Design flood hydrographs: a regional analysis based on flood reduction functions

Daniele Ganora[1], Giulia Evangelista[1], Silvia Cordero[2], and Pierluigi Claps[1]

[1]Politecnico di Torino
[2]Agenzia Interregionale Po, Ufficio periferico di Torino

November 22, 2022

Abstract

Flood hazard mapping and the design of many water infrastructures are commonly based on the use of a single hydrologic variable, the design instantaneous peak flow. However, the entire flood hydrograph (or at least the flood volume) is needed in many circumstances, including the evaluation of potential risks in dam safety analysis, the design of detention basins, the application of inundation methods or of river levee failure. While many efforts have been made in the last decades to improve the peak flow estimation in a generic section of a river network, procedures for the systematic estimation of flood volumes in ungauged sections over large regions are not consolidated yet. In this paper, the estimation of the flood volumes in ungauged basins is developed, based on the Flood Reduction Function (FRF), a parsimonious representation of the flood hydrograph structure. The FRF is a volume-duration relationship that allows to easily extract flood hydrographs based on few parameters, that exhibit marked regularity. Based on data from 87 basins (763 station-years of flood hydrographs) in the Northwest Italy, a two-parameter FRF has been considered to build a regionalized estimation procedure. The estimation of FRF parameters in ungauged basins has been obtained with different procedures. Results suggest that the multiple linear regression model can be an effective method to estimate the FRF in ungauged basins, producing nearly unbiased predictions and design hydrographs which reasonably resemble the observed ones.

Design flood hydrographs: a regional analysis based on flood reduction functions

D. Ganora[1], G. Evangelista[1], S. Cordero[2], P. Claps[1]

[1]Politecnico di Torino – Department of Environment, Land and Infrastructure Engineering. Corso Duca degli Abruzzi 24, 10129, Torino, Italy.

[2]Agenzia Interregionale Po, Ufficio periferico di Torino. Via Pastrengo 2/ter., 10024, Moncalieri (TO), Italy.

Corresponding author: Giulia Evangelista (giulia.evangelista@polito.it)

Key Points:

- A parsimonious regional statistical model to estimate flood hydrographs in ungauged sites is proposed.
- New observations of flood volumes in 87 river sections in Italy are used, along with several attributes of the upstream watersheds.
- Design hydrograph shapes in ungauged basins can be estimated from hydro-geomorphologic basin features through linear multiple regressions.

Abstract

Flood hazard mapping and the design of many water infrastructures are commonly based on the use of a single hydrologic variable, the design instantaneous peak flow. However, the entire flood hydrograph (or at least the flood volume) is needed in many circumstances, including the evaluation of potential risks in dam safety analysis, the design of detention basins, the application of inundation methods or of river levee failure. While many efforts have been made in the last decades to improve the peak flow estimation in a generic section of a river network, procedures for the systematic estimation of flood volumes in ungauged sections over large regions are not consolidated yet.

In this paper, the estimation of the flood volumes in ungauged basins is developed, based on the Flood Reduction Function (FRF), a parsimonious representation of the flood hydrograph structure. The FRF is a volume-duration relationship that allows to easily extract flood hydrographs based on few parameters, that exhibit marked regularity. Based on data from 87 basins (763 station-years of flood hydrographs) in the Northwest Italy, a two-parameter FRF has been considered to build a regionalized estimation procedure. The estimation of FRF parameters in ungauged basins has been obtained with different procedures. Results suggest that the multiple linear regression model can be an effective method to estimate the FRF in ungauged basins, producing nearly unbiased predictions and design hydrographs which reasonably resemble the observed ones.

Plain Language Summary

Regional statistical analyses allow to obtain estimates of hydrologic variables related to ungauged sites or basins. We suggest here to study the maximum flood volumes for given durations and reach a state-of-the-art regional estimation method, starting from a new dataset of flood hydrographs in NW Italy. We use a representation of flood volumes called flood reduction function (FRF), that resembles the intensity-duration curve built for the extreme rainfall and is computed similarly from the raw data. As for the design rainfall events, the design hydrographs can analytically derive from an analytical form of the FRF. Local empirical 2-parameter curves are built on 87 stations and different statistical frameworks are applied, to try to relate the FRF parameters to hydro-geomorphological features of the basins. A rather simple multiple linear regression model is the preferred one, which connects FRF curves to indices of extreme rainfall and to basin features connected to the watershed lag time. Applying the regional model selected, we obtain reasonably shaped design hydrographs, as compared to the observed ones.

1 Introduction

Modern hazard and flood risk analysis, used in the design of flood mitigation infrastructures, must rely on reliable information about hydrograph volume and shape, in addition to the peak flow value. While the estimation of design flood peaks in ungauged basins has a long history (Gumbel, 1945; Cunnane, 1988; Castellarin et al., 2012) and many operational models are currently available, methods to estimate flood volume (and hydrograph shape) are still limited and not well consolidated (see e.g. Tomirotti et a., 2017, Brunner et al., 2017). Historical records usually include only annual maximum peak flows and even if hydrograph tracks are available on paper, they require a digitalization effort. With a very limited availability of flood volume data in many countries, no wonder is the lack of regional statistical characterizations of such complex curves.

Nowadays, design hydrographs are typically determined with indirect methods (e.g., rainfall-runoff procedures), while their direct evaluation (i.e., based on discharge observations) is not so useful, since many interesting watersheds are ungauged. Since most of the available studies only focus on the volume estimation, leaving the hydrograph shape to a subjective choice (e.g., triangular, rectangular, etc.), Yue et al. (1999) underline that a multivariate joint probability distribution should be applied, due to the positive correlation between flood peak and volume. In the same direction, copula-based bivariate modelling of peak and volume have been developed: see Bacova Mitková and Halmová (2014), Salvadori and De Michele (2007), Zhang and Singh (2006) among others. Xiao et al. (2009) propose a different approach, i.e., a Multi-characteristic Synthesis Index, with the aim of simultaneously accounting for different hydrograph features (e.g., peak and volume for 1, 3, 7, etc. days of duration), in order to define an "overall" design hydrograph. Yue et al. (2002) describe the design hydrograph with a 2-parameter Beta function, defined by the centroid and the variance of the flood hydrograph. On the simulation side, Mediero et al. (2010) developed a Monte-Carlo methodology for generating flood hydrographs, consistent with the available statistical properties of the peak, volume and duration. More recently, Requena et al. (2016) presented a bivariate procedure to extend flood series through the generation of peak-volume samples, while Brunner et al. (2017) estimate the synthetic hydrograph using a dimensionless lognormal probability distribution, scaled according to flood peak and volume values, modelled through a bivariate copula. In the field of urban drainage, a derived distribution approach had been proposed by Guo and Adams (1998), who obtained the distribution of peaks and volumes in analytical form starting from the distribution of rainfall events. This approach has not been substantially followed on natural basins.

A different and less common approach to characterize flood volumes relies on the use of the flood reduction function (FRF), which describes, for a given return period, the maximum flood volume in a given duration regardless of the hydrograph shape (e.g. Bacchi et al., 1992). The FRF curve is

conceptually similar to the average intensity-duration-frequency (IDF) function used to represent rainfall depths for a given duration (Grimaldi et al., 2011). Analogously to the design hyetograph in the "Chicago" method, the design hydrograph shape can be constrained to present partial volumes compatible with the form of the FRF. However, the shape of the curves over time (hydrograph and hyetograph) are both undetermined.

Whatever the adopted model, to devise a design flood hydrograph in ungauged basins (i.e., where few or no observations are available) specific procedures to estimate the model's parameters are required. These approaches are commonly referred to as Regional Statistical Models and have been developed for many hydrological variables, e.g. mean annual flow, seasonal flow regime, flow duration curves and low flows, flood peaks (Blöschl et al., 2013). Application of regional models to flood volumes are indeed quite rare, as the only available approach, to our knowledge, is that of Tomirotti and Mignosa (2017). In this paper, we address the regionalization of the flood hydrograph volume via the Flood Reduction Function schematization, using data of empirical FRF obtained in 87 watersheds in the upper Po river basin, in the North West of Italy.

2 Methodology

2.1 Flood Reduction Function (FRF)

Flood reduction function (FRF) is a curve representing, for a given duration D, the maximum value of the average discharge Q_D computed on all the possible time windows of duration D over a period of interest. The curve is usually normalized by the instantaneous maximum discharge, Q_{PEAK} (e.g., $Q_{D=0}$), of the same period as:

$$\varepsilon_D = \frac{Q_D}{Q_{PEAK}} = \frac{1}{Q_{PEAK}} \max\left(\frac{1}{D}\int_t^{t+D} Q(\tau)\, d\tau\right) \qquad (1)$$

In the practical applications the period of interest for the selection of maxima is usually the year and the observations of Q are recorded at discrete time steps (e.g., 10 minutes). For a specified value of D, ε_D is also referred to as the "flood reduction ratio". Following the classical approach used for flow duration curves and intensity-duration-frequency curves (Chow, 1951), the empirical FRF is computed for all the available years, normalized by the corresponding annual maximum. Subsequently, they are averaged over the years, to obtain a basin representative $\overline{\varepsilon_D}$ curve (see fig. 1b) to be used in practical applications.

Over the years, for each duration D one computes a sample of ε_{D_j} values (j=1…N) so that a statistical treatment would enable to estimate a quantile $\varepsilon_{D,T}$. This can be done for each duration and, in

principle, the probability distribution may vary among durations. However, it has been shown (Franchini and Galeati, 2000) that it is generally acceptable to use a unique probability distribution for all durations D. This enables to use a simple expression for the estimation of the quantile of the flood volume for design purposes, as:

$$W_{D,T} = \overline{\varepsilon_D}\, Q_T\, D \qquad (2)$$

In (2) Q_T is the quantile of the flood peak for a given return period, that can be represented as $Q_T = \bar{Q} \cdot K_T$ according to the index method (Darlymple, 1960). This entails that the non-dimensional probability distribution of flood peaks, K_T, is adopted as the non-dimensional distribution of the flood volumes, regardless of the duration D of interest.

Eq. (2) can be used to build synthetic design hydrographs by derivation of the volume function:

$$\hat{Q}(t) = \frac{dW}{dt} = Q_T \cdot \overline{\varepsilon_D} + t \cdot Q_T \cdot \overline{\varepsilon_D}' \qquad (3)$$

where $\hat{Q}(t)$ is the synthetic hydrograph and $\overline{\varepsilon_D}'$ indicates the first derivative of ε with respect to the duration *D*. It is important to stress the conceptual difference between the chronological time *t* and the duration *D*, the latter being essentially a moving time window.

According to the UK National Environmental Research Council (NERC, 1975), the average FRF can be represented by a 2-parameter curve:

$$\varepsilon_D = \frac{Q_{D,T}}{Q_T} = (1 + b \cdot D)^{-c}\ , \qquad (4)$$

where b and c are parameters to be determined. The dependence on the return period is, again, only concentrated in the peak flow Q_T, so that the form of the curve (4) is independent on the return period. In the scientific literature, other analytical forms of the FRF have been proposed: Franchini and Galeati (2000) compared different analytical models (the NERC; the geomorphoclimatic model by Fiorentino et al., 1987; the stochastic model by Bacchi et al., 1992) against the empirical FRFs of 12 basins in central Italy. All the models showed a reliable fitting to the observed FRF, although the geomorphoclimatic model was more complex to apply. As the NERC function allows double curvature and it is compatible with a conceptual framework (see Section 2.2), we have considered the NERC as the best choice for an analytical FRF to be regionalized, although the results obtained in this work can be easily generalized to other kind of functions.

Summarizing, one can use the empirical FRF at a gauged location and eq. (3) to obtain a single representative form as a design hydrograph. The procedure can be summarized with a few steps sketched in Figure 1:

1. starting from a discharge time series (fig. 1a) for each year, the empirical FRF of eq. (1) can be computed by considering different time windows (fig. 1b, thin grey lines);
2. the average empirical FRF is then obtained by averaging the individual values for each duration (fig. 1b, black dots);
3. the analytical NERC model (eq. 4) is fitted to the average FRF, providing an analytical representation of the FRF (fig. 1b, solid blue line);
4. a synthetic hydrograph consistent with eq. (4) can be built from the fitted FRF.

To apply step 4, it is first necessary to define the peak position as, with respect to the duration D, if the peak is at t=0, the hydrograph shape is that represented as a red solid line in Figure 1c. To obtain a symmetric hydrograph, as the red dashed line of Figure 1c, the equation must be rewritten as:

$$\hat{Q}(t) = Q_T \cdot [(1 + b\mathrm{t})^{-c} - \mathrm{cbt}(1 + bt)^{-c-1}] \qquad (5)$$

Of course, both the hydrograph forms are consistent with the same FRF, (i.e., eq. 3 and eq. 5 lead to the same ε_D values) when recomputing the volumes over moving time windows.

Considering two asymmetrical limbs, a more general analytical form of the hydrograph shape can be written as:

$$\hat{Q}(t) \begin{cases} \left(1 + \frac{b}{1-r}|t|\right)^{-c} - \frac{b \cdot c}{1-r}|t|\left(1 + \frac{b}{1-r}|t|\right)^{-c-1} & t < 0 \\ \left(1 + \frac{b}{r}t\right)^{-c} - \frac{b \cdot c}{r}t\left(1 + \frac{b}{r}t\right)^{-c-1} & t \geq 0 \end{cases} \qquad (6)$$

where the shape depends on the "skew" parameter r ($0 \leq r \leq 1$). The symmetric hydrograph (with central peak) is generated from eq. (6) with r=0.5. The “initial peak” hydrograph is obtained with with r=0. An example of asymmetrical shape is the blue curve shown in Figure 1c, obtained with r = 0.68.

Some authors (Tomirotti and Mignosa, 2017) let the *r* parameter vary with the hydrograph duration, after considering various real hydrographs on large rivers. In this study, the parameter *r* is considered constant for each basin, i.e., it is independent of *D*. The reasons of this choice are discussed later. It should also be clarified that here, unlike in the work by Tomirotti and Mignosa (2017), the skew parameter r is the meaning of the ratio between the time after the peak and the duration D.

A specific analysis on the hydrograph shape is offered in Section 5.1, as an additional validation to the proposed method.

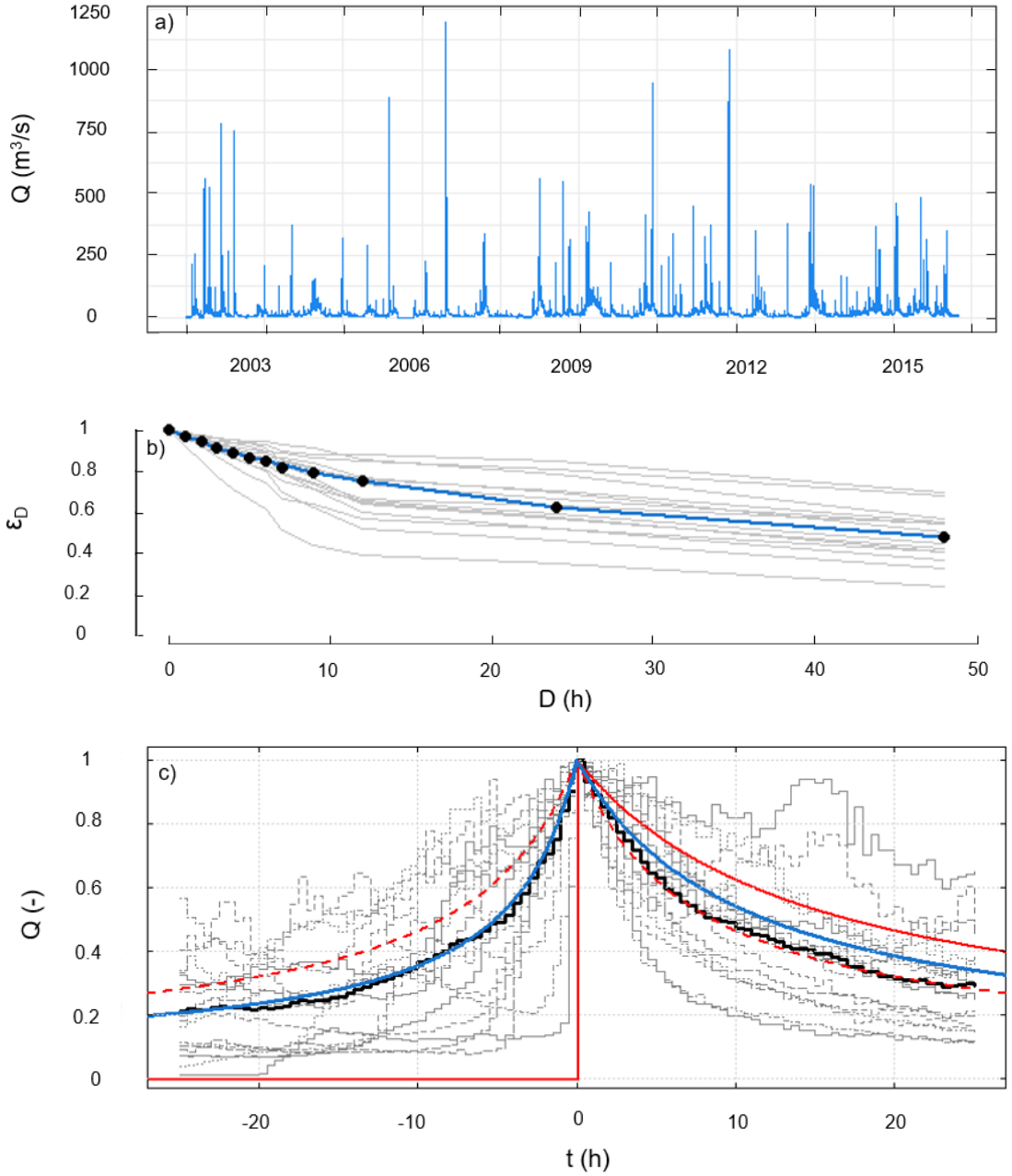

Figure 1: Example of FRF hydrograph analysis for the Stura di Lanzo at Torino basin. a) Time series of discharge values recorded at 30' resolution. b) Empirical annual FRFs (thin-grey lines), empirical average FRF (black dots) and NERC analytical FRF (b=0.05627, c=0.55830). c) NERC hydrograph obtained from FRF of panel b) with r = 0.68 (solid blue line) compared to the reference hydrograph (black line) and the single event observed hydrographs (grey lines). Red dashed and solid curves show symmetrical NERC hydrograph (r=0.5) and NERC hydrograph with instantaneous initial peak (r=1), respectively.

Basically, the regionalized methods presented in this paper address the step 3 of the above procedure in ungauged basins. The following section provides the theory and in Section 3 the application in North-West of Italy demonstrates the feasibility.

2.2 Estimation of the Flood Reduction Function in ungauged basins

As mentioned in the Introduction, while regionalization of peak flow values is a consolidated practice, with many procedures available, much less can be found in literature as regards the regionalization of other hydrograph-related characteristics. Something different from the regionalization of peaks can be found in NERC (1975), where the non-dimensional (with respect to the mean annual daily flood) FRF values at 3 and 10 days (AR3 and AR10 respectively) were related to catchment characteristics through linear regressions. The analysis was based on a sample of 64 stations and an initial set of 4 catchment descriptors; the final regional models to estimate AR3 and AR10 resulted both function of the stream slope. Much later, another approach was proposed by Maione et al. (2003) and later followed by Tomirotti and Mignosa (2017), for the regional estimation of the FRF (named Flood Duration Frequency in the original paper). Maione et al. (2003) used a single-parameter FRF (Bacchi et al., 1992) to correlate this parameter to the watershed area by a linear regression fitted on 8 gauged basins in the Po basin (Italy) with 46 years of average record length. More recently, Brunner et al. (2018) tested different approaches for regionalization of the a synthetic normalised hydrograph shape. A total of 24 different approaches were tested to estimate the 10 parameters of a synthetic design hydrograph form proposed in the paper. They were linear regression techniques, nonlinear regression models, i.e., random forest, bagging and boosting, spatial proximity approaches and methods based on homogeneous regions. Strictly speaking, the FRF concept was not used.

The foundations of the procedures proposed here lie in a conceptual interpretation of the NERC Flood Reduction Function proposed by Silvagni (1984) who connected parameters *b* and *c* of the NERC curve to the parameters of the rational formula (Mulvaney, 1851). In practice, assuming that the peak Q_T can be estimated considering a rectangular design rainfall over the basin, with duration equal to the time of concentration t_c, Silvagni (1984) suggested that also $Q_{D,T}$ could be estimated through the rational formula but using a rainfall event having duration t_c+D. Assuming that the design rainfall intensity $i_T(d)$ for a given duration d and return period T can be expressed using a two-parameter IDF curve $i_T(d)=a_T d^{n-1}$ (e.g. Koutsoyiannis et al., 1998), the author obtained a 'rational' FRF expression as:

$$\varepsilon_D = \frac{Q_{D,T}}{Q_T} = \frac{i_T(t_c + D)A\varphi}{3.6}\left(\frac{i_T(t_c)A\varphi}{3.6}\right)^{-1} = \\ = \frac{a_T \cdot (t_c + D)^{n-1}}{a_T \cdot {t_c}^{n-1}} = \left(1 + \frac{D}{t_c}\right)^{n-1}, \quad (7)$$

where A is the basin area and φ is the runoff coefficient. Comparing eq. (7) with eq. (4) one can recognize that the parameters of the NERC FRF can assume the meaning of:

$$b = \frac{1}{t_c} \quad e \quad c = 1 - n. \tag{8}$$

Eqs. (4) and (8) can be used, in principle, to estimate *b* and *c* only from the IDF parameter *n* and the time of concentration t_c. By inverting the procedure, Franchini e Galeati (2000) observed that using several empirical FRFs to estimate t_c with the conceptual analogy of Silvagni (1984) the results were significantly different from those obtained estimating t_c with the most common equations in literature. They suggested that the parameter *b* of the NERC equation cannot be directly linked to the usual basin time of concentration. Rather, 1/*b* "should be interpreted as a more general critical time, characteristic of the basin response" in the FRF framework. In the following, we will then refer to the parameter 1/*b*, whose values can be referred to an intuitive meaning of "t_c".

The method of Regional Analysis proposed here is essentially built through the institution of relations between the two parameters of the NERC FRF and several basin characteristics. Three different regional statistical approaches are applied to an extensive dataset of hydrographs and Flood Reduction curves, to allow the estimation of the FRF parameters in ungauged basins. In particular, the methods considered are: the multiple linear regression (LR; e.g., Montgomery et al., 2001), the Canonical Correlation Analysis (CCA; e.g. Ouarda et al., 2000) and the Alternating Conditional Expectation algorithm (ACE; e.g. Breiman and Friedman, 1985). In all the techniques the basin characteristics, referred to as descriptors, include geographical, morphological and climatic basin attributes. These are related to the basins upstream of the available gauging stations and can be easily computed in any ungauged basin by means of GIS procedures.

For each regionalization approach tested, several alternative models, based on different subsets of descriptors, have been implemented, and subsequently ranked, according to their prediction performances, e.g., by the Adjusted R-square (R^2_{adj}). The most significant models are further validated with a visual checking of the results, and with a leave-one-out cross-validation procedure (see Hastie et al., 2009). In the following, the details of the applications are presented, and a final assessment of the most convenient method is discussed.

2.2.1 Multiple linear regression

Multiple linear regressions have been widely used to regionalize hydrological variables. An example of a prediction equation is:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p \tag{9}$$

where $\hat{y}$ is the (FRF) parameter to regionalize, *x* is a basin descriptor and β is its corresponding regression coefficient. In this work, the ordinary least squares method (e.g., Montgomery et al., 2001)

is used and the model considers both the NERC parameters *b* and *c* as the regionalized variable $\hat{y}$. Different possible transformations (log, Box-Cox (Box and Cox, 1964)) have been considered for both the variables set *x* and *y*, e.g. considering *y=c* and *y=b* or *y=ln(b)* or *y=*1/*b* or *y=ln(1/b)*. Only the most significant results are reported here, and, for instance, no transformation of *c* has provided satisfactory results.

Regarding the covariates *x*, a preliminary analysis of their frequency distribution showed that some of them are markedly skewed, and that the logarithmic and Box-Cox transformation can be effective to correct the skewness. For each set of transformations (on *y* and *x*), all the possible combinations of 2 and 3 descriptors have been computed, obtaining about 6,000 combinations. On the obtained results, the regression models are tested for significance (*t*-Student test at 5%), multicollinearity (VIF test, see Montgomery et al., 2001) and residual analysis (normal probability plot and homoscedasticity). The subset of the results passing all the tests is then ranked according to the R^2_{adj} computed on the variables back transformed to their original units. The Section 3.2 of this paper thoroughly describes the results of application of this procedure.

2.2.2 Canonical Correlation Analysis (CCA)

Canonical correlation analysis (CCA) is a method to explore relationships between two multivariate sets of variables, that are here represented by FRF parameters (b^{-1} and *c*) and by the basin descriptors. The CCA allows one to determine which is the linear combination of the variables of the latter group most correlated to a linear combination of the variables of the former group. CCA is widely used in statistic: e.g. multivariate regression and factorial discriminant analysis are special cases of the CCA method (Ouarda et al., 2001). Approaches belonging to this family are commonly applied in hydrology since the works of Snyder (1962) and Wong (1963). More recently Ouarda et al. (2000) developed a CCA-based procedure to assess the joint regional estimation of spring flood peaks and volume for Northern Canadian basins.

To resume the functioning of the CCA, let **X** be the *n* x *p* matrix of basins descriptors, where *n* is the number of basins in the dataset and *p* is the number of the considered descriptors, and let **Y** be the *n* x 2 matrix of the FRF parameters. The predicted parameters $\widehat{\mathbf{Y}} = \left[\frac{1}{\hat{b}} \quad \hat{c}\right]$ can be computed as

$$\widehat{\mathbf{Y}} = \varrho \cdot [\boldsymbol{x} - \overline{\boldsymbol{x}}]\boldsymbol{\Lambda} \cdot \mathbf{B^{-1}} + \overline{\mathbf{Y}} \qquad (10)$$

where the descriptors of the ungauged basin are included in vector **x** while each column of $\overline{\boldsymbol{x}}$ is the mean value of the corresponding descriptors in **X,** computed from the *n* gauged basins of the calibration dataset. Similarly, $\overline{\mathbf{Y}}$ is the vector of mean values of the FRF parameters computed from

the *n* gauged basins of the dataset. Two matrices of canonical variables are defined: $\mathbf{U}=[\boldsymbol{x}-\overline{\boldsymbol{x}}]\boldsymbol{\Lambda}$ and $\mathbf{V}=[\boldsymbol{Y}-\overline{\boldsymbol{Y}}]\mathbf{B}$. The canonical correlation between the j[th] pair of canonical variables is then:

$$\varrho = \frac{cov(u_j, v_j)}{\sqrt{var(u_j)var(v_j)}} \tag{11}$$

Matrices Λ and B contain the canonical coefficients, scaled to make the covariance matrices of the canonical variables the identity matrix, and ϱ is the square root of the corresponding eigenvalue (Ouarda et al., 2001).

The aim of the CCA is thus to find the coefficients a and b that maximize ϱ.

The results of the CCA application are reported later, in Section 3.3.

2.2.3 Alternating Conditional Expectation algorithm (ACE)

The Alternating Conditional Expectation (ACE) algorithm has been proposed by Breiman and Friedman (1985a, 1985b) as a non-parametric model to find those transformations that produce the best-fitting additive model. Considering *y* and $x_1,\ldots,x_p$ respectively as the response and the predictors random variables, the ACE algorithm provides a mapping function *t* for each variable which defines a set of non-parametric transformations. The prediction variable is then obtained as

$$\hat{y} = t_y^{-1}\left[t_{x_1}(x_1) + t_{x_2}(x_2) + \cdots + t_{x_p}(x_p)\right] \tag{12}$$

where t_y^{-1} is the inverse of the mapping function of the variable y, and t_{x_i} is the mapping function of the i[th] descriptor. The optimal transformations are achieved through an iterative series of optimizations. While the reader can refer to Breiman and Friedman (1985) for the algorithm details, it is worth recalling the practical procedure through a graphical example reported in Figure 2: after the mapping functions have been computed, the two descriptors x_1 and x_2 of the ungauged basin are entered in the respective mapping functions to obtain their transformed values; their sum is then back-transformed with the t_y mapping function to obtain the final estimate.

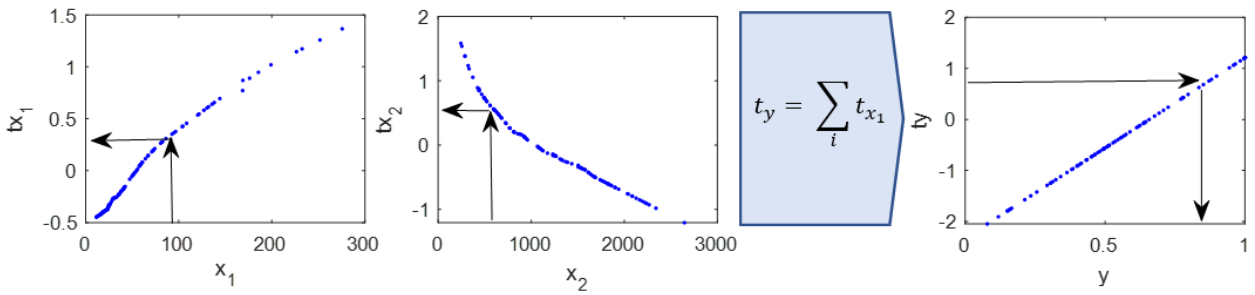

Figure 2: Example of ACE application: x_1 and x_2 are 2 independent variables and tx_1 and tx_2 their non-parametric transformations; y is the dependent variable obtained as back-transformation of t_y. The mapping functions are represented in blue dots as they are computed for each sample value.

With respect to the linear regression, the ACE approach can automatically detect possible efficient non-linear transformation of the variables (both *x* and *y*), so that no preliminary transformations are applied. The mapping function t_y has been forced here to be linear, to ensure a more robust inversion of t_y; no constraints are instead applied to the descriptors. All the numerical analyses have been performed with the R package "acepack" (Spector et al., 2016).
The results of the ACE application are reported in Section 3.4, later in this paper.

3 Case study and regional model building

3.1 Case study and data preparation

The methodologies presented in the previous section have been used to build regional models of the FRF curves in an area of about 25.000 km^2 in the North-West of Italy. The case study has been organized by assembling a new dataset of flood hydrographs, extracting flood waves from the continuous discharge time series originally recorded in 87 gauging stations of the Regional Agency for Environmental Protection (ARPA Piemonte). The dataset has been initially compiled using information available from previous, partially unpublished, studies that reported data manually collected by the former Italian Hydrographic Service. In particular, they consist in:

- 26 time series of hydrometric levels obtained from digitalization (at 15') of data recorded by analogic gauges during the period 1928-1994. These water levels have been transformed into discharge values using previously obtained rating curves for these stations (Claps et al., 2010);
- FRFs obtained by considering the three major events in a single year (analogic procedure) for 18 gauges between 1928 and 1994.
- New data, from 2000 to 2015, available in digital format with a time resolution of 10' to 30'.

In all cases, annual records with more than 30% of missing values in a single year were discarded. However, incomplete years (with less than 30% missing values) were further investigated: if no significant precipitation was found during or before the gap periods, the river was considered in low-flow conditions during these gaps and the record was considered reliable for the flood hydrograph extraction.

Altogether, the dataset reaches a total of 87 gauging stations with at least 6 years of record over the period 1928 – 2015, resulting in a total of 763 station-year records, with average length of 15 years and maximum length of 64 years. The spatial distribution of the gauges is shown in Figure 3a, while data availability over time is summarized in Figure 3b. All the data used in terms of annual FRFs, as well as the main characteristics of the 87 basins, are available in a web-gis (www.resba.it).

As a preparatory step for the regional statistical analyses, we computed the empirical mean annual FRFs in all of the 87 stations and then we fitted the empirical curves with the NERC model of eq. (4). The best-fitting curve was obtained by numerical least square minimization, using the MATLAB® function "fit" (with the default "trust-region" algorithm; Moré & Sorensen, 1983). Parameters *b* and *c* are then jointly estimated and constrained to be non-negative. The final fitting of the NERC curve to the empirical average FRF resulted adequate for all the basins, with a mean coefficient of determination R^2_{adj} equal to 0.995.

For all 87 watersheds, almost 100 basin attributes were available, as already published in the work by Gallo et al. (2013). The set of geomorphological descriptors was obtained by processing the NASA SRTM (Shuttle Radar Topography Mission) Digital Elevation Model (Farr et al., 2007), sampled at a 100 meters spatial resolution. A subset of descriptors to be used in the regional analyses has been selected, as described in Table A.1 in Appendix. The procedure for selecting subsets of descriptors is widely detailed in Cordero (2019).

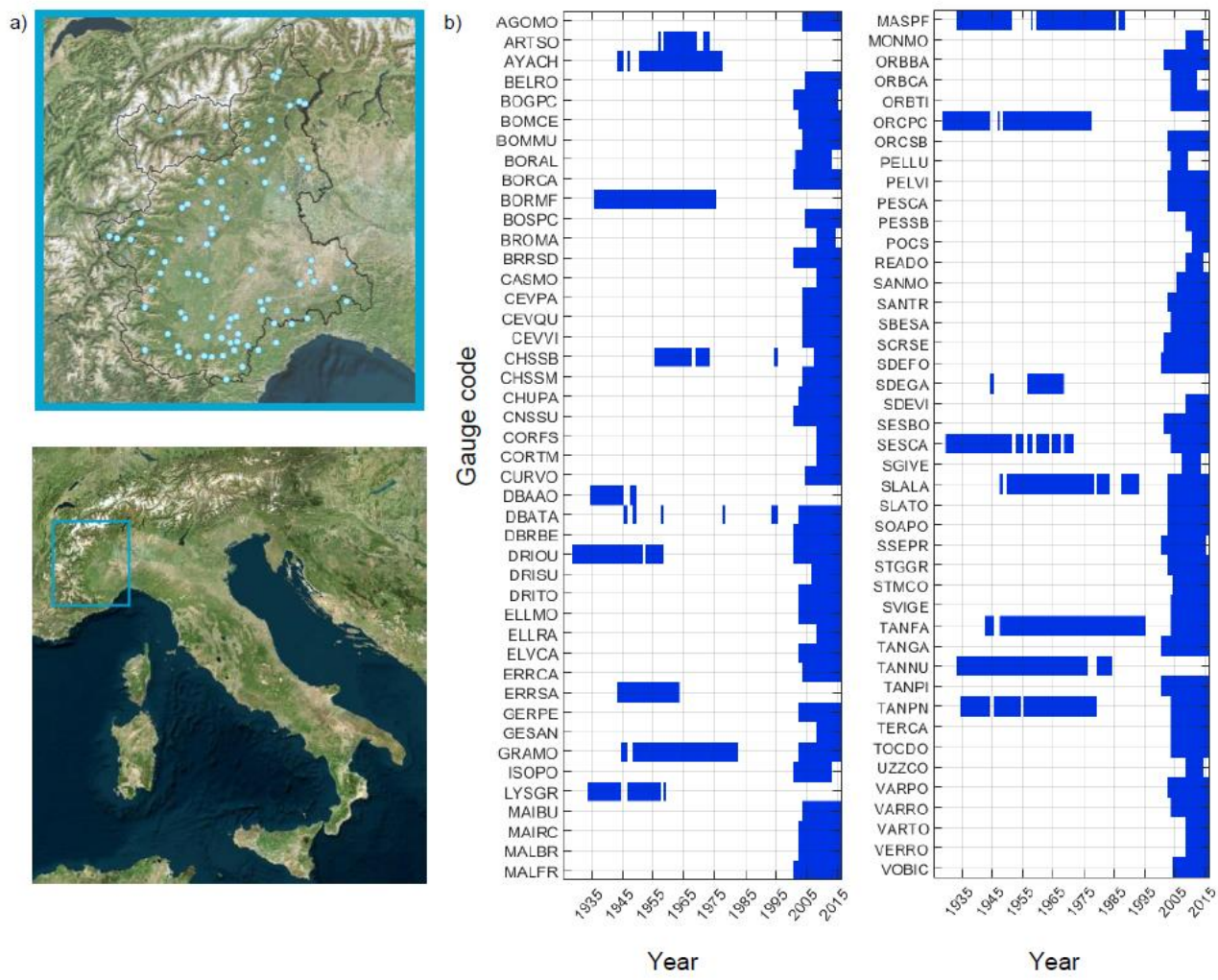

Figure 3: a) Geographical location of the 87 gauging stations in the database. b) Years with available data (after quality checks) for each gauge. Gauge codes allow to access the watershed descriptors in Gallo et al., 2013 and in the above mentioned web-gis.

3.2 Regional model calibration 1: Multiple linear regression

In pursuing the goal to reconstruct the NERC parameters 1/*b* and *c* for ungauged basins, both were considered as prediction variables in multiple linear regressions. Observations of *b* and *c* were the values fitted to the average FRF curve for each station in the data preparation step discussed above. All the possible combinations of 2 and 3 basin attributes (Table A.1 in Appendix) were used as set of covariates in the multiple regressions, including some variants where 1/*b*, *c* and the covariates were transformed. The best-performing models obtained are reported in Table **1**, together with some goodness of fit indicators, i.e. the adjusted coefficient of determination, R^2_{adj}, and the relative root mean squared error, $RRMSE = \frac{\sqrt{\sum(\hat{y}_i - y_i)^2 n^{-1}}}{\bar{y}}$. For operational purposes, when different models reached similar performances the preferred one has been that based on "simpler" descriptors (i.e., easier to compute). This is the case of models 1 and 2 in

Table **1**. Despite the high performance of the models subjected to Box-Cox transformation in limiting the skewness of residuals (see Cordero, 2019), considering the overall performances we suggest concluding with the choice of the models ID 1 and ID 5 of
Table **1**, where the log-transformation of the *x* variable is applied.

For both *c* and *1/b*, the descriptors emerging in the best regressions are substantially the same. It is also clear that, despite the high correlation between b^{-1} and *c* (see ID=9 in Table 1), the most robust way to estimate *b* is not as a function of *c*, as shown in the results of ID = 10.

Table 1: best regionalization linear models for 1/b e c. From left to right: model identification, transformation applied to independent variables, number of independent variables, dependent variables (y), independent variables (x), coefficients (β), R^2_{adj} and RRMSE. The last 2 models refer to 1/b estimated directly from c. For the meaning of symbols, see the Appendix.

ID	Transformation	N. descriptors	y	x	β	R^2_{adj}	RRMSE
1	Natural logarithm	3	c	ln (H_{avg}) ln (LDP) ln (IDFn)	4.8403 -0.58869 0.13813 0.80202	0.5879	0.40
2	Box-Cox	3	$c^{0.1635}$	$H_{avg}^{0.6938}$ $LDP^{-0.2929}$ $IDFn^{1.9984}$	1.1436 -0.0019 -0.3648 0.6423	0.5487	0.41
3	Natural logarithm	2	c	ln (H_{avg}) ln (LDP)	2.9990 -0.4190 0.1421	0.5626	0.42
4	Box-Cox	2	$c^{0.1635}$	$H_{avg}^{0.6938}$ $LDP^{-0.2929}$	1.1943 -0.0013 -0.3609	0.5417	0.41
5	Natural logarithm	3	ln(1/b)	ln (ku_{fa}) ln (LDP) ln (LDPs)	3.7285 -2.0775 0.5006 -0.7715	0.4287	0.74
6	Box-Cox	3	$(1/b)^{0.0851}$	$H_{avg}^{0.6938}$ $ku_{fa}^{0.5423}$ $LDP^{-0.2929}$	1.9611 -0.0008 -0.2048 -0.8146	0.3797	0.76
7	Natural logarithm	2	ln(1/b)	ln (H_{avg}) ln (LDP)	4.7943 -0.6859 0.6986	0.3491	0.79

ID	Transformation	N. descriptors	y	x	β	R^2_{adj}	RRMSE
8	Box-Cox	2	$(1/b)^{0.0851}$	$H_{avg}^{0.6938}$ $LDP^{-0.2929}$	1.6236 -0.0008 -0.7506	0.3311	0.80
9	b estimated from observed c		ln(1/b)	ln (c)	3.4589 1.2908	0.6190	0.61
10	b estimated from c estimated by model n.1		ln(1/b)	ln ($\hat{c}$)	3.4589 1.2908	0.2856	0.84

The best models found (ID=1 for c and ID=5 for b^{-1}) have the expressions:

$$ln\left(\frac{1}{b}\right) = 3.7285 - 2.0775 \cdot ln(ku_{fa}) + 0.5006 \cdot ln(LDP) - 0.7715 \cdot ln(LDP_s) \quad (13)$$

$$c = 4.8403 - 0.58869 \cdot ln(H_{avg}) + 0.13813 \cdot ln(LDP) + 0.80202 \cdot ln(IDFn) \quad (14)$$

Again, the descriptors definition is reported in Table A.1 in Appendix. A graphical representation of the performances of eq. (13) and (14) is shown in Figure 4, that reports a comparison between the observed and predicted FRF parameters.

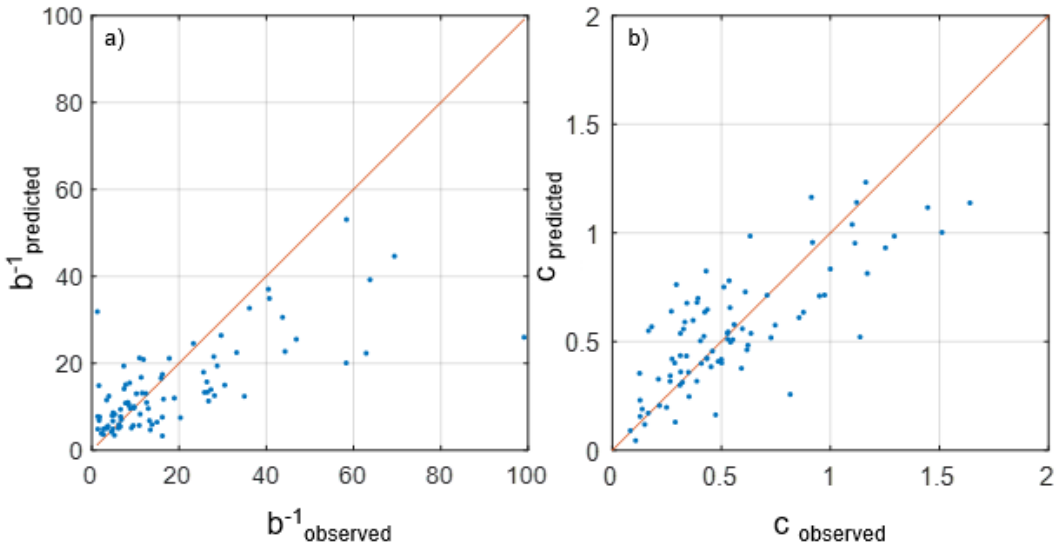

Figure 4: Estimated versus empirical values of parameter 1/b (a) and c (b) based on the linear regionalization model of eq. (13)-(14).

3.3 Regional model calibration 2: CCA

The CCA method is applied considering again all the possible combinations of 2 and 3 descriptors. This time they have been reduced to a benchmark set that includes the 10 most robust and easy to compute descriptors selected by an iterative pruning procedure. This procedure deletes, at each iteration, the descriptors most correlated to each other. The set of 10 descriptors selected is reported in Table 2.

Figure 5 shows, for both parameters b^{-1} and c, the values obtained from the local estimates versus those obtained from the regional CCA estimate built using a 10-descriptors model. All the model coefficients (a_j, b_j and the mean values $\bar{x}_j$ and $\bar{y}_j$) are reported in Table 2. The fitting performances are: R^2_{adj} = 0.4432 and RRMSE = 0.7314 for b^{-1}, and R^2_{adj} =0.58 and RRMSE = 0.3993 for c.

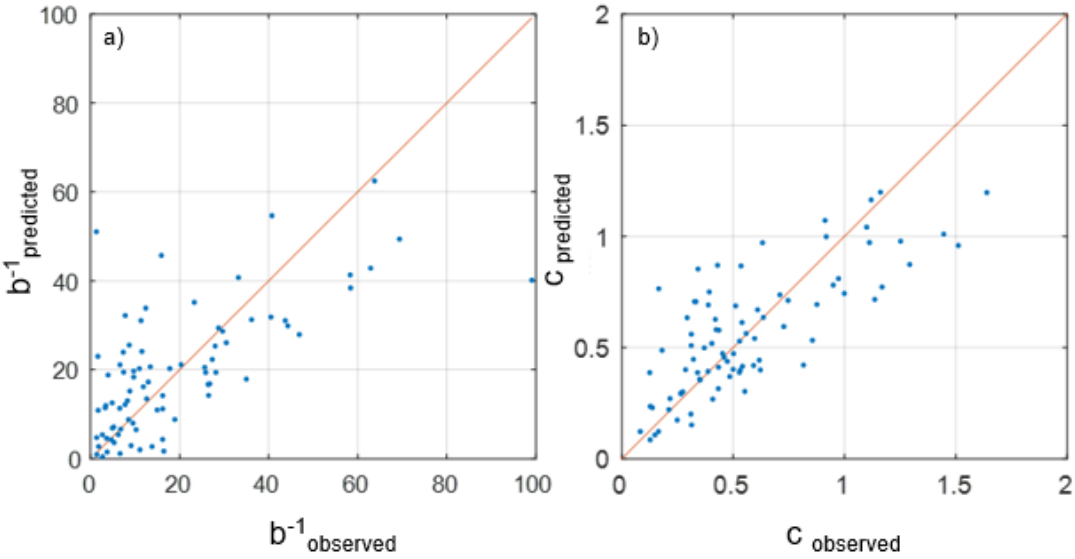

Figure 5: values obtained from local estimates of 1/b (a) and c (b) versus those obtained from the CCA regional model based on 10 descriptors (see Table 2).

Table 2: coefficients of the best CCA model based on 10 descriptors. Descriptors meaning is reported in Appendix.

Descriptors	$\bar{x}_j$	a_1	a_2
A	430.68	-5.1894·10-4	-5.8926·10-5
H_{avg}	1290.50	-1.4485·10-3	2.3539·10-4
X_b	403985.71	6.2250·10-6	2.9468·10-6
Y_b	4976076.19	3.4813·10-6	1.7074·10-6

D_d	0.64	1.3540	2.9862
LDP	44.04	8.0066·10-3	2.2960·10-2
MAP	1239.83	-2.7431·10-3	-5.9138·10-4
IDFa	24.03	0.1003	-4.2900·10-3
IDFn	0.46	11.9365	0.4314
c_f	0.44	-0.7036	-3.4828
	$\bar{y}_j$	**b_1**	**b_2**
1/b	18.5718	-0,0264	0,0865
c	0.5628	3,8355	-2,7990
ϱ		0.7781	

Among the top 10 combinations that use only 2 or 3 descriptors (see Table), similarly to the number of independent variables used in the previous method, the most significant model from a hydrological and practical point of view is ranked eighth, based on ϱ. For this model, the coefficients a_j, b_j, the mean values $\bar{x}_j$, $\bar{y}_j$ and the canonical correlation ϱ are summarized in Table 4. However, R^2_{adj} and RRMSE are respectively -0.1 and 1.03 for b^{-1} and 0.39 and 0.48 for c and the model is therefore not explanatory.

Table 3: top 10 combinations (ranked by ϱ values) that use only 2 or 3 descriptors.

Ranking	**Descriptors**	***ϱ***
1	H_{avg}, ku_{fa}, $Fourier_{B2}$	0.74699
2	H_{avg}, Lca_{12h}, cv[rr]	0.74650
3	H_{avg}, ku_{fa}, cv[MAP]	0.74586
4	H_{avg}, Lca_{3h}, cv[rr]	0.74253
5	H_{avg}, ku_{fa}, cv[rr]	0.74153
6	H_{avg}, ku_{fa}, IDFn	0.74109
7	H_{avg}, R_s, cv[rr]	0.74026
8	***H_{avg}, R_s, IDFn***	***0.73923***
9	H_{avg}, ku_{fa}, Lca_{12h}	0.73903
10	H_{avg}, ku_{fa}, Lca_{3h}	0.73852

Figure 6 shows the fitting performances of model n. 8 in Table 3.

Table 4: coefficients of the best CCA model based on 3 descriptors. Descriptors meaning is reported in Appendix.

Descriptors	$\bar{x}_j$	a_1	a_2
H_{avg}	1290.50	-0.0023	0.0018
R_s	2.1834	-0.3847	-0.2578
IDFn	0.46	7.5196	-26.0262
	$\bar{y}_j$	b_1	b_2
1/b	18.5718	-0.0379	0.0821
c	0.5628	4.1778	-2.2565
ϱ		0.7392	

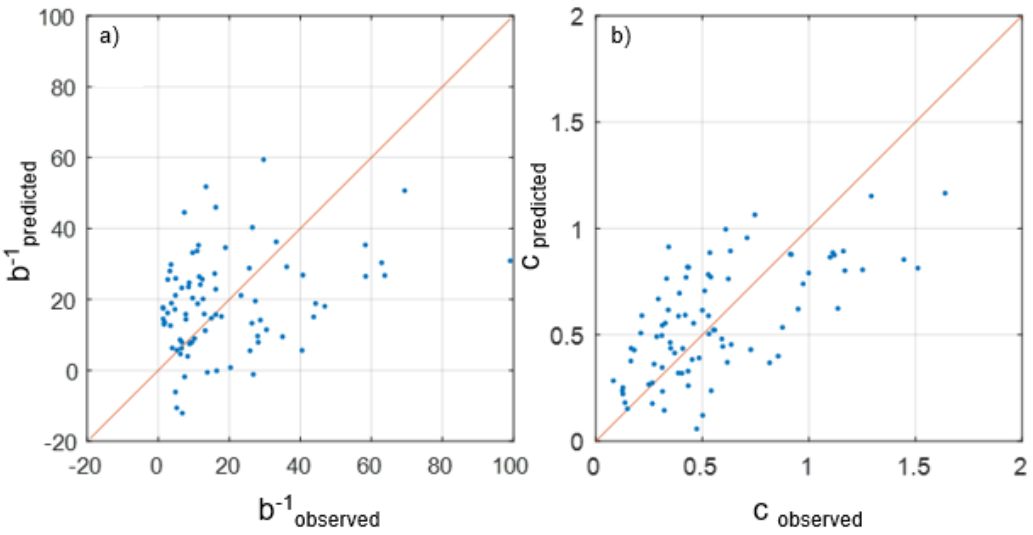

Figure 6: values obtained from local estimates of 1/b (a) and c (b) versus those obtained from the CCA regional model based on 3 descriptors (Table 4).

3.4 Regional model calibration 3: ACE

For the ACE algorithm application, as in the case of multiple regressions, all the possible combinations of 2 and 3 descriptors have been considered. Preliminary data transformations are not applied in this case, since the ACE algorithm already searches for an optimal transformation. The best models found for c and b^{-1}, ranked by $R^2_{adj,}$ are listed in Table 5. The mapping functions of the highest performing models are shown in Figure 7. Estimated values of b^{-1} and c are finally compared to the observed ones in Figure 8.

Results are quite interesting and are fully commented on in the Discussion Section.

Table 5: best ACE models among all possible combination of 2 and 3 descriptors ranked by R^2_{adj}.

ID	n. descriptors	y	x	mapping functions	R^2_{adj}	RRMSE
1	3	c	H_{avg}, LDP,IDFn	*	0.6483	0.36
2	3	b^{-1}	A,H_{avg}, F_f	*	0.4881	0.70
3	2	c	H_{avg}, LDP	Figure 7a	0.6115	0.39
4	2	b^{-1}	H_{avg}, LDP	Figure 7b	0.4306	0.75

*for the sake of brevity, the mapping functions are not reported. They are available in Cordero (2019).

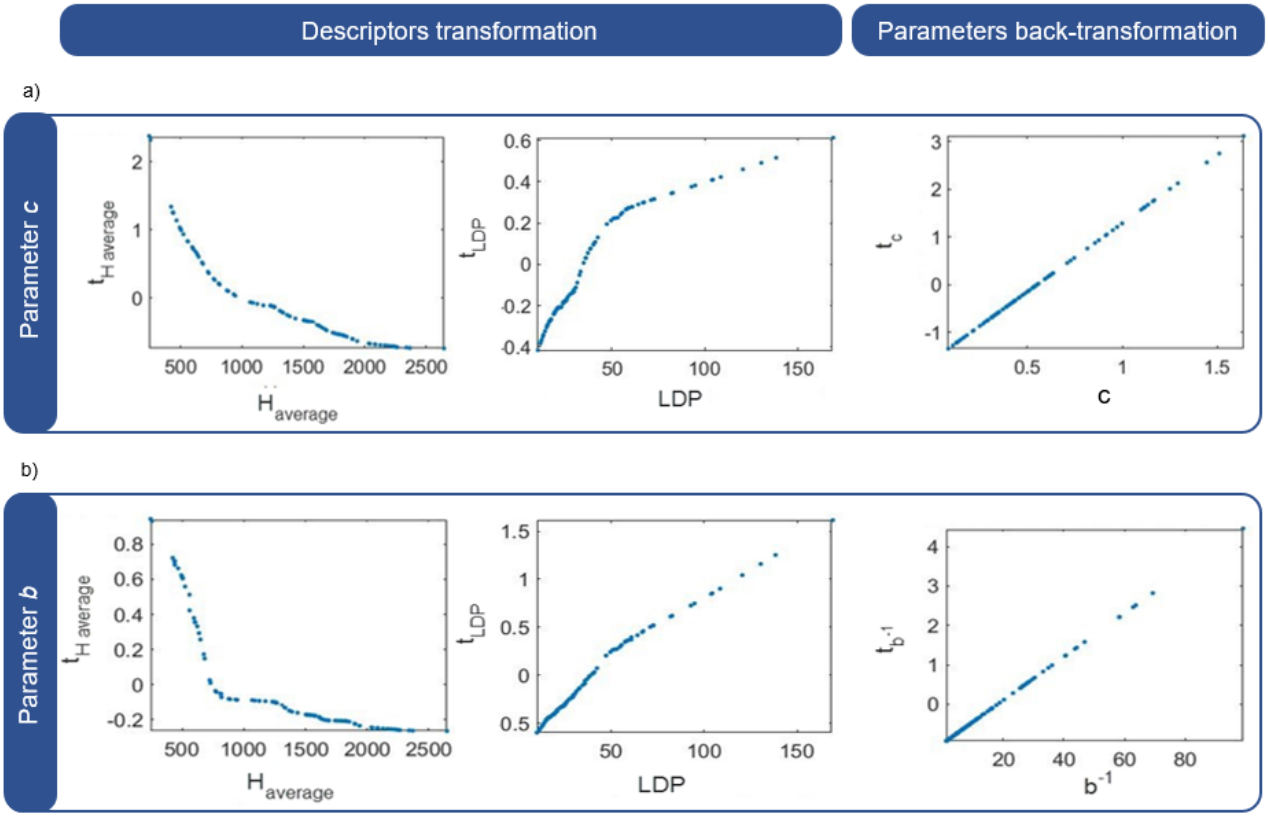

Figure 7: a) Best ACE model to estimate FRF c parameter among all possible combination of 2 descriptors (ID 3 of Table 5). b) Best ACE model to estimate FRF b^{-1} parameter among all possible combination of 2 descriptors (ID 4 of Table 5).

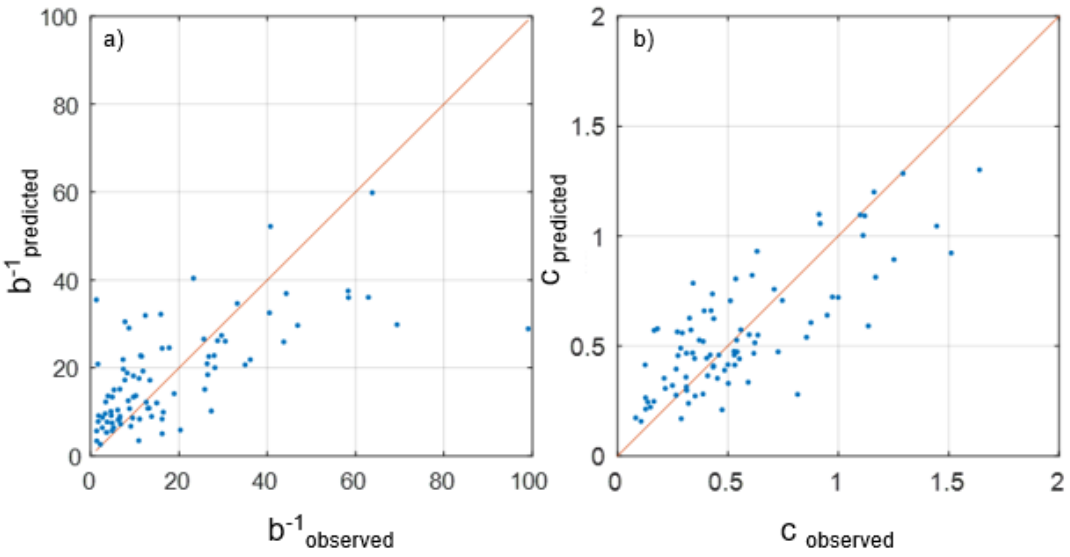

Figure 8: Estimated versus empirical values of parameter 1/b (a) and c (b). On panel a) ACE regional model for FRF parameter 1/b (ID 4 in Table 5); on panel b) ACE regional model for FRF parameter c (ID 3 in Table 5). See Table 2 for R^2_{adj} and RRMSE.

5 Discussion

Before undertaking comparative analyses among the result obtained with the three methods applied, we point out that for all models the leave-one-out cross-validation procedure has been applied. The main goal is to check the correct reproduction of the observed mean Flood Reduction Functions, but some considerations is also applied to the form of the hydrographs.

First of all, the cross-validation has been applied to check the model performances in the reproduction of the FRF parameter, i.e., *1/b* and *c*. Considering the multiple Linear Regression models of eqs. (13) – (14) and applying the leave-one-out cross-validation procedure, prediction performances worsen, as expected. For the parameter *1/b*, R^2_{adj} drops from 0.4287 to 0.3669 and the RRMSE rises from 0.74 to 0.78; for the parameter *c*, the R^2_{adj} changes from 0.5879 to 0.5374 and the RRMSE increases from 0.40 to 0.42. Overall, the performance degradation looks not very significant.

To better inform the comparisons, we have plotted the variations between the whole observed curves and the estimated ones. In Figure 9 each line represents the difference, $\Delta\varepsilon_D$, between the predicted (regional in panel a; cross-validated in panel b) and observed ε_D for all durations D. Each curve refers to a specific station. Figure 9 shows that the performances in cross-validation are basically indistinguishable from those obtained with the pure regional model, in which the data of the "prediction" station are also used to fit the model. Panels (c) and (d) of the same Figure 9 reports the relative errors obtained, that are bounded within +/-10% for most of the basins and also for the longer durations. A slight underestimation bias must also be acknowledged.

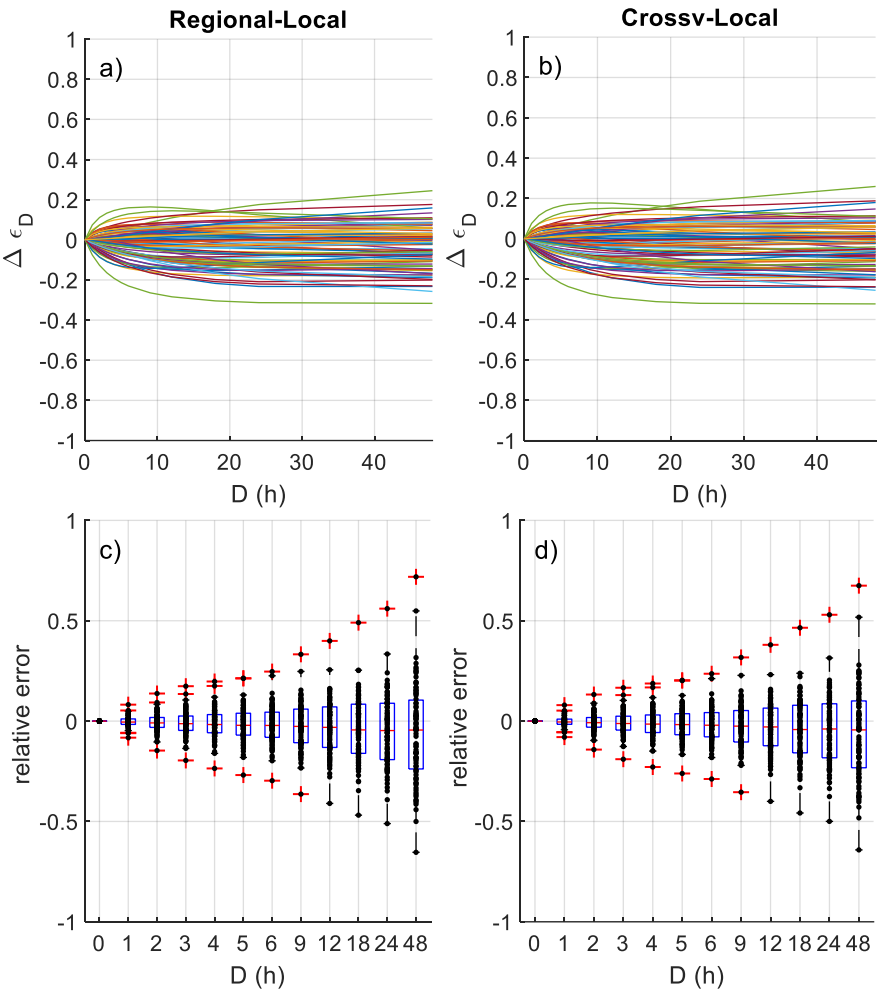

Figure 9: regional estimation expressed by eq. 13 and 14 for FRF parameter 1/b and c, respectively. a) difference between regionalised FRF and observed FRF. b) difference between regionalised FRF after cross-validation and observed FRF. c) and d) box plot of the relative error $(\boldsymbol{\varepsilon_{D,model}} - \boldsymbol{\varepsilon_{D,observed}})/\boldsymbol{\varepsilon_{D,observed}}$ respectively for fitted FRF and regionalised FRF after cross-validation.

The application of the CCA method produces average differences, in terms of FRF, that are even smaller, if we consider the 10-parameter model (results not shown). However, for some basins there are severe underestimates of the observed curve, exceeding 30% in a few cases, and which go over the 40% in cross-validation. For this reason, also considering the high number of parameters required, this method is deemed not so efficient in the domain of this regional analysis.

The application of the non-linear ACE models produces interesting results. In cross-validation, results confirm that the models with two descriptors are more robust than those with three independent variables. The passage from 3 to 2 descriptors has also a positive effect on the R^2_{adj} values. After cross-validation, the best model for parameter *c* is the 2-descriptors one with R^2_{adj} = 0.5377 (for the

3 descriptors the R^2_{adj} is 0.525). The RRMSE remains almost constant between the 3-variable and the 2-variable models (from 0.43 to 0.42 for c and from 0.82 to 0.83 for b^{-1}).

The results of the FRF regional estimation for ACE models is shown in Figure 10, where the panels (a) and (b) refer to the $\Delta\varepsilon_D$ curves computed before and after the cross-validation, respectively. Despite the high RRMSE of the individual parameter estimates, the overall errors on the FRF curves estimation remain limited.

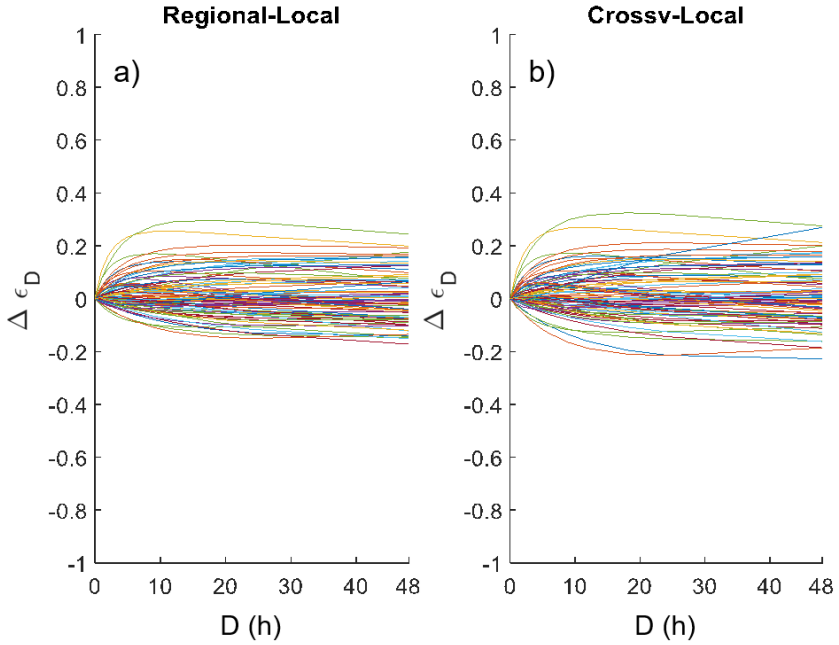

Figure 10: results of the ACE Regional model: ID=3 and ID =4 from Table 5. The graphs report the differences, over the duration D, between: (a) regionalised FRF and observed FRF, and (b) between regionalised FRF after cross-validation and observed FRF.

Summarizing the results obtained, the multiple linear regression globally leads to errors slightly larger than those obtained by ACE (compare fig. 9 and 10). However, the linear regression model is much easier to apply and less sensitive to extrapolation. To safely apply ACE models in extrapolation, the transformation curves should be first approximated with a polynomial of degree higher than 4, hence leading to a degradation of the model performances. The recommended models are, in conclusion, linear regressions with three descriptors.

5.1 Design hydrograph shape

As a final step of the analysis, we have evaluated the impact of the FRF estimation errors on the design hydrograph. We have compared the “regional” hydrograph (i.e., the one based on regionalized parameters) to a “reference” hydrograph, obtained from observations. As there is no established procedure to define what a “reference hydrograph shape” is, we have overlapped a sequence of

suitable observed high-flow hydrographs, normalized by their peak value and centred around their time to peak. On this sample of standardized hydrograph shapes we have computed an average shape. The high-flow hydrographs were selected from the full-length discharge time series using a threshold so that the local peak is equal or greater than the 50% of the mean annual maximum of flow.

Using this approach, we have estimated the (constant) skew parameter r, by numerically minimising the deviations between the FRF-based hydrograph and the empirical average shape.

After application of the regional analysis, two examples of comparisons are computed, as reported in Figure 11. The regional hydrograph (red line) is plotted against the reference hydrograph (bold black line); the thin black lines represent the real recorded hydrographs (after standardization). In both cases the fitting is reasonably good, although in panel a) one can notice that the regional hydrograph does not fit the quasi-convex shapes of the rising and falling limbs. However, if we consider the local NERC hydrograph (i.e., based on local parameters; blue line), we can notice that the FRF model itself is not fully adequate to represent the average hydrograph shapes. Sticking on the NERC model and fitting it to the observations, one can recognize that the regional hydrograph has very good performances in reproducing the 'local' NERC FRF function. Overall, almost all the investigated basins show a good fitting, and the maximum relative error of the regional estimates, computed in terms of differences between the area under the reference hydrograph and the one under the regional curve, does not exceed 30%.

As regard the skew parameter *r*, we have computed it for all stations and we have observed that it assumes rather constant values over the case study area. Slightly larger values of *r* just occurred for basins with higher average elevation. In a concrete application we then suggest that the value of r can be taken from a neighbour gauged basin, at least until a specific regional procedure is built for this parameter, which can be matter for future investigations.

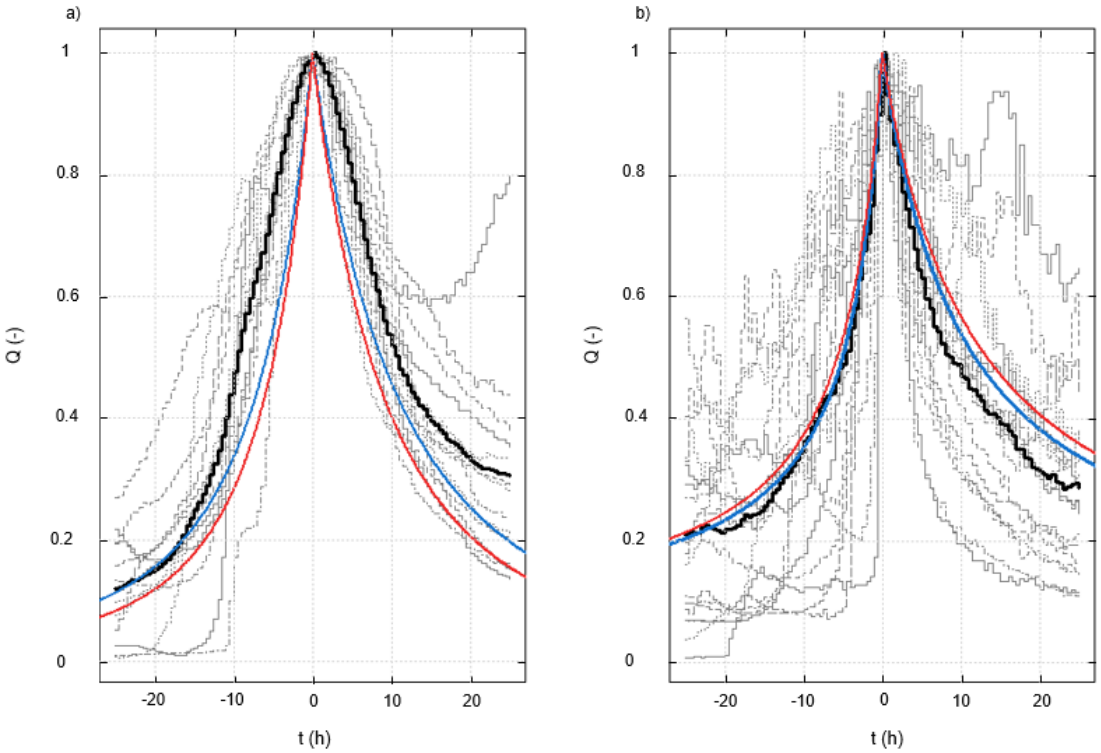

Figure 11: Comparison between NERC synthetic flood hydrographs built using the proposed regional model (red curve) and the analytical FRF (blue curve), compared to the empirical average hydrograph (black curve). Bormida a Cassine watershed (code: BORCA, area 1516.25 km^2, mean elevation 493 m asl, r=0.68) on panel a), Stura di Lanzo at Torino watershed (code: SLATO, area 879.97 km^2, mean elevation 1368 m asl, r=0.60) on panel b).

6 Conclusions

Flood hazard management and particularly the design of mitigation infrastructures requires to account for the flood volume, in addition to the flood peak design value. However, statistical methods to estimate the flood volume or the shape of the flood hydrograph are still not consolidated, due to the conceptual difficulty in representing the hydrograph shape in a simple way and to the scarcity of data. These difficulties are exacerbated in ungauged basins. This paper addresses this problem adopting the Flood Reduction Function (FRF) as a powerful and parsimonious representation of the link existing between hydrograph volume and duration. The FRF can be used to "summarize" the hydrographs characteristics in a few parameters, easy to be estimated also in ungauged basins, that are then used to build synthetic design hydrographs with minimal assumptions of their shape. The FRF approach can also be used in gauged basins to "regularise" a sequence of observed hydrographs and to allow one to compute, in a systematic and reproducible way, a single representative mean hydrograph shape.

The work presented here shows that a simple parametrization of the FRF function, known as NERC function, can be regionalized, using a set of basin attributes derived from terrain analysis, land use features and climatic indexes. Different regionalization methods (multiple linear regression, canonical correlation analysis, alternating conditional expectation algorithms) were tested here, with the result that a rather simple multiple linear regression model can provide satisfactory estimation performances for the set of basins considered. The use of the NERC regionalized parameters has also allowed us to assess the model capability to build synthetic hydrographs for each of the investigated basins, that have been compared to the average empirical hydrograph observed in the same gauging station.

In conclusion, with the reasonably good results obtained we have shown that the estimation of flood hydrograph in ungauged basins can be performed through regionalization techniques like those used for the frequency analysis of flood peaks, and with minimal additional assumptions. However, as the records of flood hydrographs are much shorter than the corresponding record of flow maxima, an effort to both collect new data and made available existing records is required to properly support all the practical applications that involve the management of flood volumes.

Acknowledgements

Authors acknowledge support from EU Territorial co-operation Program INTERREG (Alcotra), project RESBA-1729. ARPA Piemonte kindly provided the data and the original stream-gauge records.

Data availability statement

Data in support of this manuscript are available in a web-gis (www.resba.it).

Appendix

Table A.1: List of the geomorphological, climatological and soil descriptors used in the regional analyses.

Attribute category	Attribute	Notation	Units	Description
Geomorphological	East coordinate of the basin's centroid (WGS84 UTM32N)	X_b	m	Reference System: WGS84 (EPSG: 4326).
	North coordinate of the basin's centroid (WGS84 UTM32N)	Y_b	m	Reference System: WGS84 (EPSG: 4326).
	Basin area	A	Km^2	The area required for channel initiation has been set to 1 Km^2.
	Basin mean elevation	H_{avg}	m a.s.l.	-
	Length of longest drainage path	LDP	Km	Path included between the outlet and the furthest point from it, placed on the edge of the basin watershed and identified by following the drainage directions.
	Mean slope of longest drainage path	LDPs	-	Ratio between the difference between basin maximum and minimum elevation of the basin to the LDP.
	Drainage density	D_d	Km^{-1}	Ratio between the total length of the river network to the basin area.
	Shape factor	F_f	-	Ratio of the basin area to the square of the length of the main channel.
	Width function kurtosis	Ku_{fa}	-	The width function is defined by counting the number of pixels having equal distance from the gauging station. This distance is measured following the drainage path. The 4^{th} statistical moment of the width function has been computed.
	Slope ratio	R_s	-	Ratio of average slope of streams of two adjacent orders *u* and *u+1*. Streams are numbered according to the Horton's criterion.
Climatological	Mean a parameter of the IDF curve	IDF_a	$mm\ h^{-1}$	Scale factor of the IDF curve. The average value over the basin area has been calculated.
	Mean n parameter of the IDF curve	IDF_n	-	Scaling exponent of the IDF curve. The average value over the basin area has been calculated.
	Mean coefficient of L-skewness for 3-hours duration	Lca3h	-	Coefficient of L-skewness for 3-hours duration. The average value over the basin area has been calculated.
	Mean coefficient of L-skewness for 12-hours duration	Lca12h	-	Coefficient of L-skewness for 12-hours duration. The average value over the basin area has been calculated.
	Mean annual precipitation over the basin	MAP	mm	Total mean annual precipitation (Bartolini et al., 2011).
	Spatial coefficient of variation of the mean annual precipitation	cv [MAP]	mm	-
	Coefficient of variation of the rainfall regime over the basin	cv [rr]	-	Calculated from the 12 mean monthly rainfall depths, computed using monthly data from (Bartolini et al., 2011).
	Mean Fourier coefficient B2 of the rainfall regime	$Fourier_{B2}$	-	Mean values of the B2 coefficient of the Fourier series representation of the rainfall regime. The reader can refer to Gallo et al. (2013) for details about the meaning of the B2 coefficient.
Soil	Mean permeability index	c_f	%	Permeability index used in the Vapi project (Villani, 2003). This coefficient has been obtained by classification of permeability value in flood conditions by balancing the rational formula. The average value over the basin area has been calculated.

References

Bacchi, B., Brath, A., & Kottegoda, N.T. (1992), Analysis of the Relationships Between Flood Peaks and Flood Volumes Based on Crossing Properties of River Flow Processes. *Water Resources Research*, *28*(10), 2773–2782.

Bacova Mitková, V., & Halmová, D. (2014), Joint modeling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava. *Journal of Hydrology and Hydromechanics 62*(3), 186-196. https://doi.org/10.2478/johh-2014-0026

Bartolini, E., Allamano, P., Laio, F., & Claps, P. (2011), Analisi spaziale delle precipitazioni medie ed intense su Piemonte e Valle d'Aosta. Working Paper 02, Politecnico di Torino.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., & Savenije, H., (Eds.). (2013), Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales. Cambridge University Press, New York, https://doi.org/10.1017/CBO9781139235761

Box, G.E.P., & Cox, D.R. (1964), An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological) 26*(2), 211–252.

Breiman, L., & Friedman, J.H. (1985a), Estimating Optimal Transformations for Multiple Regression and Correlation. *Journal of the American Statistical Association 80*(391), 580–598. https://doi.org/10.2307/2288473

Breiman, L., & Friedman, J.H. (1985b), Estimating Optimal Transformations for Multiple Regression and Correlation - rejoinder. *Journal of the American Statistical Association 80*(391), 614–619. https://doi.org/10.2307/2288477

Brunner, M.I., Furrer, R., Sikorska, A.E., Viviroli, D., Seibert, J., & Favre, A.C. (2018), Synthetic design hydrographs for ungauged catchments: a comparison of regionalization methods, *Stochastic Environmental Research and Risk Assessment 32*(7), 1993–2023. https://doi.org/10.1007/s00477-018-1523-3

Brunner, M.I., Viviroli, D., Sikorska, A.E., Vannier, O., Favre, A.-C., & Seibert, J. (2017), Flood type specific construction of synthetic design hydrographs. *Water Resources Research 53*(2), 1390–1406. https://doi.org/10.1002/2016WR019535

Castellarin, A., Kohnova, S., Gaal, L., Fleig, A., Salinas, J.L., Toumazis, A., Kjeldsen, T.R., & Macdonald, N. (2012), Review of applied-statistical methods for flood-frequency analysis in Europe, NERC/Centre for Ecology & Hydrology (ESSEM COST Action ES0901).

Chow, V.T. (1951), A general formula for hydrologic frequency analysis. *Eos, Transactions American Geophysical Union 32*(2), 231–237. https://doi.org/10.1029/TR032i002p00231

Claps, P., Ganora, D., Laio, F., & Radice, R. (2010), Riesame ed integrazione di serie di portate al colmo mediante scale di deflusso di piena, paper presented at 32nd Convegno Nazionale Di

Idraulica e Costruzioni Idrauliche, Palermo, Italy.

Cordero, S. (2019), Metodologie statistiche e sperimentali per il supporto ai piani di emergenza in presenza di invasi artificiali, (Doctoral dissertation). Retrieved from Iris (https://iris.polito.it/handle/11583/2744152#.X8CzZC9aZZg). Politecnico di Torino.

Cunnane, C. (1988), Methods And Merits Of Regional Flood Frequency-Analysis. *Journal of Hydrology 100*(1-3), 269–290.

Dalrymple, T. (1960), Flood-Frequency Analyses. Manual of Hydrology: Part 3. Flood-Flow Techniques. Usgpo 1543-A: 80. http://pubs.usgs.gov/wsp/1543a/report.pdf.

Farr, T. G., Rosen, P. A., Caro, E., (Eds.). (2007), The Shuttle Radar Topography Mission. *Reviews of Geophysics 45*(2), RG2004, doi:10.1029/2005RG00018

Fiorentino, M., Rossi, F., & Villani, P. (1987), Effect of the basin geomorphoclimatic characteristics on the mean annual flood reduction curve, in: Proceedings of the 18th Annual Pittsburgh IASTED International Conference. p. Vol.18, part 5, pp.1777–1784.

Franchini, M., & Galeati, G. (2000), Comparative analysis of some methods for deriving the expected flood reduction curve in the frequency domain. *Hydrology and Earth System Sciences 4*(1), 155–172. https://doi.org/10.5194/hess-4-155-2000

Gallo, E., Ganora, D., Laio, F., Masoero, A., & Claps, P. (2013), Atlas of the Piedmont watersheds (in italian). Renerfor-Alcotra project, Piedmont Region, 978-88-96046-06-7 (http://www.idrologia.polito.it/web2/open-data/Renerfor/atlante_bacini_piemontesi_LR.pdf).

Grimaldi, S., Kao, S.C., Castellarin, A., Papalexiou, S.M., Viglione, A., Laio, F., Aksoy, H., & Gedikli, A. (2011), Statistical Hydrology, in: *Peter Wilderer (ed.) (Ed.), Treatise on Water Science*. Oxford: Academic Press, pp. 479–517. https://doi.org/10.1016/B978-0-444-53199-5.00046-4

Gumbel, E.J. (1945), Simplified plotting of statistical observations. *Eos, Transactions American Geophysical Union 26*(1), 69–82. https://doi.org/10.1029/TR026i001p00069

Guo, Y., & Adams, B.J. (1998), Hydrologic analysis of urban catchments with event-based probabilistic models. *Water Resources Research 34*(12), 3421–3431.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), The Elements of Statistical Learning: Data Mining, Inference and Prediction. 2nd edn Springer, New York.

Koutsoyiannis, D., Kozonis, D., & Manetas, A. (1998), A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology 206*(1-2), 118–135.

Libertino, A., Allamano, P., Laio, F., & Claps, P. (2018), Regional-scale analysis of extreme precipitation from short and fragmented records. *Advances in Water Resources 112*, 147–159. https://doi.org/https://doi.org/10.1016/j.advwatres.2017.12.015

Maione, U., Mignosa, P., & Tomirotti, M. (2003), Regional estimation of synthetic design hydrographs. *International Journal of River Basin Management 1*(2), 151–163. https://doi.org/10.1080/15715124.2003.9635202

Mediero, L., Jiménez-Álvarez, A., & Garrote, L.(2010), Design flood hydrographs from the relationship between flood peak and volume. *Hydrology and Earth System Sciences 14*(12), 2495–2505. https://doi.org/10.5194/hess-14-2495-2010

Montgomery, D., Peck, E., & Vining, G. (2001), Introduction to linear regression analysis, third ed. ed. *Wiley Series Probability and Statistics*,Wiley, New York.

Moré, J.J., & Sorensen, D.C. (1983), Computing a Trust Region Step. *Journal on Scientific and Statistical Computing 4*(3), 553–572. https://doi.org/https://doi.org/10.1137/0904038

Mulvaney, T.J. (1851), On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and flood discharges in a given catchment. Proceedings of the Institution of Civil Engineers of Ireland, 4, 19–31.

Natural Environmental Research Council (1975), Estimation of flood volumes over different duration. *Flood Studies Report* 1, 352–373.

Ouarda, T., Girard, C., Cavadias, G.S., & Bobee, B. (2001), Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology 254*(1-4), 157–173. https://doi.org/10.1016/S0022-1694(01)00488-7

Ouarda, T.B.M.J., Hache, M., Bruneau, P., & Bobe, B. (2000), Regional flood peak and Volume estimation in northern canadian basin. *Journal of Cold Regions Engineering 14*(4), 176–191.

Requena, A.I., Flores, I., Mediero, L., & Garrote, L. (2016), Extension of observed flood series by combining a distributed hydro-meteorological model and a copula-based model. *Stochastic Environmental Research and Risk Assessment* 30, 1363–1378. https://doi.org/10.1007/s00477-015-1138-x

Salvadori, G., & De Michele, C. (2007), On the Use of Copulas in Hydrology: Theory and Practice. *Journal of Hydrologic Engineering 12*(4), 369–380. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(369)

Serinaldi, F., & Kilsby, C.G. (2013), The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs. *Water Resources Research 49*(6), 3423–3442. https://doi.org/10.1002/wrcr.20221

Silvagni, G. (1984), Valutazione dei massimi deflussi di piena. Pubb. n. 489, Pubb. Ist. Idraulica, Univ. di Napoli.

Snyder, W.M. (1962), Some possibilities for multivariate analysis in hydrologic studies. *Journal of Geophysical Research 67*(2), 721–729. https://doi.org/10.1029/JZ067i002p00721

Spector, P., Friedman, J., Tibshirani, R., Lumley, T., Garbett, S., & Baron, J. (2016), ACE and AVAS for Selecting Multiple Regression Transformations. [Software]. CRAN. https://cran.r-project.org/package=acepack.

Tomirotti, M., & Mignosa, P. (2017), A methodology to derive Synthetic Design Hydrographs for river flood management. *Journal of Hydrology 555*, 736–743. https://doi.org/10.1016/j.jhydrol.2017.10.036

Villani, P. (2003). Rapporto sulla valutazione delle piene in Piemonte, 89–118, Del Paguro, Fisciano.

Volpi, E., & Fiori, A. (2012), Design event selection in bivariate hydrological frequency analysis. *Hydrological Sciences Journal 57*(8), 1506–1515. https://doi.org/10.1080/02626667.2012.726357

Wong, T.S. (1963), A multivariate statistical model for predicting mean annual flood in New England. *Annals of the Association of American Geographers 53*(3), 298–311. https://doi.org/10.1111/j.1467-8306.1963.tb00451.x

Xiao, Y., Guo, S., Liu, P., Yan, B., & Chen, L. (2009), Design flood hydrograph based on multicharacteristic synthesis index method. *Journal of Hydrologic Engineering 14*(12), 1359–1364. https://doi.org/10.1061/(ASCE)1084-0699(2009)14:12(1359)

Yue, S., Ouarda, T.B.M.., Bobée, B., Legendre, P., & Bruneau, P. (1999), The Gumbel mixed model for flood frequency analysis. *Journal of Hydrology 226*(1-2), 88–100. https://doi.org/https://doi.org/10.1016/S0022-1694(99)00168-7

Yue, S., Ouarda, T.B.M.J., Bobée, B., Legendre, P., & Bruneau, P. (2002), Approach for Describing Statistical Properties of Flood Hydrograph. *Journal of Hydrologic Engineering 7*(2), 147–153. https://doi.org/10.1061/(ASCE)1084-0699(2002)7:2(147)

Zhang, L., & Singh, V.P. (2006), Bivariate Flood Frequency Analysis Using the Copula Method. *Journal of Hydrologic Engineering 11*(2), 150–164. https://doi.org/10.1061/(ASCE)1084-0699(2006)11:2(150)