Citation for published version:
Mediero, L & Kjeldsen, TR 2014, 'Regional flood hydrology in a semi-arid catchment using a GLS regression model', Journal of Hydrology, vol. 514, pp. 158-171. https://doi.org/10.1016/j.jhydrol.2014.04.007

DOI:
10.1016/j.jhydrol.2014.04.007

Publication date:
2014

Document Version
Early version, also known as pre-print

Link to publication

Regional flood hydrology in a semi-arid catchment using a GLS regression model

L. Mediero[1] and T. R. Kjeldsen[2]

[1]{Technical University of Madrid, Department of Civil Engineering: Hydraulic and Energy, 28040 Madrid, Spain}

[2]{Department of Architecture and Civil Engineering, University of Bath, BA2 7AY, Bath, UK}

Correspondence to: L. Mediero (luis.mediero@upm.es)

Abstract

The regional flood frequency hydrology of the 86,000 km^2 and semi-arid Ebro catchment is investigated using an extended generalised least square model that includes separate descriptions for sampling errors and model errors. The Ebro catchment is characterised by large hydro-climatic heterogeneities among sub-regions. However, differences in flood processes among sites are better explained by a set of new catchment descriptors introduced into hydrological regression models, such as new characteristics derived from the slope of flow duration curves, the ratio of mean annual precipitation to extreme precipitations and the aridity index. These additions enabled a more direct link to be established between the general flow regime and the extreme flood characteristics through-out the entire catchment. The new regression models developed in this study were compared to a set of existing models recommended for flood frequency estimation in Spain. It was found that the generalised least squares model developed in this study improves the existing ordinary least squares models

[1] Corresponding author. Tel.: +34 91 336 5261; fax: +34 91 336 6764
E-mail address: luis.mediero@upm.es

both at regional and trans-regional scales. An adequate description of flood processes is obtained and, as a direct consequence, more reliable flood predictions in ungauged catchments are achieved.

Keywords: Regional flood hydrology; GLS regression model; Ebro catchment; Catchment descriptors; Prediction in ungauged basins

1 Introduction

The prediction of flood frequencies in ungauged catchments is essential for both designing hydraulic infrastructures and effective flood risk management, as floods are one of the most important causes of economic losses in most parts of the world and most catchments are ungauged. To be better prepared for future floods, the European Union has recently established a framework for the assessment and management of flood risks, with the aim of reducing its adverse consequences by knowing flood levels for given probabilities at any stream point (EU, 2007).

The flood level for a given probability at any stream section is usually calculated by a hydraulic model that takes flood quantile estimations as input, which can be obtained from observed data. However, most stream points are ungauged. Thus, spatial information expansion is required to extend the known information in a few gauged catchments to these unguaged sites (Merz and Blöschl, 2008). This expansion usually entails two steps: (i) estimation of regional quantiles at gauged sites for the probability of interest; (ii) use of a regional method to transfer the known information at gauged sites to ungauged catchments.

Several regional flood frequency analyses have been developed in past years. Most of them are based on the use of the index flood method as regional model to estimate flood frequency curves (e.g. Robson and Reed, 1999; Bocchiola et al., 2003; Laio et al., 2011; Dawdy et al., 2012). Regions are assumed to be composed of a set of sites that are homogeneous, which can be grouped by different methods, such as geographical boundaries, cluster analysis and pooling methods. Homogeneity of proposed regions is confirmed by passing a statistical heterogeneity test (Hosking and Wallis, 1997; Castellarin et al., 2008).

The prediction at ungauged sites can be conducted by means of either statistical methods that use series of discharge records or process-based methods that use climate data to run rainfall-runoff models. A comparison between them in Austria can be found in Viglione et al. (2013). Statistical methods are usually based on a regression model that tries to explain differences among flood generation processes through a set of physiographic variables. Catchment response can be characterised in regression models by either the T-year quantile or the index-flood (so-called index-flood indirect estimation methods) (Brath et al., 2001). Other methods exist, such as regional envelope and multivariate probabilistic regional envelope curves (Castellarin et al., 2007) and regional analysis that incorporates historical and palaeoflood information at ungauged sites (Gaume et al., 2010), among many others. A complete review of methods for predicting floods in ungauged basins can be found in Blöschl et al. (2013). In Spain, a regional flood frequency analysis has been conducted recently to improve flood frequency estimations at both gauged and ungauged sites, within the Floods Directive framework (Jiménez-Álvarez et al., 2012). Mainland Spain was divided into 36 homogeneous regions defined by geographical boundaries. Regional quantiles at gauged sites in most regions were estimated by a Generalised Extreme Value (GEV) distribution fitted by the L-moments method with a regional shape parameter, which is estimated by the regional value of the L-coefficient of skewness (L-CS). An ordinary least squares (OLS) regression model was developed to estimate quantiles at ungauged sites in each region.

The main strength of an OLS model is its simplicity, as the estimation of the model uncertainty is straightforward. However, OLS assumes that the uncertainty of quantile estimates at each site are identical, which is not the case as record-lengths vary from site to site. The OLS also neglects both the correlation between quantiles and the correlation between regression model errors. In addition, the existing OLS models in Spain use a reduced set of explanatory variables, usually basin area, precipitation quantiles and mean basin elevation (CEDEX, 2011; Jiménez-Álvarez et al., 2012). More variables could be added to the regression model to account for differences in processes that generate floods. To improve the OLS model currently applied in Spain and overcome its weaknesses, a new regression model is proposed.

In this paper, a regional flood frequency hydrology analysis was carried out in the Ebro River catchment in Spain focusing on the spatial expansion of information to improve the existing regression models. The generalised least squares (GLS) technique that includes the clustering tendency of residuals (Kjeldsen and Jones, 2010) was adapted to the recommendations given in Spain to estimate the frequency distribution, suggesting the use of a GEV distribution fitted through the L-moments method with a given regional shape parameter (Jiménez-Álvarez et al., 2012; MARM, 2011). The semi-arid Ebro River catchment was selected as case study because it shows a significant heterogeneity of climate drivers, rainfall patterns and soil characteristics among homogeneous sub-regions. In addition, a limitation of the existing analysis consists of applying an OLS regression model to each of the five homogenous regions in which the catchment was divided. This paper also addresses the development of a united regression model in the whole Ebro River catchment to avoid undesirable overfitted regression models to a reduced set of gauging stations. Summarising, an exploratory analysis was conducted to investigate how catchment descriptors explain the differences in flood processes among catchments.

2 Hydrological regression models

Regression models are commonly used to describe the between-catchment variation in the at-site estimates of T-year flood quantiles (x_T) at gauged sites by relating the hydrological response to different physiographic variables (so called catchment descriptors), which then take on the role of simplified surrogates of drivers of the flood generation processes. Having estimated a regression model, the T-year event can then be predicted in ungauged catchments where only the catchment descriptors are available. Denoting the vector of at-site log-transformed flood quantiles from N sites as $\boldsymbol{y}$ (Eq. 1), the associated matrix of m different catchment descriptors with a first column of unity as **X,** i.e. the dimension of this matrix is N×(m+1), and the vector of m+1 regression model parameters as $\boldsymbol{\theta}$, a regression model can be formulated by Eq. (2).

$$\boldsymbol{y} = log_{10}(\boldsymbol{x_T}) \quad (1)$$

$$\boldsymbol{y} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\theta} + \boldsymbol{\eta} + \boldsymbol{\varepsilon} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\theta} + \boldsymbol{\omega} \quad (2)$$

where $\boldsymbol{\varepsilon}$ is the vector of sampling errors of the log-transformed at-site quantile, $\boldsymbol{\eta}$ is the vector of regression model errors, and $\boldsymbol{\omega}$ is the vector of total regression errors ($\boldsymbol{\omega}$=$\boldsymbol{\varepsilon}$+$\boldsymbol{\eta}$).

The formulation in Eq. (2) shows that the regression model error can be split into the sampling estimate error and the modelling error. The sampling error represents differences between the quantile estimation from observed data and its true value (ξ), which is unknown, as we would need a record length of an infinite number of years to know it exactly (Eq. 3). This error only depends on the observed data at each site, the probability distribution used to estimate quantiles and the method used to estimate the distribution parameters. In contrast, the modelling error represents the difference between the regression model estimation and its true value (ξ)(Eq. 4). The model error can be also interpreted as the inability of the regression model to explain the catchment behaviour perfectly when only lumped catchment descriptors are used as surrogate explanatory variables for the more complex, and often non-linear, catchment scale hydrological processes. In contrast to the sampling error, the model error depends on the structure of the regression model and, thus, on the selection of catchment descriptors.

$$\boldsymbol{y} = \boldsymbol{\xi} + \boldsymbol{\varepsilon} \tag{3}$$

$$\boldsymbol{\xi} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\theta} + \boldsymbol{\eta} \tag{4}$$

The two errors represent fundamentally different aspects of the modelling process, and in the following the covariance structure of each error type will be discussed. The covariance matrix of the regression errors ($\boldsymbol{\Sigma}_{\boldsymbol{\omega}}$) is defined as the sum of the covariance matrix of the sampling errors ($\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$) plus the covariance matrix of the modelling errors ($\boldsymbol{\Sigma}_{\boldsymbol{\eta}}$) (Eq. 5). It is assumed that the two errors are mutually independent.

$$\boldsymbol{\Sigma}_{\mathbf{w}} = \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} + \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \tag{5}$$

Typically, the parameters of the regression model, $\boldsymbol{\theta}$, are estimated by the least squares method. Different sub-methods exist depending on the complexity of the covariance structure of the errors adopted in the regression model. They are classified, in an increasing complexity order, as: ordinary least square (OLS), weighted least squares (WLS), and generalised least square (GLS). A more in-depth review of regression models can be found in Rosbjerg et al. (2013). Other methods of estimating the model parameters include maximum-likelihood and Bayesian methods.

The GLS technique was developed for application in hydrology by Stedinger and Tasker (1985) to account for the heteroscedasticity and cross-correlation of residuals. Specifically, the GLS model assumes that estimates of flood quantiles at different sites are correlated, as they have been estimated using correlated flood data. This then leads to a $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ matrix with diagonal elements equal to the estimation variance of quantiles (σ^2_ε) and off-diagonal elements equal to the covariance between quantiles across pairs of sites (Eq. 6).

$$\mathbf{\Sigma}_\varepsilon = \begin{cases} \mathbf{\Sigma}_{\varepsilon,\mathbf{ii}} = \sigma^2_\varepsilon\left(y_i\right) & i = j \\ \mathbf{\Sigma}_{\varepsilon,\mathbf{ij}} = cov\left(y_i\ ,\ y_j\right) & i \neq j \end{cases} \tag{6}$$

In the GLS model formulation presented by Stedinger and Tasker (1985), the model error matrix ($\mathbf{\Sigma_\eta}$) only includes non-zero elements along the diagonal, as it assumes that the modelling errors are uncorrelated between sites. Based on the observation that localised clusters of positive and negative residuals were prevalent among neighbouring catchments when modelling a large set of annual maximum series (AMS) of peak flows in the UK, Kjeldsen and Jones (2009) extended the GLS model to include off-diagonal elements larger than zero into the $\mathbf{\Sigma_\eta}$ matrix to describe inter-site correlations of modelling errors (Eq. 7 and 8).

$$\mathbf{\Sigma_\eta} = \sigma^2_\eta \mathbf{R_\eta} \tag{7}$$

$$\mathbf{R_\eta} = \begin{cases} 1 & i = j \\ \boldsymbol{\rho}_{\boldsymbol{\eta},\mathbf{ij}} & i \neq j \end{cases} \tag{8}$$

where σ^2_η is the variance of modelling errors, $\mathbf{R_\eta}$ is a matrix describing inter-site correlations and $\boldsymbol{\rho}_{\boldsymbol{\eta},\mathbf{ij}}$ is the correlation of model errors between sites *i* and *j*. The split between a model error variance, assumed constant across all catchments, and a correlation matrix, $\mathbf{R_\eta}$, is convenient for subsequent model development is the next sections.

In the following sections this GLS regression model framework is developed and tested using hydrological flood data from a large semi-arid catchment situated in North-East Spain.

3 Case study: the Ebro River catchment

The Ebro River catchment is located in the Northeast of Spain covering an area of 84,000 km^2 (see Fig. 1). The regional hydrology shows significant spatial heterogeneities because of i) abrupt changes in orography, as terrain elevation ranges from sea level at the Ebro Delta to 3,404 m.o.s.l at the Aneto peak in the central Pyrenees, which is the highest point in the catchment; ii) heterogeneities in precipitation patterns, as the Southeast part of the catchment has a mean annual rainfall of 450 mm, while in some regions of the Pyrenees a mean annual rainfall of 2,500 mm is observed, and iii) a great variability in quantiles of maximum daily precipitation, as the 100-year rainfall quantile ranges from 80 mm in the central South part of the catchment and up to 160 mm in some parts of the Pyrenees.

Observed AMS of instantaneous peak flow from 93 gauging stations located in natural or near-natural catchments were used in the study (Fig. 1). Regional L-moment values for the five homogeneous regions used in the Ebro River catchment can be seen in Table 1. Eight different catchment descriptors were readily available for each of the 93 catchments, including: 1) catchment area in km^2 (A); 2) mean elevation of the catchment over the mean sea level in m (H); 3) maximum daily precipitation with a *T*-year return period in mm (P_T); 4) mean annual precipitation in mm (P_m); 5) mean infiltration rate in mm (t_{inf}), which was calculated from a national gridded map obtained previously by the kriging method applied to a set of site values estimated from either field measurements or a function that simulates the water transference in a soil; 6) mean catchment slope (S); 7) initial abstraction in mm (P_0), defined as the precipitation needed before runoff begins, which was calculated from a national gridded map obtained previously using information provided by maps of t_{inf} and land use from the CORINE Land cover; and, finally, 8) catchment area (again measured in km^2) located at elevations in excess of 1,500 m (A_{1500}).

A further three catchment descriptors were developed as part of this study to better capture climatic differences between sites: i) the mean potential evapotranspiration in mm (PET), which was obtained from temperature series in the period 1940-1995 through the Thornthwaite and Penman equations; ii) the aridity index (I_a), defined as the ratio of P_m to PET; and iii) the extremity index (I_e), defined as the ratio of P_m to P_T.

Two additional catchment descriptors were used to capture differences in flood response from the information given by flow duration curves (FDC). Specifically, a concavity index (*IC*) was adopted, which gives information about the relationship between low-flow and high-flow regimes (Eq. 9) (Sauquet and Catalogne, 2011). A coefficient was defined to measure the slope of the upper part of the FDC for the highest flows ($SFDC_p$) (Eq. 10),

$$IC = \frac{Q_{0.1} - Q_{0.99}}{Q_{0.01} - Q_{0.99}} \quad (9)$$

$$SFDC_p = \frac{Q_{\max} - Q_p}{100\ p} \quad (10)$$

where Q_p is the daily runoff for an exceedance probability of p and Q_{max} is the maximum daily runoff. Both Q_p and Q_{max} are calculated from a FDC standardised by the mean daily runoff to enable the comparison between catchments.

All these descriptors can be obtained easily from digital terrain models and other gridded dataset of climate, such as rainfall and evapotranspiration, except for the case of those descriptors that capture the properties of the FDC. In this case, a further analysis should be carried out to establish relationships between these indexes and different soil descriptors to enable estimation in ungauged catchments. However, this additional step is beyond the scope of this paper.

The following sub-section addresses how these catchment descriptors can explain the differences in flood generation processes among catchments.

3.1 Explaining flood processes by catchment descriptors

Catchment area, A, and the respective T-year rainfall quantile, P_T, are the two first catchment descriptors usually introduced into a regression model. As expected, catchment area always exerts the largest influence on the magnitude of floods, as generally larger catchments lead to larger floods. The inclusion of the rainfall quantile gives additional information about differences in flood magnitude between similar sized catchments, as larger values of P_T will usually result in larger floods being generated.

The mean catchment slope, S, explains differences among catchments due to their topography. Catchments with steeper slopes are expected to have faster runoff velocity in hillslopes which reduces the concentration time, and consequently lead to higher peak flow values.

The concavity index, IC, characterises the upper part of the FDC, explaining differences in catchment hydrological responses. Larger values of IC are obtained at sites where the hydrological response is more smoothed due to the existence of aquifers or the influence of snowmelt. In contrast, smaller values of IC are found in catchments with fast runoff responses due to the existence of impermeable soils or extreme climate conditions, as is often the case in arid and semi-arid regions (Castellarin et al., 2013).

The extremity index, I_e, explains how large P_T is in comparison to P_m. This descriptor gives information about the variability of extreme rainfall events compared to the mean annual rainfall. Smaller values of I_e will typically be observed in more arid regions, where larger year to year variability in extreme rainfall events is observed.

P_0 is related to the potential maximum water retention of a soil. Therefore, this descriptor gives information about the portion of precipitation transformed into surface runoff in the catchment. In fact, P_0 supplies different information than the *IC* index. The latter explains the probability distribution of daily runoffs, capturing the relationship between surface runoff and subsurface flow, without accounting for the precipitation. However, P_0 gives information about the hydrologic abstraction process to transform precipitation into surface runoff.

Potential evapotranspiration, *PET*, gives information about the initial moisture content. A catchment with wetter soil moisture content will drive a larger flood than a catchment with dryer soil moisture content, for the case of a similar rainfall event. The aridity index, $I_{a,}$ also accounts for the likely initial soil moisture content before a flood begins.

4 Methodology

The methodology section describes the GLS regression model used in this study. In the following four sub-sections, the necessary developments of different aspects of the GLS model are described in more detail. Firstly, the estimation of the covariance matrix of sampling errors is presented based on Taylor series approximations (so-called the delta method). Next, the estimation of the covariance matrix of the modelling errors is addressed. Then, the estimation of the regression model parameters by the maximum likelihood technique is described. Finally, three measures to assess the quality of the GLS regression model are presented.

4.1 Covariance matrix of sampling errors

The diagonal elements of $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ contain the sampling variance of the log-transformed T-year quantile of the at-site estimates (Eq. 6), which primarily depends on the frequency distribution used, the record-length, and the procedure to estimate its parameters. In this paper, Taylor series expansions were used to obtain approximate analytical solutions of these uncertainties, but other methods could also have been adopted such as jackknife resampling (Liu and Singh, 1992) or bootstrapping (Efron and Tibshirani, 1993). In the case of the GEV distribution, which is the frequency distribution recommended in the Ebro River catchment by Jiménez-

Álvarez et al. (2012), the asymptotic variance of the log-transformed quantile is given by Rao and Hamed (2000) and shown in Eq. (11).

$$\sigma_\varepsilon^2(y)=\left(\frac{log_{10}(e)}{x_T}\right)^2\left[\left(\frac{\partial x_T}{\partial u}\right)^2\sigma^2(u)+\left(\frac{\partial x_T}{\partial \alpha}\right)^2\sigma^2(\alpha)+\left(\frac{\partial x_T}{\partial k}\right)^2\sigma^2(k)\right.$$
$$+2\left(\frac{\partial x_T}{\partial u}\right)\left(\frac{\partial x_T}{\partial \alpha}\right)cov(u,\alpha)+2\left(\frac{\partial x_T}{\partial u}\right)\left(\frac{\partial x_T}{\partial k}\right)cov(u,k) \tag{11}$$
$$\left.+2\left(\frac{\partial x_T}{\partial \alpha}\right)\left(\frac{\partial x_T}{\partial k}\right)cov(\alpha,k)\right]$$

where y is the log-transformed quantile defined in Eq. (1), u, α and k are the location, scale and shape parameters, respectively, of the GEV distribution and e is Euler's number. The T-year flood quantile, x_T, in the case of a GEV distribution is given by Eq. (12).

$$x_T = u + \frac{\alpha}{k}\left[1-\left(-\ln\left(1-\frac{1}{T}\right)\right)^k\right] \tag{12}$$

In the case when the shape parameter is estimated by a regional estimate of the L-coefficient of skewness, L-CS, and considered a constant, Eq. (11) can be reduced to only three terms, as k is a constant (Eq. 13) (Lu and Stedinger, 1992). Further details on the analytical expressions of the individual terms in Eq. (13) can be found in Appendix A.

$$\sigma_\varepsilon^2(y)=\left(\frac{\log_{10}(e)}{x_T}\right)^2\left[\left(\frac{\partial x_T}{\partial u}\right)^2\sigma^2(u)+\left(\frac{\partial x_T}{\partial \alpha}\right)^2\sigma^2(\alpha)+2\left(\frac{\partial x_T}{\partial u}\right)\left(\frac{\partial x_T}{\partial \alpha}\right)\mathrm{cov}(u,\alpha)\right] \tag{13}$$

The off-diagonal elements of $\mathbf{\Sigma_\varepsilon}$ describe the covariance between at-site estimates at different sites to account for the fact that individual storms are more likely to affect neighbour catchments than catchments located further apart. The covariance between log-transformed quantiles at different sites is estimated using Eq. (14). Further details on the analytical evaluation of this covariance term can be found in Appendix B.

$$\mathbf{\Sigma_{\varepsilon,ij}} = cov(\boldsymbol{y_i}, \boldsymbol{y_j}) = \frac{(log_{10}(e))^2}{\boldsymbol{x_{T,i}}\ \boldsymbol{x_{T,j}}}cov(\boldsymbol{x_{T,i}}, \boldsymbol{x_{T,j}}) \tag{14}$$

When the L-moment method is used, correlations between probability weighted moments (PWM) at two different sites are needed in order to estimate the off-diagonal elements of $\mathbf{\Sigma_\varepsilon}$ (Eq, B9-B11). As in previous studies, this correlation is assumed to be related to the correlation between AMS by a power function as suggested by Eq. (15) (Stedinger, 1983; Madsen and Rosbjerg, 1997; Martins and Stedinger, 2002).

$$\rho_{b_{ri},b_{rj}} = \rho_{ij}^{\delta} \qquad (15)$$

where b_{ri} is the rth order PWM at site i, $\rho_{b_{ri},b_{rj}}$ is the correlation between two rth order PWMs at sites i and j, ρ_{ij} is the correlation between AMS of peak flows at sites i and j, and δ is the exponent of ρ_{ij}, which is unknown.

A bootstrap experiment was carried out to estimate the values of δ from the properties of correlations between PWMs following the methodology used by Kjeldsen and Jones (2006). For each pair of sites, the overlapping period was identified and a new sample was generated by means of a bootstrap technique. A year is selected randomly with replacement from the overlapped record. For each selected year the pair of associated annual maximum peak flow observations is transferred to the bootstrap sample in order to keep the inter-site correlation. The procedure is repeated until the synthetic sample length equals the overlapping length, and finally, the PWMs are calculated from the synthetic samples. The procedure is repeated 1,000 times to estimate the correlation between PWM at different sites.

The final step involves the correlation between logarithmic values of AMS at different sites, ρ_{ij}, used to estimate $\rho_{b_{ri},b_{rj}}$, which was smoothed by a double exponential expression (Eq. 16) proposed by Kjeldsen and Jones (2009). .

$$\rho_{\varepsilon,ij} = \varphi_{\varepsilon,1}\, e^{-\varphi_{\varepsilon,2} d_{ij}} + \left(1-\varphi_{\varepsilon,1}\right) e^{-\varphi_{\varepsilon,3} d_{ij}} \qquad (16)$$

where $\rho_{\varepsilon,ij}$ is the smoothed correlation with distance between sites i and j, d_{ij} is the distance between centroids of catchments i and j (in km) and $\varphi_{\varepsilon,1}$, $\varphi_{\varepsilon,2}$ and $\varphi_{\varepsilon,3}$ are coefficients estimated using the least squares technique.

4.2 Covariance matrix of modelling errors

The covariance matrix of the modelling errors, $\mathbf{\Sigma_\eta}$, equals a matrix describing inter-site correlations ($\mathbf{R_\eta}$) scaled by the variance of modelling errors, σ^2_η, (Eq. 7-8). Therefore, the diagonal elements of $\mathbf{\Sigma_\eta}$ describe the uncertainty in model estimations and are equal to the variance of modelling errors (σ_η^2). The off-diagonal elements of $\mathbf{\Sigma_\eta}$ describe the cross-correlation of model errors between sites by $\rho_{\eta,ij}$, which is smoothed with distance between sites following an expression similar to Eq. (16) with parameters $\varphi_{\eta,1}$, $\varphi_{\eta,2}$ and $\varphi_{\eta,3}$.

$$\rho_{\eta.ij} = \varphi_{\eta,1} \exp(\varphi_{\eta,2} d) + (1 - \varphi_{\eta,1}) \exp(\varphi_{\eta,3} d)$$

4.3 Estimation of regression model parameters

The proposed model has several unknown parameters: the *m+1* parameters of the regression model ($\boldsymbol{\theta}$), the variance of the model errors (σ_η^2) and the three parameters describing the model error correlation with distance ($\varphi_{\eta,1}$, $\varphi_{\eta,2}$ and $\varphi_{\eta,3}$). All these parameters can be estimated by the maximum likelihood technique, assuming that regression residuals follow a normal distribution with mean equal to zero and variance given by the covariance matrix $\boldsymbol{\Sigma_\omega}$ (Kjeldsen and Jones, 2009) (Eq. 17). The negative log-likelihood function, *-ln(L)*, for the regression model is given by Eq. (18), and is minimised to estimate the model parameter values.

$$\boldsymbol{\Sigma_w} = \boldsymbol{\Sigma_\varepsilon} + \boldsymbol{\Sigma_\eta} = \boldsymbol{\Sigma_\varepsilon} + \sigma_\eta^2 \mathbf{R_\eta} = \sigma_\eta^2 \left(\boldsymbol{\Sigma_\varepsilon} / \sigma_\eta^2 + \mathbf{R_\eta} \right) = \sigma_\eta^2 \mathbf{G} \tag{17}$$

$$-2ln(L) = ln\left[det\left(\sigma_\eta^2 \mathbf{G} \right) \right] + \left(\boldsymbol{y} - \mathbf{X}\boldsymbol{\theta} \right)^T \left(\sigma_\eta^2 \mathbf{G} \right)^{-1} \left(\boldsymbol{y} - \mathbf{X}\boldsymbol{\theta} \right) \tag{18}$$

In practice, the number of unknown parameters can be reduced, as for given values of σ_η^2 and $\varphi_{\eta,1}$, $\varphi_{\eta,2}$ and $\varphi_{\eta,3}$ the regression model parameters that minimise the negative log-likelihood function are given by the GLS estimator (Eq. 19). Therefore, the unknown parameters of the log-likelihood function are reduced to four: σ_η^2, $\varphi_{\eta,1}$, $\varphi_{\eta,2}$ and $\varphi_{\eta,3}$.

$$\hat{\boldsymbol{\theta}} = \left(\mathbf{X^T G^{-1} X} \right)^{-1} \mathbf{X^T G^{-1}} \boldsymbol{y} \tag{19}$$

4.4 Measures to select the regression model

Once a regression model with *m* catchment descriptors is fitted to the observations, a multicollinearity test should be applied to avoid the inclusion of linear related covariates. The variance inflation factor (*VIF*) was used, as it is a common test of multicollinearity (Eq. 20).

$$VIF_j = \frac{1}{\left(1 - R^2{}_j \right)} \tag{20}$$

where R^2_j is the determination coefficient between the *j*th catchment descriptor and the remaining *m-1* catchment descriptors used in the regression model. Multicollinearity arises when *VIF* exceeds a value of five (Montgomery et al., 2012).

Griffis and Stedinger (2007) suggested the standard error of prediction (*SEP*) of the true flood quantiles as a useful tool to compare regression models (Eq. 21).

$$SEP = \sqrt{10^{\ln(10) AVP_{GLS}} - 1} \tag{21}$$

$$AVP_{GLS} = \sigma_\eta^2 + \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{x_i} \left(\mathbf{X}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{X} \right)^{-1} \boldsymbol{x_i^T} \tag{22}$$

where AVP_{GLS} is the average variance of prediction for a *GLS* regression model (Eq. 22) across all the *N* gauging stations used in the regression model and $\boldsymbol{x_i}$ is a row vector with the catchment descriptors used in the regression model at site *i*. Lower values of AVP_{GLS} and *SEP* suggest a more accurate regression model.

In addition, the improvement of a more complex GLS model when compared to a simpler OLS model should be quantified to decide when the more complex model can be accepted. For this purpose, the error variance ratio (*EVR*) was adopted to quantify the relationship between the magnitude of the average sampling variance and the magnitude of the model error variance (Eq. 23). Griffis and Stedinger (2007) argue that an OLS model should be used when *EVR* is greater than 0.2, as the sampling error is negligible compared to the modelling error .

$$EVR = \frac{tr\left(\hat{\boldsymbol{\Sigma}}\right)}{N\, \sigma_\eta^2 (m)} \tag{23}$$

where $tr\left(\hat{\boldsymbol{\Sigma}}\right)$ is the trace of the covariance matrix of the sampling errors and $\sigma_\eta^2(m)$ is the variance of modelling errors for the regression model with *m* catchment descriptors.

5 Results

This section is composed of two sub-sections. Firstly, the results about the implementation of the proposed GLS technique with a view to the existing recommendations given in Spain to estimate the frequency distribution is presented. Then, the application of the GLS regression model to the semi-arid Ebro River catchment is documented.

5.1 Specification of sampling and model error structures

5.1.1 Assessment of sampling variance based on Taylor series

The accuracy of the analytical expressions of the variance of the flood quantile estimates based on the Taylor series approximations (Appendix A) was assessed through a Monte Carlo experiment. A set of random synthetic series with varying sample lengths from 10 to 100 was generated from a GEV distribution. Five experiments were conducted, one for each homogeneous region in the Ebro River catchment. The regional growth curve was used in each homogeneous region, with L-mean equal to one and the regional values of L-CV and L-CS given in Table 1. A total of 10,000 random realisations were generated for each case.

The results of the Monte Carlo experiment (Fig. 2) show the Taylor series approximation fits the sampling variance estimated by Monte Carlo simulations almost perfectly for the three return periods in Regions 91, 92, 93 and 95, except some slight deviations for shorter record-lengths in Regions 92 and 94. In Region 93, the analytical expressions overestimate the sampling variance, mainly for the case of smaller record lengths. These deviations can be explained by the sharp curvature of the frequency distribution in this region, given by a low shape parameter (Table 1) that leads to large uncertainties in quantile estimates. All the gauging stations used in the Ebro River catchment exceed 20 years of record-length. As the main purpose of the variance-covariance estimates is to give relative weight to the different sites in the GLS model framework, the performances of the Taylor series approximations were considered adequate for the purpose of this study.

5.1.2 Correlation of sampling errors

The off-diagonal elements of $\mathbf{\Sigma}_{\boldsymbol{\varepsilon}}$ represent the sampling covariance between quantiles at different sites (Eq. 14). Evaluation of these non-diagonal elements requires a functional relationship between the correlation of the observed flood series and the corresponding correlation between the PWMs as expressed in Eq. (15). The bootstrap experiment described in Section 3.1 was executed on the set of 93 gauging stations selected in the Ebro River catchment. The procedure was repeated 1,000 times to estimate the correlation at different sites. Figure 3 shows the correlation between AMF series at each pair of sites against the correlation between PWMs. The results suggest a linear relationship for the case of the first two order PWMs. Consequently, it is concluded that the value of the power δ used in Eq. 15 is equal to one for all the combinations between the first two order PWM.

Next, the three coefficients ($\varphi_{\varepsilon,i}$) of the double exponential expression (Eq. 16) were estimated from the AMF data at the 93 observed sites in the Ebro River catchment using a simple least squares approach. Pairs of gauge stations with an overlapping record exceeding 30 years were selected to fit the model. The results are reported in Table 2 and the fitted model is shown in Fig. 4.

5.2 Development of a GLS regression model in the Ebro River catchment

Once the covariance matrices of the sampling were obtained, the parameters of a number of alternative regression models were estimated for the Ebro River catchment.

Firstly, the results of the GLS regression model were compared to the results of the existing OLS regression models developed by Jiménez-Álvarez et al. (2012). This comparison was conducted on the homogeneous regions 91 and 92. Following on, the results of applying the GLS regression model in these regions were improved using additional catchment descriptors. Finally, an exploratory analysis was carried out to obtain a GLS regression model of the entire Ebro River catchment, aiming to capture its great heterogeneities by a single model.

5.2.1 GLS regression model applied to the Region 91

The Region 91 has observed data from 34 gauging stations. Firstly, a GLS regression model was compared to the existing OLS model using the same catchment descriptors: A, P_T, H and t_{inf}. Adopting the GLS model leads to a decrease of 3-5% in the *SEP* (Table 3). However, the regression parameters are very similar. The benefits of the GLS model from the OLS model were quantified by the *EVR* measure. The three GLS models selected improve the existing OLS models, as *EVR* is positive for the three return periods. However, *EVR* is smaller than 20%, showing that the sampling error is negligible compared to the GLS modelling error. Consequently, the OLS could be preferred in this case, as the use of a more complex GLS model does not lead to a sufficient improvement from the more simple OLS model. Nevertheless, the developed GLS regression model is a powerful tool that takes into account the sampling variance of quantile estimations, the spatial correlation of quantiles between sites, the error of the regression model and the spatial correlation of residuals. Additional catchment descriptors were introduced in the analysis to improve the initial results of the GLS regression model.

In this region, climatic differences among catchments are almost negligible. On one hand, *IC* provides information about the soil storage capacity and the existence of aquifers. On the other, as P_T shows a small variability, I_e gives information about the initial moisture content before the flood event. In addition, PET was also included in the two-year return period regression model.

The results of the GLS regression model in Region 91 are shown in Table 4. Small modelling errors are achieved for the three return periods (Fig. 5). *SEP* was reduced to 15-20% from the 30-40% obtained by the OLS model. This is a significant improvement of the GLS regression model. Furthermore, the *EVR* results show values around 40%, which indicates that the sampling error variance cannot be neglected compared to the modelling error variance. Consequently, the improved GLS regression model is preferred to the existing OLS model in the Region 91 of the Ebro River catchment. In addition, no linear related covariates were found in this region, as *VIF* values are smaller than five in all the models (Table 5).

The evolution of spatial correlation of modelling errors is also shown in Fig. 6. The introduction of P_T into the regression model leads to a significant reduction of the spatial correlation between residuals, thus suggesting a more complete description of the processes controlling the between-sites variation in flood quantiles. Furthermore, the introduction of the last descriptor into the regression models leads to the lowest spatial correlation. It should be noted the inclusion of *H* was considered worthwhile, as its introduction removes almost completely the correlation of residuals with distance.

5.2.2 GLS regression model applied to the Region 92

The Region 92 has observed data from 25 gauging stations. Firstly, a GLS regression model was constructed and compared to the existing OLS model using the same set of catchment descriptors: A, P_T and A_{1500}. The results show that the GLS model leads to a reduction of 6-9% in the *SEP* for the return periods of 25 and 100 years when compared to the benchmark performance of the existing OLS model (Table 3). In addition, the GLS models for 25 and 100 years are preferred to the existing OLS models in terms of *EVR*, as sampling errors are more than 100% greater than modelling errors. However, for the case of a return period of two years, the OLS model is preferred, as the GLS model worsens the *SEP*.

In this region, the existing regression models were improved by introducing the initial abstraction, P_0, to explain differences in runoff production. Once the portion of precipitation transformed into runoff is considered in the regression model, the concavity index, *IC*, was introduced to account for the relationship between surface and subsurface processes in the catchment.

The results of the improved GLS regression model in the Region 92 are presented in Table 6. A reduction of modelling errors is obtained as additional catchment descriptors are included in the model (Fig. 5). The *SEP* values obtained using the GLS model halved those of the existing OLS regression models, obtaining values around 20-25%. In this region, adopting the GLS regression model leads to a significant improvement compared to the OLS model for return periods of 25 and 100 years. The GLS model is clearly preferred to the OLS for all combinations of catchment descriptors, even for the model with only the catchment area. Furthermore, regression models with six parameters lead to small *EVR* values, i.e., small modelling variances are achieved compared to the mean sampling variance. However, the GLS model for the two-year return period requires at least five descriptors to be preferred to the OLS model, as the regression model errors show a slight increase with respect to the rest of return periods. Nevertheless, the GLS regression model improves significantly the results of the OLS model and SEP is reduced to 25%. In addition, no linear related covariates were found in this region, as *VIF* values are smaller than five in all the models (Table 7).

The introduction of the initial abstraction, P_0, leads to an almost complete eradication of the cross correlation between model error residuals (Fig. 6), suggesting that this descriptor effectively explains the local differences between flood series not otherwise captured by the scale and climate descriptors.

5.2.3 A GLS regression model for the entire Ebro River catchment

A GLS regression model was fitted to the 93 gauging stations of the Ebro River catchment, with the aim of capturing its great heterogeneities by a single model. In this case, the aridity index (I_a) was found to explain much of the remaining spatial clustering of the regression residuals when the effects of both catchment area (A) and extreme rainfall (P_T) have been taken into account. In the central part of the catchment there exists a large area characterised as being semi-arid, while sub-humid climate areas can be found at the catchment boundaries, and small humid climate areas are observed in the Pyrenees. However, for the case of the two-year return period, P_m explains better the differences in the magnitude of floods.

The results of the GLS regression model for the entire Ebro River catchment are shown in Table 8. The GLS model gives *SEP* values around 30%, which means that the GLS model for the entire Ebro River catchment captures its spatial heterogeneities in the regional hydrology and leads to a good description of the flood processes. However, the results are slightly larger than those of the GLS models applied individually to the homogeneous regions 91 and 92. Consequently, a GLS model fitted to a given homogeneous region with a reduced number of sites leads to more accurate results, as it was expected. Nevertheless, the GLS model for the entire basin also improves the results of the OLS model. In addition, no linear related covariates were found in this region, as *VIF* values are smaller than five in all the models (Table 9).

6 Conclusions

A regional flood frequency hydrology analysis was carried out focusing on the spatial expansion of information by a regression model based on the generalised least squares technique, where inter-site correlations of both sampling and modelling errors were explicitly accounted for in the error structure of the regression model. The regression model was developed following the existing recommendations in Spain for estimating flood frequency curves: (i) a Generalised Extreme Value distribution fitted by the L-moment estimation method; (ii) at-site estimations of both location and scale parameters and regional estimation of the shape parameter. The covariance matrix of sampling errors was adapted to reflect these assumptions case.

The semi-arid Ebro River catchment located in Spain was selected as case study because previous studies encountered great heterogeneities of climate drivers, rainfall patterns and soil characteristics among sub-regions.

An exploratory analysis on catchment descriptors was conducted to explain differences in flood processes among catchments. The results showed that differences in T-year peak flow estimates between catchments were mainly explained by: (i) catchment area, which is the main driver of the flood magnitude; (ii) One day T-year design rainfall, which is the main driver of the differences in flood magnitude between catchments with similar catchment area; (iii) the concavity index, which characterises the split between fast surface runoff and slow subsurface flow based on the FDC; (iv) mean catchment slope, which explains differences due to the topography that have influence on the runoff velocity in hillslopes; (v) the extremity index, which in the Ebro River catchment gives information about the influence of antecedent precipitation on probable initial moisture content before the onset of flood events; (vi) potential evapotranspiration, which gives a better description of the probable initial moisture content; (vii) the precipitation depth absorbed by the soil before runoff begins, which explains differences caused by the hydrologic abstraction process.

Summarising, the use of these catchment descriptors in a generalised least squares regression model improved the results of the existing ordinary least squares regression models, in terms of variance of modelling errors and standard error of prediction. In addition, most of the regression models removed almost completely the spatial correlation of residuals, which suggests a satisfactory description of the flood processes that controls quantile variations between sites. Consequently, the generalised least squares regression model developed in this paper can be used for making more reliable predictions in ungauged catchments with the purpose of both designing hydraulic infrastructures at sites without observed information, and thus improving flood risk management.

Appendix A: Variance and covariance of the GEV parameters for the case of a constant shape parameter

In the case of a GEV distribution, the asymptotic variance of x_T for a constant shape parameter can be simplified by Eq. 11. In terms of L-moments, the remaining two parameters of the GEV distribution can be estimated by Equations A1 and A2.

$$\alpha = \frac{\lambda_2 \, k}{\left(1-2^{-k}\right)\left(\Gamma(1+k)\right)} = \lambda_2 \, K_1 \qquad \text{(A1)}$$

$$u = \lambda_1 - \frac{\alpha}{k}\left(1-\Gamma(1+k)\right) = \lambda_1 - \alpha \, K_2 = \lambda_1 - \lambda_2 \, K_1 K_2 \qquad \text{(A2)}$$

where λ_1 and λ_2 are the first two L-moments, Γ is the gamma function and K_1 and K_2 are constants for a given k parameter (Equations A3 and A4).

$$K_1 = \frac{k}{\left(1-2^{-k}\right)\Gamma(1+k)} \qquad \text{(A3)}$$

$$K_2 = \frac{1-\Gamma(1+k)}{k} \qquad \text{(A4)}$$

Therefore, the variance and covariance of u and α parameters, for a given k parameter in Eq. (11), are derived as follows in terms of L-moments:

$$\begin{aligned}\sigma^2(u) = var(u) &= \left(\frac{\partial u}{\partial \lambda_1}\right)^2 var(\lambda_1) + \left(\frac{\partial u}{\partial \lambda_2}\right)^2 var(\lambda_2) + 2\left(\frac{\partial u}{\partial \lambda_1}\right)\left(\frac{\partial u}{\partial \lambda_2}\right) cov(\lambda_1, \lambda_2) \\ &= var(\lambda_1) + \left(\frac{\partial u}{\partial \lambda_2}\right)^2 var(\lambda_2) + 2\left(\frac{\partial u}{\partial \lambda_2}\right) cov(\lambda_1, \lambda_2)\end{aligned} \tag{A5}$$

$$\begin{aligned}\sigma^2(\alpha) = \mathrm{var}(\alpha) &= \left(\frac{\partial \alpha}{\partial \lambda_1}\right)^2 \mathrm{var}(\lambda_1) + \left(\frac{\partial \alpha}{\partial \lambda_2}\right)^2 \mathrm{var}(\lambda_2) \\ &+ 2\left(\frac{\partial \alpha}{\partial \lambda_1}\right)\left(\frac{\partial \alpha}{\partial \lambda_2}\right) \mathrm{cov}(\lambda_1, \lambda_2) = \left(\frac{\partial \alpha}{\partial \lambda_2}\right)^2 \mathrm{var}(\lambda_2)\end{aligned} \tag{A6}$$

$$\begin{aligned}\mathrm{cov}(u, \alpha) &= \left(\frac{\partial u}{\partial \lambda_1}\right)\left(\frac{\partial \alpha}{\partial \lambda_1}\right) \mathrm{var}(\lambda_1) + \left(\frac{\partial u}{\partial \lambda_2}\right)\left(\frac{\partial \alpha}{\partial \lambda_2}\right) \mathrm{var}(\lambda_2) \\ &+ \left(\frac{\partial u}{\partial \lambda_1}\right)\left(\frac{\partial \alpha}{\partial \lambda_2}\right) \mathrm{cov}(\lambda_1, \lambda_2) + \left(\frac{\partial u}{\partial \lambda_2}\right)\left(\frac{\partial \alpha}{\partial \lambda_1}\right) \mathrm{cov}(\lambda_1, \lambda_2) \\ &= \left(\frac{\partial u}{\partial \lambda_2}\right)\left(\frac{\partial \alpha}{\partial \lambda_2}\right) \mathrm{var}(\lambda_2) + \left(\frac{\partial \alpha}{\partial \lambda_2}\right) \mathrm{cov}(\lambda_1, \lambda_2)\end{aligned} \tag{A7}$$

The variance and covariance of the first two L-moments can be obtained by Eq. A8 (Elamir and Seheult, 2004).

$$\mathrm{var}(\lambda) = \begin{bmatrix} \mathrm{var}(\lambda_1) & \mathrm{cov}(\lambda_1, \lambda_2) \\ \mathrm{cov}(\lambda_1, \lambda_2) & \mathrm{var}(\lambda_2) \end{bmatrix} = C \, \Theta \, C^T \tag{A8}$$

where:

$$\Theta = \begin{bmatrix} \mathrm{var}(b_0) & \mathrm{cov}(b_0, b_1) \\ \mathrm{cov}(b_0, b_1) & \mathrm{var}(b_1) \end{bmatrix} \tag{A9}$$

$$C = \begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} \tag{A10}$$

where $var(b_0)$, $var(b_1)$ and $cov(b_0,b_1)$ can be estimated as follows (Hosking et al., 1985):

$$\mathrm{var}(b_0) = \frac{\alpha^2}{n\,k^2}\left[\Gamma(1+2k) - \Gamma^2(1+k)\right] \tag{A11}$$

$$\mathrm{var}(b_1) = \frac{2^{-2k}\,\alpha^2}{n\,k^2}\left[\Gamma(1+2k)\,q(k) - \Gamma^2(1+k)\right] \quad \text{(A12)}$$

$$\mathrm{cov}(b_0, b_1) = \frac{\alpha^2}{2\,n\,k^2}\left[2^{-2k}\,\Gamma(1+2k) + \left(1-2^{1-k}\right)\Gamma^2(1+k)\right] \quad \text{(A13)}$$

where:

$$q(k) = 1 + \frac{2\,k^2}{\Gamma(1+2k)} \sum_{i=1}^{\infty}\left(\frac{\Gamma(2k+i)}{k+i}\,\frac{(-1/2)^i}{i!}\right) \quad \text{(A14)}$$

Appendix B: Covariance between GEV quantiles at different sites

The covariance between GEV quantiles at different sites with a constant shape parameter can be obtained by Eq. B1.

$$\begin{aligned}\mathrm{cov}(x_{T,i},\, x_{T,j}) = {} & \left(\frac{\partial x_{T,i}}{\partial u_i}\right)\left(\frac{\partial x_{T,j}}{\partial u_j}\right)\mathrm{cov}(u_i, u_j) + \left(\frac{\partial x_{T,i}}{\partial u_i}\right)\left(\frac{\partial x_{T,j}}{\partial \alpha_j}\right)\mathrm{cov}(u_i, \alpha_j) \\ & + \left(\frac{\partial x_{T,i}}{\partial \alpha_i}\right)\left(\frac{\partial x_{T,j}}{\partial u_j}\right)\mathrm{cov}(\alpha_i, u_j) + \left(\frac{\partial x_{T,i}}{\partial \alpha_i}\right)\left(\frac{\partial x_{T,j}}{\partial \alpha_j}\right)\mathrm{cov}(\alpha_i, \alpha_j)\end{aligned} \quad \text{(B1)}$$

As the partial derivative of x_T with respect to the location parameter equals one, Eq. B1 can be simplified to Eq. B2.

$$\begin{aligned}\mathrm{cov}(x_{T,i},\, x_{T,j}) = {} & \mathrm{cov}(u_i, u_j) + \left(\frac{\partial x_{T,j}}{\partial \alpha_j}\right)\mathrm{cov}(u_i, \alpha_j) + \left(\frac{\partial x_{T,i}}{\partial \alpha_i}\right)\mathrm{cov}(\alpha_i, u_j) \\ & + \left(\frac{\partial x_{T,i}}{\partial \alpha_i}\right)\left(\frac{\partial x_{T,j}}{\partial \alpha_j}\right)\mathrm{cov}(\alpha_i, \alpha_j)\end{aligned} \quad \text{(B2)}$$

The covariance between u and α parameters for a given k parameter can be obtaiend as follows, in terms of L-moments:

$$\begin{aligned}\mathrm{cov}(u_i, u_j) &= \mathrm{cov}(\lambda_{1,i} - \lambda_{2,i}\;K_{1,i}K_{2,i},\, \lambda_{1,j} - \lambda_{2,j}\;K_{1,j}K_{2,j}) \\ &= \mathrm{cov}(\lambda_{1,i},\, \lambda_{1,j}) + K_{1,i}K_{2,i}\;K_{1,j}K_{2,j}\;\mathrm{cov}(\lambda_{2,i},\, \lambda_{2,j}) \\ &\quad - K_{1,j}K_{2,j}\;\mathrm{cov}(\lambda_{1,i},\, \lambda_{2,j}) - K_{1,i}K_{2,i}\;\mathrm{cov}(\lambda_{2,i},\, \lambda_{1,j})\end{aligned} \quad \text{(B3)}$$

$$\mathrm{cov}(\alpha_i, \alpha_j) = \mathrm{cov}(\lambda_{2,i}K_{1,i},\, \lambda_{2,j}K_{1,j}) = K_{1,i}\;K_{1,j}\;\mathrm{cov}(\lambda_{2,i},\, \lambda_{2,j}) \quad \text{(B4)}$$

$$\begin{aligned}\operatorname{cov}(u_i,\alpha_j) &= \operatorname{cov}(\lambda_{1,i}-\lambda_{2,i}\;K_{1,i}K_{2,i},\,\lambda_{2,j}K_{1,j})\\ &= K_{1,j}\operatorname{cov}(\lambda_{1,i},\,\lambda_{2,j}) - K_{1,i}K_{2,i}\;K_{1,j}\operatorname{cov}(\lambda_{2,i},\,\lambda_{2,j})\end{aligned} \tag{B5}$$

where K_1 and K_2 are given in Equations A3 and A4. Covariance between L-moments can be obtained in terms of PWM as follows:

$$\operatorname{cov}(\lambda_{1,i},\,\lambda_{1,j}) = \operatorname{cov}(b_{0,i},\,b_{0,j}) \tag{B6}$$

$$\begin{aligned}\operatorname{cov}(\lambda_{2,i},\,\lambda_{2,j}) &= \operatorname{cov}(2b_{1,i}-b_{0,i},\,2b_{1,j}-b_{0,j})\\ &= 4\operatorname{cov}(b_{1,i},\,b_{1,j}) - 2\operatorname{cov}(b_{1,i},\,b_{0,j}) - 2\operatorname{cov}(b_{0,i},\,b_{1,j}) + \operatorname{cov}(b_{0,i},\,b_{0,j})\end{aligned} \tag{B7}$$

$$\operatorname{cov}(\lambda_{1,i},\,\lambda_{2,j}) = \operatorname{cov}(b_{0,i},\,2b_{1,j}-b_{0,j}) = 2\operatorname{cov}(b_{0,i},\,b_{1,j}) - \operatorname{cov}(b_{0,i},\,b_{0,j}) \tag{B8}$$

where covariance between PWM can be obtained by the following expressions:

$$\operatorname{cov}(b_{0,i},\,b_{0,j}) = \sqrt{\operatorname{var}(b_{0,i})}\,\sqrt{\operatorname{var}(b_{0,j})}\,\frac{m_{ij}}{n_i\;n_j}\,\rho_{b_{0i},b_{0j}} \tag{B9}$$

$$\operatorname{cov}(b_{1,i},\,b_{1,j}) = \sqrt{\operatorname{var}(b_{1,i})}\,\sqrt{\operatorname{var}(b_{1,j})}\,\frac{m_{ij}}{n_i\;n_j}\,\rho_{b_{1i},b_{1j}} \tag{B10}$$

$$\operatorname{cov}(b_{0,i},\,b_{1,j}) = \sqrt{\operatorname{var}(b_{0,i})}\,\sqrt{\operatorname{var}(b_{1,j})}\,\frac{m_{ij}}{n_i\;n_j}\,\rho_{b_{0i},b_{1j}} \tag{B11}$$

where m_{ij} is the number of overlapping years between sites i and j, n_i and n_j are record-lengths at sites i and j respectively, and $\rho_{b_{ri},b_{rj}}$ is the correlation between the rth order PWMs at sites i and j given by Eq. 15.

Acknowledgements

The authors thank the financial contribution made by the COST Office grant ES0901 "European procedures for flood frequency estimation (FloodFreq)" by means of the Short Term Scientific Mission COST-STSM-ECOST-STSM-ES0901-020712-018289.

References

Blöschl, G., Sivalapan, M., Wagener, T., Viglione, A. and Savenije, H.: Runoff prediction in ungauged basin. Synthesis across processes, places and scales. Cambridge University Press, New York, USA, 2013.

Bocchiola, D., De Michele, C. and Rosso, R.: Review of advances in index flood estimation. Hydrol. Earth Syst. Sci., 7 (3), 283-296, 2013.

Brath, A., Castellarin, A., Franchini, M. and Galeati, G.: Estimating the index flood using indirect methods. Hydrolog. Sci. J., 46 (3), 399-418, 2001.

Castellarin, A., Botter, G., Hughes, D. A., Liu, S., Ouarda, T. B. M. J., Parajka, J., Post, D. A., Sivapalan, M., Spence, C., Viglione, A. and Vogel, R. M.: Prediction of flow duration curves in ungauged basins. In: Runoff prediction in ungauged basins. Synthesis across processes, places and scales. G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione and H. Savenije (Eds.), Chapter 7, 135-162. Cambridge University Press, Cambridge, UK, 2013.

Castellarin, A., Burn, D. H. and Brath, A.: Homogeneity testing: how homogeneous do heterogeneous cross-correlated regions seem? J. Hydrol., 360, 67-76, 2008.

Castellarin, A., Vogel, R. M. and Matalas, N. C.: Multivariate probabilistic regional envelopes of extreme floods. J. Hydrol., 336, 376-390, 2007.

CEDEX.: Mapa de caudales máximos. Memoria Técnica (Map of maximum flows. Technical Report). Centre for Hydrographic Studies of CEDEX, Madrid, Spain, 2011.

Dawdy, D. R., Griffis, V. W. and Gupta, V. K.: Regional flood-frequency analysis: how we got here and where we are going. J. Hydrol. Eng., 17, 953-959, 2012.

Efron, B. and Tibshirani, R. J.: An introduction to the Bootstrap. Monographs on statistics and applied probability, 57. Chapman & Hall / CRC Press, Boca Raton, Florida, USA, 1993.

Elamir, E. A. H. and Seheult, A. H.: Exact variance structure of sample L-moments. J. Stat. Plan. Inf., 124, 337-359, 2004.

European Union: Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. Official Journal of the European Union, L 288, 27-34, 2007.

Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S. and Blöschl, G.: Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. J. Hydrol., 394, 101-117, 2010.

Griffis, V. W. and Stedinger, J. R.: The use of GLS regression in regional hydrologic analyses. J. Hydrol., 344, 82-95, 2007.

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

Hosking, J. R. M., Wallis, J. R. and Wood, E. F.: Estimation of the Generalized Extreme-Value distribution by the method of Probability-Weighted Moments. Technometrics, 27 (3), 251-261, 1985.

Jiménez-Álvarez, A., García-Montañés, C., Mediero, L., Inicio, L. and Garrote, J.: Map of maximum flows of intercommunity basins, Revista de Obras Públicas, 3533, 7-32, 2012.

Kjeldsen, T.R. and Jones, D.A. Predicting the index flood in ungauged UK catchments: on the link between data-transfer and spatial model structure. J. Hydrol., 387, 1-10, 2010.

Kjeldsen, T. R. and Jones, D. A.: An exploratory analysis of error components in hydrological regression modelling. Water Resour. Res., 45, W02407, 2009.

Kjeldsen, T. R. and Jones, D. A.: Prediction uncertainty in a median-based index flood method using L-moments. Water Resour. Res., 42, W07414, 2006.

Laio, F., Ganora, D., Claps, P. and Galeati, G.: Spatially smooth regional estimation of the flood frequency curve (with uncertainty). J. Hydrol., 408, 67-77, 2011.

Liu, R. Y. and Singh, K.: Efficiency and robustness in resampling. Ann. Stat., 20 (1), 370-384, 1992.

Lu, L. and Stedinger, J. R.: Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison. J Hydrol, 138, 247-267, 1992.

Madsen, H. and Rosbjerg, D.: The partial duration series method in regional index-flood modelling. Water Resour. Res., 33 (4), 771-781, 1997.

MARM.: Guía metodológica para el desarrollo del Sistema Nacional de Cartografía de Zonas Inundables (Handbook of flood mapping in Spain). Ministerio de Medio Ambiente y Medio Rural y Marino, Madrid, Spain, 2011.

Martins, E. S. and Stedinger, J. R.: Cross correlations among estimators of shape. Water Resour. Res., 38 (11), 1252, 2002.

Merz, R. and Blöschl, G.: Flood frequency hydrology. 1. Temporal, spatial and causal expansion of information, Water Resour. Res., 44, W08423, 2008.

Montgomery, D. C., Peck, E. A. and Vining, G. G.: Introduction to linear regression analysis. John Wiley & Sons, Hoboken, New Jersey, USA, 2012.

Rao, A. R. and Hamed, K. H.: Flood frequency analysis. CRC Press, Boca Raton, Florida, USA, 2000.

Robson, A. and Reed, D.: Flood Estimation Handbook. 3: Statistical procedures for flood frequency estimation. Centre for Ecology & Hydrology, Wallingford, UK, 1999.

Rosbjerg, D., Blöschl, G., Burn, D. H., Castellarin, A., Croke, B., Di Baldassare, G., Iacobellis, V., Kjeldsen, T. R., Kuczera, G., Merz, R., Montanari, A., Morris, D., Ouarda, T. B. M. J., Ren, L., Rogger, M., Salinas, J. L., Toth, E. and Viglione, A. Prediction of floods in ungauged basins. In: Runoff prediction in ungauged basins. Synthesis across processes, places and scales. G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione and H. Savenije (Eds.), Chapter 9, 189-226. Cambridge University Press, Cambridge, UK, 2013.

Sauquet, E. and Catalogne, C.: Comparison of catchment grouping methods for flow duration curve estimation at ungauged sites in France. Hydrol. Earth Syst. Sci., 15, 2421-2435, 2011.

Stedinger, J. R.: Estimating a regional flood frequency distribution. Water Resour. Res., 19 (2), 503-510, 1983.

Stedinger, J. R. and Tasker, G. D.: Regional hydrologic analysis 1. Ordinary, weighted and generalized least squares compared, Water Resour. Res., 21 (9), 1421-1432, 1985.

Viglione, A., Parajka, J., Rogger, M., Salinas, J.L., Laaha, G., Sivalapan, M. and Blöschl, G.: Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. Hydrol. Earth Syst. Sci., 17, 2263-2279, 2013.

Figure 1. Location of the Ebro River catchment. Solid points show location of the gauging stations used in the study.

Figure 2. Comparison between sampling variance estimated by Monte Carlo simulations and the analytical solution estimated by Taylor series approximation. Regions by rows: a) Region 91; b) Region 92; c) Region 93; d) Region 94; e) Region 95. Return period by columns: 1) two years; 2) 25 years; 3) 100 years

Figure 3. Correlation between AMF series and PWMs. a) Between first-order PWMs (b_0); b) between second-order PWMs (b_1); c) between first-order and second-order PWMs (b_0 and b_1)

Figure 4. Correlation between AMF series and distance between catchment centroids for the 93 flood series from the Ebro catchment. Solid line shows the double exponential function fitted using the least square technique.

Figure 5. Evolution of the variance of modelling errors, σ^2_η. Regions by rows: a) Region 91; b) Region 92; c) Entire Ebro River catchment.

Figure 6. Evolution of correlation of residuals with distance between sites in km^2. Regions by rows: a) Region 91; b) Region 92; c) Entire Ebro River catchment. Return period by column: 1) two years; 2) 25 years; 3) 100 years.

Figure 1
Click here to download high resolution image

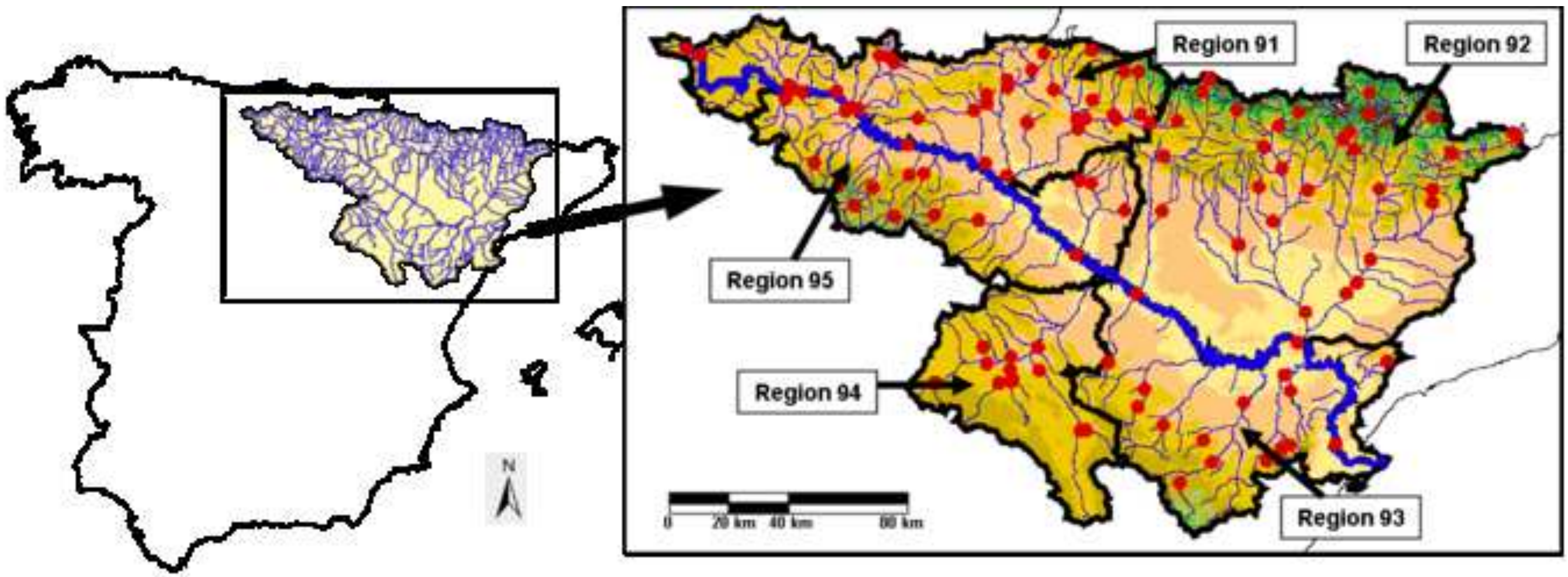

Figure 2
Click here to download high resolution image

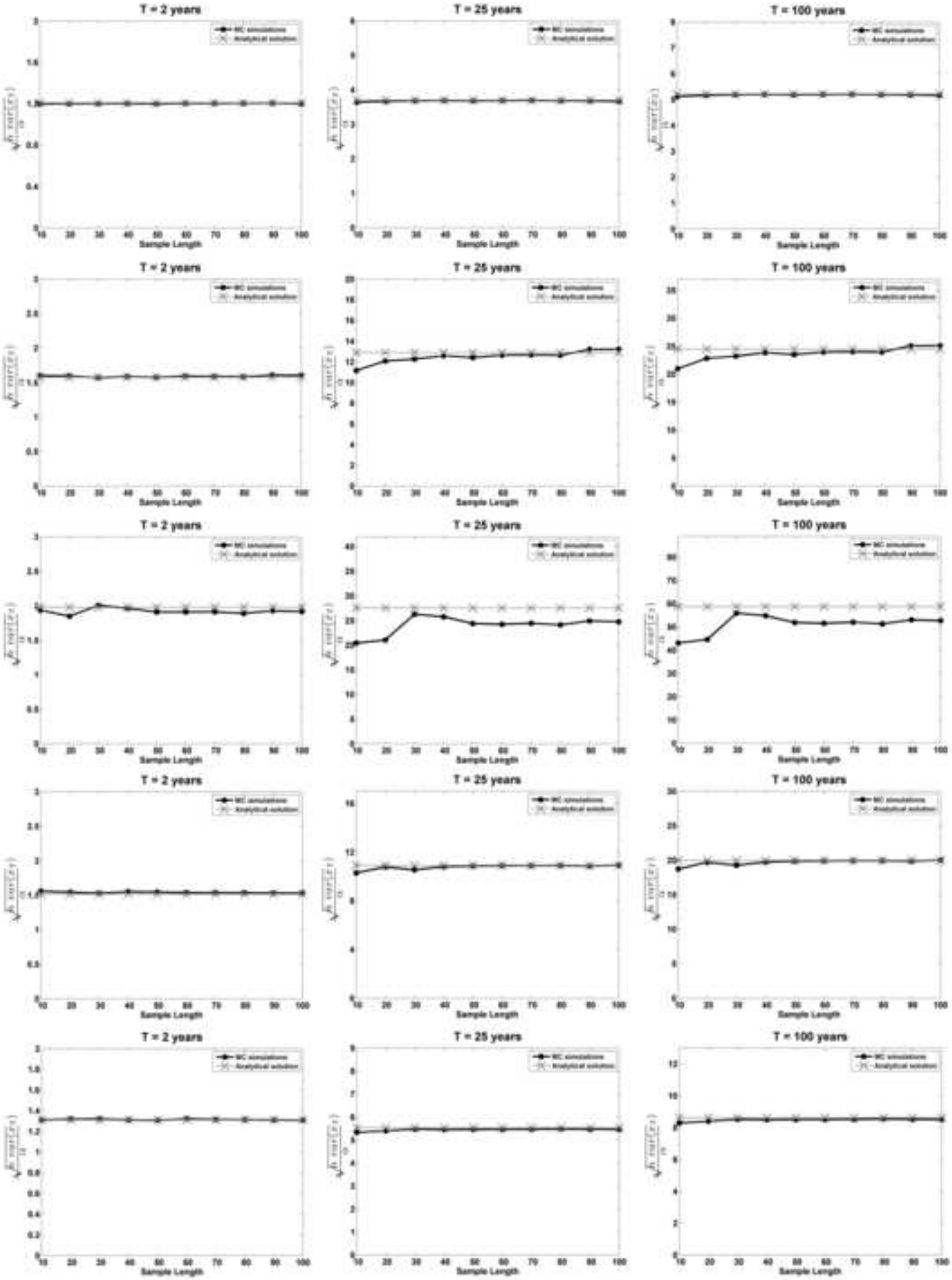

Figure 3
Click here to download high resolution image

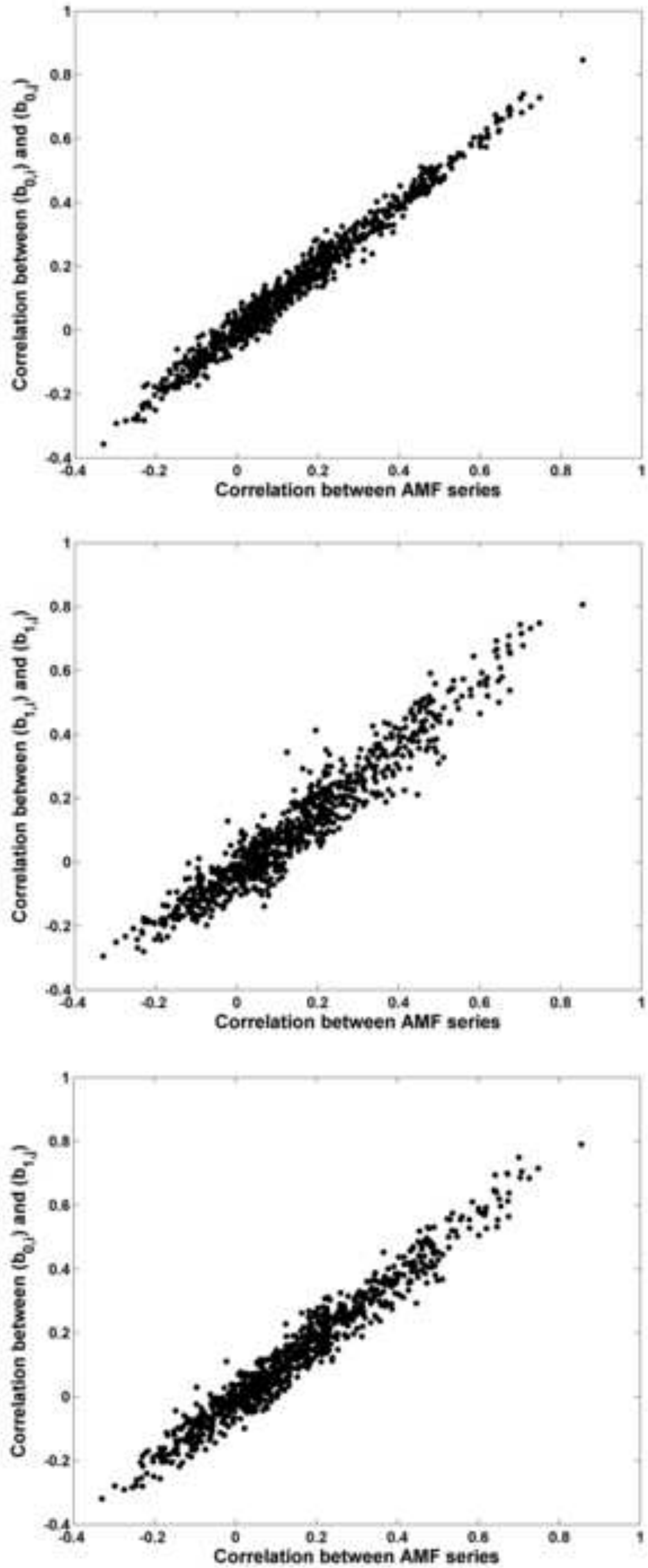

Figure 4
Click here to download high resolution image

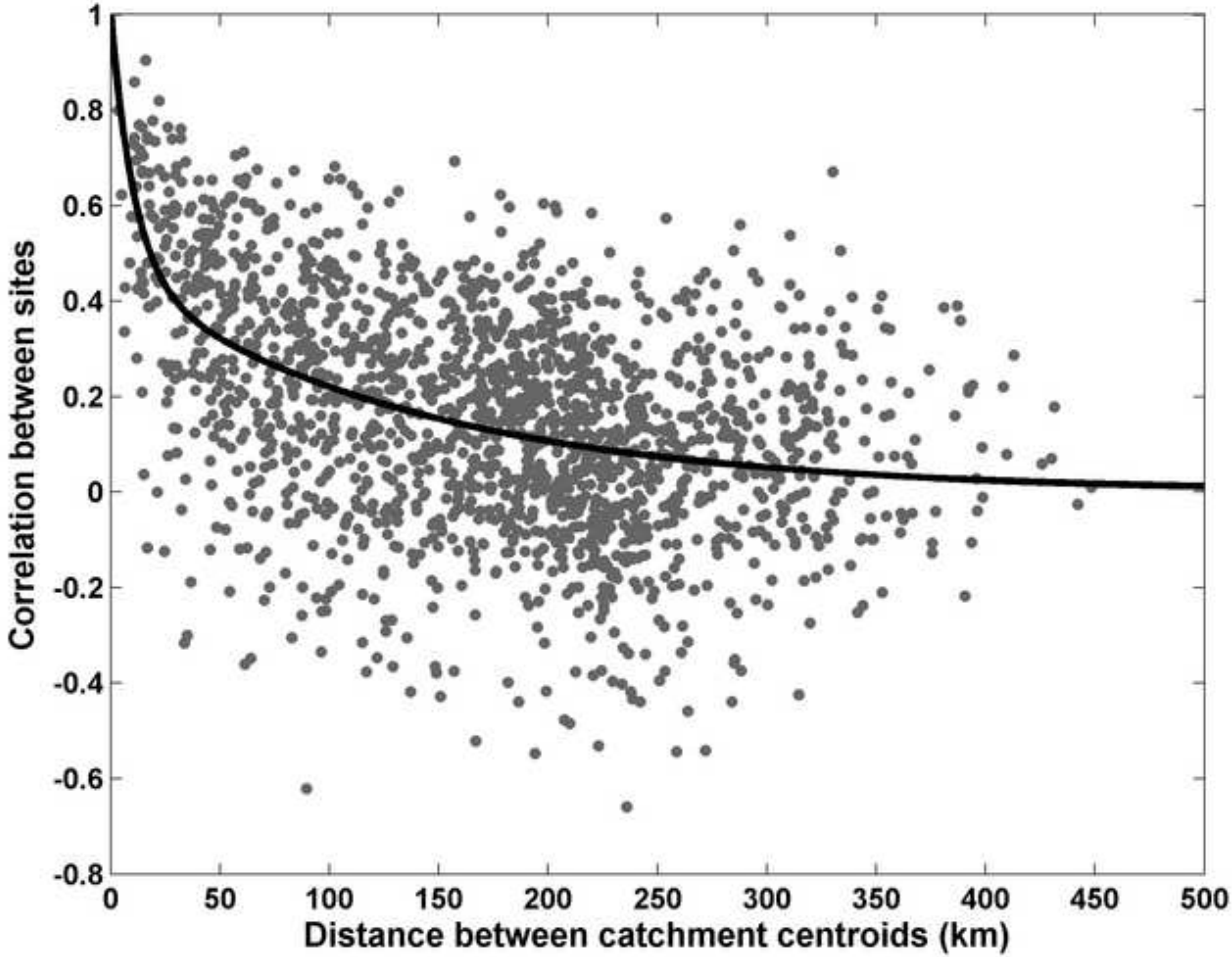

Figure 5
Click here to download high resolution image

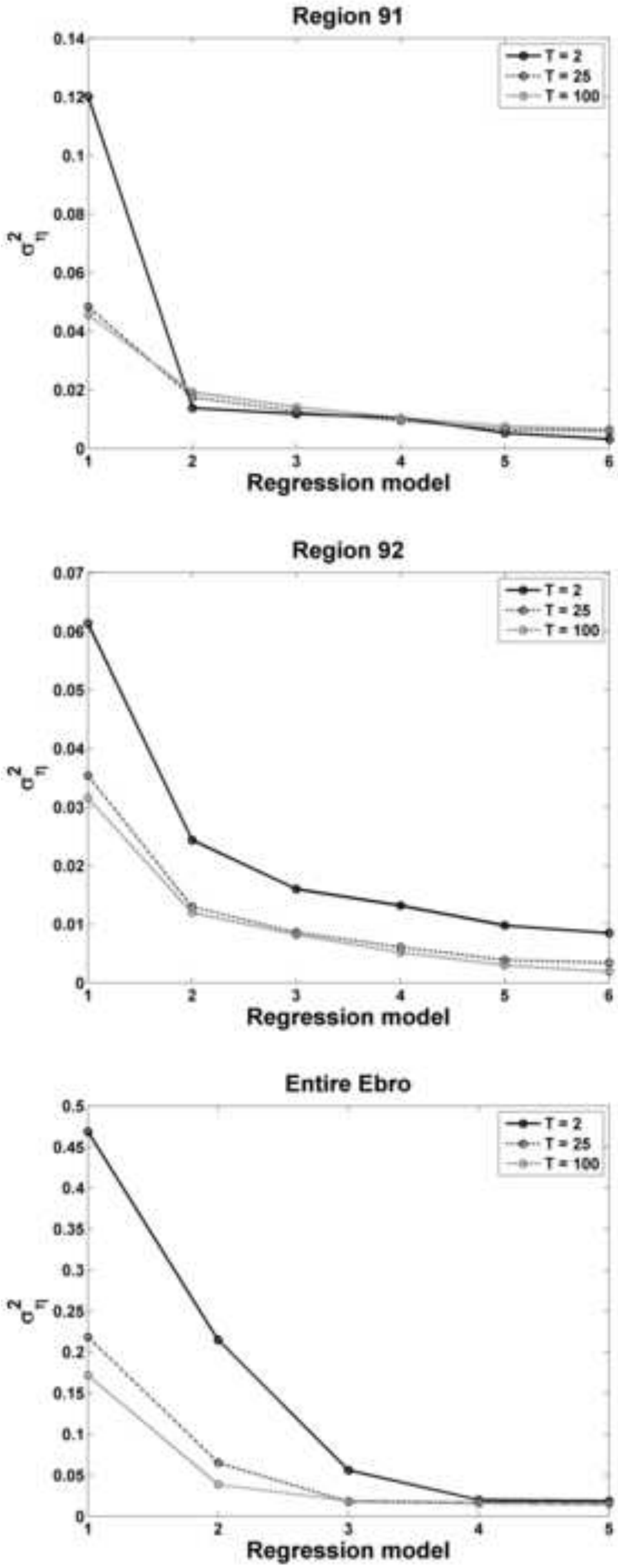

Figure 6
Click here to download high resolution image

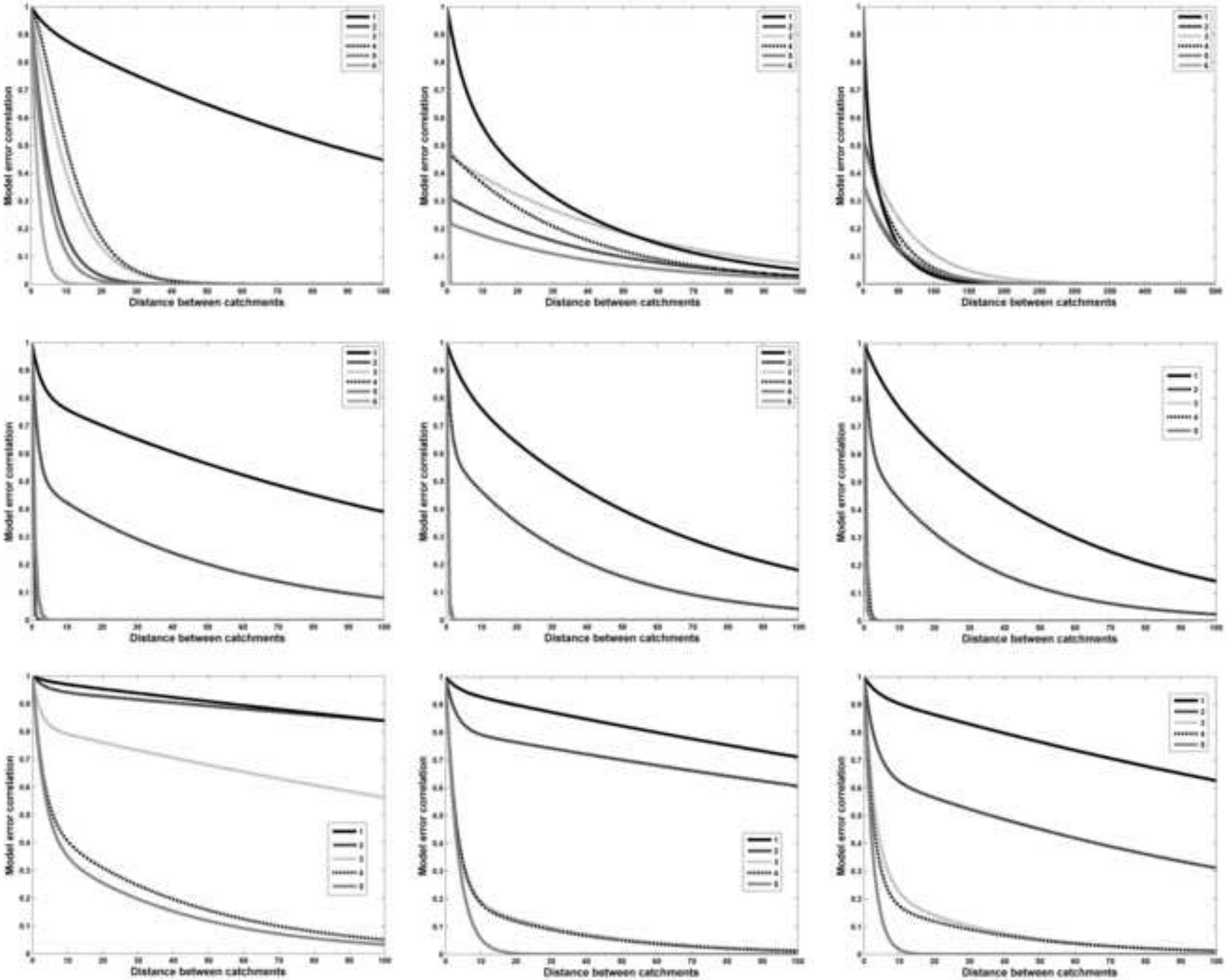

Table 1. Regional values of the L-CS and L-coefficient of variation (L-CV), number of gauging stations, N, and regional shape parameter of the GEV distribution, k, in the five homogeneous regions of the Ebro River catchment.

Region	L-CS	L-CV	N	k
91	0.194	0.257	34	-0.037
92	0.410	0.343	25	-0.343
93	0.489	0.569	10	-0.444
94	0.386	0.497	12	-0.312
95	0.272	0.357	12	-0.154

Table 2. Coefficients ($\varphi_{\varepsilon,i}$) and results of the root mean squared error (*RMSE*) and coefficient of determination (R^2) for the double exponential function (Eq. 14) fitted to the observed data.

$\varphi_{\varepsilon,1}$	$\varphi_{\varepsilon,2}$	$\varphi_{\varepsilon,3}$	*RMSE*	R^2
0.5406	0.0952	0.0073	0.210	0.370

Table 3
Click here to download Table: Table_3.docx

Table 3. Comparison between OLS and GLS regression models for return periods, T, of two, 25 and 100 years.

	Region 91					
	OLS			GLS		
Coefficient	T = 2	T = 25	T = 100	T = 2	T = 25	T = 100
Intercept (θ_0)	-4.8949	-5.5541	-5.7549	-5.0833	-5.2676	-5.7015
$\log_{10}(A)$	0.7753	0.7733	0.7738	0.7822	0.7743	0.7732
$\log_{10}(P_T)$	2.9029	2.6320	2.5530	3.0057	2.7855	2.5547
$\log_{10}(H)$	0.0296	0.2758	0.3441	0.0337	0.0951	0.3311
$\log_{10}(t_{inf})$	-0.0480	-0.0179	-0.0200	-0.0655	-0.0695	-0.0382
σ^2_η	0.0196	0.0247	0.0268	0.0136	0.0168	0.0176
SEP (%)	33.12	37.40	39.05	29.81	34.15	34.08
EVR	-	-	-	0.100	0.135	0.150
	Region 92					
	OLS			GLS		
Coefficient	T = 2	T = 25	T = 100	T = 2	T = 25	T = 100
Intercept (θ_0)	-4.2161	-5.7193	-6.0179	-2.9825	-4.7874	-4.9560
$\log_{10}(A)$	0.7025	0.6616	0.6445	0.5987	0.5877	0.5667
$\log_{10}(P_T)$	2.4689	3.1736	3.2754	1.9132	2.8129	2.8796
$\log_{10}(A_{1500})$	0.0555	0.0525	0.0576	0.0888	0.0790	0.0861
σ^2_η	0.0258	0.0253	0.0270	0.0229	0.0117	0.0100
SEP (%)	38.34	37.88	39.23	42.60	31.90	30.26
EVR	-	-	-	0.113	1.194	1.802

Table 4. Parameters and statistics of the GLS regression models fitted in the Region 91 for return periods, T, of two, 25 and 100 years.

T	Model	θ_0	log(A)	log(P_T)	log(IC)^9	log(S)$^{-1}$	I_e	log(PET)	σ^2_η	AVP_{GLS}	*SEP (%)*	*MLE*	*EVR*
2	1	0.1178	0.7845	-	-	-	-	-	0.1202	0.1751	123.7	-81.1	0.011
	2	-5.3561	0.7747	3.1911	-	-	-	-	0.0138	0.0153	29.12	-107.3	0.098
	3	-5.5245	0.7760	3.2620	-9.8808	-	-	-	0.0118	0.0137	27.48	-116.0	0.115
	4	-5.7334	0.8185	3.3343	-10.7554	-0.0246	-	-	0.0105	0.0126	26.28	-122.0	0.129
	5	-5.1518	0.8234	2.7761	-10.2848	-0.0288	0.0169	-	0.0053	0.0066	18.92	-135.9	0.254
	6	-10.468	0.7933	2.8021	-10.1947	-0.0275	0.0271	1.8255	0.0032	0.0042	14.97	-148.2	0.427
25		θ_0	log(A)	log(P_T)	log(IC)^9	log(S)$^{-1}$	I_e	log(H)					
	1	0.5563	0.7433	-	-	-	-	-	0.0484	0.0565	59.08	-77.24	0.047
	2	-5.4299	0.7648	2.9819	-	-	-	-	0.0175	0.0204	33.81	-99.36	0.129
	3	-5.2504	0.7598	2.8763	-12.0162	-	-	-	0.0126	0.0158	29.58	-112.6	0.179
	4	-5.3600	0.7939	2.8953	-12.6452	-0.0309	-	-	0.0095	0.0120	25.68	-119.4	0.238
	5	-4.9145	0.8075	2.5148	-13.4199	-0.0342	0.0125	-	0.0065	0.0084	21.39	-126.2	0.348
	6	-5.7168	0.8177	2.5009	-14.2157	-0.0335	0.0103	0.2875	0.0059	0.0076	20.34	-128.12	0.383
100		θ_0	log(A)	log(P_T)	log(IC)^9	log(S)$^{-1}$	I_e	log(H)					
	1	0.6834	0.7375	-	-	-	-	-	0.0456	0.0522	56.47	-76.38	0.058
	2	-5.5355	0.7617	2.9404	-	-	-	-	0.0193	0.0228	35.86	-96.58	0.137
	3	-5.2784	0.7564	2.8039	-12.5680	-	-	-	0.0141	0.0181	31.73	-110.18	0.187
	4	-5.3863	0.7916	2.8192	-13.1645	-0.0322	-	-	0.0105	0.0136	27.31	-117.09	0.253
	5	-4.9152	0.8034	2.4591	-14.0366	-0.0349	0.0114	-	0.0076	0.0102	23.58	-122.29	0.346
	6	-6.0353	0.8188	2.4423	-15.0828	-0.0346	0.0088	0.3956	0.0065	0.0084	21.36	-124.54	0.409

Table 5. Results of the *VIF* coefficient for the GLS regression models fitted in the Region 91 for return periods, T, of two, 25 and 100 years.

T	Model	log(A)	log(P_T)	log(IC)^9	log(S)$^{-1}$	I_e	log(PET)
2	1	-	-	-	-	-	-
	2	1.0060	1.0060	-	-	-	-
	3	1.0085	1.0066	1.0033	-	-	-
	4	1.1735	1.0102	1.0211	1.1878	-	-
	5	1.2226	1.1117	1.0262	1.1922	1.1625	-
	6	1.3373	1.1140	1.0337	1.1971	2.2382	2.3499
25		log(A)	log(P_T)	log(IC)^9	log(S)$^{-1}$	I_e	log(H)
	1	-	-	-	-	-	-
	2	1.0106	1.0106	-	-	-	-
	3	1.0133	1.0106	1.0027	-	-	-
	4	1.1816	1.0169	1.0202	1.1909	-	-
	5	1.2282	1.0771	1.0262	1.1950	1.1190	-
	6	1.2379	1.0812	1.1070	1.2227	1.4795	1.5862
100		log(A)	log(P_T)	log(IC)^9	log(S)$^{-1}$	I_e	log(H)
	1	-	-	-	-	-	-
	2	1.0109	1.0109	-	-	-	-
	3	1.0137	1.0111	1.0029	-	-	-
	4	1.1835	1.0195	1.0201	1.1934	-	-
	5	1.2303	1.0685	1.0265	1.1974	1.1072	-
	6	1.2409	1.0686	1.1066	1.2237	1.4835	1.5802

Table 6

Table 6. Parameters and statistics of the GLS regression models fitted in the Region 92 for return periods, T, of two, 25 and 100 years.

T	Model	θ_0	log(A)	log(P_T)	log(P_0)	log(S)$^{-4}$	log(*IC*)$^{-2}$	log(PET)	σ^2_η	AVP_{GLS}	*SEP (%)*	*MLE*	*EVR*
2	1	0.2124	0.6948	-	-	-	-	-	0.0614	0.0910	78.74	-63.31	0.042
	2	-3.8397	0.6907	2.3191	-	-	-	-	0.0244	0.0313	42.50	-69.96	0.106
	3	-3.4487	0.7295	2.6532	-0.8035	-	-	-	0.0160	0.0190	32.55	-75.24	0.162
	4	-3.2008	0.7334	2.5334	-0.8262	-0.0042	-	-	0.0132	0.0164	30.15	-79.37	0.196
	5	-4.2909	0.7631	3.0036	-0.6297	-0.0052	20.7162	-	0.0098	0.0128	26.48	-85.19	0.265
	6	-2.3569	0.7795	2.7762	-0.3597	-0.0042	25.6163	-0.6854	0.0085	0.0116	25.22	-87.92	0.305
25		θ_0	log(A)	log(P_T)	log(P_0)	log(S)$^{-2}$	log(*IC*)$^{-2}$	log(H)					
	1	0.7975	0.6658	-	-	-	-	-	0.0354	0.0482	53.97	-59.05	0.395
	2	-5.5499	0.6670	3.1263	-	-	-	-	0.0130	0.0181	31.72	-68.22	1.071
	3	-5.2879	0.7048	3.4322	-0.7303	-	-	-	0.0086	0.0124	26.02	-72.50	1.616
	4	-4.9948	0.7058	3.3175	-0.7436	-0.0342	-	-	0.0061	0.0101	23.42	-75.76	2.296
	5	-5.2986	0.7235	3.5572	-0.8263	-0.0467	-0.0165	-	0.0039	0.0082	21.05	-78.72	3.566
	6	-6.0505	0.7380	3.5820	-0.7889	-0.0375	-0.0213	0.1997	0.0034	0.0081	20.99	-79.15	4.149
100		θ_0	log(A)	log(P_T)	log(P_0)	log(S)$^{-2}$	log(*IC*)$^{-2}$	log(H)					
	1	1.0508	0.6568	-	-	-	-	-	0.0316	0.0427	50.40	-57.02	0.571
	2	-5.7504	0.6513	3.1881	-	-	-	-	0.0120	0.0170	30.69	-65.60	1.506
	3	-5.4909	0.6891	3.4835	-0.7355	-	-	-	0.0082	0.0124	26.05	-69.80	2.208
	4	-5.1896	0.6899	3.3698	-0.7445	-0.0362	-	-	0.0052	0.0096	22.90	-73.13	3.488
	5	-5.6423	0.7114	3.6705	-0.8368	-0.0491	-0.0171	-	0.0030	0.0079	20.70	-75.77	5.973
	6	-6.7731	0.7324	3.7300	-0.7801	-0.0363	-0.0243	0.2851	0.0019	0.0073	19.86	-76.53	9.279

Table 7

Click here to download Table: Table_7.docx

Table 7. Results of the *VIF* coefficient for the GLS regression models fitted in the Region 92 for return periods, T, of two, 25 and 100 years.

T	Model	log(A)	log(P_T)	log(P_0)	log(S)$^{-4}$	log(*IC*)$^{-2}$	log(PET)
2	1	-	-	-	-	-	-
	2	1.0135	1.0135	-	-	-	-
	3	1.2079	1.0142	1.1977	-	-	-
	4	1.2106	1.0327	1.2008	1.0225	-	-
	5	1.2780	1.2812	1.3827	1.0685	1.6025	-
	6	1.3348	1.4027	2.1386	1.2074	1.8078	2.3525
25		log(A)	log(P_T)	log(P_0)	log(S)$^{-2}$	log(*IC*)$^{-2}$	log(H)
	1	-	-	-	-	-	-
	2	1.0087	1.0087	-	-	-	-
	3	1.2122	1.0134	1.2025	-	-	-
	4	1.2168	1.0273	1.2068	1.0209	-	-
	5	1.2461	1.0532	1.2453	1.1967	1.2383	-
	6	1.5315	1.0582	1.2676	1.9033	1.9632	3.4105
100		log(A)	log(P_T)	log(P_0)	log(S)$^{-2}$	log(*IC*)$^{-2}$	log(H)
	1	-	-	-	-	-	-
	2	1.0052	1.0052	-	-	-	-
	3	1.2109	1.0159	1.2096	-	-	-
	4	1.2160	1.0239	1.2137	1.0151	-	-
	5	1.2455	1.0524	1.2552	1.1945	1.2414	-
	6	1.5329	1.0554	1.2776	1.9039	1.9690	3.4037

Table 8

Click here to download Table: Table_8.docx

Table 8. GLS regression models in the entire Ebro River catchment for return periods, T, of two, 25 and 100 years.

T	Model	θ_0	log(A)	log(P_T)	log(P_m)$^{-5}$	log(H)	log(PET)	σ^2_η	AVP_{GLS}	*SEP (%)*	*MLE*	*EVR*
2	1	-0.0472	0.7429	-	-	-	-	0.4685	0.7817	788.0	-212.5	0.008
	2	-5.1623	0.7519	2.9652	-	-	-	0.2147	0.3676	245.4	-243.1	0.018
	3	-2.7361	0.7607	1.8430	-122.95	-	-	0.0561	0.0801	72.73	-258.8	0.069
	4	0.9896	0.7305	1.3611	-211.59	-0.7869	-	0.0199	0.0229	35.90	-267.9	0.196
	5	3.1365	0.7308	1.4349	-195.29	-0.9033	-0.7030	0.0184	0.0211	34.40	-271.0	0.212
25		θ_0	log(A)	log(P_T)	(I_a)$^{-5}$	log(IC)$^{-2}$	log(PET)					
	1	0.5265	0.7012	-	-	-	-	0.2182	0.3287	217.1	-196.1	0.060
	2	-5.3890	0.7135	2.9550	-	-	-	0.0653	0.0962	81.57	-230.2	0.199
	3	-4.5950	0.7181	2.6027	-0.0318	-	-	0.0176	0.0195	33.04	-240.1	0.742
	4	-4.2863	0.7097	2.4894	-0.0320	-0.0119	-	0.0159	0.0180	31.64	-247.5	0.821
	5	-1.6877	0.7175	2.4654	-0.0264	-0.0132	-0.9109	0.0148	0.0167	30.46	-253.5	0.881
100		θ_0	log(A)	log(P_T)	(I_a)$^{-5}$	log($FDCS_1$)$^{-1}$	log(H)					
	1	0.7484	0.6881	-	-	-	-	0.1718	0.2434	162.3	-188.9	0.089
	2	-5.4774	0.6993	2.9552	-	-	-	0.0386	0.0478	53.72	-222.5	0.397
	3	-4.7826	0.7055	2.6522	-0.0255	-	-	0.0188	0.0209	34.24	-233.3	0.816
	4	-4.4789	0.7103	2.6070	-0.0280	-0.3016	-	0.0167	0.0189	32.49	-240.4	0.918
	5	-5.4395	0.7250	2.6234	-0.0285	-0.3731	0.3117	0.0149	0.0168	30.52	-245.7	1.031

Table 9

Click here to download Table: Table_9.docx

Table 9. Results of the *VIF* coefficient for the GLS regression models fitted in the entire Ebro River catchment for return periods, T, of two, 25 and 100 years.

T	Model	log(A)	log(P_T)	log(P_m)$^{-5}$	log(H)	log(PET)
2	1	-	-	-	-	-
	2	1.0619	1.0619	-	-	-
	3	1.0782	1.8467	1.8555	-	-
	4	1.0976	1.9108	2.1729	1.2312	-
	5	1.1006	2.1182	3.8694	1.2924	2.4152
25		log(A)	log(P_T)	(I_a)$^{-5}$	log(*IC*)$^{-2}$	log(PET)
	1	-	-	-	-	-
	2	1.0636	1.0636	-	-	-
	3	1.0886	1.2022	1.1924	-	-
	4	1.0987	1.2855	1.1984	1.0719	-
	5	1.1075	1.2857	1.6745	1.0938	1.5301
100		log(A)	log(P_T)	(I_a)$^{-5}$	log($FDCS_1$)$^{-1}$	log(H)
	1	-	-	-	-	-
	2	1.0627	1.0627	-	-	-
	3	1.0899	1.1756	1.1671	-	-
	4	1.1020	1.2281	1.3182	1.1464	-
	5	1.1483	1.2313	1.3185	1.2396	1.1411