Spatially smooth regional estimation of the flood frequency curve (with uncertainty)

F. Laio[a], D. Ganora[a], P. Claps[a], G. Galeati[b]

[a] *Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, Politecnico di Torino, Torino, ITALY.*
[b] *ENEL Produzione s.p.a., ITALY.*

Abstract

Identification of the flood frequency curve in ungauged basins is usually performed by means of regional models based on the grouping of data recorded at various gauging stations. The present work aims at implementing a regional procedure that overcomes some of the limitations of the standard approaches and adds a clearer representation of the uncertainty components of the estimation.

The information in the sample records is summarized in a set of sample L-moments, that become the variables to be regionalized. To transfer the information to ungauged basins we adopt a regional model for each of the L-moments, based on a comprehensive multiple regression approach. The independent variables of the regression are selected among a large number of geomorpholoclimatic catchment descriptors. Each model is calibrated on the entire dataset of stations using non-standard least-squares techniques accounting for the sample variability of L-moments, without resorting to

Email addresses: francesco.laio@polito.it (F. Laio), daniele.ganora@polito.it (D. Ganora), pierluigi.claps@polito.it (P. Claps), giorgo.galeati@enel.com (G. Galeati)

any grouping procedure to create sub-regions. In this way, L-moments are allowed to vary smoothly from site to site, following the variation of the descriptors selected in the regression models. This approach overcomes the subjectivity affecting the techniques for the definition and verification of the homogeneous regions. In addition, the method provides accurate confidence bands for the frequency curves estimated in ungauged basins.

The procedure has been applied to a vast region in North-Western Italy (about 30,000 km^2). Cross-validation techniques are used to assess the efficiency of this approach in reconstructing the flood frequency curves, demonstrating the feasibility and the robustness of the approach.
Keywords: Regional flood frequency analysis, L-moments, Uncertainty, Ungauged basins, Short records

1. Introduction

The evaluation of the frequency of flood events in ungauged catchments is usually approached by building suitable statistical relationships (models) between flood statistics and basins characteristics, calibrated on a set of records of annual maxima. These models are used to transfer the information available at the gauged sites to the target basin, where only morphoclimatic catchment's characteristics are available. This type of procedure is called a *regional model*, because it identifies a subset of basins, called *region*, that is used as a pooling set where the information to be transferred to ungauged site resides. In standard regional models, the basins, which are assumed to belong to a homogeneous region, donate their (common) statistical properties of the flood frequency curve to the ungauged basins that are assumed to fall in the same region.

Various methods to achieve this goal have been proposed in the literature (see for example the review by Cunnane (1988) and Grimaldi et al. (2011)), differing to each other mainly on the basis of the distribution used to describe the at-site data (see e.g. Hosking and Wallis, 1997, for a bouquet of distributions), and on the pooling criterion used for the delineation of regions. Several techniques have been proposed for region delineation. Among others, we can mention: cluster analysis and proximity pooling (Burn, 1990), hierarchical approaches (Fiorentino et al., 1987; Gabriele and Arnell, 1991), neural network classifiers (Hall and Minns, 1999) and mixed approaches (Merz and Bloschl, 2005). For any of these techniques the check for statistical homogeneity within the regions is an important issue (Viglione et al., 2007; Castellarin et al., 2008).

However, most of the standard statistical tools for the estimation of the flood frequency curve in ungauged basins present limitations. In particular, (i) the subdivision of the domain of interest in homogeneous regions, and (ii) the choice of an a priori probability distribution to describe the sample data, can be considered as limiting factors, due to the difficulties of managing estimations where abrupt changes occur across regions, or distributions demonstrate not to keep their properties inside and across regions.

Regarding the point (i), different approaches exist to create homogeneous regions. For instance, regions can be created by splitting in separated areas the geographical space or a multi-dimensional space of the physiographic basin's characteristics (e.g. Ouarda et al., 2001, fig.1). The regions can be defined by means of fixed boundaries (e.g. cluster analysis procedures) or by means of a pooling technique that does not define fixed regions, as in the region-of-influence (ROI) approach (Burn, 1990). The ROI approach is more flexible than the fixed-regions approach because it creates site-dependent regions. However, the estimates are not smooth (both in geographic or physiographic spaces) due to possible discontinuities at the border between one ROI and another.

The main limitation of the approaches that use a subdivision in separate regions is the difficulty to assess a reliable and stable configuration of the regions (e.g. which catchments to include or not in a particular region). In fact, since there is no prior information about the regions configuration, any algorithm used for regions delineation induces some errors. Then, the regions must be tested for their statistical homogeneity, although the related tests can be rather weak in the estimation of statistical heterogeneity (Viglione et al., 2007). A few papers have tried to overcome this problem proposing methods based on the interpolation of the hydrological variable in the descriptors space (Chokmani and Ouarda, 2004; Chebana and Ouarda, 2008), or based on the so-called top-kriging (Skoien et al., 2006). The first technique presents problems in the definition of the descriptors used for the interpolation, while the top-kriging is heavily dependent on the availability of large datasets that would support a reliable construction of a "objective" variogram. The idea not to resort to a grouping procedure to form the regions has been also developed by Stedinger and Tasker (1985), and recently improved by Griffis and Stedinger (2007), where the advantages of using this approach are underlined. Using no regions there is no longer the need for an homogeneity test: the statistical characteristics of the floods can vary from site to site and the model will try to reproduce this variability.

All the above approaches require, at the initial stage, an hypothesis on the at-site frequency distribution chosen to describe the data CDF (cumulative distribution function) and to estimate flood quantiles. In fact, these methods basically perform more or less refined interpolation techniques on the flood quantiles estimated at site. This bring us back to point (ii) above, which is related to the choice of an a-priori CDF to represent the data. However, different probability distributions can fit equally well the data for low return periods, while they may produce diverging estimates when extrapolated to high return periods (an example will be given in the following figure 6). This effect becomes even more evident in the case of short records, which are particularly important in data-scarce regions.

In this paper, we followed the idea of transferring hydrological information assuming no regions nor pooling groups, and we use the L-moments and their dimensionless ratios as statistical variables to be transferred to the ungauged sites. In particular, we select the sample L-moment of order one (the mean), the coefficient of L-variation (L_{CV}) and the L-skewness (L_{CA}) of the record. Regionalizing these three L-moments allows one to reconstruct the whole flood frequency curve, at least if three-parameter curves are selected. The choice of the mean, L_{CV} and L_{CA} as hydrological signatures in a regional framework can be interpreted in an index-flood framework (Dalrymple, 1960) considering the mean as the scale factor and the L-moments ratios as the descriptors of the dimensionless growth curve. A similar approach has been applied by Vogel et al. (1999) to the annual streamflow, who regionalized the first two moments instead of the L-moments.

The use of the mean, L_{CV} and L_{CA} instead of a quantile or distribution-parameter is also helpful, for both calibration and prediction purposes, when catchments with short sample records are used in the analysis. In fact, during the model calibration phase, sample L-moments are computed even if their sample variability is high (but known or quantifiable), without resorting to often inefficient estimates of the at-site parent distribution. This avoids information loss due to the elimination of short records. On the other hand, if one is interested in the local quantile prediction at a gauged site with a short record, it is still possible to compute, for instance, the index-flood (Q_{ind}) and the L_{CV} directly on the sample record, leaving to the regional procedure the estimation of L_{CA}. From this point of view, this approach extends the original index-flood method, in which Q_{ind} is often estimated locally, based on even few at-site measurements, while the growth curve is derived by a regional model.

The relationships built to transfer the information to the ungauged sites are based on multiple regressions and are discussed in section 2.2. The choice of the probability distribution used for the final quantile estimation is based on a model averaging approach and is reported in section 3. The proposed methodology is applied to an area of about 30.000 km^2 located in North-Western Italy, including 70 gauging stations. The application is presented in section 4 and final remarks are reported in the conclusions section.

2. Model Definition

2.1. At-Site Estimates: Systematic and Non-Systematic Information

The first step of the procedure is to check the available data and use them to compute suitable statistical indicators at the gauged sites. Among the possible types of data which can be used in the statistical analysis (e.g. Stedinger et al., 1993), common procedures implicitly assume a record of n systematic measures. Sometimes, however, systematic records of data can be integrated with additional data, derived from measurements of significant occasional events. This can be particularly useful when the original systematic record is short. When a number of occasional additional measurements is available, one can merge them with the systematic ones to improve the robustness of the final estimates (e.g. Bayliss and Reed, 2001). This is done producing a new time series of "equivalent" length m, that is the number of years between the first and the last measurement of both the systematic and the occasional record, merged together.

To calculate the probability weighted moments (PWMs) of the extended record, we use a method suggested by Wang (1990): the merged sample is arranged in increasing order

$$x_{(1)} \leq x_{(2)} \leq \ldots x_{(m-l+1)} \leq x_{(m-l+2)} \leq \ldots \leq x_{(m)} \tag{1}$$

where the subscript in round brackets indicates the sorted position; the l largest events, exceeding a threshold x_0, are considered as a censored sample, whose elements can be either systematic or occasional data. When working with censored samples, the theoretical formula for the PWM of order r

of a random variable X with distribution function $F(x) = P(X \leq x)$, as $\beta_r = \int_0^1 x(F)F^r dF$, must be split in two components (Wang, 1990),

$$\beta_r = \int_0^{F_0} x(F)F^r dF + \int_{F_0}^1 x(F)F^r dF = \beta_r'' + \beta_r' \qquad (2)$$

where $F_0 = F(x_0)$ is the non-exceedance probability relative to the censoring threshold x_0. The unbiased estimator of β_r'' is then (Wang, 1990):

$$b_r'' = \frac{1}{n} \sum_{i=1}^{n} \frac{(i-1)(i-2)\dots(i-r)}{(n-1)(n-2)\dots(n-r)} x_{(i)}'' \qquad (3)$$

where $x_{(i)}''$ is deducted from the sorted sample as

$$x_{(i)}'' = \begin{cases} x_{(i)} & \text{if } x_{(i)} < x_0, \\ 0 & \text{if } x_{(i)} \geq x_0. \end{cases}$$

On the other hand, the estimator of β_r' is (Wang, 1990)

$$b_r' = \frac{1}{m} \sum_{i=1}^{m} \frac{(i-1)(i-2)\dots(i-r)}{(m-1)(m-2)\dots(m-r)} x_{(i)}' \qquad (4)$$

where $x_{(i)}'$ is the above-threshold sample, i.e.

$$x_{(i)}' = \begin{cases} 0 & \text{if } x_{(i)} < x_0, \\ x_{(i)} & \text{if } x_{(i)} \geq x_0. \end{cases}$$

Still following Wang (1990), the unbiased estimator of β_r is $b_r = b_r' + b_r''$.

The censoring threshold x_0 represents the level above which the non-systematic flood values are assumed as deserving to be recorded. x_0 can be assumed equal to the smallest non-systematic measure (Bayliss and Reed, 2001). In the absence of non-systematic information, the above formulas reduce to the usual definitions of PWMs.

L-moments are then obtained as linear combination of PWMs (e.g. Hosking and Wallis, 1997). The first statistic of interest is usually the index-flood, that corresponds to the sample average,

$$Q_{ind} = b_0, \tag{5}$$

while the L-moments ratios L_{CV} and L_{CA} are computed as:

$$L_{CV} = \frac{2b_1 - b_0}{b_0}, \tag{6}$$

$$L_{CA} = \frac{6b_2 - 6b_1 + b_0}{2b_1 - b_0}. \tag{7}$$

Also the coefficient of L-kurtosis,

$$L_{kur} = \frac{20b_3 - 30b_2 + 12b_1 - b_0}{2b_1 - b_0}, \tag{8}$$

can be used in some cases, for example to estimate a four-parameter probability distribution.

The estimates of sample L-moments are integrated with an estimate of their sample variance, which is a key element of our model because of the particular regression approach adopted in the regionalization procedure. Elmir and Seheult (2004) proposed a method for the computation of variances and covariances of sample L-moments and of the ratios of sample L-moments;

however, their formulation appears to be inconsistent when applied to short samples, producing in some cases negative estimates of the variance. Instead, we start defining the standard deviation of the index-flood, following the Bulletin 17B Appendix 6 (Interagency Advisory Committee on Water Data, 1982), as

$$\sigma_{Q_{ind}} = \sqrt{\frac{1}{n^2}\sum_{x_i<x_0}(x_i - Q_{ind})^2 + \frac{1}{m^2}\sum_{x_i\geq x_0}(x_i - Q_{ind})^2} \tag{9}$$

where Q_{ind} is calculated with equation (5). It is easy to see that, in the absence of non-systematic data, equation (9) reduces to the usual standard deviation of the mean $\sigma_{Q_{ind}} = \sigma_Q/\sqrt{n}$.

The uncertainty of estimates of L_{CV} and L_{CA} is more difficult to assess. Due to the presence of short samples, equations reported by Elmir and Seheult (2004) cannot be applied, so we resort to a set of simplified formulas obtained by Viglione (2007) on the basis of Monte Carlo simulations. The standard deviation of the L_{CV} and L_{CA} are:

$$\sigma_{L_{CV}} = \frac{0.9 \cdot L_{CV}}{\sqrt{n}}, \tag{10}$$

and

$$\sigma_{L_{CA}} = \frac{0.45 + 0.6 \cdot |L_{CA}|}{\sqrt{n}}, \tag{11}$$

respectively. Moreover, sample L_{CV} and L_{CA} are found to be correlated, with a cross-correlation coefficient

$$\rho = \frac{1 - \exp(-5 \cdot L_{CA})}{1 + \exp(-5 \cdot L_{CA})}. \tag{12}$$

Equations (10)-(12) are approximations, and cannot be easily modified to deal with samples extended by mean of occasional information. Consequently, we use $\sigma_{L_{CV}}$ and $\sigma_{L_{CA}}$ calculated only on the systematic sample.

2.2. Regression Models

After the definition of the statistics of interest at the gauged stations, a model to transfer the information to the ungauged sites is needed. In this work, the regional model is intended as a set of relations that allows one to estimate the first three L-moments in an ungauged basin on the basis of a number of basins descriptors. These relationships, defined by means of linear regressions, are built considering the whole descriptors domain, without using any subregion or any limitation. Consequently, homogeneity tests are no longer necessary, because the flood frequency curves are allowed to change site by site.

We define $\mathbf{Y}_T$ as the vector containing the true values of the statistics of interest, in turn index-flood, coefficient of L-variation and coefficient of L-skewness. Any transformation of these variables can also be considered. The basic hypothesis is that $\mathbf{Y}_T$ can be described through the linear relation:

$$\mathbf{Y}_T = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\delta} \tag{13}$$

where the $(N \times p)$ matrix $\mathbf{X}$ contains p suitable descriptors relative to N basins, $\boldsymbol{\beta}$ is the vector of regression coefficients and $\boldsymbol{\delta}$ is the vector of the residuals due to the incorrectness of the linear model approximation, i.e. the model error. Moreover, in regional flood frequency analysis applications, the true statistics $\mathbf{Y}_T$ are not known, and should be replaced by their sample estimators in the whole calibration phase:

$$\mathbf{Y} = \mathbf{Y}_T + \boldsymbol{\eta} \tag{14}$$

where $\boldsymbol{\eta}$ represents the vector of the sampling errors, built up by considering the relations (9)-(11).

Combining equation (13) and (14), the regression model becomes:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{15}$$

where $\boldsymbol{\varepsilon} = \boldsymbol{\delta} + \boldsymbol{\eta}$ is the vector of the residuals that contains both the model and the sampling errors.

The simplest method to estimate the regression coefficients is based on the ordinary least squares (OLS) procedure, that, however, is usually not appropriate in hydrological analyses. In fact, due to the presence of records of different length and of cross-correlation among different records (e.g. Stedinger and Tasker, 1985), the requirements of homoscedasticity and independence of the residuals are often violated. To deal with these limitations, the weighted and the generalized least squares (WLS and GLS respectively) methods have been developed, which require the definition of the covariance matrix of the observations (Montgomery et al., 2001).

In a GLS framework, the vector containing unbiased estimators $\hat{\boldsymbol{\beta}}$ of the regression coefficients $\boldsymbol{\beta}$ can be computed as (Montgomery et al., 2001)

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{Y}, \tag{16}$$

where $\boldsymbol{\Lambda}$ is the sampling covariance matrix of the at-site estimators $\mathbf{Y}$. In particular, the ordinary least squares (OLS) are the special case in which $\boldsymbol{\Lambda}$ is the identity matrix, whereas the weighted least squares (WLS) involve a generic diagonal matrix (the diagonal elements of $\boldsymbol{\Lambda}$ are the sample variances of each at site estimator). $\boldsymbol{\Lambda}$ has positive values also out of diagonal in the GLS case, i.e. when cross-correlations between sample estimates cannot be neglected.

If one considers a non-exact model (Stedinger and Tasker, 1985; Griffis and Stedinger, 2007), i.e. the model as an approximation of a real unknown functional relation, the variance term relative to the model error also has to be accounted for. In this case, the covariance matrix $\boldsymbol{\Lambda}$ is computed by Stedinger and Tasker (1985) combining two terms: the (unknown) model variance and the (estimable) sample variance. The method used in this work is based on this approach, where the two uncertainty components are separated and the model variance is also used as a quality index. Note that in the literature the terms WLS and GLS usually refer to covariance matrices containing only the sample variance; then, to avoid misunderstandings due to the notation, here we will refer to this approach as iGLS (or iWLS), where the "i" stands for "iterative", since equation (18) requires an iterative solution. In this case $\boldsymbol{\Lambda}$ is approximated by its estimator, defined as

$$\hat{\boldsymbol{\Lambda}}\left(\sigma_\delta^2\right) = \sigma_\delta^2 \mathbf{I}_N + \hat{\boldsymbol{\Sigma}} \tag{17}$$

where $\hat{\mathbf{\Sigma}}$ is the sample covariance matrix of $\mathbf{Y}$, $\mathbf{I}_N$ is the identity matrix and σ_δ^2 is the model error variance. The regression coefficients $\hat{\boldsymbol{\beta}}$, computed with equation (16), and σ_δ^2 are (jointly) estimated (Griffis and Stedinger, 2007) searching for nonnegative solution to the equation

$$\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)^T \left[\hat{\sigma}_\delta^2 \mathbf{I}_N + \hat{\mathbf{\Sigma}}\right]^{-1} \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = N - p \tag{18}$$

where $\hat{\sigma}_\delta^2$ is the estimate of σ_δ^2, N is the number of catchments and p is the number of independent variables used in the regression (including the intercept).

In the paper by Stedinger and Tasker (1985) and related works, a complete covariance matrix $\hat{\mathbf{\Sigma}}$ is used, that includes covariances in the off-diagonal elements. In our study, the basins are assumed to be independent of each other, because of the high climatic heterogeneity of the area: thus $\hat{\mathbf{\Sigma}}$ reduces to a diagonal matrix containing the sample variance of the i-th at site estimate of Q_{ind}, L_{CV} and L_{CA} as the i-th diagonal element. Strictly speaking, our model therefore follows an iWLS approach.

2.3. Regression Model Selection

In regional analyses a great number of physical descriptors at the basin scale can be used nowadays, thanks to the availability of accurate digital terrain models and remotely sensed data. Despite this, it is necessary to define a suitable subset of descriptors to be used in the regression, in order to obtain a robust model. Each model should be tested for significance and against multicollinearity before application (Montgomery et al., 2001). The statistical significance of the descriptors used in the model is tested through standard Student t-test, applied to each estimated regression coefficient $\hat{\beta}_j$. The null hypothesis $H_0 : \beta_j = 0$ is tested using the statistic

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\mathrm{var}(\hat{\beta}_j)}} \tag{19}$$

(e.g. Montgomery et al., 2001) where the variance of the regression coefficient is taken from the diagonal of the sampling covariance matrix $\left(\mathbf{X}^T\hat{\mathbf{\Lambda}}\mathbf{X}\right)^{-1}$ (Reis et al., 2005).

The t_0 statistic is compared against its limit value and the null hypothesis is rejected, i.e. the coefficient is considered significantly different from zero, if $|t_0| > t_{\alpha/2,n-p}$, where t is the quantile of the (two-tailed) Student distribution with a confidence level α and $n - p$ degrees of freedom.

The regression is also checked against multicollinearity, in order to avoid to select descriptors that are mutually near-linearly related, that would lead to misleading results. The test used for this purpose is the variance inflation factor (VIF) test (e.g. Montgomery et al., 2001) with a limit value of 5, that is commonly accepted as an indicator of absence of multicollinearity. The VIF value is calculated for each descriptor j of a selected model as $\mathrm{VIF}_j = (1 - R_j^2)^{-1}$, where R_j^2 is the coefficient of determination obtained when the vector of values of the j-th descriptor is regressed on the remaining $p - 1$ descriptors. The test is passed if all the VIF values are lower than the selected limit.

The models passing the t-Student and VIF tests are retained and the model choice within this subset is based on the analysis of the regression residuals: models with the lowest model variance are favored. After the choice of the most appropriate model, we use this model to calculate the predicted value of the variable of interest (Q_{ind}, L_{CV} and L_{CA}) in an ungauged basin. Hence forward we will use the "$\hat{\ }$" symbol to refer to the value predicted by the regression, while the symbol without any mark will denote the sample estimate. One therefore has

$$\hat{Y} = \mathbf{x}\hat{\boldsymbol{\beta}}, \tag{20}$$

where $\mathbf{x}$ is the row-vector of descriptors relative to the ungauged basin and $\hat{\boldsymbol{\beta}}$ the regional regression coefficients vector (equation (16)); the variance of $\hat{Y}$ is (Reis et al., 2005)

$$\sigma^2_{\hat{Y}} = \hat{\sigma}^2_{\delta} + \mathbf{x}\left(\mathbf{X}^T\hat{\mathbf{\Lambda}}^{-1}\mathbf{X}\right)^{-1}\mathbf{x}^T, \tag{21}$$

with $\mathbf{X}$ taken from the calibration dataset and $\hat{\mathbf{\Lambda}}$ from equation (17).

The method proposed here allows one to easily estimate the variance of the regional Q_{ind}, L_{CV} and L_{CA} estimators. This is a relevant advantage over standard regional methods, also because it allows one to decide, in gauged stations, whether to use regional or sample estimators. In fact, in these cases, it is possible to compute both the sample (at-site) and the regional estimators and then choose the one with the lowest variance. To this end, the standard deviation of the sample estimates, computed on the available data through equations (9), (10) or (11), is compared to the standard deviation of the corresponding estimates obtained by the regional model by means of equation (21).

3. Selection Of The Probability Distribution

The final aim of a regional procedure is to estimate the flood quantile and its uncertainty for a specific return period at an ungauged site. So far, however, the procedure focused only on modelling Q_{ind}, L_{CV} and L_{CA} leaving aside the problem of the selection of the distribution. The necessity of defining a probability distribution function introduces an additional source of uncertainty, due to the inherent ambiguity in this choice at the regional scale, particularly when one deals with short samples. Indeed, for low return periods there is more than one distribution that fits well the data, and the selection of a suitable distribution for the regional model is not trivial (Laio et al., 2009). A reasonable solution to this problem, when there are no prior knowledge about a suitable distribution to use, is to define the quantile for a specific return period following a model-averaging approach.

The model averaging approach follows the idea that more than one distribution may be suitable for the quantile estimation. Instead of choosing only one distribution (among those that behave well in the fitting range), it is suggested to evaluate many of them and to take their average for each quantile. The different distributions will share the same three lower-order L-moments, but of course the quantile estimators will be different due to the different shape of the distributions.

After the computation of the frequency curve, the uncertainty of the quantile estimates is also assessed. Since regional L-moments are estimated with their variance, we can use a Monte Carlo simulation to define the confidence limits of the frequency curve adopted. The method is summarized as follow: (i) for each basin the regional Q_{ind}, L_{CV} and L_{CV} are computed as well as their variances; (ii) a set of fictitious values of Q'_{ind}, L'_{CV} and L'_{CV} is randomly extracted from the specific distribution of each L-moment; (iii) the parameters of any selected distribution are computed on the basis of the L-moments sampled in (ii), and the quantile is estimated for the required return period; (iv) points (ii) and (iii) are repeated for a great number of times, so that the distribution of the quantile can be empirically estimated; (v) confidence bands are defined based on the quantile distribution built in point (iv).

Note that, when dealing with regional estimates, Q_{ind}, L_{CV} and L_{CV} are assumed to be independent, so that one can consider three univariate distributions. For the index-flood a lognormal distribution $Q'_{ind} \sim logN\left(\hat{Q}_{ind}, \sigma^2_{\hat{Q}ind}\right)$ is appropriate when the regionalized variable is $\log Q_{ind}$ instead of Q_{ind}, as in our case study (see section 4.2 for details); while two independent normal distributions are used for L-moments ratios: $L'_{CV} \sim N\left(\hat{L}_{CV}, \sigma^2_{L\hat{C}V}\right)$ and $L'_{CA} \sim N\left(\hat{L}_{CA}, \sigma^2_{L\hat{C}A}\right)$. The normality (or log-normality) of $\hat{L}_{CV}$ and $\hat{L}_{CA}$ (or Q_{ind}) distributions results from normality of residuals of the linear (or multiplicative) regression.

Differently, the uncertainty of a quantile estimation based on sample data depends on the mutually correlated L_{CV} and L_{CA} (equation (12)). Therefore, the index-flood is sampled from a normal distribution $Q'_{ind} \sim N\left(Q_{ind}, \sigma^2_{Qind}\right)$ while the L-moments ratios are jointly extracted from a multinormal distribution $(L'_{CV}, L'_{CA}) \sim N\left(L_{CV}, \sigma^2_{LCV}, L_{CA}, \sigma^2_{LCA}, \rho\right)$. Normality or joint normality of the average and L-moments estimators is asymptotically obtained, with a rather fast convergence for small sample sizes (Hosking and Wallis, 1997).

4. Case Study

4.1. Data Availability

The methods described above are applied to a set of 70 catchments located in the North-Western part of Italy (see figure 1, which refers to the database used in Claps et al. (2008, p.56)). The analysis is carried out on basins belonging mainly to mountainous areas, with area ranging between 22 and 3,320 km^2 and mean elevation from 471 to 2,719 m a.s.l. To reduce any effect of upstream lakes and/or reservoirs, we discarded basins whose catchment area is covered by lakes in a percentage beyond 10%. The investigated region presents basins subjected to various climate regimes, from purely nivo-glacial to almost temperate-mediterranean.

The first step in the model building is the analysis of available data of annual streamflow maxima, increased, in some cases, by including non-systematic information about large floods occurred in the area. Occasional values are retrieved from reports issued by the national or regional environment agencies and refer to unusually intense events occurred when no systematic measurements were available. The method for inclusion of occasional information allowed us to extend the time series length of 18 basins using a total of 36 non-systematic measurements. The equivalent time series are, on average, 20 years longer than those without non-systematic information.

After the data checking, the sample index-flow, L_{CV} and L_{CA} coefficients and their standard deviations are computed using the equations in section 2.1. A short summary of the sample coefficients is shown in figure 2 (panel (a) for the index-flood and panel (b) for L_{CV} and L_{CA}), where the filled circles highlight the values related to the stations presenting non-systematic information.

A set of 40 basins descriptors (a detailed description can be found in Claps et al. (2008, p.65)) has been built for the group of catchments involved in this analysis, using geomorphologic and climatic characteristics available in the CUBIST database (CUBIST Team, 2007), with procedures developed in the CUBIST project (`www.cubist.polito.it`). The digital terrain model used for the calculation (about 90 m cell grid) comes from the Shuttle Radar Topography Mission (SRTM) of the NASA and it is freely available (see `http://www2.jpl.nasa.gov/srtm/index.html`).

4.2. Regional Model Definition

The model structure adopted in this work for regional estimation of Q_{ind}, L_{CV} and L_{CA} is linear, and parameters are determined with an improved least squares procedure, as discussed in detail in section 2. Although this model has an additive structure (see equation (13)), in hydrology it is common to use also multiplicative models (Griffis and Stedinger, 2007, among others) in the form

$$Y = \alpha_1 X_2^{\alpha_2} X_3^{\alpha_3} \dots X_p^{\alpha_p} \varepsilon \qquad (22)$$

that can be reduced to the linear additive form by means of a log-transformation of both sides of the equation.

We examine additive and multiplicative model structures for each of the cited statistics; details on the descriptors involved and on the transformations applied are summarized in table 1. In particular, concerning the index-flood, two additive and two multiplicative models are considered, with the dependent variable equal either to Q_{ind} or to Q_{ind}/A, where A is the catchment area. These models will be referred as Qind, QindA, lnQind and lnQindA, respectively. The regional model for L_{CV} is still based on an additive model (named LCV) and a multiplicative one (lnLCV), while the L_{CA} is reproduced through an additive model only (LCA). A direct application of the multiplicative model to L_{CA} is not possible due to the non-positiveness of the variable, that is incompatible with a logarithmic transformation.

The best models to be used for the regional estimation are identified among all the possible combinations of a number of descriptors ranging from 1 to 4, plus the intercept. The limit of 4 descriptors is mainly due to the computational efforts required in exploring all the descriptors combinations ($\sim 102,000$ combinations with 40 descriptors). Moreover, additional tests showed that using more than 4 descriptors does not consistently improve the efficiency and the robustness of the final estimates.

All of the above combinations of models are then tested for significance and multicollinearity, and the ones passing the Student and the VIF tests are ranked on the basis of their model variance ($\hat{\sigma}^2_\delta$). Models that emerge as the most efficient are finally checked in order to verify the basic regression hypotheses (see diagnostic plots in figures 3-5). Finally, the best model for each independent variable is selected, as reported in table 2.

When the dependent variable of interest is log-transformed, equations

(20) and (21) yield estimates that are not directly usable and need to be back-transformed to their original space. In this case, if the regression residuals are normally distributed, $\hat{Y}$ is also normally distributed, and its back-transformation leads to a lognormal variable. Therefore, we evaluate the mean of the estimate as

$$\mu = \exp\left(\mu_{\hat{Y}} + \frac{1}{2}\sigma^2_{\hat{Y}}\right) \tag{23}$$

with $\mu_{\hat{Y}}$ equal to $\hat{Y}$, estimated with the regression in the logarithmic space (equation (20)), and $\sigma^2_{\hat{Y}}$ coming from equation (21). The variance of the estimate is obtained as

$$\sigma^2_{\mu} = \mu^2 \cdot \left[\exp\left(\sigma^2_{\hat{Y}}\right) - 1\right]. \tag{24}$$

This back-transformation can be important to avoid estimation bias (e.g. Seber and Wild, 1989, 2.8.7), although very often the simple exponential transformation

$$\mu' = \exp(\hat{Y}). \tag{25}$$

is used to reconstruct the variable in its original space.

4.3. Regression Results

Solutions obtained after sorting the models are reported in table 2, together with a short summary of the prediction errors, i.e. the root mean squared error

$$RMSE = \sqrt{\frac{1}{N}\sum_{s=1}^{N}\left(\hat{Y}_s - Y_s\right)^2}, \tag{26}$$

the mean absolute error

$$MAE = \frac{1}{N}\sum_{s=1}^{N}\left|\hat{Y}_s - Y_s\right|, \tag{27}$$

and the Nash-Sutcliffe efficiency

$$NS = 1 - \frac{\sum_{s=1}^{N}\left(Y_s - \hat{Y}_s\right)^2}{\sum_{s=1}^{N}\left(Y_s - \frac{1}{N}\sum_{s=1}^{N} Y_s\right)}, \tag{28}$$

computed after a cross-validation procedure (table 4), where N is the total number of the gauged stations. Cross-validation is a procedure to validate models and can be easily implemented as follow: (i) one station, in turn, is removed from the database; (ii) the model coefficients are re-calibrated on the basis of the remaining data; (iii) the variable of interest is reconstructed in the site removed and (iv) the residual is computed by comparing the estimate with the sample value.

The model selected for Q_{ind} leads to a rather efficient estimation of the variable. Among the possible transformations (linear or log-transformed; normalized or not by the catchment area), our analysis showed that the most suitable model is lnQind. The selected best model with four descriptors include: the catchment area A, the mean annual precipitation (MAP), a permeability index c_f and the coefficient a of the IDF curve (Intensity-Duration curve of the average of maximum annual rainfall, as expressed in the form

$h = ad^n$, with h as the cumulative precipitation for a duration d, and a and n as catchment-averaged regression parameters). This model passes the Student test with a level of significance of 1% and the VIF test with a limit value of 5. Figure 3 shows the regression diagnostic plots, demonstrating the good qualities of the model.

The regional model of L_{CV} is investigated through an additive structure as well as a multiplicative one. These approaches are respectively referred to as LCV and lnLCV. Regarding LCV, only a few models pass the Student test with a 1% confidence level. Therefore the 2% level is also considered, that correspond to a greater probability of rejecting the null hypothesis that the regression coefficient is equal to zero. The first-ranked model (see table 2 and 4) has four descriptors: the mean elevation (H), the length of the longest drainage path ($LLDP$), the length of the vector linking the centroid to the basin outlet (LOV) and the coefficient n of the IDF curve already introduced for the lnQind model. Diagnostic plots for this latter models are shown in figure 4.

From observation of figure 4 (panel (a)), where the regional prediction are compared against the sample values, we note that the model is not able to catch the whole sample variability. This behavior can be traced back to two main factors: i) intrinsic limitations of the multiple (linear) regression approach based on a set of simple descriptors (reality is certainly more complex than this); ii) uncertainty affecting sample estimates used for model calibration, especially when they are estimated on short samples. An intuitive representation of ii) can be seen in figure 4 panel (a), looking at empty and filled circles, as a function of the sample length: it is apparent that the model is rather efficient at describing the L_{CV} of larger samples, while the large sample variability of L_{CV} in small samples decreases the quality of estimation for small samples.

The last statistic needed for flood regionalization is the coefficient of L-skewness (L_{CA}), that is investigated using only the additive model. The best model we obtain is characterized by three descriptors: longitude and latitude (X_s and Y_s) and the L_{CV} estimate obtained from the previous step. The results are shown in greater detail in figure 5; in this case, similar considerations apply as those already discussed for the L_{CV}.

4.4. Estimation of Quantiles

As already mentioned in section 3, the final aim of the procedure is the estimation of the flood quantiles corresponding to assigned return periods (with uncertainty). Our work applies regional regression models to distribution-free statistics to avoid certain arbitrary constraints induced by the preliminary choice of a distribution probability.

In this section, we discuss about: (i) the estimation of a flood quantiles by means of the model averaging approach and (ii) the assessment of the quantile estimates uncertainty by means of Monte Carlo simulations. To address the first point, we evaluated six different distributions commonly used in hydrology, fitting each of them to the sample data relative to each of the 70 basins under analysis. The distributions considered are: the Pearson type III or Gamma (GAM), the generalized extreme value (GEV), the three-parameters lognormal (LN3), the Gumbel (G), the generalized logistic (GL) and the generalized Pareto (GP) (see Claps and Laio (2008, p.265) for the adopted parameterization). The frequency curves fitted on the sample data can be plotted together with the sample data. For this purpose, we assign a non-exceedance probability to each sample value by means of a plotting position. In this work we use the Hazen plotting position as defined by Hirsch (1987) to include the non-systematic information.

An example is shown in figure 6 for the river Chisone at S. Martino. This example shows that all the distributions have a similar behavior up to a 100-years return period, except for the Gumbel, that is a less flexible distribution, having only two parameters. A similar behavior is obtained for most of the basins (see Claps and Laio, 2008, p.285). The Gumbel distribution is reported only for comparison in these graphs, but is not considered in the model-averaging procedure, because it has only 2 parameters. It is rather clear that all other models are almost equally suitable to represent the sample data; as a consequence we propose to take their average as the frequency curve to consider for quantile estimation (thicker line in figure 6).

For the assessment of the uncertainty of the quantile estimates we use the Monte Carlo procedure described in section 3. An example of the obtained results is in figure 7 (see Claps and Laio (2008, p.190) for a complete report).

4.5. L-moments Estimation in Data-Scarce Stations

Strictly speaking, an ungauged catchment has no data records; thus one needs to use regional models to obtain the estimates of all the three L-moments under consideration. However, if only few measurements are available, it is sometimes possible to estimate at least the lower-order L-moments from the sample with an acceptable degree of robustness. The choice between the regional and the sample estimation method depends on the variance of the corresponding estimators.

An example is shown in figure 8, where a simple tool to decide if it is more reliable a sample L-moment rather than a regional one is reported. Each panel of figure 8 represents the sample standard deviation of each L-moment as a function of a sample coefficient (abscissa) and the record length (ordinate). The thicker iso-lines correspond the average standard deviation of the model predictions, and represent the limits that divide the area where is more suitable the sample estimator to the area where the regional one is preferable. When a sample is available, one can enter in the plots and check if the point falls in the shaded area (sample standard deviation lower that the regional one): in this case it is suggested to use the sample estimate. Circles reported in figure 8 represent the calibration set and put in evidence as, increasing the L-moment order, the regional approaches become more reliable for short records, due to increased variance of sample L-moments estimators with increasing L-moment order. For instance, the Ayasse basin at Champorcher (which have a 29-years record, $\sigma_Q = 9.9$, $L_{CV} = 0.266$ and $L_{CA} = 0.274$), has a sample $\sigma_{Q_{ind}}$ equal to 1.8 and a sample $\sigma_{L_{CV}}$ equal to 0.05, which implies that the corresponding point falls in the grey area in figure 8a and 8b, i.e.for both Q_{ind} and L_{CV} it is preferable to use the sample estimates. Instead, the sample $\sigma_{L_{CA}}$, equal to 0.19, falls in the white area in figure 8c, i.e. the regional L_{CA} is more appropriate, because the (averaged) regional standard deviation is 0.094.

In the light of the results of figure 8, one could take advantage of the regional model to improve the local estimation of the flood frequency curve, replacing sample L-moments with regionally-estimated values whenever the regional estimates have smaller variance. For the present case study, this applies to about 30% of the L_{CV} and about 80% of the L_{CA} values.

5. Conclusions

The approach to the regional flood frequency analysis proposed in this work aims at overcoming some limitations of the classical methods based on (pooling) regions. Although some features of our model already appeared in the scientific literature, the overall conceptual framework is novel and useful for facilitating flood frequency analysis where non-systematic or limited measurement are available.

The method does not require to build up an at-site probability distribution. The sample record is characterized by its L-moments, that are used as the statistics necessary to reconstruct the complete flood frequency curve, and that become the statistics to be regionalized. The use of regression models against a set of basins descriptors allows the predicted L-moments to vary smoothly over the whole descriptors domain, without any subdivision in sub-regions.

Although for higher-order L-moments a unique linear regression is still not able to completely describe the sample variability, this is a step forward with respect to other approaches (for example the "hierarchical" models) in which the higher-order moments or L-moments are typically kept constant over large regions. By avoiding the subjectivity of procedures that create regions and estimate their homogeneity the model provides a "global" optimization rather than a "local" one.

The representation of sample data by L-moments avoids to force the user to accept possible bad fittings related to the preemptive choice of a probability distribution, and allows one to preserve information contained in short samples, that otherwise would be discarded. In the present work, eight stations out of 70 present less than 20 data, and would probably be discarded in a traditional approach. Even though the importance of these short samples in the whole data set is low for the higher-order L-moments, due to their high variance, their preservation is important for "local" estimation. In fact, our approach allows one to combine sample and regional predictions for the estimation of on-site frequency curve.

A final remark can be devoted to the inclusion of non-systematic measurements in flood time series. In literature, non-systematic data are commonly referred to historical flood, occurred before the beginning of the measurement period. However, in the Italian context, we often found time series with large gaps and with some large events measured during this "ungauged" period. In our procedure, these information can be interpreted as non-systematic data and can be used as valuable additional measurements.

Acknowledgments

The authors are grateful to Marta Zanetta for the support in the first part of the work. The study has been supported by the Italian Ministry of Education through the grants 2008KXN4K8 and 2007HBTS85. Comments by Attilio Castellarin and an anonymous reviewer are acknowledged and appreciated.

References

Bayliss, A., Reed, D., 2001. The use of historical data in flood frequency estimation. Technical Report. Centre for Ecology and Hydrology.

Burn, D., 1990. Evaluation of regional flood frequency analysis with a region of influence approach. Water Resources Research 26, 2257–2265.

Castellarin, A., Burn, D., Brath, A., 2008. Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? Journal Of Hydrology 360, 67–76.

Chebana, F., Ouarda, T., 2008. Depth and homogeneity in regional flood frequency analysis. Water Resources Research 44.

Chokmani, K., Ouarda, T., 2004. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. Water Resources Research 40.

Claps, P., Laio, F., 2008. Updating of the procedures for flood evaluation in Piemonte - vol. I (in Italian). Technical Report. Politecnico di Torino. Aggiornamento delle procedure di valutazione delle piene in Piemonte - vol. I. http://www.idrologia.polito.it/piene/PienePiemonte08_Volume1.pdf.

Claps, P., Laio, F., Zanetta, M., 2008. Updating of the procedures for flood evaluation in Piemonte - vol. II (in Italian). Technical Report. Politecnico di Torino. Aggiornamento delle procedure di valutazione delle piene in Piemonte - vol. II. http://www.idrologia.polito.it/piene/PienePiemonte08_Volume2.pdf.

CUBIST Team, 2007. Cubist project: Characterisation of ungauged basins by integrated use of hydrological techniques. Geophysical Research Abstracts, Vol. 10, EGU2008-A-12048, 2008 SRef-ID: 1607-7962/gra/EGU2008-A-12048 EGU General Assembly 2008.

Cunnane, C., 1988. Methods And Merits Of Regional Flood Frequency-Analysis. Journal Of Hydrology 100, 269–290.

Dalrymple, T., 1960. Flood frequency analyses. volume 1543-A of *Water Supply Paper*. U.S. Geological Survey, Reston, Va.

Elmir, E., Seheult, A., 2004. Exact variance structure of sample L-moments. Journal Of Statistical Planning And Inference 124, 337–359.

Fiorentino, M., Gabriele, A., Rossi, F., Versace, P., 1987. Hierarchical approach for regional flood frequency analysis, in: Singh, V.P. (Ed.), Regional flood frequency analysis. D. Reidel, Norwell, Mass, pp. 35–49.

Gabriele, S., Arnell, N., 1991. A hierarchical approach to regional flood frequency-analysis. Water Resources Research 27, 1281–1289.

Griffis, V., Stedinger, J., 2007. The use of GLS regression in regional hydrologic analyses. Journal Of Hydrology 344, 82–95.

Grimaldi, S., Kao, S.C., Castellarin, A., Papalexiou, S.M., Viglione, A., Laio, F., Aksoy, H., Gedikli, A., 2011. Regional frequency analysis, in: Wilderer, P. (Ed.), Statistical Hydrology. Elsevier. volume 2 of *Treatise on Water Science*. chapter 2.18, pp. 479–517. Doi:10.1016/B978-0-444-53199-5.00046-4.

Hall, M., Minns, A., 1999. The classification of hydrologically homogeneous regions. Hydrological Sciences Journal-Journal Des Sciences Hydrologiques 44, 693–704.

Hirsch, R., 1987. Probability plotting position formulas for flood records with historical information. Journal Of Hydrology 96, 185–199.

Hosking, J., Wallis, J., 1997. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press.

Interagency Advisory Committee on Water Data, 1982. Bulletin 17B - Guidelines for determining flood flow frequency. US Department of the Interior Geological Survey.

Laio, F., Di Baldassarre, G., Montanari, A., 2009. Model selection techniques for the frequency analysis of hydrological extremes. Water Resources Research 45.

Merz, R., Bloschl, G., 2005. Flood frequency regionalisation-spatial proximity vs. catchment attributes. Journal Of Hydrology 302, 283–306.

Montgomery, D., Peck, E., Vining, G., 2001. Introduction to linear regression analysis. Wiley Series in Probability and Statistics. third edition.

Ouarda, T., Girard, C., Cavadias, G., Bobee, B., 2001. Regional flood frequency estimation with canonical correlation analysis. Journal of Hydrology 254, 157–173.

Reis, D., Stedinger, J., Martins, E., 2005. Bayesian generalized least squares regression with application to log Pearson type 3 regional skew estimation. Water Resources Research 41.

Seber, G., Wild, C., 1989. Nonlinear Regression. Series in probability and mathematical statistics, Wiley, New York.

Skoien, J., Merz, R., Bloschl, G., 2006. Top-kriging - geostatistics on stream networks. Hydrology And Earth System Sciences 10, 277–287.

Stedinger, J., Tasker, G., 1985. Regional Hydrologic Analysis .1. Ordinary, Weighted, And Generalized Least-Squares Compared. Water Resources Research 21, 1421–1432.

Stedinger, J.R., Vogel, R.M., Foufoula-Georgiou, E., 1993. Frequency analysis of extreme events, in: Maidment, D.R. (Ed.), Handbook of Hydrology. McGraw-Hill. chapter 18.

Viglione, A., 2007. A simple method to estimate variance and covariance of sample L-CV and L-CA. `http://www.idrologia.polito.it/~alviglio/lavori/varcov\_t\_t3.pdf`.

Viglione, A., Laio, F., Claps, P., 2007. A comparison of homogeneity tests for regional frequency analysis. Water Resources Research 43, W03428.

Vogel, R., Wilson, I., Daly, C., 1999. Regional regression models of annual streamflow for the United States. Journal Of Irrigation And Drainage Engineering-ASCE 125, 148–157.

Wang, Q., 1990. Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution. Journal Of Hydrology 120, 115–124.

Table 1: Different model structures used in the analysis. The last column provides the matrix of independent variables $\mathbf{X}$ to be used in the linear regression, that depends on the descriptors matrix $\mathbf{X}_d$ in which each column is a different descriptor and each row a different catchment. The symbol $\mathbf{1}$ indicates an unitary column vector introduced to account for the intercept coefficient in equation (15).

Model denomination	Original variable	Transformation	Sample standard deviation	Descriptors
Qind	Q_{ind}	none	from eq. (9)	$\mathbf{X} = [\mathbf{1}, \mathbf{X}_d]$
QindA	Q_{ind}	Q_{ind}/A	$\sigma_{Q_{ind}}/A$	$\mathbf{X} = [\mathbf{1}, \mathbf{X}_d]$
lnQind	Q_{ind}	$\log(Q_{ind})$	$\sigma_{Q_{ind}}/Q_{ind}$	$\mathbf{X} = [\mathbf{1}, \log \mathbf{X}_d]$
lnQindA	Q_{ind}	$\log(Q_{ind}/A)$	$\sigma_{Q_{ind}}/Q_{ind}$	$\mathbf{X} = [\mathbf{1}, \log \mathbf{X}_d]$
LCV	L_{CV}	none	from eq. (10)	$\mathbf{X} = [\mathbf{1}, \mathbf{X}_d]$
lnLCV	L_{CV}	$\log(L_{CV})$	$\sigma_{L_{CV}}/L_{CV}$	$\mathbf{X} = [\mathbf{1}, \log \mathbf{X}_d]$
LCA	L_{CA}	none	from eq. (11)	$\mathbf{X} = [\mathbf{1}, \mathbf{X}_d]$

Table 2: Regional models for the estimation of Q_{ind}, L_{CV} and L_{CA}. For a short description of the independent variables see table 3.

Model	Equation
lnQind	$\log \hat{Q}_{ind} = -8.76 + 7.99 \cdot 10^{-1} \log A + 1.09 \log a + 9.53 \cdot 10^{-1} \log MAP + 7.85 \cdot 10^{-1} \log c_f$
LCV	$\hat{L}_{CV} = 1.58 \cdot 10^{-1} - 9.79 \cdot 10^{-5} H - 3.19 \cdot 10^{-3} LLDP + 9.67 \cdot 10^{-3} LOV + 6.07 \cdot 10^{-1} n$
LCA	$\hat{L}_{CA} = 3.92 - 6.16 \cdot 10^{-7} X_s - 6.94 \cdot 10^{-7} Y_s + 3.59 \cdot 10^{-1} \hat{L}_{CV}$

Table 3: Descriptors involved in the regional models of table 2. More details in Claps et al. (2008, p.66).

A	Catchment area
H	Mean catchment elevation
$LLDP$	Length of the longest drainage path
LOV	Length of orientation vector
X_s, Y_s	Basin outlet coordinates
c_f	Permeability index
MAP	Mean Annual Precipitation
a, n	Coefficients of the precipitation IDF curve in the form $h = ad^n$
$\hat{L}_{CV}$	Estimated L_{CV}

Table 4: Summary statistics for the selected models, computed in cross-validation mode.

Model	σ_δ^2	σ_δ	NS	RMSE	MAE
lnQind	0.1153	0.340	0.89	101.2	60.1
LCV	0.0054	0.074	0.05	0.105	0.08
LCA	0.0085	0.092	0.09	0.165	0.14

Figure 1: Geographical location of the gauging stations used for the calibration and validation of the model. The area is located in northwestern Italy, the names of the stations are found in Claps et al. (2008, p.56).

Figure 2: Summary of sample estimates for the 70 basins located in Northwestern Italy. Panel (a) shows the index-flood values related to the correspondent basin area, while panel (b) reports sample L_{CV} versus L_{CA}. Panel (c) reports the diagnostic plot of Hosking and Wallis (1997) in which sample L_{CA}- L_{kur} pairs are compared to those of some probability distributions: Gamma (GAM), generalized extreme value (GEV), lognormal (LN3), Gumbel (G), generalized logistic (GL) and generalized Pareto (GP). For all the panels, filled circles indicates the basins where non-systematic information have been included in the analysis.

Figure 3: Diagnostic diagram for index-flood estimation, model lnQind. Panel (a) reports the results in the log-transformed space. Panel (b) shows the comparison between sample and estimated values in the original index-flow space. Empty and filled circles differ for the back-transformation used. Panel (c) and (d) report the check plots for residual normality and homoschedasticity.

Figure 4: Diagnostic plots for L_{CV} estimation, model LCV. Panel (a) shows the comparison between regional and sample estimates. Panel (b) reports the normalplot of the residuals.

Figure 5: Diagnostic plots for L_{CA} estimation, model LCA. Panel (a) shows the comparison between regional and sample estimates, highlighting the effect of sample length by different circles size. Panel (b) reports the normalplot of the residuals.

Figure 6: Example of sample flood data for the river Chisone at S. Martino and superposition of different theoretical frequency distributions. The thicker line is obtained by averaging the theoretical curves. Black dots represents empirical data, circled ones correspond to non-systematic events.

Figure 7: Example of quantiles confidence bands for the river Chisone at S. Martino obtained with a Monte Carlo simulation. Panel (a) reports the bands when the three L-moments are all obtained from sample data; while the curve in panel (b) is based only on a set of regional L-moments obtained after cross-validation.

Figure 8: Comparison between regional and sample standard deviations for the index-flood (panel a), L_{CV} (panel b) and L_{CA} (panel c). In each panel the thinner iso-lines represent the standard deviation of sample estimators (in abscissa, based on the sample of σ_Q, L_{CV} and L_{CV} respectively) and sample lengths n (in ordinate). Thicker line represents the average of the regional standard deviation obtained in the case study, and separate the area of the plot in which the (mean) regional variance is lower than the sample one. For basins falling in the shaded area it is suggested to used the sample estimate instead of the regional one and viceversa.

Figure 1

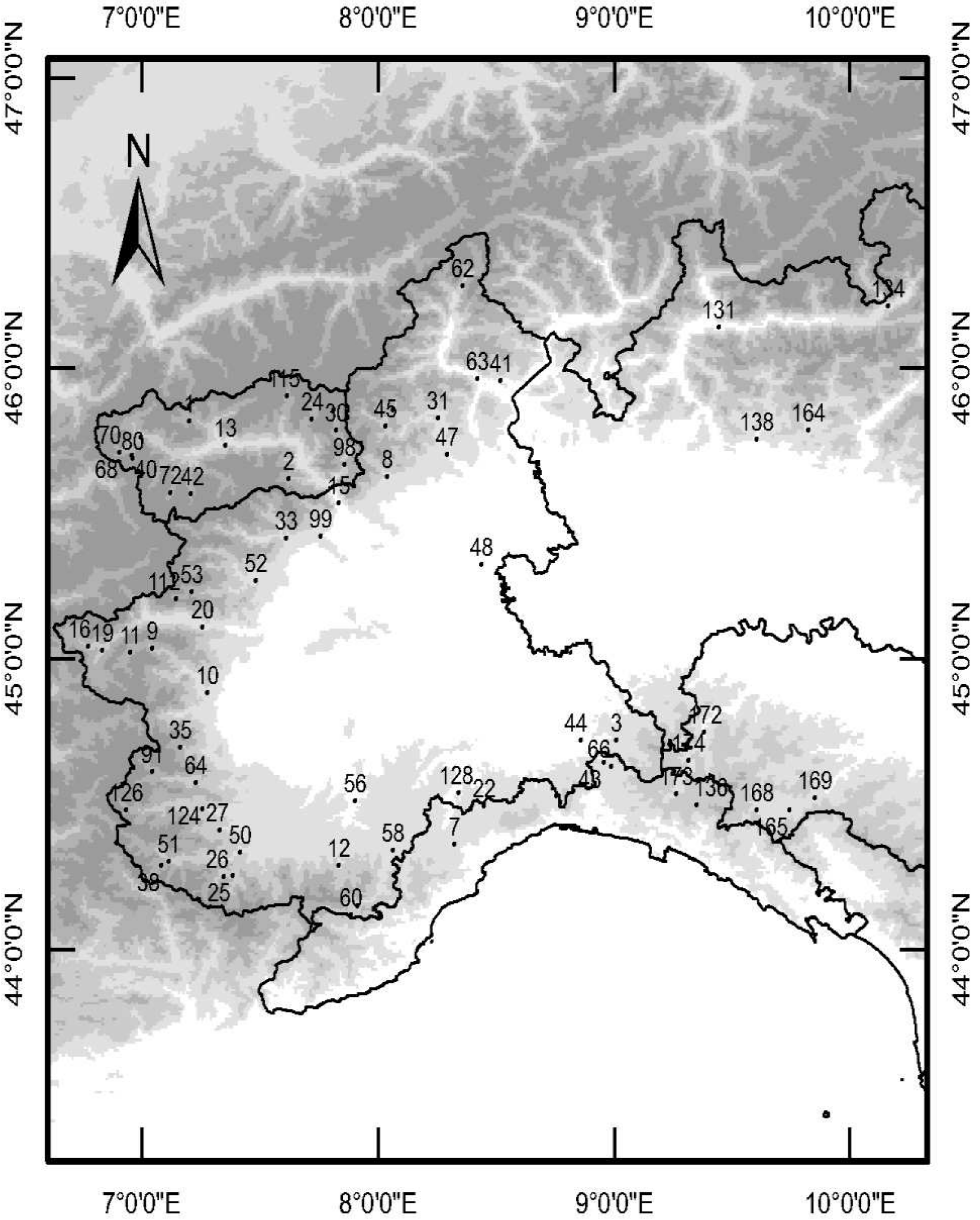
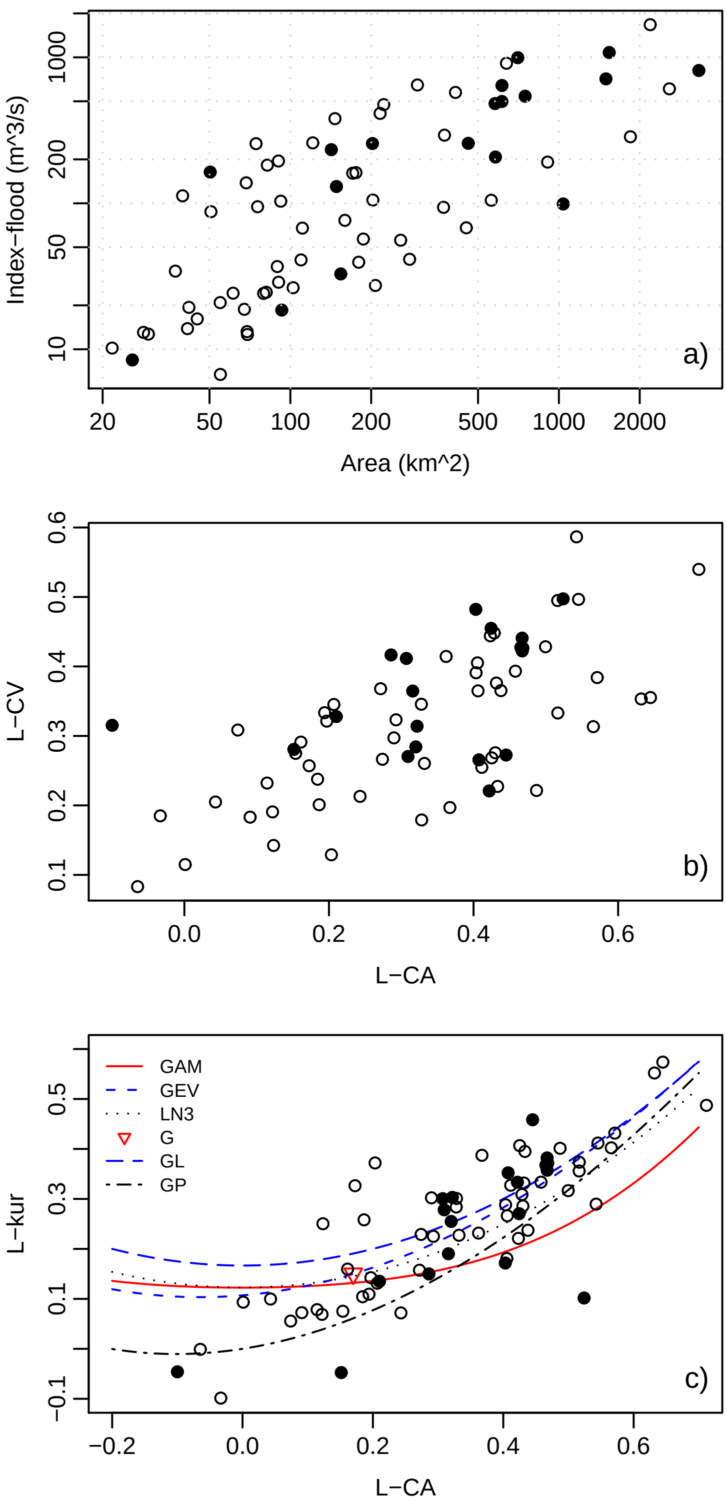
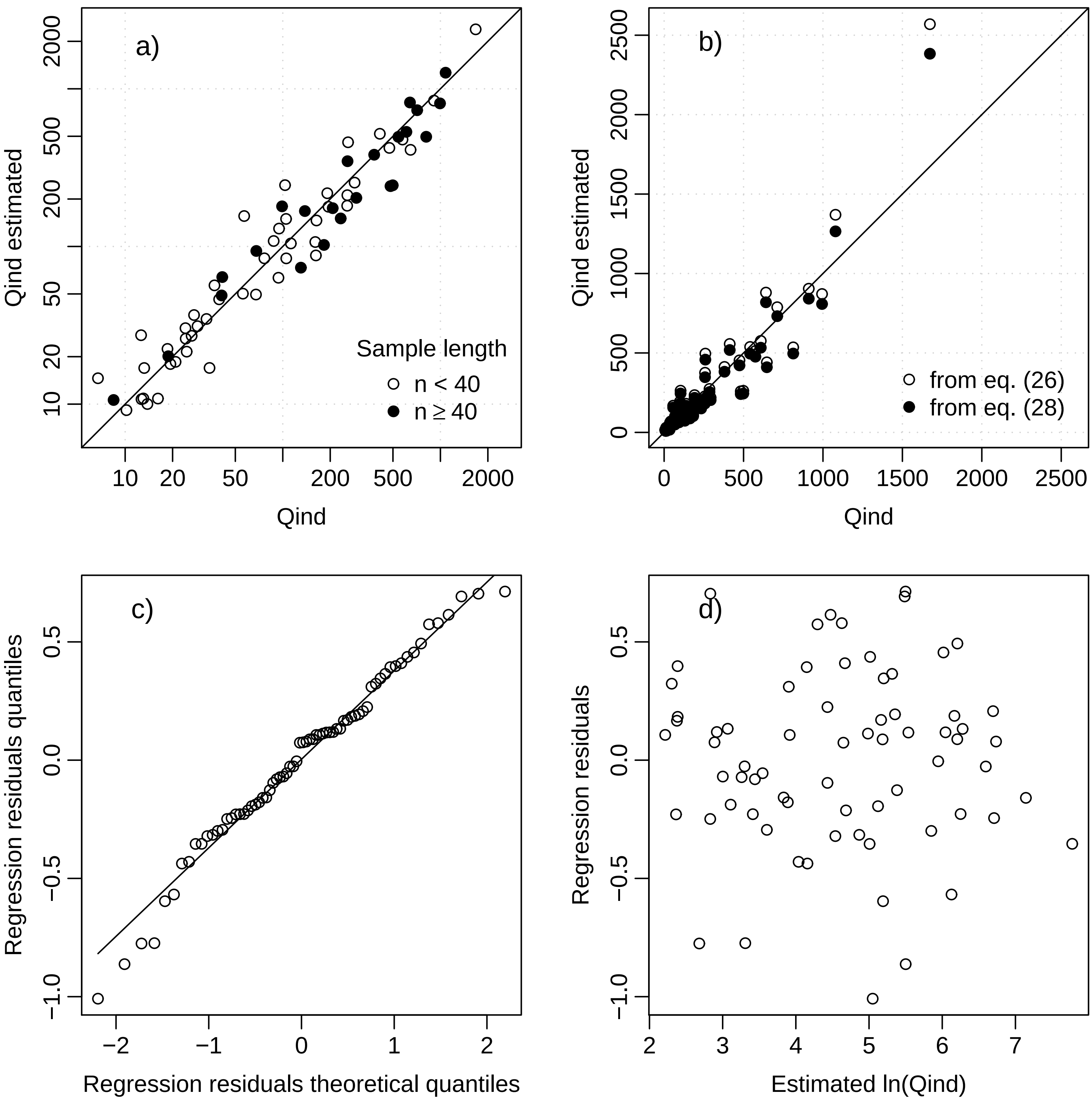

7°0'0"E
8°0'0"E
9°0'0"E
10°0'0"E
47°0'0"N
46°0'0"N
45°0'0"N
44°0'0"N
N
62
134
131
6341
115
24
30
45
31
138
164
1
13
70
80
98
47
8
68
40
7242
2
15
33
99
48
52
112
53
20
16
19
11
9
10
35
44
3
172
66
174
91
64
43
173
136
168
169
126
56
128
22
124
27
50
58
7
165
51
26
12
38
25
60

Figure 2



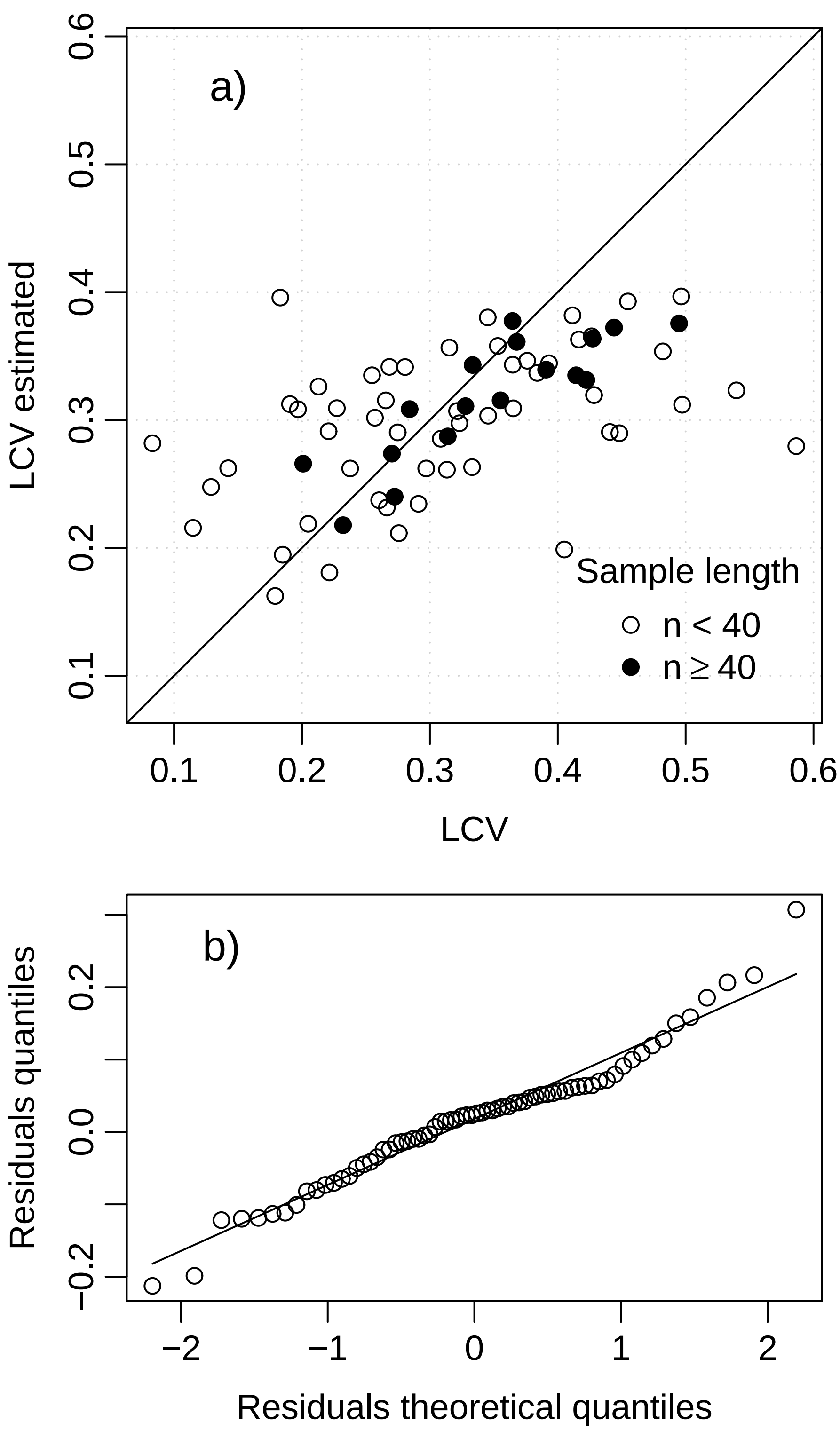
Index−flood (m^3/s)
10
50
200
1000
20
50
100
200
500
1000
2000
Area (km^2)
a)
L−CV
0.1
0.2
0.3
0.4
0.5
0.6
0.0
0.2
0.4
0.6
L−CA
b)
L−kur
−0.1
0.1
0.3
0.5
−0.2
0.0
0.2
0.4
0.6
L−CA
GAM
GEV
LN3
G
GL
GP
c)

Figure 3



a)
Qind estimated
Qind
Sample length
n < 40
n ≥ 40
10
20
50
200
500
2000
b)
Qind estimated
Qind
from eq. (26)
from eq. (28)
0
500
1000
1500
2000
2500
c)
Regression residuals quantiles
Regression residuals theoretical quantiles
−2
−1
0
1
2
−1.0
−0.5
0.0
0.5
d)
Regression residuals
Estimated ln(Qind)
3
4
5
6
7

Figure 4



a)
Sample length
n < 40
n ≥ 40
LCV estimated
LCV
0.1
0.2
0.3
0.4
0.5
0.6
b)
Residuals quantiles
Residuals theoretical quantiles
−0.2
0.0
0.2
−2
−1
0
1
2

Figure 5

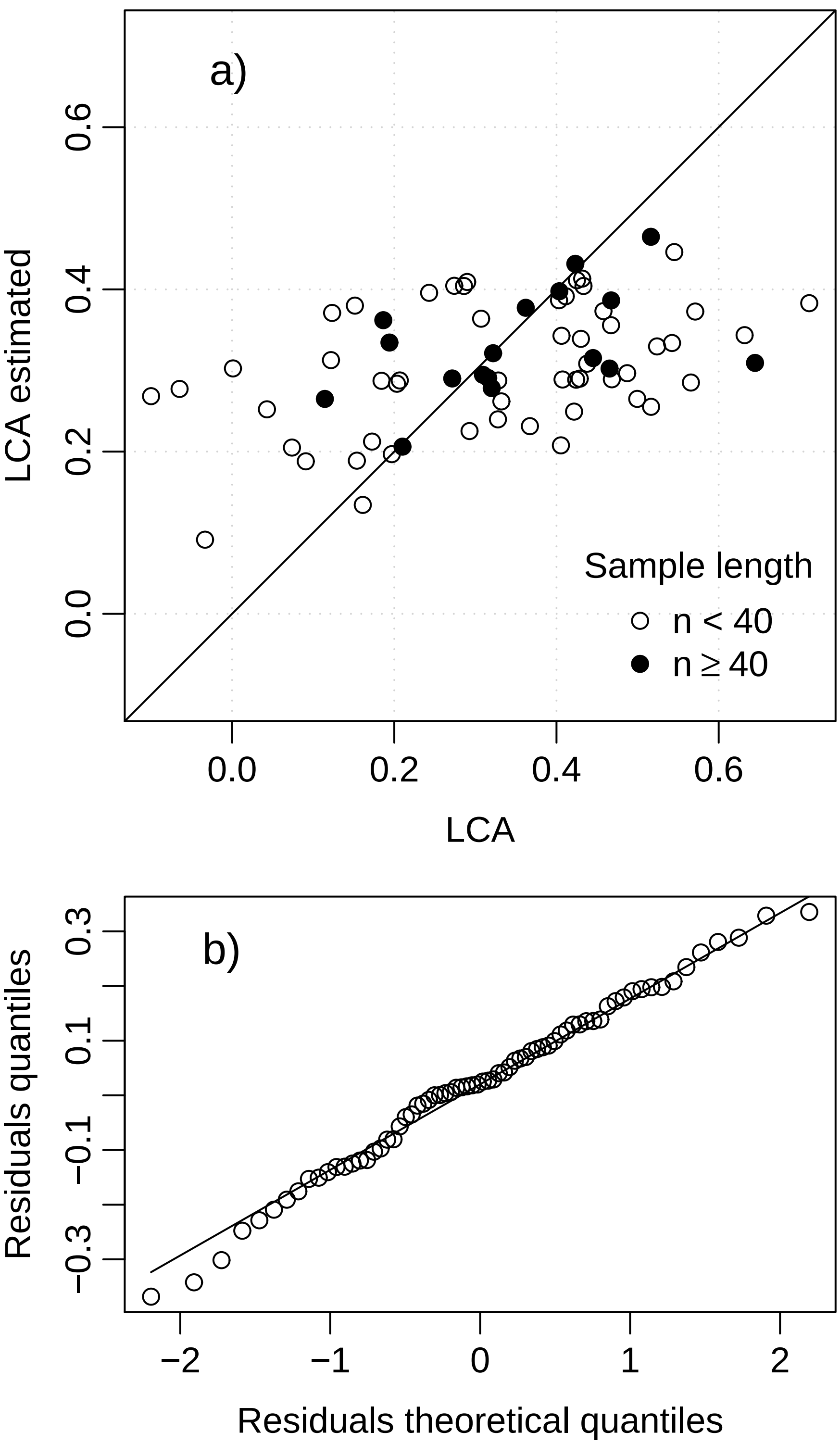
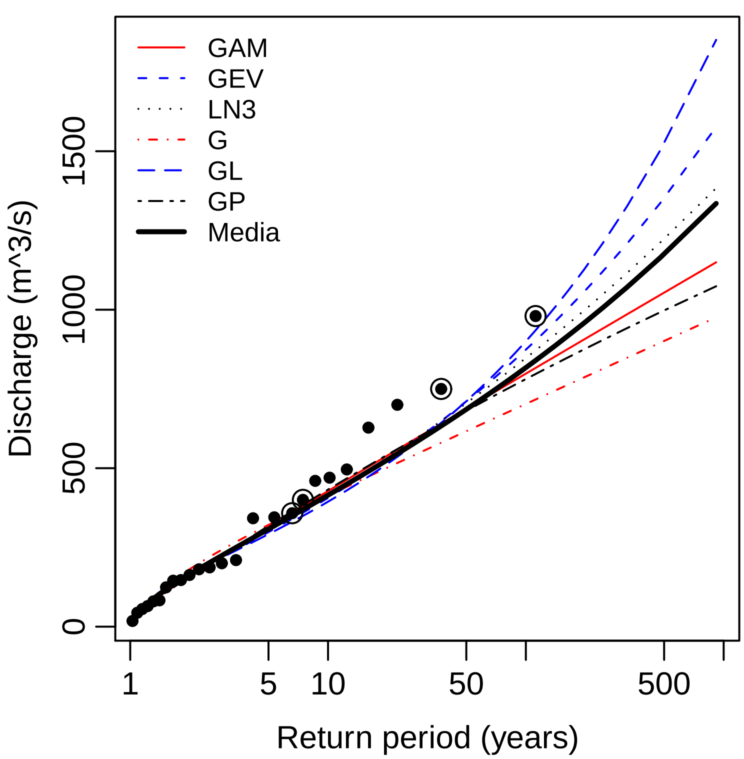
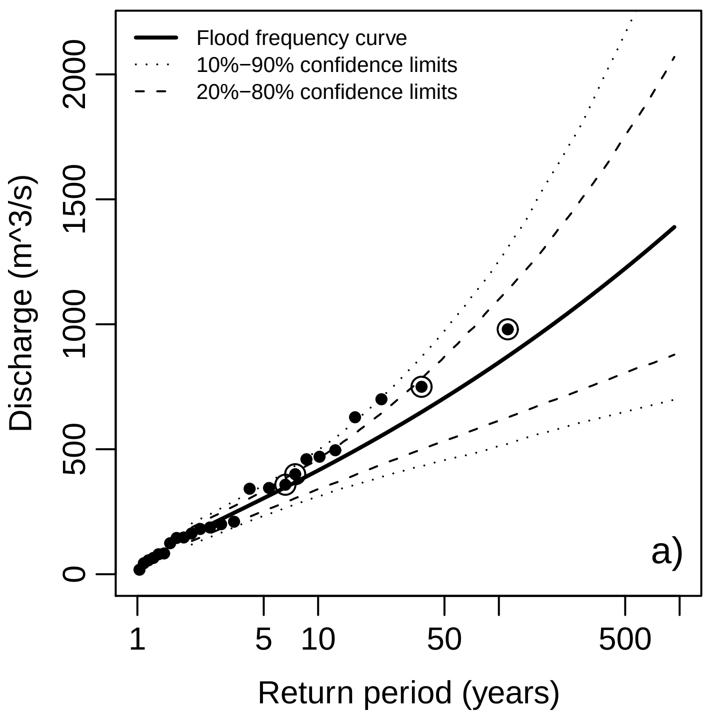

a)
Sample length
n < 40
n ≥ 40
0.0
0.2
0.4
0.6
LCA
LCA estimated
b)
−2
−1
0
1
2
−0.3
−0.1
0.1
0.3
Residuals theoretical quantiles
Residuals quantiles

Figure 6

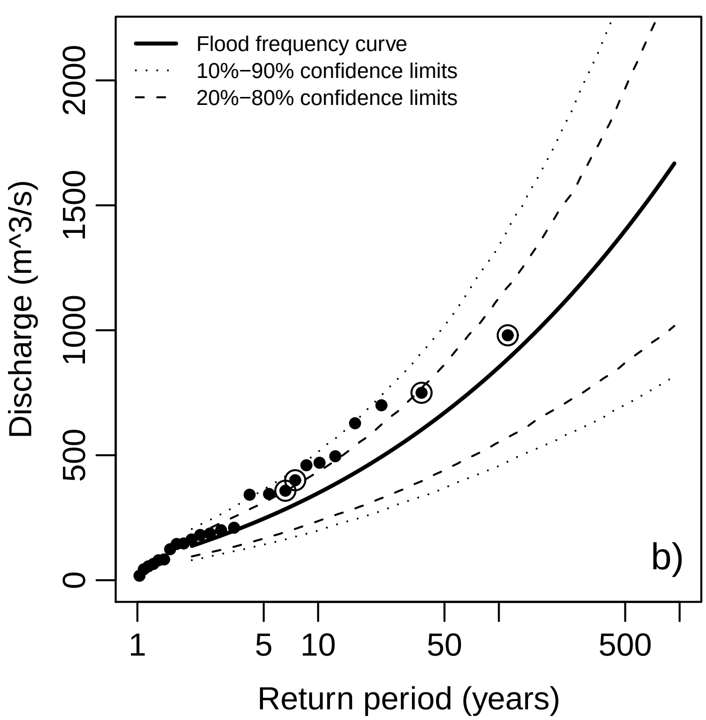


GAM
GEV
LN3
G
GL
GP
Media
Discharge (m^3/s)
0
500
1000
1500
1
5
10
50
500
Return period (years)

Figure 7

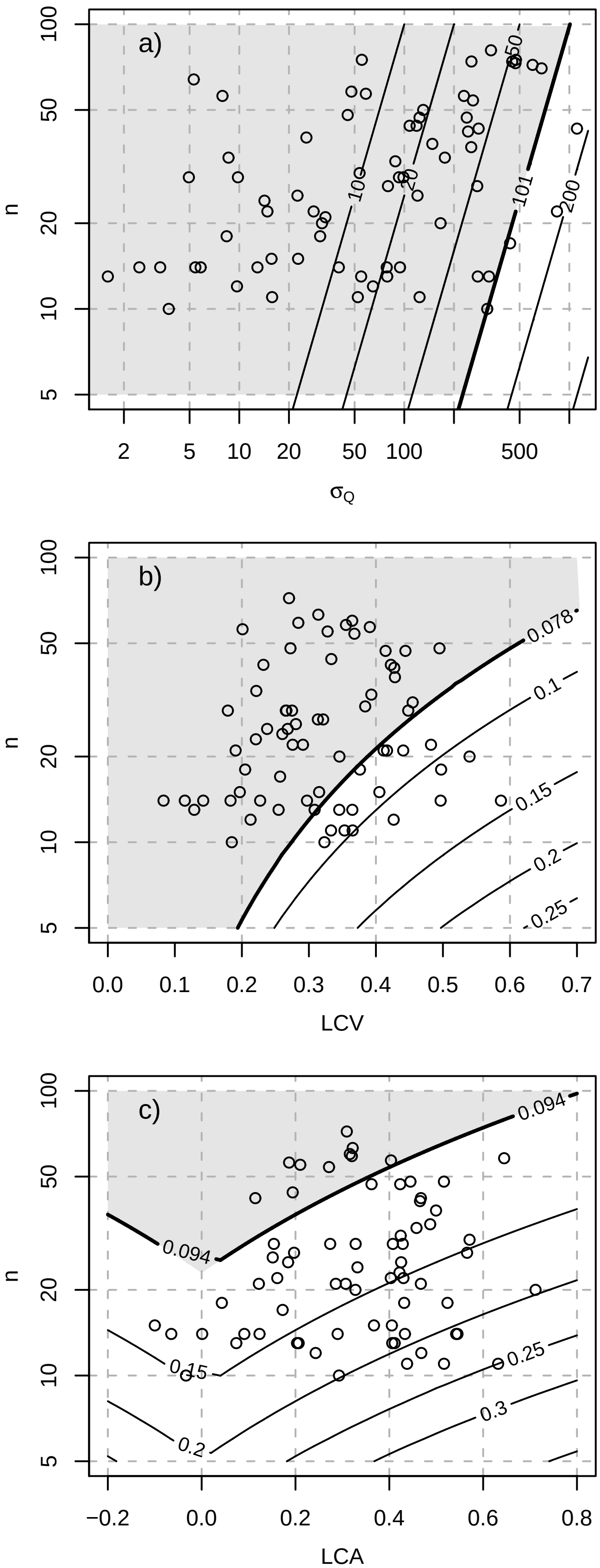


Flood frequency curve
10%−90% confidence limits
20%−80% confidence limits
Discharge (m^3/s)
0
500
1000
1500
2000
a)
1
5
10
50
500
Return period (years)



Flood frequency curve
10%−90% confidence limits
20%−80% confidence limits
Discharge (m^3/s)
0
500
1000
1500
2000
b)
1
5
10
50
500
Return period (years)

Figure 8



a)
100
50
20
10
5
n
2
5
10
20
50
100
500
σQ
10
50
101
200
b)
0.078
0.1
0.15
0.2
0.25
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
LCV
c)
0.094
0.094
0.15
0.2
0.25
0.3
−0.2
0.0
0.2
0.4
0.6
0.8
LCA