application to the Ceppo Morelli dam located in the Anza river basin in northern Italy is presented.

Calculating Flood Hydrograph

The flood peak has a fundamental role in both assessing the hydrologic safety of dams and checking adequacy of the dam spillway. However, the flood volume can play an important role in the definition of the spillway design flood, and, consequently, may significantly influence the hydrologic safety of the dam. Generally, flood peak and flood volume are two statistically dependent random variables. A joint analysis of flood peak and flood volume can be used to determine the design flood hydrograph. Below, a statistical procedure for the evaluation of flood hydrograph is pro-

[1]Assistant Professor, DIIAR-Politecnico di Milano, 32 Piazza Leonardo da Vinci, Milano I-20133, Italy. E-mail: carlo.demichele@polimi.it

[2]Assistant Professor, Dip. di Matematica, Univ. di Lecce, Provinciale Lecce-Arnesano, P.O. Box 193, I-73100 Lecce, Italy. E-mail: gianfausto.salvadori@unile.it

[3]Graduate Student, DIIAR-Politecnico di Milano, 32 Piazza Leonardo da Vinci, Milano I-20133, Italy. E-mail: michele.canossi@polimi.it

[4]Director of Hydraulic Section-RID, Via Curtatone 3, I-00185 Roma, Italy. E-mail: alberto.petaccia@registroitalianodighe.it

[5]Professor, DIIAR-Politecnico di Milano, 32 Piazza Leonardo da Vinci, Milano I-20133, Italy. E-mail: renzo.rosso@polimi.it

Note. Discussion open until June 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on March 17, 2003; approved on May 6, 2004. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 10, No. 1, January 1, 2005. ©ASCE, ISSN 1084-0699/2005/1-50–57/$25.00.

posed based on: (1) a bivariate analysis of the maximum annual values of flood peak and flood volume and (2) a lumped model to calculate the flood hydrograph from the values of peak and volume.

Peak–Volume Analysis

In general, the bivariate analysis of two random variables can be carried out exploiting the mathematical tool of "2-Copulas," which is outlined below. Then, an application to the analysis of the statistical dependence between flood peak and flood volume is illustrated.

Overview of 2–Copulas

The problem of specifying a probability model for dependent bivariate observations $(X_1,Y_1),\ldots,(X_n,Y_n)$ from a population with a non-normal distribution function F_{XY} can be simplified by expressing F_{XY} in terms of its marginals F_X and F_Y and an associated dependence function $\mathbf{C}$, implicitly defined through the functional identity $F_{XY}=\mathbf{C}(F_X,F_Y)$. A natural way of studying bivariate data thus consists of separately estimating the dependence function and the marginals. This two-step approach to stochastic modeling is often convenient, since many tractable models are readily available for the marginal distributions. It is clearly appropriate when the marginals are known, and it is invaluable as a general strategy for data analysis in that it enables the dependence structure to be investigated independently of marginal effects.

Let $\mathbf{I}=[0,1]$. A two-dimensional Copula (or 2-Copula) is a bivariate function $\mathbf{C}:\mathbf{I}\times\mathbf{I}\to\mathbf{I}$ such that

(1) for all u, $z\in\mathbf{I}$ it holds

$$\mathbf{C}(u,0)=0$$

$$\mathbf{C}(u,1)=u$$

$$\mathbf{C}(0,z)=0$$

and

$$\mathbf{C}(1,z)=z$$

(2) for all u_1, u_2, z_1, $z_2\in\mathbf{I}$ such that $u_1\leqslant u_2$ and $z_1\leqslant z_2$ it holds

$$\mathbf{C}(u_2,z_2)-\mathbf{C}(u_2,z_1)-\mathbf{C}(u_1,z_2)+\mathbf{C}(u_1,z_1)\geqslant 0$$

For all the mathematical details omitted see Joe (1997) and Nelsen (1999); for an application to rainfall, see De Michele and Salvadori (2003) and Salvadori and De Michele (2004). The link between 2-Copulas and bivariate distributions is provided by Sklar's theorem:

Let X, Y be continuous random variables and let F_{XY} be their joint distribution function with marginals F_X and F_Y. Then there exists a unique 2-Copula $\mathbf{C}$ such that

$$F_{XY}(x,y)=\mathbf{C}(F_X(x),F_Y(y)) \tag{1}$$

for all x, y. Conversely, if $\mathbf{C}$ is a 2-Copula and F_X and F_Y are distribution functions, then F_{XY} is a joint distribution function with marginals F_X andF_Y.

The interesting point is that the properties of F_{XY} can be discussed in terms of the structure of $\mathbf{C}$: in fact, it is precisely the 2-Copula that captures many of the features of a joint distribution, and measures of association and dependence properties between random variables can be investigated in terms of 2-Copulas. Actually, a 2-Copula exactly describes and models the dependence structure between random variables, independently of the marginal laws of the variables involved. Clearly, this provides a large freedom in choosing the univariate marginal distributions once the desired dependence framework has been selected, and it usually makes it easier to formulate bivariate (and/or multivariate) models. Incidentally, we observe that all the bivariate models present in the literature can easily be described in terms of proper 2-Copulas.

In the sequel we shall refer to a particular class of 2-Copulas, i.e., the Gumbel's family (Gumbel 1960)—see also Joe (1997) and Nelsen (1999). The analytical expression is

$$\mathbf{C}_\delta(u,z)=\exp\{-[(-\ln u)^\delta+(-\ln z)^\delta]^{1/\delta}\} \tag{2}$$

where u, $z\in\mathbf{I}$ and $\delta\in[1,\infty[$. Here δ represents the dependence parameter. The (limit) case $\delta=1$ corresponds to independent variables, with $\mathbf{C}_1(u,z)=uz$; the (limit) case $\delta\to\infty$ corresponds to complete dependence between the variables. Note that this family of 2-Copulas models positively dependent variables, as is of interest for the random variables considered here.

An interesting property of Gumbel's 2-Copulas is that these are Archimedean (Nelsen 1999); this means that a 2-Copula $\mathbf{C}_\delta$ is the solution of the functional equation $\gamma(\mathbf{C}_\delta(u,z))=\gamma(u)+\gamma(z)$, where the generator γ: $\mathbf{I}\to[0,\infty[$ is a continuous, convex, strictly decreasing function such that $\gamma(1)=0$. In the present case we have (Nelsen 1999), $\gamma(t)=(-\ln t)^\delta$, where $t\in\mathbf{I}$. Kendall's τ rank correlation coefficient can be expressed as a one-to-one function of δ as

$$\tau(\delta)=1+4\int_{\mathbf{I}}\frac{\gamma(l)}{\gamma'(l)}dl \tag{3}$$

which yields, after some algebra

$$\tau(\delta)=\frac{\delta-1}{\delta} \tag{4}$$

Clearly the above relation between δ and τ confers a natural interpretation to δ as an association parameter. Indeed, since τ is a measure of association based on the ranks, this suggests how δ might be estimated in situations where the marginals are unknown (see, e.g., Genest and Rivest 1993; Carriere 1994; and Nelsen 1999).

Let (u_l,z_l), $l=1,\ldots,m$, denote a sample of size m from a continuous bivariate distribution. The empirical 2-Copula $\mathbf{c}_m$ is the function given by, for $i,j=1,\ldots,m$

$$\mathbf{c}_m\left(\frac{i}{m},\frac{j}{m}\right)=\frac{m_{ij}}{m} \tag{5}$$

where m_{ij}=number of sample pairs (u,z) such that $u\leqslant u_{(i)}$ and $z\leqslant z_{(j)}$, with the $u_{(i)}$s and the $z_{(j)}$s denoting the order statistics from the sample.

Using the empirical 2-Copula frequency $\mathbf{c}_m$, an estimator $\hat{\tau}$ of τ is given by

$$\hat{\tau}=\frac{2m}{m-1}\sum_{i=2}^{m}\sum_{j=2}^{m}\sum_{p=1}^{i-1}\sum_{q=1}^{j-1}\left[\mathbf{c}_m\left(\frac{i}{m},\frac{j}{m}\right)\mathbf{c}_m\left(\frac{p}{m},\frac{q}{m}\right)-\mathbf{c}_m\left(\frac{i}{m},\frac{q}{m}\right)\mathbf{c}_m\left(\frac{p}{m},\frac{j}{m}\right)\right] \tag{6}$$

Once an estimate $\hat{\tau}$ of τ is obtained, it is then possible to calculate an estimate $\hat{\delta}$ of δ using Eq. (4), and thus select a well-defined 2-Copula from Gumbel's family. Most importantly, we observe that such a procedure does not depend upon the marginal laws involved, which then need not be known or estimated in advance;

indeed, in practical applications, empirical 2-Copulas may represent a fundamental tool for fitting a given dependence structure to the available data (see, e.g., Deheuvels 1979; Genest and Rivest 1993; Nelsen 1999; De Michele and Salvadori 2003; Salvadori and De Michele 2004).

Below we show how Gumbel's 2-Copulas are suitable candidates to model the dependence between flood peak and flood volume.

Peak–Volume Model

The random variables of interest here are the maximum annual values of flood peak, Q_{max}, and flood volume, V; as a marginal distribution function for both Q_{max} and V we use the generalized extreme value (GEV) distribution

$$F_{Q_{max}}(q) = \begin{cases} \exp\left[-\left(1-\kappa_Q \dfrac{q-\varepsilon_Q}{\alpha_Q}\right)^{1/\kappa_Q}\right] & q > \varepsilon_Q + \dfrac{\alpha_Q}{\kappa_Q} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

$$F_V(v) = \begin{cases} \exp\left[-\left(1-\kappa_V \dfrac{v-\varepsilon_V}{\alpha_V}\right)^{1/\kappa_V}\right] & v > \varepsilon_V + \dfrac{\alpha_V}{\kappa_V} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $\varepsilon_Q, \varepsilon_V$ are location parameters, $\alpha_Q, \alpha_V > 0$ are scale parameters, and $\kappa_Q, \kappa_V < 0$ are shape parameters. We only consider negative shape parameters in order to deal with upper-unbounded random variables. In addition we have $\varepsilon_Q \geqslant -\alpha_Q/\kappa_Q$ and $\varepsilon_V \geqslant -\alpha_V/\kappa_V$ because Q_{max} and V are positive defined. The constraint $\kappa_Q, \kappa_V < 0$ yields (asymptotic) excess probability functions with an algebraic falloff, i.e., for $x \gg 1$

$$1 - F_{GEV}(x) \approx x^{1/\kappa} \tag{9}$$

and hence only the moments of order less than $-1/\kappa$ exist. It is just the presence of such heavy tails that makes GEV laws useful for describing extreme phenomena.

In order to model the dependence between Q_{max} and V we use a 2-Copula from Gumbel's family and Sklar's theorem. Thus the joint distribution $F_{Q_{max}V}$ of Q_{max} and V is given by

$$\begin{aligned} F_{Q_{max},V}(q,v) &= \mathbf{C}_\delta(F_{Q_{max}}(q), F_V(v)) \\ &= \exp\{-[(-\ln F_{Q_{max}}(q))^\delta + (-\ln F_V(v))^\delta]^{1/\delta}\} \end{aligned} \tag{10}$$

It is then easy to obtain the joint density function, $f_{Q_{max},V}$ by differentiating $F_{Q_{max},V}$ with respect to Q_{max} and V. Then, inverting Eq. (10), it is possible to determine the flood peak and volume with a given frequency of occurrence or return period.

Return Period of Pair (Q_{max}, V) and Comparison between Univariate and Bivariate Analysis

The determination of the design flood is commonly based on the quantile q_T of the maximum annual flood peak Q_{max} with return period $T_{Q_{max}}$; alternatively, the quantile v_T of the maximum annual flood volume V with the same return period $T_V = T_{Q_{max}}$ is considered. Thus, essentially, an univariate approach is adopted in practice. In particular, the spillway design flood is usually calculated choosing a return period $T = T_{Q_{max}} = T_V = 1{,}000$ years (or, sometimes, $T = 5{,}000$ or 10,000 years, depending upon the country considered).

However, since in general flood peak and flood volume have a positive (strong or weak) dependence, the design flood should be determined considering the return period of the pair (Q_{max}, V); in other words, the flood event would be correctly identified only by considering the joint variability of the relevant pair of random variables.

Now, at least two different "design events" can be defined in the bivariate case. In fact, having fixed a return period T, we may consider as critical events the following ones:

1. (OR case) either $Q_{max} > q_T$, or $V > v_T$, or both, i.e.

$$E_{or} = \{Q_{max} > q_T \text{ or } V > v_T\} \tag{11}$$

2. (AND case) both $Q_{max} > q_T$ and $V > v_T$, i.e.

$$E_{and} = \{Q_{max} > q_T \text{ and } V > v_T\} \tag{12}$$

In simple words: for E_{or} to happen it is sufficient that either Q_{max} or V (or both) exceed given thresholds; instead, for E_{and} to happen it is necessary that both Q_{max} and V are larger than prescribed values. Thus, two different joint return periods can be defined accordingly

$$T_{or} = \frac{1}{P[Q_{max} > q_T \text{ or } V > v_T]} = \frac{1}{1 - \mathbf{C}_\delta(u_T, z_T)} \tag{13}$$

$$T_{and} = \frac{1}{P[Q_{max} > q_T \text{ and } V > v_T]} = \frac{1}{1 - u_T - z_T + \mathbf{C}_\delta(u_T, z_T)} \tag{14}$$

where $u_T = F_{Q_{max}}(q_T)$ and $z_T = F_V(v_T)$. Since for Archimedean copulas $\mathbf{C}_\delta(x,x) < x$, then necessarily $T_{or} < T < T_{and}$. This last inequality has important consequences. In fact, if a return period T is fixed, then E_{or} appears more frequently than expected (since $T_{or} < T$); thus, E_{or} is not a T-years bivariate event, and if E_{or} were used as a "critical" design event, in order to have $T_{or} = T$ the marginal quantiles q_T, v_T should be increased; in other words, if q_T, v_T were used as "critical" design quantiles, then the resulting work would be underdimensioned, and would be at risk! On the contrary, E_{and} appears less frequently than expected (since $T_{and} > T$); again, E_{and} is not a T-years bivariate event, and if E_{and} were used as a "critical" design event, in order to have $T_{and} = T$ the marginal quantiles q_T, v_T should be decreased; thus, if q_T, v_T were used as "critical" design quantiles, then the resulting work would be overdimensioned, yielding a waste of money!

Note that the above bivariate analysis also includes, as a very particular case, the univariate one: in fact, considering the copula used in this paper, as the dependence parameter $\delta \to \infty$ (i.e., as V becomes almost surely a function of Q_{max}, and vice versa), both $T_{or} \to T$ and $T_{and} \to T$. As an example, in the present case $\delta \approx 3.055$; then, fixing $T = 1{,}000$ years, we obtain $T_{or} \approx 798$ years and $T_{and} \approx 1{,}341$ years, i.e., $T_{or} \approx 79\%\ T$ and $T_{and} \approx 134\%\ T$, in agreement with the previous explanation. In general, the same conclusions remain valid for other families of Archimedean copulas.

Evidently, the analysis of bivariate return periods is greatly facilitated by using copulas; furthermore, many other kinds of joint events can be considered, and all the corresponding pairs (Q_{max}, V) having a prescribed return period can easily be calculated: this gives the hydrologists a precise understanding of the stochastic joint dynamics of the variables of interest, as well as the possibility of correctly sizing the works.

Monte Carlo Generation of Flood Peak and Flood Volume

An algorithm to generate random variables (Q_{max}, V) with 2-Copula $\mathbf{C}_\delta$ is as follows (Nelsen 1999). Let

$$s_u(z) = \frac{\partial}{\partial u}\mathbf{C}_\delta(u,z) = P\{Z \leqslant z | U = u\} \tag{15}$$

which exists and is nondecreasing almost everywhere in **I**, then:

1. Generate two independent random variables, r_1 and r_2, both uniform on **I**;
2. Set $u=r_1$ and $z=s_u^{-1}(r_2)$; and
3. From the pair (u,z) it is possible to generate a pair $(q_{\max},v)$ extracted from the joint law $F_{Q\max V}(q,v) = \mathbf{C}(F_{Q\max}(q),F_V(v))$ setting $q=F^{-1}_{Q_{\max}}(u)$ and $v=F_V^{-1}(z)$.

Shape of Flood Hydrograph

Here we consider the hypothesis that the maximum annual flood peak and the maximum annual flood volume are generated by the same flood event. This hypothesis will be tested against observed data for the case study considered in the following section.

The determination of flood hydrograph given peak $Q_{\max}$ and volume V requires the knowledge of the shape of the hydrograph. A first-order approximation is to consider a triangular flood hydrograph, where the base time T_b is equal to $T_b=2V/Q_{\max}$, the time of rise equals $T_p=T_b/2.67$, and the time of recession is equal to $1.67T_p$, (Soil Conservation Service 1972; Chow et al. 1988, p. 229). Then the flood hydrograph $q(t)$ is

$$q(t) = \begin{cases} 1.335\dfrac{Q_{\max}^2}{V}t & 0 \leqslant t \leqslant T_p \\ 1.6Q_{\max} - 0.8\dfrac{Q_{\max}^2}{V}t & T_p \leqslant t \leqslant T_b \end{cases} \tag{16}$$

Another possibility is to build the flood hydrograph with fixed flood peak $Q_{\max}$ and flood volume V through the convolution of an instantaneous unit hydrograph of a linear reservoir. In this case the flood hydrograph $q(t)$ is

$$q(t) = \begin{cases} \dfrac{V}{t_0}(1 - e^{-t/k}) & 0 \leqslant t \leqslant t_0 \\ \dfrac{V}{t_0}[e^{-(t-t_0)/k} - e^{-t/k}] & t_0 \leqslant t \end{cases} \tag{17}$$

where t_0=function of $Q_{\max}$ and V via the *transcendent* equation $t_0Q_{\max}/V=1-e^{-(t_0/k)}$, and k=time constant of the linear reservoir model, easy to estimate using the method of moments (see Bras 1990, p. 445).

Alternatively, a flood hydrograph with fixed flood peak $Q_{\max}$ and flood volume V can be obtained through the convolution of an instantaneous unit hydrograph of a cascade of n equal linear reservoirs (Nash 1957), see also Bras (1990, pp. 446–447). In this case $q(t)$ is

$$q(t) = \begin{cases} \dfrac{V}{t_0}\displaystyle\int_0^t \frac{1}{k\Gamma(n)}\left(\frac{t-\zeta}{k}\right)^{n-1} e^{-(t-\zeta)/k}d\zeta & 0 \leqslant t \leqslant t_0 \\ \dfrac{V}{t_0}\displaystyle\int_0^{t_0} \frac{1}{k\Gamma(n)}\left(\frac{t-\zeta}{k}\right)^{n-1} e^{-(t-\zeta)/k}d\zeta & t_0 \leqslant t \end{cases} \tag{18}$$

where t_0=again a function of $Q_{\max}$ and V; and t_0 can be determined numerically assuming that the maximum of $q(t)$ is equal to $Q_{\max}$. Here n and k are the model parameters: the first one represents the number of linear reservoirs and the second one a time constant; n and k are easy to estimate using the method of moments (Bras 1990, pp. 446–447).

Simulation of Reservoir Behavior and Adequacy of Dam Spillway

A simulation of reservoir behavior can be carried out considering a long synthetic series of maximum annual flood events (representing the expected design life of the reservoir), and thus it is possible to check adequacy of the dam spillway. The framework of simulation can be outlined as follow:

1. Generate a synthetic couple of maximum annual values of flood peak and flood volume ($Q_{\max}$, V);
2. Transform the couple ($Q_{\max}$, V) into the flood hydrograph $q(t)$;
3. Associate an initial reservoir level (e.g., taken as outcome of the empirical distribution of initial levels recorded in several years); and
4. Operate the reservoir routing of flood hydrograph [for more details see Bras (1990, pp. 475–478) and Zoppou (1999)] considering as an outlet only the uncontrolled spillway, and check that the reservoir level does not pass the crest level of the dam. The procedure is iterated to consider a long series

of maximum annual flood events (1,000 years) representing the expected design life of the dam. Thus it is possible to check adequacy of dam spillway during the life of the reservoir.

Case Study

The procedure given in the previous sections is illustrated here for the Ceppo Morelli dam. The dam with a hydroelectric power plant was built in 1929 on the Anza catchment, a subbasin of the Toce river basin, located in Northern Italy. The catchment area is 125 km^2. The maximum water storage is small, about 0.47×10^6 m^3. The maximum water level is at 782.5 m.a.s.l., and the dam crest level is 784 m.a.s.l. The dam has 84 m of uncontrolled spillway at 780.75 m.a.s.l. The dam has also intermediate outlets and bottom outlets. The last ones are obstructed by river sediments. Hourly observations of reservoir level and operations to the controlled outlets are available from 1937. This information yields the flood hydrograph downstream from the reservoir. Operating an inverse reservoir routing it is possible to reconstitute the historical flood hydrograph for the inflow to reservoir (see Zoppou 1999), and from this obtain the series of maximum annual flood peak and maximum annual flood volume (49 years). In addition, checking the dates of occurrence of maximum annual values of flood peaks with those of maximum annual flood volumes, it turns out that in almost all of the cases these occurred during the same flood event, for this reason, in this paper we consider these variables generated by the same storm. The 49 pairs of maximum annual flood peak and volume are given in Fig. 1. A strong positive dependence between the two variables is present.

We estimate the statistical dependence between the maximum annual flood peak $Q_{\max}$, and maximum annual flood volume V, considering as measures of association the canonical (Pearson's) coefficient of linear correlation ρ_P, and the Kendall's τ. The estimated values are given in Table 1. The dependence between $Q_{\max}$ and V is positive. For the latter measure, shown is the corresponding estimate of the dependence parameter δ of the 2-Copula, calculated using Eq. (4). For the sake of comparison, also shown is the estimate of δ calculated via the maximum likelihood method. Finally, we compute the average estimate of δ equal to 3.055; this indicates a strong association between the variables considered. The estimates reported are consistent with one another, and pre-

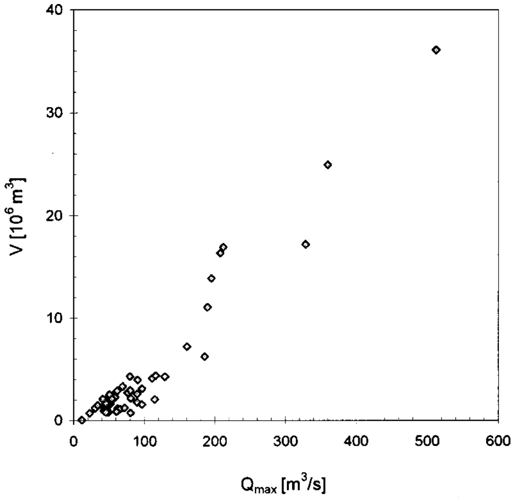

Fig. 1. Maximum annual flood peak versus maximum annual flood volume

cisely fix the degree of dependence between Q_{max} and V independently of the marginal laws of such variables. As a further analysis, in Fig. 2 we plot the Gumbel's 2-Copula, and compare it with the corresponding empirical 2-Copula function $\mathbf{c}_m$ calculated via Eq. (5). The agreement between the data and the model is visually good, and is objectively checked via a standard χ^2 test: as a result, the hypothesis that the theoretical model is consistent with the data can be accepted at all the standard significance levels (1, 5, and 10%).

In Table 2 we show the estimates of the parameters of the GEV marginal laws of Q_{max} and V; here the L moments technique is used (Hosking 1990). The shape parameter κ is always negative for both variables, which then exhibit asymptotic heavy tail behavior. It is important to note that, as a consequence of Eq. (9), the skewness of Q_{max} does not exist at all, as well as the variance of V; thus, apparently, the traditional method of moments technique could not be used to estimate the parameters of interest (as opposed to the L moments technique adopted here). For the sake of comparison, the joint estimates of the parameters calculated using the maximum likelihood method are also shown in Table 2. From this, it is evident that the estimates of the marginal parameters obtained with the two methods are consistent with one another. In Figs. 3 and 4 the empirical distributions of Q_{max} and V are shown, as well as the corresponding GEV fits, using the parameter's estimates given in Table 2. The agreement between the empirical and theoretical distribution (considering the two sets of estimates) is good for both variables; in particular, the use of both the Kolmogorov–Smirnov and Anderson–Darling goodness-of-fit tests (see, e.g., Kottegoda and Rosso 1997, pp. 285–293) shows that the theoretical model is consistent with the data at all the standard significance levels (1, 5, and 10%).

In passing, we observe a further important feature of the present approach: the possibility of selecting a suitable 2-Copula via the estimate of the Kendall's τ, as explained above, may not be affected at all by the possible nonexistence of lower-order moments. On the contrary, the use of other measures of association such as the canonical Pearson's coefficient of linear correlation ρ_P (which, instead, is often used in common practice) necessarily requires the existence of (at least) the second-order

Table 1. Estimated Values of Two Measures of Association: Pearson's ρ_p and Kendall's τ; for Last Measure, Shown is Corresponding Estimate of Dependence Parameter δ of Gumbel's Copula

Dependence parameter		δ
ρ_p	0.964	—
τ	0.651	2.868
ML	—	3.241

Note: For sake of comparison, also shown is estimate of δ calculated via maximum likelihood (ML) method.

Table 2. Estimated Values of Generalized Extreme Value Parameters for Both Flood Peak and Volume Using L Moments Technique and Also Joint Estimates Using Maximum Likelihood (ML) Method

Parameter	L moments	ML
ε_Q (m^3/s)	57.972	59.221
α_Q (m^3/s)	33.689	35.676
κ_Q (−)	−0.420	−0.338
ε_V (10^6 m^3)	1.744	1.774
α_V (10^6 m^3)	1.620	1.544
κ_V (−)	−0.564	−0.570

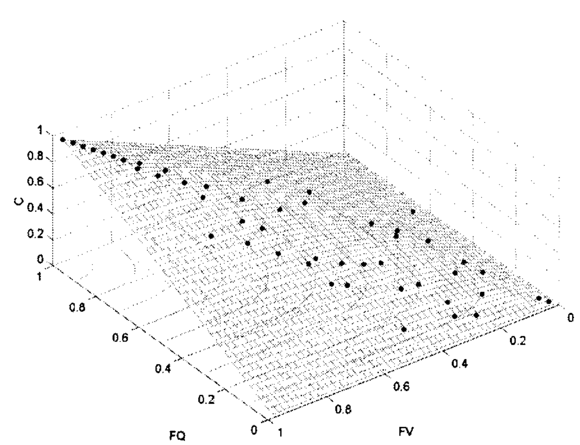

Fig. 2. Comparison between empirical and theoretical copula

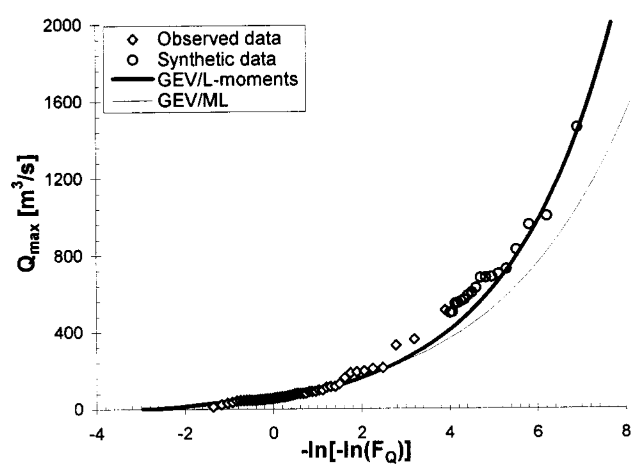

Fig. 3. Comparison between empirical distributions of observed and synthetic data (only upper part $-\ln[-\ln(F_Q)]>4$) and generalized extreme value fits using L moments and maximum likelihood parameters estimations for Q_{max}.

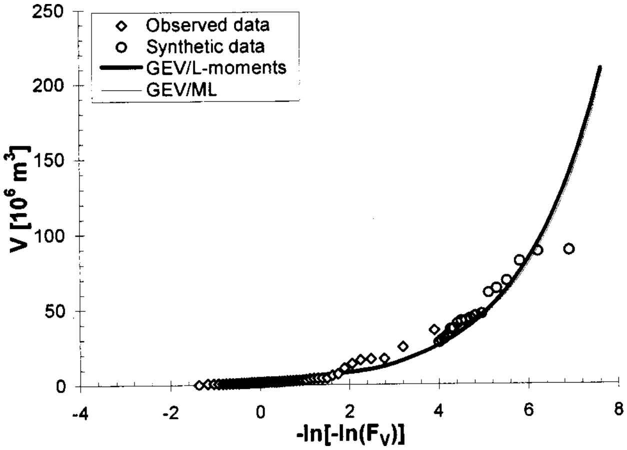

Fig. 4. Comparison between empirical distributions of observed and synthetic data (only upper part $-\ln[-\ln(F_V)]>4$) and generalized extreme value fits using L moments and maximum likelihood parameters estimations for V

moment. However, it is clear that meaningless estimates of ρ_P may be obtained if, as in the considered case, the existence of the variance is questionable: in fact, $\kappa_V \approx -0.564 < -1/2$; nevertheless, the "suspicious" value of ρ_P is also reported in Table 1 for the sake of completeness.

From the couple flood peak $Q_{\max}$ and flood volume V the corresponding flood hydrograph is calculated using the instantaneous unit hydrograph of a cascade of n equal linear reservoirs (Nash model) [Eq. (18)]. The model parameters, n and k, are estimated applying the method of moments to each of the observed flood hydrographs. The rainfall precipitation was transformed into runoff using the Soil Conservation Service (SCS)–CN method with an estimated value of the CN parameter, at basin scale, equal to 55 (this was obtained by fitting over the observed hyetograph and hydrograph data).

For four flood events, Fig. 5 gives a comparison between the observed flood hydrograph and the estimated one, considering the sample estimate of n and k obtained on the single flood event. Fig. 5 also includes the rainfall hyetograph and the estimated direct runoff.

The expected value and the standard deviation of estimate of the two parameters n and k are as follows: $E[n]=2.2$, $E[k]=2.1$ h, $\sigma[\hat{n}]=1.35$, $\sigma[\hat{k}]=0.63$ h. Note that $E[n]$ and $E[k]$ are used in the long hydrologic simulation to assess the safety of the dam in terms of adequacy of the dam spillway.

A long synthetic series of 1,000 flood-peak–flood-volume pairs is generated using Monte Carlo simulation. The upper part of the empirical distribution of synthetic data $\{-\ln[-\ln(F)]>4\}$ is reported, respectively, in Fig. 3 for $Q_{\max}$ and in Fig. 4 for V. Using the series of peak–volume pairs, the corresponding set of 1,000 synthetic hydrographs was generated (covering the expected design life of the dam), using the Nash model and Eq. (18). To each flood hydrograph is associated a reservoir level at the start of the flood event. Also available is the series of reservoir levels antecedent to the maximum annual flood event. Fig. 6 shows the cumulative distribution of the reservoir level. Note that several distributions were considered to fit such data, but none provided acceptable agreements. This is due to the fact that the levels are influenced not only by the precipitation inflows but also by the operation policy of the dam manager. Accordingly, in the present analysis we used the empirical distribution. From this we

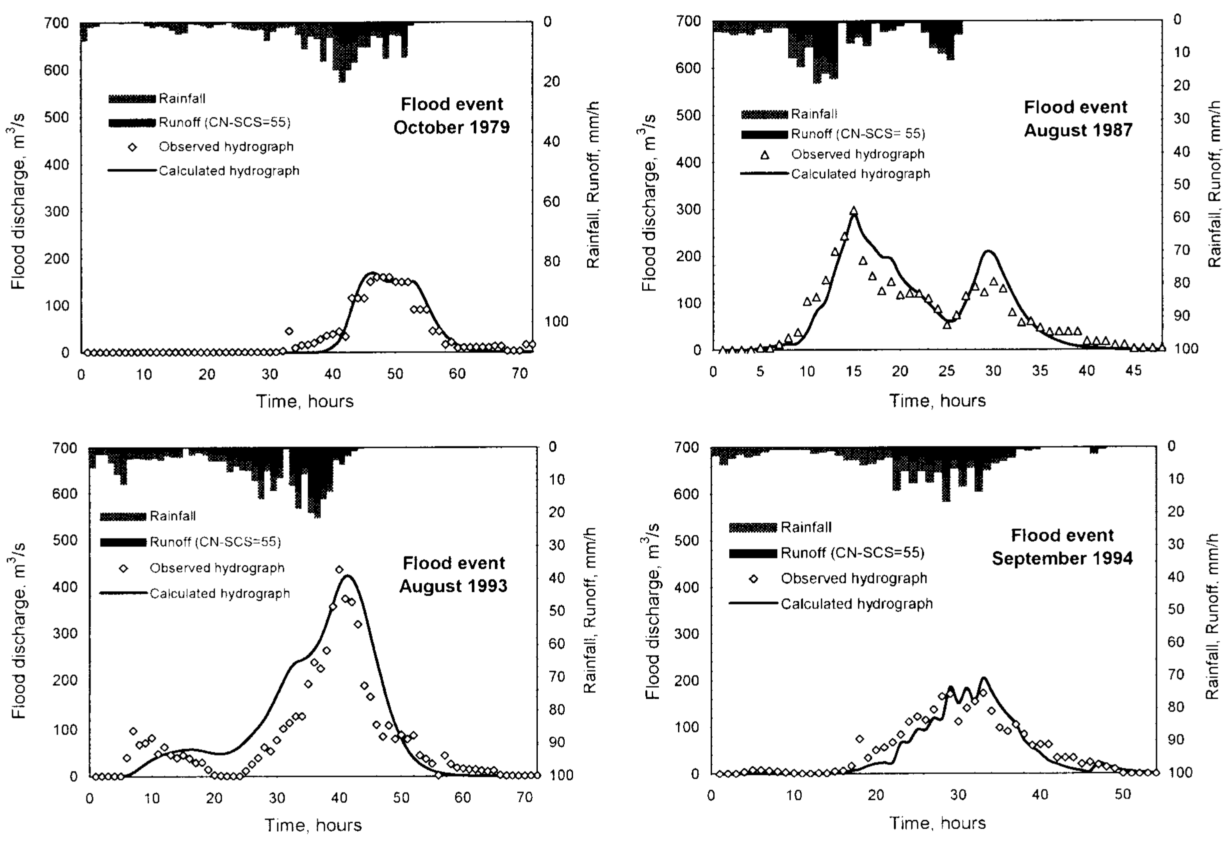

Fig. 5. Comparison between observed and calculated flood hydrographs

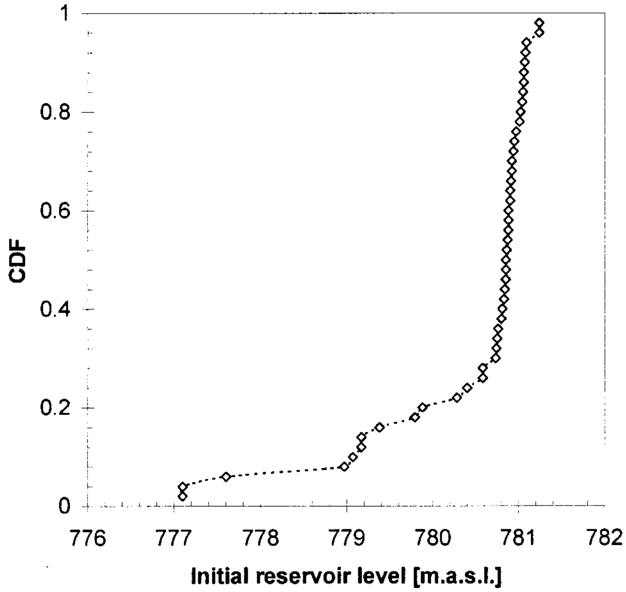

Fig. 6. Empirical distribution of initial reservoir level before flood event

extracted a series of 1,000 reservoir levels and associated each of these to a synthetic hydrograph. Operating the reservoir routing of the synthetic series of hydrographs we assessed the safety of the dam in terms of adequacy of dam spillway.

Fig. 7 shows the empirical distribution of the maximum reservoir level reached during the simulated flood event. The analyses show that the maximum water level (782.5 m) is exceeded only in ≈2% of the cases, whereas the dam crest level (784 m) was never exceeded. Thus, the overall result indicates the hydrological safety of the Ceppo Morelli dam.

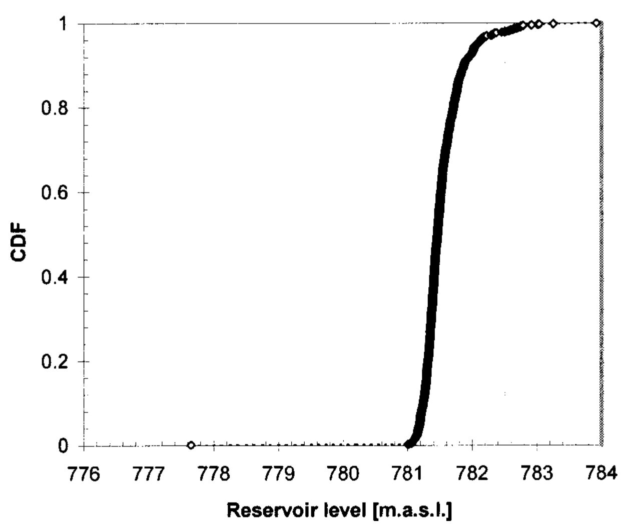

Fig. 7. Empirical distribution of maximum reservoir level during flood event as result of simulation of 1,000 flood events. Dashed vertical line indicates maximum water level (782.5 m) and tick solid vertical line dam crest level (784 m).

Conclusions

The main purpose of the paper is to outline a very general model that describes the possible bivariate behavior of the random variables flood peak and flood volume, which are of primary interest in hydrological practice. The paper presents a bivariate statistical procedure for the evaluation of flood hydrograph and to check the adequacy of spillway, an important issue for dam engineering, and in general for the hydrological practice, where the variables of interest usually show some (weak or strong) degree of dependence requiring a bivariate model. Indeed, from a mathematical point of view, the use of 2-Copulas represents the easiest approach to problems involving couples of variables. Many of the difficulties in the bivariate approaches present in the literature (having a theoretical and/or a practical origin), can now be bypassed through a proper formulation of the problem in terms of 2-Copulas, which represent the most recent and most promising mathematical tool for investigating bivariate problems: this represents a significant improvement in the field of statistical hydrology and modeling.

The proposed procedure is based on the bivariate probabilistic analysis of maximum annual values of flood peak and flood volume. A bivariate extreme value distribution is considered using 2-Copulas. In particular the Gumbel's 2-Copula is adopted to represent the positive dependence between the flood peak and flood volume, and a GEV law is considered as a marginal distribution for both variables. From the pair peak–volume, the flood hydrograph is obtained through the convolution of an instantaneous unit hydrograph of a linear model. Successively, a long series of flood hydrographs (1,000 years) is generated, and operating the reservoir routing on this, it is possible to check the behavior of the reservoir during the expected design life of the dam and thus adequacy of dam spillway. The Ceppo Morelli dam is considered. The results show the adequacy of spillway and consequently the safety of the dam. The proposed methodology does not rely upon the specific climate conditions, or the size of the catchment area, or the dimension of the dam considered. The model presented is fully general, and in principle it could be applied to scenarios different from the one considered here.

Note how the approach proposed could be used both as an evaluation method of an existing design and as a design mode of dam spillway or river polder spillway. In other words, if a dam already exists, the approach can be used to check the adequacy of dam spillway; at the same time, if flood hydrographs are available (as usually provided by flood gages), the approach provides a clear methodology for the design of dam spillways.

Acknowledgments

The research was partially supported by MURST via the project "Hydrological Safety of Impounded Rivers." The support of "Progetto Giovani Ricercatori" is also acknowledged.

References

American Society of Civil Engineers (ASCE). (2000). "Dam safety." *Policy Statement 280*, New York.

Berga, L. (1998). "New trends in hydrological safety." *Dam safety*, L. Berga, ed., Balkema, Rotterdam, The Netherlands, 1099–1110.

Bras, R. L. (1990). *Hydrology: An introduction to hydrologic science*, Addison-Wesley, Reading, Mass.

Carriere, J. F. (1994). "A large sample test for one-parameter families of copulas." *Commun. Stat: Theory Meth.*, 23(5), 1311–1317.

Chow, V. T., Maidment, D. R., and Mays, L. W. (1998). *Applied hydrology*, McGraw–Hill, Singapore.

Committee on Failures and Accidents to Large Dams of the United States Committeee on Large Dams. (1975). "Lessons from dam incidents, USA." ASCE/USCOLD, New York.

De Almeida, A. B., and Viseu, T. (1997). "Dams and valley: A present and future challenge." *Proc., Int. NATO Workshop on Dams and Safety Management at Downstream Valleys*, Lisbon, Portugal, A. B. De Almeida, and T. Viseu, eds., Balkema, Rotterdam, The Netherlands, 3–25.

Deheuvels, P. (1979). "La fonction de dependence empirique et ses proprietes. Un test non parametrique d'independence." *Acad. Roy. Belg. Bull. Cl. Sci.*, 65(5), 274–292.

De Michele, C., and Salvadori, G. (2003). "A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas." *J. Geophys. Res., [Atmos.]*, 108(D2), ACL 15-1–ACL 15-11.

Genest, C., and Rivest, L. (1993). "Statistical inference procedures for bivariate archimedean copulas." *J. Am. Stat. Assoc.*, 88(423), 1034–1043.

Gumbel, M. E. J. (1960). "Distributions des valeurs extrêmes en plusieurs dimensions." *Publ. Inst. Stat. Univ. Paris*, 9, 171–173.

Hosking, J. R. M. (1990). "*L*-moments: Analysis and estimation of distributions using linear combinations of order statistics." *J. R. Stat. Soc. Ser. B. Methodol.*, 52, 105–124.

Joe, H. (1997). *Multivariate models and dependence concepts*, Chapman and Hall, London.

Kottegoda, N. T., and Rosso, R. (1997). *Statistics, probability and reliability for civil and environmental engineers*, McGraw–Hill, New York.

Nash, J. E. (1957). "The form of the instantaneous unit hydrograph." *IAHS Publication*, 45(3–4), 114–121.

Nelsen, R. B. (1999) *An introduction to copulas*, Springler, New York.

Salvadori, G., and De Michele, C. (2004). "Analytical calculation of storm volume statistics involving Pareto-like intensity-duration marginals." *Geophys. Res. Lett.*, 31, L04502.

Soil Conservation Service. (1972). "Hydrology." *National engineering handbook*, Sec. 4, Soil Conservation Service, U.S. Department of Agriculture, Washington, D.C.

Zoppou, C. (1999). "Reverse routing of flood hydrographs using level pool routing." *J. Hydraul. Eng.*, 4(2), 184–188.