**Hydrology and Earth System Sciences**

# Design flood hydrographs from the relationship between flood peak and volume

**L. Mediero[1], A. Jiménez-Álvarez[2], and L. Garrote[1]**

[1]Department of Civil Engineering: Hydraulics and Energetics, Technical University of Madrid, Spain
[2]Centre for Hydrographic Studies of CEDEX, Madrid, Spain

**Abstract.** Hydrological frequency analyses are usually focused on flood peaks. Flood volumes and durations have not been studied as extensively, although there are many practical situations, such as when designing a dam, in which the full hydrograph is of interest. A flood hydrograph may be described by a multivariate function of the peak, volume and duration. Most standard bivariate and trivariate functions do not produce univariate three-parameter functions as marginal distributions, however, three-parameter functions are required to fit highly skewed data, such as flood peak and flood volume series. In this paper, the relationship between flood peak and hydrograph volume is analysed to overcome this problem. A Monte Carlo experiment was conducted to generate an ensemble of hydrographs that maintain the statistical properties of marginal distributions of the peaks, volumes and durations. This ensemble can be applied to determine the Design Flood Hydrograph (DFH) for a reservoir, which is not a unique hydrograph, but rather a curve in the peak-volume space. All hydrographs on that curve have the same return period, which can be understood as the inverse of the probability to exceed a certain water level in the reservoir in any given year. The procedure can also be applied to design the length of the spillway crest in terms of the risk of exceeding a given water level in the reservoir.

## 1 Introduction

Hydrological frequency analyses are usually focused on flood peaks, for example, culverts, bridges and river channel defences are designed by considering the peak flow for a given return period. There are many studies about how to estimate the flood peak frequency curve (Cunnane, 1988, 1989; GREHYS, 1996), but flood volumes have not been studied as extensively, despite the fact that they are needed to design some structures like dams, where the entire flood hydrograph is of interest.

The spillway length of a dam is designed by considering the peak of the outflow hydrograph. The inflow hydrograph must be routed through the reservoir, and its peak is lowered by storage and releases. Knowledge of the flood peak is not sufficient to design the dam spillway; the entire flood hydrograph must be utilised. The univariate flood frequency analysis on peaks should be extended to a multivariate analysis on other variables to estimate not only the peak for that return period, but also other variables to construct an entire hydrograph.

A flood event may be described by a multivariate function of the peak, volume and duration, as a joint distribution of their marginal distributions. Some attempts at describing floods in this way have been conducted. Goel et al. (1998) employed a bivariate normal distribution of the peak and volume, after the normalisation of a data series by two Box-Cox transformations, to lower the skewness coefficient to a value nearly equal to zero and to correct the coefficient of kurtosis to a value of nearly three. Other studies were based on the bivariate normal distribution (Krstanovic and Singh, 1987; Sackl and Bergmann, 1987), but, as flood peaks and volumes are highly skewed, prior transformations in data series are required. In the case where statistical behaviours of peak and volume data are represented by Gumbel distributions, a bivariate extreme value distribution can be used (Yue et al., 1999). In addition, a bivariate lognormal distribution was developed by Yue (2001). All these attempts assume that flood variables can be represented by the same distribution. To relax the restriction of a unique distribution function to represent peak and volume, bivariate and trivariate distributions have been derived using the Copula method. Different
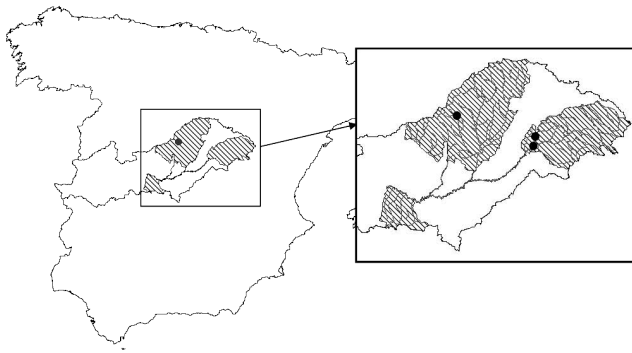
**Fig. 1.** Location of the three case studies. The lined area corresponds to the 32nd homogeneous region in the Tagus basin.

Copula families have been used. Favre et al. (2004) considered the Farlie-Gumbel-Morgenestern, Frank and Clayton families and no significant differences were shown among them. De Michele et al. (2005) considered an Archimedean Gumbel's 2-Copulas and simulated the dependence between peak and flood volume with Kendall's $\tau$ rank correlation coefficient. Grimaldi and Serenaldi (2006) developed an asymmetric Archimedean Copula that is more flexible than symmetric Copulas. Zhang and Singh (2007) utilised the Gumbel-Hougaard Copula to simulate the trivariate distribution of the peak, volume and duration.

A dam can be designed with a DFH, which is a hydrograph adopted according to design standards to ensure the safety of a structure (Xiao et al., 2009). Design standards for dams are based either on the Probable Maximum Flood (PMF) or on a given return period. Some attempts have been made to estimate the return period of a hydrograph as the inverse of its probability of occurrence, which is estimated by the joint probability of a bivariate distribution. This joint probability is not explicit when the variables are correlated and the conditional return period, given a maximum value of the other variable, must be calculated (Zhang and Singh, 2006). The joint return period has a lower probability of occurrence than the inclusive probability of both events, known as the primary return period, and a higher probability of occurrence than the exclusive probability of both events, known as the secondary return period. This means that a structure could be under-dimensioned if it is designed with the primary return period and over-dimensioned if it is designed with the secondary return period (Salvadori and De Michele, 2004).

But the return period is the average time elapsed between two successive events that exceed a given threshold (Ponce, 1989), which must be defined in terms of the acceptable risk to the structure. The hydrological risk at a bridge or a culvert is related to the maximum water level in the reach, which mainly depends on the peak discharge. Therefore, the threshold can be defined as a given discharge. However, the hydrological risk at the dam is related to the maximum reservoir level and maximum released flow during the event, which de-

pends on more than the maximum inflow discharge, as there can be several floods with different combinations of volumes and peaks that yield the same level and release. At first, a greater peak will be worse for dams with smaller reservoir areas and a greater volume will be worse for dams with larger reservoir areas, but the crest length of the spillway must be considered and could modify this statement. Therefore, depending on the reservoir area, the crest length of the spillway and whether the spillway is controlled or uncontrolled, either the peak or the volume could be the more influential parameter in determining the risk. The problem is complex and a set of hydrographs can have the same design return period. In addition, a pair of peak and volume values will have a different return period than that of their marginal distributions. Therefore, peaks and volumes cannot be utilized independently as thresholds to assess dam risk. The threshold must be defined as a given water level in the reservoir, so that the return period is the inverse of the probability of exceeding that reservoir water elevation in any given year.

In this paper, a methodology is presented to obtain the DFH for designing dams in Spanish basins. The peaks and volumes in most Spanish basins are highly skewed and are best described by the Generalised Extreme Value distribution (GEV). As a suitable bivariate distribution from three-parameter distributions has not been developed yet, the relationship between the peak discharge and hydrograph volume has been analysed from recorded data to generate a large set of annual maxima synthetic hydrographs that preserve the marginal distributions of the peaks, volumes and durations. Each hydrograph is routed through the dam to compute the maximum water level in the reservoir. As the return period assigned to a flood is the inverse of the probability of exceeding a particular water level, it is calculated as the total number of hydrographs divided by the number of hydrographs that reached a maximum water level higher than the threshold. With this procedure, the DFH for a given return period is not a unique hydrograph, but rather a curve in the peak-volume domain, so that there will be a set of hydrographs with the same return period and the same risk to the dam.

## 2 Case studies

The Santillana, Entrepeñas and Buendia reservoirs were selected as case studies. The three reservoirs are located on the Tagus basin, in the central west of Spain, and belong to the 32nd homogeneous region (Fig. 1). There are no recorded data of the inflow discharges to the reservoirs, but they can be estimated from the recorded mean daily water levels and releases at the 93033 (Santillana), 93001 (Entrepeñas) and 93087 (Buendía) reservoir stations.

The Santillana reservoir is located on the Manzanares River, near the city of Madrid. The dam is an earthfill embankment with a height of 40 m and a crest length of 1355 m. Flood flows over the spillway are controlled by a 5.25 m by

**Table 1.** Main variables of reservoirs: drainage area ($A_d$), volume up to the spillway crest ($V$), flooded area at the spillway crest height ($A_f$), elevation of the spillway crest ($H_s$).

| Reservoir | $A_d$ (km$^2$) | V (hm$^3$) | $A_f$ (km$^2$) | $H_s$ (m a.s.l.) |
|-----------|-----------|-----------|-----------|-------------|
| 93001 | 4060 | 710.1 | 29.56 | 715 |
| 93033 | 3256 | 48.9 | 5.35 | 889 |
| 93087 | 247 | 1519.2 | 77.5 | 710.5 |

12 m gate. The Entrepeñas reservoir is located on the Tagus River. The dam has a concrete cross-section with a height of 87.35 m and a length of 383 m. Flood flows over the spillway are controlled by five 10.76 m by 5.50 m gates. The Buendia reservoir is located on the Guadiela River. The concrete dam has a height of 78.73 m and a length of 315 m. Flood flows are controlled by five 12.20 m by 1.50 m gates. Further details of their main characteristics are included in Table 1.

## 3 Marginal distributions

The marginal distributions of Annual Maximum Discharges (AMD) and Annual Maximum Volumes (AMV) were estimated from recorded data. Identifying the AMV in a year is the main purpose for determining the marginal distribution of the maximum volumes. An AMV could be obtained from a long hydrograph with a low peak discharge, but it would not necessarily imply a high risk for the dam. As the study begins from the AMD frequency curve, the volumes linked to these peaks should be identified so that the methodology is consistent.

### 3.1 Flood peak frequency distribution

A regional study was conducted in Spain to improve local estimations of flood frequency curves and continental Spain was divided into 30 homogeneous regions. Spanish geography shows a high climatic variability, so regions were identified by means of their geographical characteristics. The index-flood is the most common regional method (Bocchiola et al., 2003; Kjeldsen and Jones, 2007; Noto and La Loggia, 2009), and it supplies regional values of the L-coefficient of skewness (L-CS) and the L-coefficient of variation (L-CV) in a homogeneous region. There is an agreement about the regionalisation of the L-CS, as its estimation uncertainty from local data is high, even for long record lengths, however, the regionalisation of the L-CV is widely debated. First, its estimation uncertainty is lower than that of the L-CS and is similar to that of the mean, which cannot be regionalised. In addition, the relationship between L-CV and the basin area seems to be extremely complex, as it depends on the interaction between different runoff processes; it has been seen that L-CV increases with basin area, until a threshold, and

then decreases (Blöschl and Sivalapan, 1997; Iacobellis et al., 2002). As this L-CV pattern has been seen in Spanish regions, a regional shape estimation procedure was selected to relax the restriction of a regional value of L-CV. A comparison between the two methods showed that the regional shape estimation improves the estimation of quantiles in the upper tail of the frequency distribution, as is observed in this paper (Hosking and Wallis, 1997, p. 150).

The three reservoirs belong to the 32nd region, which has a regional L-CS value equal to 0.253. The mean daily discharges at the reservoir stations were transformed into instantaneous maximum discharges by Fuller's formula (Fill and Steiner, 2003). A GEV distribution (Eq. 1) was fitted to the AMD series with the regional value of the L-CS (Table 2).

$$F(x) = \exp\left\{-\left[1 - k\left(\frac{x-u}{\alpha}\right)\right]^{1/k}\right\} \tag{1}$$

where, $u$ $\alpha$ and $k$ are the parameters of the GEV distribution.

### 3.2 Flood volume frequency distribution

The regionalisation results of the AMD were extended to the AMV data series. The volumes of the hydrographs linked to the AMD were identified. The start and the end of the hydrograph were assumed to be the start and the end of the surface runoff. The start was identified as an abrupt rise of the discharge by more than 20%. The end was identified as the point from which the receding limb is described by an exponential function (Eq. 2). The $\beta$ coefficient was assumed to be equal to 0.0063 h$^{-1}$ in the 32nd region. The dependence between two successive peaks was identified by the independence criterion proposed by Cunnane (1979).

$$Q_r = Q_0 \cdot e^{-\beta t} \tag{2}$$

The homogeneity of the AMV was tested at the homogeneous regions previously identified by heterogeneity measures based on L-Moments (Eq. 3–4) (Hosking and Wallis, 1993). The homogeneity requirement of the AMV series was met, as can be seen in Table 3. The volume frequency curves were fitted with a GEV distribution and a regional shape parameter (Table 2).

$$W_i = \left(\frac{\sum_{j=1}^{N} n_j \cdot \left(t_{i,j} - t_i^R\right)^2}{\sum_{j=1}^{N} n_j}\right)^{1/2} \tag{3}$$

$$H_i = \frac{W_i - \mu(W_i)}{\sigma(W_i)} \tag{4}$$

$H_i$ is measured on a large number of simulated regions with $N$ sites, where each site has the same record length as their real-world counterparts.

**Table 2.** Statistics of the AMD (m$^3$/s) and the AMV (hm$^3$) series and parameters of the GEV distributions fitted with a regional shape parameter.

| | Local statistics of AMD | | | | GEV Parameters | | |
|---|---|---|---|---|---|---|---|
| Station | Mean | L-CV | L-CS | L-CK | $u$ | $\alpha$ | $k$ |
| 93033 | 74.25 | 0.374 | 0.191 | 0.092 | 48.983 | 35.135 | −0.127 |
| 93001 | 185.29 | 0.451 | 0.356 | 0.184 | 109.194 | 105.808 | −0.127 |
| 93087 | 155.41 | 0.402 | 0.266 | 0.099 | 98.535 | 79.070 | −0.127 |
| | Local statistics of AMV | | | | GEV Parameters | | |
| Station | Mean | L-CV | L-CS | L-CK | u | $\alpha$ | k |
| 93033 | 14.67 | 0.367 | 0.113 | 0.024 | 9.226 | 4.969 | −0.348 |
| 93001 | 63.80 | 0.537 | 0.417 | 0.209 | 29.150 | 31.631 | −0.348 |
| 93087 | 80.07 | 0.674 | 0.619 | 0.441 | 25.496 | 49.819 | −0.348 |

**Table 3.** Heterogeneity tests and regional statistics of the AMD and the AMV series at the 32nd region.

| AMD Series | | | | | |
|---|---|---|---|---|---|
| Heterogeneity tests | | | Regional statistics | | |
| H1 | H2 | H3 | L-CV | L-CS | L-CK |
| 1.887 | 0.459 | 0.334 | 0.416 | 0.253 | 0.129 |
| AMV Series | | | | | |
| Heterogeneity tests | | | Regional statistics | | |
| H1 | H2 | H3 | L-CV | L-CS | L-CK |
| 1.732 | 1.054 | 0.809 | 0.535 | 0.414 | 0.255 |

**Table 4.** Local and regional regression equations, $n$ is the length of the observed data, $\rho$ is the Pearson's correlation coefficient, $a$ and $b$ are the parameters of the equation and $\sigma_{\text{reg}}$ is the standard deviation of the residuals.

| Station | $n$ | $\rho$ | $a$ | $b$ | $\sigma_{\text{reg}}$ |
|---|---|---|---|---|---|
| 93033 | 43 | 0.7157 | −0.3787 | 0.9221 | 0.2081 |
| 93001 | 68 | 0.8059 | −0.8707 | 1.2236 | 0.2749 |
| 93087 | 60 | 0.5936 | −1.3315 | 1.4335 | 0.2859 |
| Regional | 919 | 0.7004 | −0.6496 | 1.1057 | 0.2525 |
| Standardized regional | 919 | 0.6487 | −0.0855 | 1.1272 | 0.2503 |

## 4 Relationship between the peak flow and hydrograph volume

The dependence of the volume on the peak discharge was analysed to estimate the joint distribution. A linear relationship in the log-log space was found, both in each station and at the regional scale in a homogeneous region and was represented by regression equations (Fig. 2).

On a local scale, the volume for a given maximum discharge was estimated by fitting a regression equation over the observed pairs (Eq. 5).

$$V_{i,j} = 10^{a_j} \cdot Q_{i,j}^{b_j} \qquad (5)$$

Then, the relationship was analysed in the regional log-log space of the real values of peaks and volumes, but a problem of scale was encountered because the regression equation cannot distinguish the greater volumes of larger basins from the smaller volumes of smaller basins. It can be seen that there are no $Q$–$V$ pairs of the Entrepeñas and Buendía reservoirs below a peak of 1.5, while there are no $Q$–$V$ pairs of the Santillana reservoir above a peak of 2 (Fig. 2a). Therefore, a

standardisation of the peaks and volumes was performed to overcome the scale problem by dividing the peaks and volumes by their means in each station (Eq. 6–7) (Fig. 2b).

$$q_{i,j} = \frac{Q_{i,j}}{\overline{Q_j}} \qquad (6)$$

$$v_{i,j} = \frac{V_{i,j}}{\overline{V_j}} \qquad (7)$$

At the regional scale, a hydrograph volume ($V$) is estimated from its hydrograph peak ($Q$), thus, destandardising the regression equation (Eq. 8).

$$V_{i,j} = \left(10^{a_R} \cdot q_{i,j}^{b_R}\right) \cdot \overline{V_j} \qquad (8)$$

Regression equations were fitted at each case study and in the whole 32nd region, as shown in Table 4. The variability of the relationship between the peaks and volumes, which can be considered as the estimation uncertainty of the regression equation, was estimated by the residual variance ($\sigma_{\text{reg}}$) (Eqs. 9 and 10).
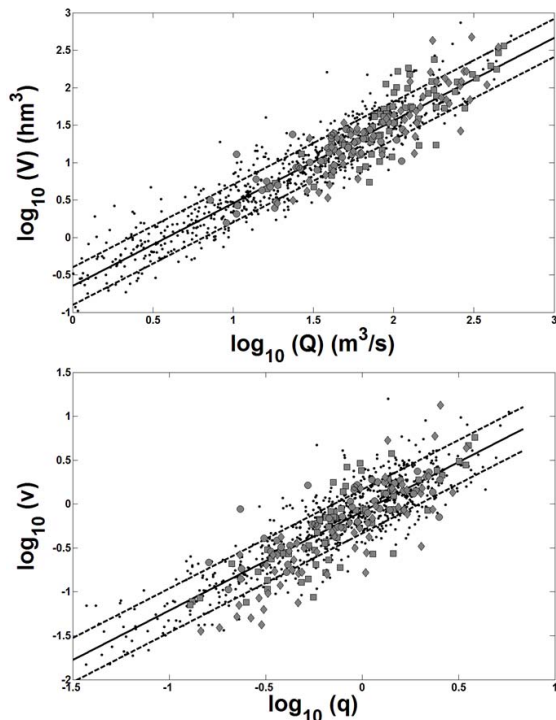
**Fig. 2.** Relationship between hydrograph volumes and peak flows. Solid lines are the regression equations and dotted lines show the confidence intervals for a confidence level of 33%. Solid points are the $Q$–$V$ pairs in the whole region, squares are the pairs in the 93001 station, circles in the 93033 station and diamonds in the 93087 station. **(a)** Observed volumes ($V$) against observed peak flows ($Q$). **(b)** Standardized volumes ($v$) against standardized peak flows ($q$).

$$\sigma_{\text{reg},j} = \sqrt{\frac{\sum_{i=1}^{n_j}\left(\log_{10}(V_{i,j}) - \log_{10}(V'_{i,j})\right)^2}{n_j - p - 1}} \tag{9}$$

$$\sigma_{\text{reg,R}} = \sqrt{\frac{\sum_{j=1}^{N}\sum_{i=1}^{n_j}\left(\log_{10}(V_{i,j}) - \log_{10}(V'_{i,j})\right)^2}{\sum_{j=1}^{N} n_j - p - 1}} \tag{10}$$

## 5 Generation of synthetic peak-volume pairs

The return period of a hydrograph is calculated as the inverse of the probability of the exceedance of the maximum water level in the reservoir that was attained while routing that hydrograph. As the probability of exceedance for high return periods is low, a large number of hydrographs is required to accurately estimate these return periods, which are used to design dams. Therefore, synthetic hydrographs must be generated to extend the observed data.

A large set of synthetic hydrographs that preserved the statistical characteristics of the observed peaks, volumes and durations was generated. The synthetic generation consists of three steps: the first is the generation of a set of synthetic peak flows, the second is the generation of a synthetic volume for each synthetic peak, comparing both the local and the regional approaches and the third is the generation of a hydrograph shape for each synthetic pair of peak and volume, which implies a certain duration.

As a first step, a random sample of probabilities with a length of 100 000 cases ($p_i$) generated from a uniform distribution in the range (0, 1) was transformed into a set of synthetic peak flows ($Q_i^s$) by an inverse GEV distribution (Eq. 11), which was fitted at each station with the regional method previously discussed. Synthetic peak flows keep the statistical properties of the fitted GEV distribution with the observed data at the stations, as shown in Fig. 3.

$$Q_i^s = u + \frac{\alpha}{k}\left[1 - (-\ln(p_i))^k\right] \tag{11}$$

The second step is the generation of synthetic volumes. A synthetic volume could be estimated from a synthetic peak with the regression equation between them, but this would lead to a perfect linear relationship that does not simulate its real variability. Therefore, as the residuals of the regression equation are normally distributed in the log-log space of variables (Fig. 4), a normal randomisation was performed for each synthetic peak flow, with a mean equal to the result of the regression equation (Eq. 5 or 8) and standard deviation equal to the residual variance of the regression ($\sigma_{\text{reg}}$) (Eq. 9 or 10).

The two first steps of the synthetic generation methodology were applied to the observed data from the three case studies. Two sets of 100 000 synthetic volumes were generated at each site from the set of synthetic peaks, one from the local regression equation and another from the regional regression equation. Both regressions were compared to assess their capability of preserving the statistical properties of the observed data (Fig. 3).

Both regressions retain the statistics of the AMV fairly well. In the Entrepeñas reservoir, the regional regression thoroughly keeps the frequency curve up to a return period of 2000 years, but, for higher return periods, the synthetic volumes are smaller than the observed ones. The local regression shows greater volumes for return periods longer than 25 years. In the case of the Santillana reservoir, the regional regression fits the frequency curve fairly well, but the local regression shows greater differences for return periods greater than 1000 years. The local regression thoroughly fits the frequency curve in the Buendía reservoir, but the regional regression shows small volumes for higher return periods. In each reservoir, both regressions must be compared to select the best one in each case.
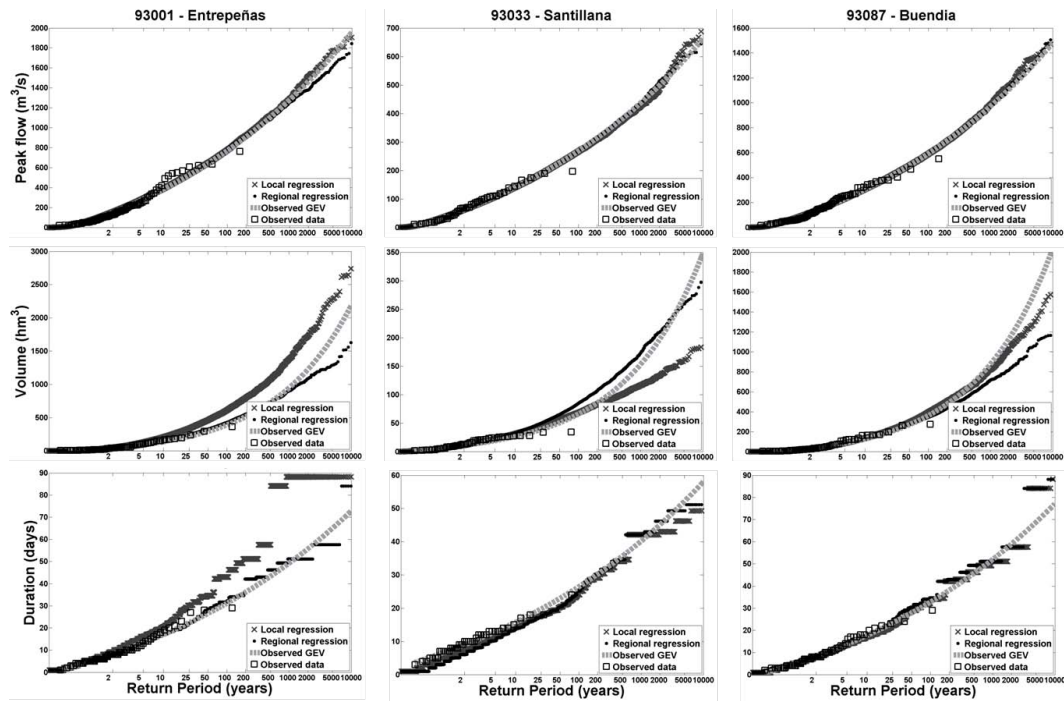
**Fig. 3.** Comparison between the observed and synthetic peaks, volumes and durations from the local and regional regressions fitted in the three case studies.
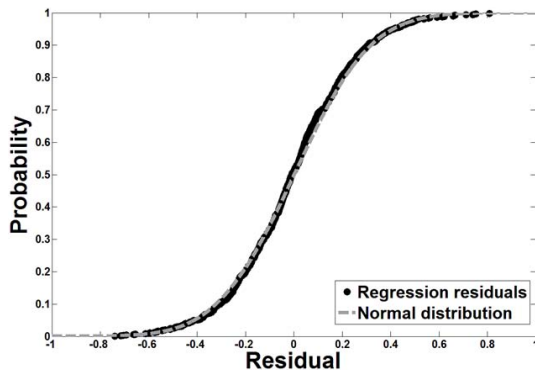


**Fig. 4.** Normality test of the residuals of the regional regression equation between the standardized peaks ($q$) and standardized volumes ($v$).

## 6 Generation of hydrographs

Each $Q$–$V$ pair must be transformed in a flood hydrograph to be routed through the reservoir. Hydrographs in a river can have multiple shapes, as different events can produce different runoff responses. Different methods have been proposed to construct a hydrograph. The selection of a method restricts the shape of the hydrograph and homogenises the results. Random shapes must be used to relax this restriction. Randomisation can be achieved coupling a stochastic rainfall generator and a hydrological model, both calibrated in the
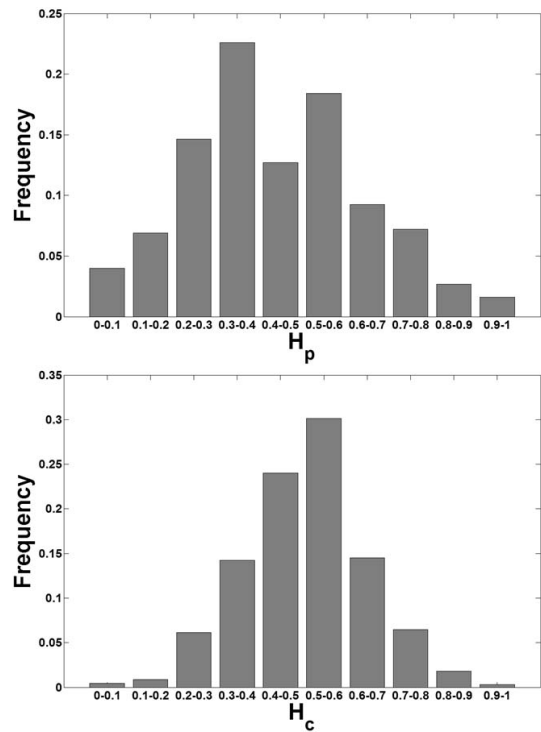


**Fig. 5.** Histograms of the shape hydrograph variables. **(a)** $H_{\mathrm{p}}$; **(b)** $H_{\mathrm{c}}$.
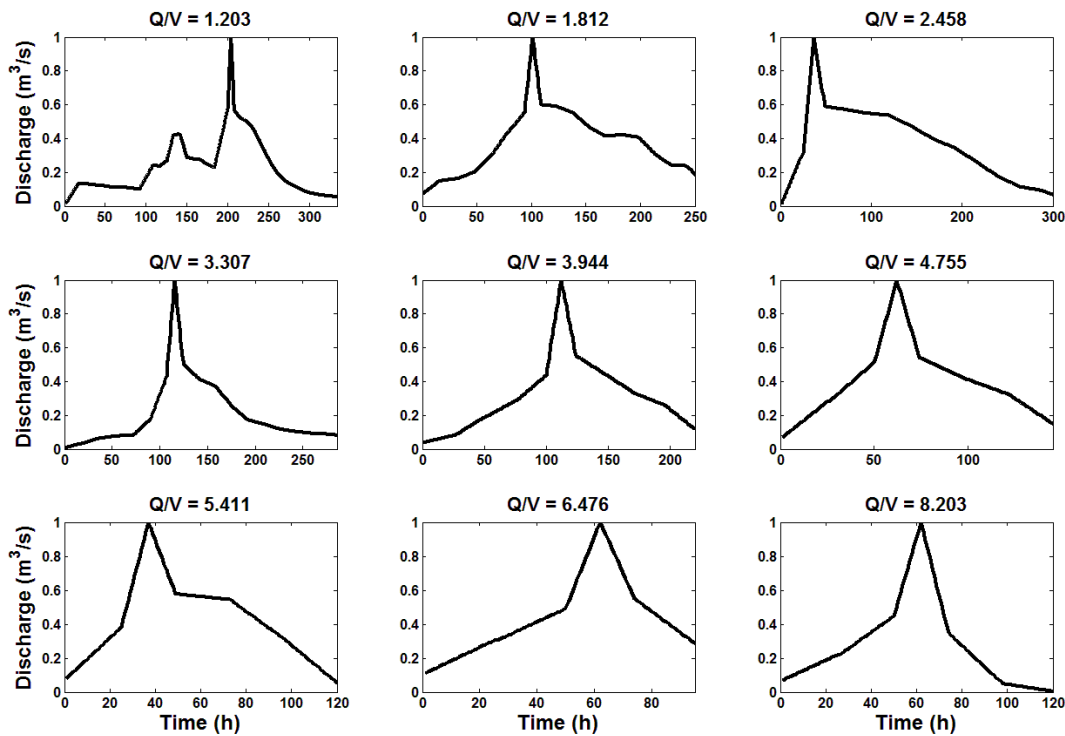
**Fig. 6.** Examples of the ratio between the peak and volume for different hydrograph shapes.

basin (Mediero et al., 2007; Garrote et al., 2008). However, if a large and sufficiently varied set of observed hydrographs is available, it can be utilised as a random sample.

A large set of 919 observed hydrographs is available in the 32nd homogeneous region. The variability of hydrograph shapes in this set was measured by two variables: the time of the peak ($H_\mathrm{p}$) and the location of the hydrograph centroid ($H_\mathrm{c}$) (Eq. 12–13). These variables were standardised to be dimensionless and comparable, with $H_\mathrm{c}$ a modification of the shape mean variable ($S_\mathrm{m}$) developed by Yue et al. (2002). It can be seen that both variables show a wide variability in the feasible space, which means that the observed hydrographs present different shapes, thus, the variability is enough so that they can be used as a random sample to generate synthetic hydrographs (Fig. 5).

$$H_\mathrm{p} = \frac{t_{\mathrm{p},i,j}}{D_{i,j}} \tag{12}$$

$$H_\mathrm{c} = \frac{1}{V_{i,j}} \sum_{k=1}^{D_{i,j}-1} \left( \frac{x_{i,j}}{D_{i,j}} \cdot V_{i,j,k} \right) \tag{13}$$

The third step of the generation was conducted as follows. First, the ratio between peak and volume is computed for each synthetic $Q$–$V$ pair, and the observed hydrograph shape with the most similar ratio is selected (Fig. 6). Then, the hydrograph is resized by the synthetic peak discharge. The synthetic hydrographs retain the statistical properties of the

hydrograph durations of the observed data for both regressions at each case study, except for the local regression in the Entrepeñas reservoir, which gives much higher durations than those observed (Fig. 3).

## 7 Design flood hydrographs

The DFH is a high magnitude flood hydrograph that ensures the dam's safety to a given level and is represented by its low probability being exceeded. In Spain, the top of the surcharge pool is fixed so that it will not be exceeded by a flood with a return period less than 1000 years. In practice, the flood hydrograph for a return period of $T$ years is constructed with the $T$-year peak flow and the output volume of a hydrological model, calibrated in the basin. In the case where the volume frequency curve is known, the $T$-year volume is used, so that the $T$-year flood hydrograph has $T$-year peak and $T$-year volume. But, the probability of occurrence of that hydrograph is unknown, as it is the joint probability of the marginal probabilities of the peak and volume.
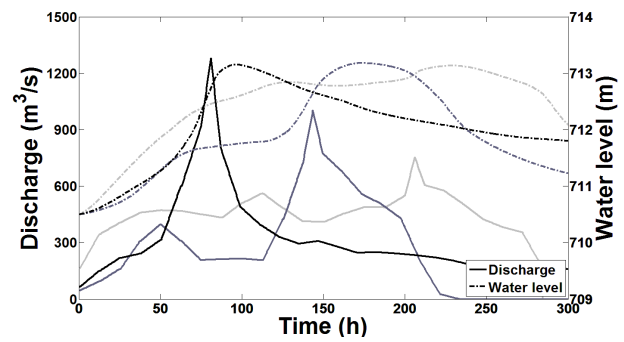
The hydrograph of a $T$-year return period must be defined in terms of risk to the dam as the inverse of its probability of exceeding a maximum water level in the reservoir or a maximum released flow, rather than estimating its probability of occurrence. Therefore, the risk of a flood can only be known by being routed through the reservoir.
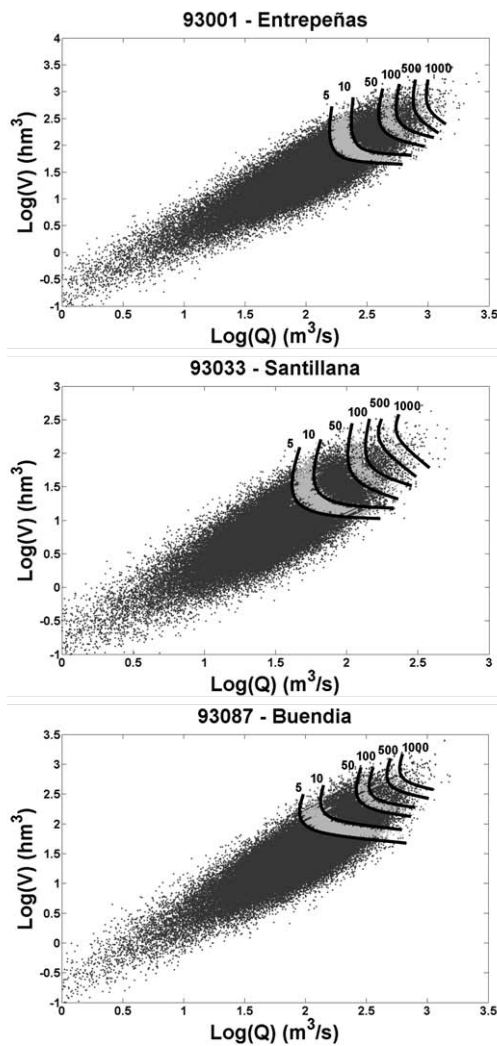
Fig. 7. Return period curves from the maximum reservoir level attained during the routing process.



**Fig. 8.** Example of three DFH that yield the same maximum reservoir level of 713.13 m, which is the level for a return period of 500 years in the Buendía reservoir.

**Table 5.** Reservoir levels for different return periods ($T$) and exceedance probabilities ($p$).

| | | Reservoir level (m) | | |
|---|---|---|---|---|
| $T$ (years) | $p$ | Santillana | Entrepeñas | Buendia |
| 5 | 0.2 | 890.55 | 716.17 | 711.24 |
| 10 | 0.1 | 890.98 | 716.52 | 711.50 |
| 50 | 0.02 | 891.96 | 717.28 | 712.13 |
| 100 | 0.01 | 892.39 | 717.62 | 712.45 |
| 500 | 0.002 | 893.42 | 718.54 | 713.13 |
| 1000 | 0.001 | 893.85 | 718.90 | 713.47 |

Each set of 100 000 synthetic hydrographs was routed through the corresponding reservoir. The reservoir level at the beginning of the flood was assumed to be at the top of the conservation pool, which is the traditional practice for dam design. For the sake of simplicity, an uncontrolled spillway was assumed, so the maximum level leads to the maximum release. Then, each set of synthetic hydrographs was sorted according to the maximum water level obtained while routing the hydrograph through the reservoir. The maximum reservoir levels for different $T$-year return periods were calculated as reservoir levels with an exceedance probability of $1/T$ over the total number of hydrographs (Table 5).

In the two-dimensional space $Q$–$V$, there will not be a unique hydrograph for a $T$-year return period, but rather a curve with a set of hydrographs that yield the same maximum reservoir level (Fig. 7). The dependence of the return period on each variable can be determined from these curves.

The milder the slope of the curve, the greater the dependence on the volume, and the steeper the slope, the greater the dependence on the peak. For a return period of 5 years, the Buendía reservoir has the mildest curve, which shows peak value ranges from 91.5 to 708.8 m$^3$/s (1.5–217 years of return period in the marginal distribution) and volume ranges from 60.1 to 191.9 hm$^3$ (2.4–9.7 years). This means that the return period of the hydrographs is mainly given by the return period of the volumes. On the other hand, the Santillana reservoir has the steepest curve. The peak ranges from 60.7 to 151.5 m$^3$/s (1.9–12.5 years) and the volume ranges from 11.2 to 99.3 hm$^3$ (2–304 years). In this case, the return period of hydrographs is mainly given by the peak discharges. The Entrepeñas reservoir is an intermediate case, with peaks ranging from 183.1 to 695.8 m$^3$/s (2.5–67 years) and a volume ranging from 43.3 to 201.5 hm$^3$ (2.1–21.7 years). Thus, the return period of the hydrographs depends on both variables.

The DFH for a given $T$-year return period is a hydrograph that yields the maximum reservoir level with an exceedance probability of $1/T$, and it has been seen that there are different hydrographs that meet that condition. In the case of the Buendía reservoir, three hydrographs for a return period of 500 years were selected (Fig. 8). The first hydrograph has a peak of 750 m$^3$/s ($T = 282$ years) and a volume of 435 hm$^3$
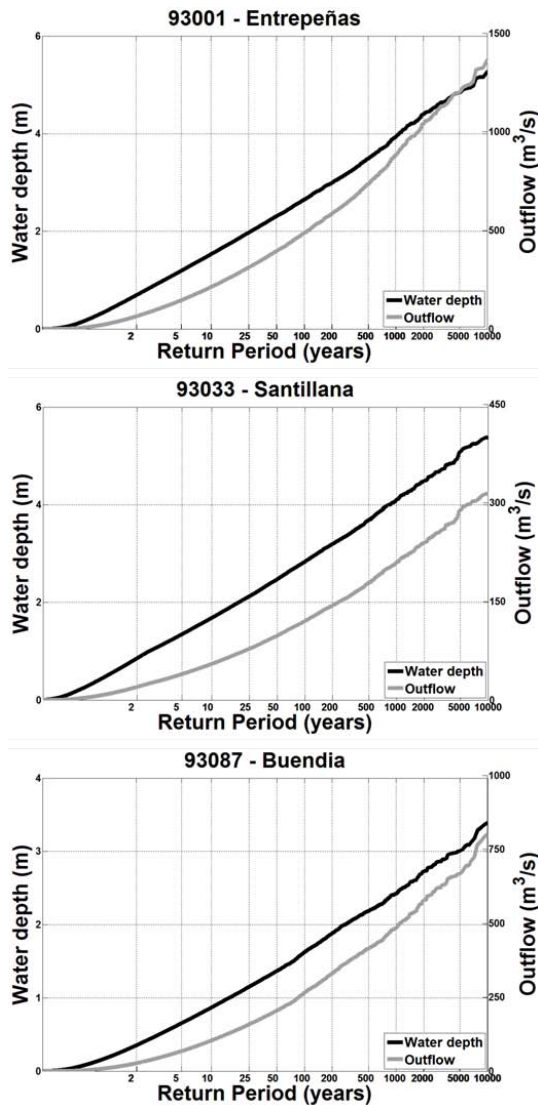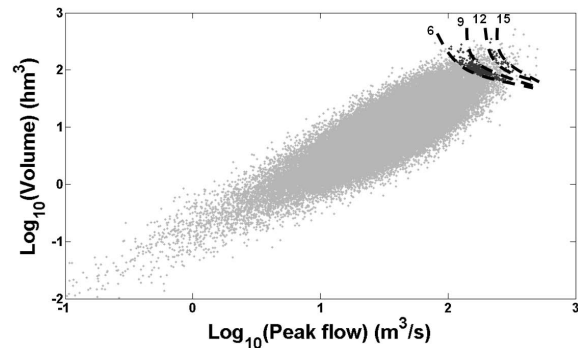
**Fig. 9.** Frequency curves of the water level over the spillway crest and release.

**Table 6.** Exceedance probabilities ($p$) of a reservoir water level of 894 m for different spillway crest lengths in the Santillana reservoir. Exceedance probabilities were transformed into return periods.

| Length (m) | $p$ | Return period (years) |
|---|---|---|
| 6 | 0.0055 | 182 |
| 9 | 0.0021 | 485 |
| 12 | 0.0008 | 1266 |
| 15 | 0.0003 | 2857 |



**Fig. 10.** Floods on the curves lead to a maximum reservoir level of 894 m for different spillway crest lengths, 6, 9, 12 and 15 m.

($T = 49$ years); the second shows a peak of $1000\,\mathrm{m^3/s}$ ($T = 1152$ years) and a volume of $579\,\mathrm{hm^3}$ ($T = 95$ years); and the third has a peak of $1280\,\mathrm{m^3/s}$ ($T = 4347$ years) and a volume of $825\,\mathrm{hm^3}$ ($T = 226$ years). These three hydrographs, with different peaks, volumes and shapes, yield a maximum reservoir level of 713.13 m, which is the level for a return period of 500 years.

The risk at the dam and in the downstream reach can be determined by the frequency curves of water levels over the spillway crest and releases (Fig. 9). An increase of the top of the dam can be deduced from the water depth frequency curve and additional river defenses could be required downstream from the dam to achieve a safety level from being flooded.

In addition, the return period curves depend on the spillway length and it can be designed from the probability of exceeding a given water level. This is particularly useful in the case where there is a maximum level that should not be exceeded, for instance, to prevent a village from being flooded. The spillway length can be selected in terms of the risk of exceeding that threshold.

Assuming that the water level at the Santillana reservoir cannot exceed an elevation of 894 m, the exceedance probability of this water level was calculated for different spillway lengths, 6, 9, 12 and 15 m, and these probabilities were transformed into return periods (Fig. 10 and Table 6). A minimum spillway length of 12 m should be selected to have a low enough probability of exceedance and risk to exceed that level, e.g., a return period greater than 1000 years or an exceedance probability lower than 0.001.

In the case where a restriction of maximum discharge downstream of the dam also exists, the spillway length can be selected from both curves, minimizing the risk of exceeding a water level and the risk of exceeding an outflow discharge downstream from the dam (Fig. 11).
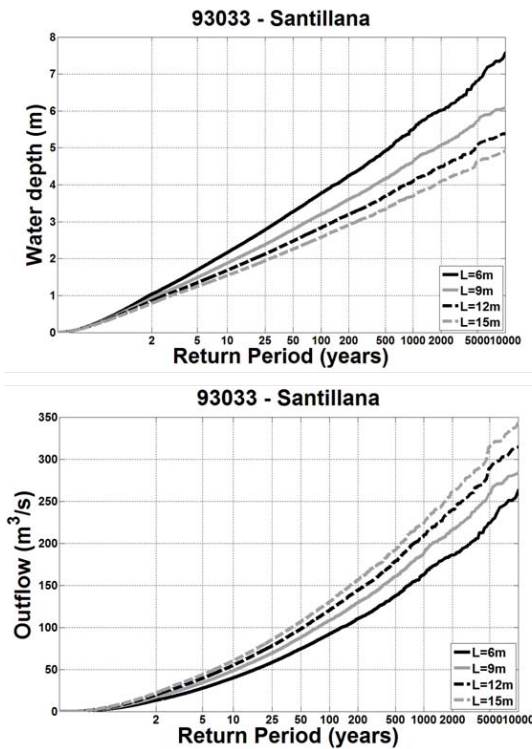
**Fig. 11.** Frequency curves of the water depth over the spillway crest and release for different lengths of the spillway crest in the Santillana reservoir.

## 8 Conclusions

A methodology for generating flood hydrographs that preserves the statistical properties of the peak, volume and duration marginal distributions has been developed. This methodology takes advantage of the regional studies of peak flows and hydrograph volumes that have been conducted recently in Spain and shows that a homogeneous region in terms of peak flow is also homogeneous in terms of the hydrograph volume. The accuracy of the peak and volume frequency curves was improved thanks to these regional analyses, which led to a regional shape parameter or regional L-CS to enhance the estimations for the higher return periods.

The relationship between peaks and volumes was analysed in the log-log space at the local and regional scales. A linear relationship exists between standardised peaks and volumes in a homogeneous region. These relationships were simulated by a regression equation and their variability was assessed by the residual variance of the regression.

A large set of synthetic peaks was generated from the peak frequency curve. The volumes linked to these peaks were generated by a regression equation and a normal randomisation to take into account the variability in the relationship between the peaks and flows. Finally, a hydrograph shape was linked to each $Q$–$V$ pair from the ratio between the peak and

volume. The synthetic sets thoroughly preserve the statistics of the peak and duration frequency curves and fairly keep the statistics of the volume frequency curve.

The set of synthetic hydrographs is particularly useful for dam design and assessing dam safety in terms of risk. Through routing the synthetic hydrographs through the reservoir, the maximum level and maximum release for each hydrograph can be known so that the return period can be fixed in terms of the maximum water level at the reservoir. There is not a unique hydrograph, but a curve with different combinations of peaks and volumes, which led to a given risk and return period. The most influential variable can be determined from the slope of these curves. The milder the slope of the curve, the greater the dependence on the volume, and the steeper the slope, the greater the dependence on the peak.

The probability distributions of water depths over the spillway crest and releases can also be determined. These distributions are useful for assessing the safety level of the dam from a hydrological point of view. Finally, the spillway length can be designed in terms of the probability of exceeding a certain water level, as the risk to the dam and the probability of exceeding an outflow discharge, as the risk of flooding a location downstream from the dam.

## Appendix A

### List of symbols used in the equations

$a_j$    vertical axis interception coefficient of the local regression equation at station $j$

$a_R$    vertical axis interception coefficient of the regional regression equation

$b_j$    slope regression line coefficient of the local regression equation at station $j$

$b_R$    slope regression line coefficient of the regional regression equation

$D_{i,j}$ duration of the hydrograph in the year $i$ at station $j$ [h]

$H_i$    $i$th heterogeneity measure

$k$    shape parameter of the GEV distribution

$N$    number of gauge stations in the region

$n_j$    length of the sample at site $j$

$p$    number of variables of the regression equation (equal to one in this paper)

$Q_0$    discharge at the starting time of the receding limb [m³/s]

$Q_{i,j}$ observed maximum peak in year $i$ at station $j$ [m³/s]

$q_{i,j}$ standardised maximum peak discharge in year $i$ at station $j$ [−]

$\overline{Q_j}$    mean value of the AMD series at station $j$ [m³/s]

$Q_r$    recession discharge at any time after the beginning of the receding limb [m³/s]

$t_{i,j}$    L-Moment ratio of the $i$th order at site $j$

$t_i^R$    regional value of the L-Moment ratio of the $i$th order

$t_{p,i,j}$ time of the peak of the hydrograph in year $i$ at station $j$ [h]

| | |
|---|---|
| $u$ | location parameter of the GEV distribution |
| $V_{i,j}$ | observed maximum volume in year $i$ at station $j$ [hm$^3$] |
| $v_{i,j}$ | standardised maximum volume in year $i$ at station $j$ [−] |
| $V'_{i,j}$ | volume estimated by the regression equation in year $i$ at station $j$ [hm$^3$] |
| $V_{i,j,k}$ | hydrograph volume between $t_k$ and $t_{k+1}$ in year $i$ at station $j$ [hm$^3$] |
| $\overline{V_j}$: | mean value of the AMV series at station $j$ [hm$^3$] |
| $W_i$ | weighted standard deviation of the at site sample L-Moment ratio of the $i$th order |
| $x_{i,j}$ | time distance from the beginning of the hydrograph to the centroid of $V_{i,j}$ [h] |
| $\alpha$ | scale parameter of the GEV distribution |
| $\beta$ | recession constant [h$^{-1}$] |
| $\mu(W_i)$ | mean of the simulated values of $W_i$ |
| $\sigma_{\mathrm{reg},j,}$ | residual variance of the regression equation at station $j$ |
| $\sigma_{\mathrm{reg,R}}$ | residual variance of the regional regression equation |
| $\sigma(V_i)$ | standard deviation of the simulated values of $W_i$ |

Edited by: S. Grimaldi

## References

Blöschl, G. and Sivalapan, M.: Process controls on regional flood frequency: coefficient of variation and basin scale, Water Resour. Res., 33(12), 2967–2980, 1997.

Bocchiola, D., De Michele, C., and Rosso, R.: Review of recent advances in index flood estimation, Hydrol. Earth Syst. Sci., 7, 283–296, doi:10.5194/hess-7-283-2003, 2003.

Cunnane, C.: A note on the Poisson assumption in partial duration series models, Water Resour. Res., 15(2), 489–494, 1979.

Cunnane, C.: Methods and merits of regional flood frequency analysis, J. Hydrol., 100, 269–290, 1988.

Cunnane, C.: Statistical distributions for flood frequency analysis, World Meteorological Organization, Geneva, Switzerland, Operational Hydrology Rep., 23–33, 1989.

De Michele, C., Salvadori, G., Canossi, M., Petaccia, A., and Rosso, R.: Bivariate statistical approach to check adequacy of dam spillway, J. Hydrol. Eng., 10(1), 50–57, 2005.

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., and Bobée, B.: Multivariate hydrological frequency analysis using copulas, Water Resour. Res., 40, W01101, doi:10.1029/2003WR002456, 2004.

Fill, H. D. and Steiner, A. A.: Estimating instantaneous peak flow from mean daily flow data, J. Hydrol. Eng., 8(6), 365–369, 2003.

Garrote, L., Molina, M., and Mediero, L.: Learning Bayesian networks from deterministic rainfall-runoff models and Monte Carlo simulation, in: Practical Hydroinformatics, Water Science and Technology Library, 68, Springer, Berlin, Germany, 375–388, 2008.

Goel, N. K., Seth, S. M., and Chandra, S.: Multivariate modeling of flood flows, J. Hydraul. Eng., 124(2), 146–155, 1998.

GREHYS: Presentation and review of some methods for regional flood frequency analysis, J. Hydrol., 186, 63–84, 1996.

Grimaldi, S. and Serinaldi, F.: Asymmetric copula in multivariate flood frequency analysis, Adv. Water Resour., 29, 1155–1167, 2006.

Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency analysis, Water Resour. Res., 29, 271–281, 1993.

Hosking, J. R. M. and Wallis, J. R.: Regional frequency analysis: An approach based on L-Moments, Cambridge University Press, New York, USA, 1997.

Iacobellis, V., Claps, P., and Fiorentino, M.: Climatic control on the variability of flood distribution, Hydrol. Earth Syst. Sci., 6, 229–238, doi:10.5194/hess-6-229-2002, 2002.

Kjeldsen, T. R. and Jones, D.: Estimation of an index flood using data transfer in the UK, Hydrolog. Sci. J., 52(1), 86–98, 2007.

Krstanovic, P. F. and Singh, V. P.: A multivariate stochastic flood analysis using entropy, in: Hydrologic Frequency Modelling, Reidel, Dordrecht, The Netherlands, 515–539, 1987.

Mediero, L., Garrote, L., and Martín-Carrasco, F.: A probabilistic model to support reservoir operation decisions during flash floods, Hydrolog. Sci. J., 52(3), 523–537, 2007.

Noto, L. V. and La Loggia, G.: Use of L-Moments Approach for regional flood frequency analysis in Sicily, Italy, Water Resour. Manag., 23(11), 2207–2229, 2009.

Ponce, V. M.: Engineering Hydrology, Principles and Practices, Prentice Hall, Upper Saddle River, New Jersey, USA, 1989.

Sackl, B. and Bergmann, H.: A bivariate flood model and its application, in: Hydrologic Frequency Modelling, Reidel, Dordrecht, The Netherlands, 571–582, 1987.

Salvadori, G. and De Michele, C.: Frequency analysis via copulas: theoretical aspects and applications to hydrological events, Water Resour. Res., 40, W12511, doi:10.1029/2004WR003133, 2004.

Xiao, Y., Guo, S., Liu, P., Yan, B., and Chen, L.: Design Flood Hydrograph based on multicharacteristic synthesis index method, J. Hydrol. Eng., 14 (12), 1359–1364, 2009.

Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P., and Bruneau, P.: The Gumbel mixed model for flood frequency analysis, J. Hydrol., 226, 88–100, 1999.

Yue, S.: The bivariate lognormal distribution to model a multivariate flood episode, Hydrol. Process., 14, 2575–2588, 2001.

Yue, S., Ouarda, T. B. M. J., Bobée, B, Legendre, P., and Bruneau, P.: Approach for describing statistical properties of flood hydrograph, J. Hydrol. Eng., 7(2), 147–153, 2002.

Zhang, L. and Singh, V. P.: Bivariate flood frequency analysis using the Copula method, J. Hydrol. Eng., 11(2), 150–164, 2006.

Zhang, L. and Singh, V. P.: Trivariate flood frequency analysis using the Gumbel-Hougaard Copula, J. Hydrol. Eng., 12(4), 431–439, 2007.