Hydrological Sciences Journal

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/thsj20

A comparison between generalized least squares regression and top-kriging for homogeneous cross-correlated flood regions

Simone Persiano, Jose Luis Salinas, Jery Russell Stedinger, William H. Farmer, David Lun, Alberto Viglione, Günter Blöschl & Attilio Castellarin

HYDROLOGICAL SCIENCES JOURNAL
2021, VOL. 66, NO. 4, 565–579
https://doi.org/10.1080/02626667.2021.1879389

A comparison between generalized least squares regression and top-kriging for homogeneous cross-correlated flood regions

Simone Persiano [a], Jose Luis Salinas [b], Jery Russell Stedinger [c], William H. Farmer [d], David Lun [b], Alberto Viglione [b,e], Günter Blöschl [b] and Attilio Castellarin [a]

[a]Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, Bologna, Italy; [b]Institute of Hydraulic Engineering and Water Resources Managament, Technische Universität Wien, Vienna, Austria; [c]School of Civil and Environmental Engineering, Cornell University, Ithaca, New York, USA; [d]Water Resources Mission Area, United States Geological Survey, Denver, Colorado, USA; [e]Department of Environment, Land and Infrastructure Engineering, Politecnico Di Torino, Turin, Italy

ABSTRACT
Spatial cross-correlation among flood sequences impacts the accuracy of regional predictors. Our study investigates this impact for two regionalization procedures, generalized least squares (GLS) regression and top-kriging (TK), which deal with cross-correlation in two fundamentally different ways and therefore might be associated with different accuracy and uncertainty of predicted flood quantiles. We perform a Monte Carlo experiment based on a dataset of annual maximum flood series for 20 catchments in a hydrologically homogeneous region. Based on a log-Pearson type III parent distribution, we generate 3000 realizations of the region with different degrees of cross-correlation. For each realization, GLS and TK are applied in leave-one-out cross-validation to predict at-site flood quantiles. Our study shows that (a) TK outperforms GLS when catchment area is the only catchment descriptor used for predicting "true" population (theoretical) flood quantiles, regardless of the level of cross-correlation, and (b) GLS and TK perform similarly when multiple catchment descriptors are used.

ARTICLE HISTORY
Received 9 September 2020
Accepted 16 December 2020

EDITOR
S. Archfield

ASSOCIATE EDITOR
Not assigned

KEYWORDS
prediction in ungauged basins (PUB); geostatistics; cross-correlation; theoretical flood quantiles

1 Introduction: state of the art and aims of the study

1.1 Prediction of flood quantiles in ungauged basins

The accurate estimation of flood quantiles (i.e. flood discharge associated with a given non-exceedance probability, usually expressed in terms of return period T– although the latter concept is often prone to misuses, especially in the presence of nonstationarity, see e.g. Serinaldi 2015) is of paramount importance in many practical engineering applications. Because gauging stations are heterogeneously and sparsely distributed in space, one of the most common tasks for hydrological engineers is to produce an accurate estimate of the design flood at ungauged or scarcely gauged river cross-sections (see the Predictions in Ungauged Basins (PUB) initiative of the International Association of Hydrological Sciences (IAHS) for the decade 2003–2012; (Sivapalan *et al.* 2003, Blöschl *et al.* 2013)). The term "ungauged" generally refers to a river cross-section where no streamflow data are available. A special case is when the ungauged river cross-section has no information available upstream or downstream within the given catchment (that is absent of nested gauged catchments, i.e. no gauges are within the watershed of other gauges); in this article, we refer to this latter case as "fully ungauged" conditions.

The estimation of flood quantiles at ungauged sites is often addressed by means of regional flood frequency analysis, or statistical regionalization, which consists of transferring the hydrological information collected at gauged sites that are supposed to be hydrologically similar to the ungauged (or scarcely gauged) target site (Hosking and Wallis 1993, 1997). In this process, it is important to consider that, being based on gathering data from a finite number of observations (i.e. events) collected at a finite number of sites (i.e. limited record size), the obtained empirical estimates of flood quantiles can differ considerably from the theoretical ones, which could be theoretically known based upon the perfect knowledge of the process (i.e. infinite observations). In regional flood frequency analysis, this difference is usually amplified by the presence of the intersite dependence of streamflow series (i.e. spatial cross-correlation), due to the temporal overlap between time series of observed streamflow at streamgauges that are close in space (i.e. concurrent flows are recorded at different streamgauges). For these reasons, the distinction between empirical and theoretical flood quantiles is crucial in regional flood frequency analysis: while empirical quantiles can be seen as estimates of what the T-year flood would have been if the target site was gauged, theoretical quantiles are the "true" population flood quantiles. The implications related to this difference are important: as observed records are realizations of a random process that will never be repeated, the knowledge of the "true" distribution (i.e. theoretical flood quantiles) from which observations arise would be fundamental for making general evaluations capable of accounting for the main properties of the (true) natural process.

CONTACT Simone Persiano simone.persiano@unibo.it Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, Bologna, Italy
Supplemental data for this article can be accessed here.

1.2 Statistical regionalization

The literature reports several regional statistical methods, which assume the hydrological variable of interest to be a random variable: they can be classified as regression-based methods, index-flood methods and geostatistical methods (see e.g. Blöschl *et al.* 2013). These methods can differ significantly in the way they identify groups of hydrologically similar catchments and account for the limited size of record and the presence of cross-correlation. Regression-based methods relate the hydrological variable of interest to observable catchment and climate characteristics; they can require the preliminary identification of a homogeneous region and may or may not account for the presence of spatial cross-correlation and of unequal record lengths from site to site (see e.g. Thomas and Benson 1970, Tasker 1980, Stedinger and Tasker 1985, Tasker and Stedinger 1989) (see also Durocher *et al.* 2019, for an application of nonparametric regression methods with spatial components). Index-flood methods (see e.g. Dalrymple 1960, Hosking and Wallis 1997) estimate the hydrological design variable of interest as the product between an index-flood (i.e. scale factor), which depends only on the specific target site, and a dimensionless quantile (i.e. growth factor), which is unique within the given homogeneous region. Note that the concept of homogeneous region has evolved significantly over the last few decades: the traditional idea of contiguous and geographically identifiable regions (see e.g. NERC 1975) has been gradually replaced with the more general idea of homogeneous pooling-groups of sites with similar hydrological behaviour, which may or may not be geographically close to each other (see e.g. Acreman and Wiltshire 1989, Burn 1990, Ouarda *et al.* 2001). Finally, geostatistical methods assume the hydrological signature of interest in the ungauged catchment to be a weighted mean of the hydrological signatures in the neighbouring catchments, where the weights account for the spatial correlation of the signatures and the mutual locations of the catchments (see e.g. De Marsily 1986, Chokmani and Ouarda 2004, Skøien *et al.* 2006, Skøien and Blöschl 2007).

1.3 Impact of intersite correlation

In regional flood frequency analysis, the presence of spatial cross-correlation among flood sequences motivates information transfer from gauged to ungauged neighbouring sites (for interpolating in space) or from long-record sites to short-record sites (see e.g. Vogel and Stedinger 1985). For instance, spatial interpolation techniques (see e.g. Skøien *et al.* 2006, Archfield and Vogel 2010) consider spatial correlation an opportunity for predicting streamflow indices in ungauged sites. At the same time, cross-correlation tends to decrease the effective record length of series for computing regional statistics (see e.g. Matalas and Langbein 1962), thereby reducing the overall information content of regional datasets. This limits the accuracy with which moments of the regional parent distribution can be estimated (see e.g. Stedinger 1983), thus hampering the identification of the theoretical regional flood statistics. Classical studies (Matalas and Langbein 1962, Stedinger 1983) theoretically derived the reduction of the hydrological information content for a region having cross-correlated flood sequences, and quantified the corresponding increase of uncertainty in regional estimators of streamflow statistics. Hosking and Wallis (1988) showed that intersite correlation increases the variance of regional flood statistics by impacting the prediction uncertainty of regional flood frequency models (not their bias). Rosbjerg (2007) demonstrated the importance of including the cross-correlation of flood peaks for properly quantifying the uncertainty of regional estimates of flood quantiles. Further studies (Hosking and Wallis 1997, Madsen and Rosbjerg 1997, Madsen *et al.* 2002) showed that cross-correlation may also hamper the identification of the actual degree of heterogeneity of the region by reducing the power of the statistical tests used for this aim (see e.g. Castellarin *et al.* 2008, for a quantification of the loss of performance of the homogeneity tests proposed by Hosking and Wallis 1993, 1997). This impacts the assessment of the homogeneity of the region (e.g. heterogeneous pooling-groups of cross-correlated sites may be considered homogeneous), which is the fundamental hypothesis of the index-flood procedures (Dalrymple 1960) and a fundamental requirement for performing an effective regional estimation of flood quantiles (see e.g. Lettenmaier *et al.* 1987, Stedinger and Lu 1995).

1.4 Regional models accounting for, or exploiting, intersite correlation

The presence of spatial correlation is an important consideration when predicting flood quantiles in ungauged basins. Several studies in the literature tackled the problem of accounting for, or exploiting, the presence of spatial correlation among concurrent streamflows when predicting flood quantiles in ungauged basins (i.e. by means of regression methods, or spatial interpolation methods). Concerning regression-based methods, one example is the generalized least squares (GLS) regression introduced by Stedinger and Tasker (1985). Specifically, Tasker and Stedinger (1989) developed a particular version of GLS explicitly for the estimation of streamflow characteristics in fully ungauged basins (i.e. no gauged sites upstream or downstream within the same river basin). The Stedinger-Tasker GLS regression, which is illustrated in detail in the Appendix, accounts for sampling variability and cross-correlation among concurrent streamflows in developing a regional (multi-)regression model, with the aim of estimating the theoretical regional flood statistics (i.e. reducing the hampering effect due to the cross-correlation structure of the study region). The Stedinger-Tasker GLS regression (hereinafter referred to as GLS, for the sake of brevity), which is adapted for a log-Pearson type III (LP3) frequency analysis (see Bulletin 17B of the Interagency Advisory Committee on Water Data 1982, and the most recent Bulletin 17C by England *et al.* 2018), is the reference procedure for estimating streamflow characteristics in ungauged catchments in the US (Eng *et al.* 2009) and has been widely used for the regionalization of flood quantiles (see e.g. Tasker *et al.* 1986, Grehys 1996, Feaster and Tasker 2002).

It is important to note that GLS regression has been formulated to perform estimates at fully ungauged sites and that the presence of nested gauged sites with considerably overlapping drainage areas can result in high cross-correlations (i.e. they share an analogous flooding experience and can be considered redundant); for this reason, Gruber and Stedinger (2008) and Veilleux *et al.* (2011) introduced some metrics to identify these redundant sites and then delete one gauged site from each pair of redundant sites. Also, a Bayesian-GLS (B-GLS) regression model has been introduced for estimating model error variance and regression parameters (see e.g. Reis *et al.* 2005, Gruber *et al.* 2007, Gruber and Stedinger 2008, Veilleux *et al.* 2011; Bulletin 17C by England *et al.* 2018); more recently, Reis *et al.* (2020) developed an operational B-GLS regression capable of providing a comprehensive framework for regional hydrological analyses, estimation of flood quantiles included.

Other significant examples are the spatial interpolation methods (or geostatistical procedures), which were developed in the last few decades as an evolution of ordinary kriging (OK) (Matheron 1971, Farmer 2016) and explicitly exploit spatial correlation (e.g. canonical kriging (CK), see Ouarda *et al.* 2001, Chokmani and Ouarda 2004; top-kriging (TK), see Skøien *et al.* 2006). Also, maintenance of variance extension (MOVE) techniques (see e.g. Hirsch *et al.* 1982, Vogel and Stedinger 1985, Grygier *et al.* 1989) can be used to fill in missing observations at scarcely gauged locations by explicitly leveraging cross-correlation.

Concerning spatial interpolation methods, they have been shown to be effective for predicting several streamflow indices and hydrological signatures in nested ungauged catchments (see e.g. Castiglioni *et al.* 2009, Archfield *et al.* 2013, Pugliese *et al.* 2014, 2016, 2018). For instance, for estimating daily streamflows at an ungauged catchment, the map-correlation method introduced by Archfield and Vogel (2010) suggests that if the correlations between a set of streamgauges and the ungauged catchment could be reliably estimated (by means of spatial methods, such as kriging), then the selection of the most correlated gauge leads to better performances than the selection of the nearest one (see also Patil and Stieglitz 2012, which show that spatial proximity alone cannot fully explain the prediction performance at a given location).

Another recently developed geostatistical procedure is top-kriging (TK) (Skøien *et al.* 2006), which interpolates the runoff signature of interest along the stream network by taking the area and the nested structure of catchments into account. The method, originally tested for the prediction of specific 100-year flood for two Austrian regions (Skøien *et al.* 2006), was shown to provide more plausible and accurate estimates than OK. TK, which is thoroughly described in the Appendix, has been shown to be particularly successful in predicting a wide spectrum of point streamflow indices and variables in various geographical and climatic contexts: low flows (Castiglioni *et al.* 2011, Laaha *et al.* 2014), high flows and floods (Merz *et al.* 2008, Archfield *et al.* 2013), flow-duration curves (Pugliese *et al.* 2014, 2016, 2018, Castellarin *et al.* 2018), stream temperature (Laaha *et al.* 2013), habitat suitability indices (Ceola *et al.* 2018), and daily streamflow series (Skøien and Blöschl 2007, Vormoor *et al.* 2011, Parajka *et al.* 2015, de Lavenne *et al.* 2016, Farmer 2016).

1.5 Open problems and aim of the study

In this context, a recent study by Archfield *et al.* (2013) compared the performances of GLS regression, CK and TK in predicting empirical estimators of flood quantiles across a set of 61 nested gauged basins (i.e. many of the gauges were within the watershed of other gauges) located in the Flint River basin in the south-eastern USA. Their study demonstrated that when the goal is the prediction of estimated empirical flood quantiles in nested ungauged sites, TK is likely to provide better predictions (i.e. smaller absolute errors and higher Nash-Sutcliffe efficiency; see Nash and Sutcliffe 1970) than CK and GLS regression models. Nevertheless, Archfield *et al.* (2013) also pointed out that, being entirely based on empirical data, their analysis cannot address the fundamental problem of understanding which technique is better suited (i.e. provides less uncertain estimates) for predicting the theoretical ("true" population) unknown flood quantiles in ungauged sites when the observed flood sequences are affected by cross-correlation.

In fact, referring to a set of correlated streamflow observations, in which the regional information content is reduced by intersite dependence, could hamper the identification of the theoretical flooding potential at a given ungauged site. On the one hand, one might expect TK to have better efficiencies in predicting the empirical estimator of the flood quantiles, weighting empirical information on the basis of the observed cross-correlation. On the other hand, one might assume that GLS, which explicitly models spatial correlation, would be more accurate in predicting the theoretical (and unknown) flood quantiles (i.e. "true" population flood quantiles). Despite its fundamental importance, this aspect has not yet been formally addressed in the literature, which motivates our study. Moreover, as recently suggested by Reis *et al.* (2020), it is important to point out that the application performed by Archfield *et al.* (2013) considered ungauged catchments that are nested with gauged ones (i.e. not fully ungauged): in such conditions, which are common in hydrological applications, TK takes advantage of flood data collected at nested gauged catchments, while GLS does not, being formulated to estimate hydrological statistics at fully ungauged sites.

The main objective of our study is to address the fundamental theoretical issue raised by Archfield *et al.* (2013), i.e. to understand which of the two ways of incorporating information on the cross-correlation structure of the data featured by GLS and TK is the most effective for estimating (a) the local empirical quantile estimate and (b) the theoretical flood quantile. To this aim, differently from Archfield *et al.* (2013), we consider a spatially limited, hydrologically homogeneous region and perform a controlled Monte Carlo experiment, which enables us to also assess the performance of the procedures in predicting the theoretical flood quantiles. In particular, we mimic the regional flood frequency regime and the spatial structure of the correlation among concurrent flows of a real-world example dataset for a homogeneous pooling-group of catchments (see the Supplementary material for a detailed description of the dataset). Based on the structure of the real-world study region, we design a Monte Carlo simulation framework to generate 1000 realizations of the homogeneous regional set of floods (hereinafter referred to as the homogeneous region, for the sake of brevity) for three levels of regional cross-correlation (i.e. 3000 realizations in total).

In this preliminary study, we disregard the representation of the possible impact of nestedness in defining the cross-correlation structure based on the evidence. This impact is shown to be negligible for the study area (see the Supplementary material) as well as for other areas (see e.g. Fig. 4 in Castellarin 2007). Future analyses will specifically address this aspect, which has been shown to be significant in other geographical areas and hydro-climatic contexts (see e.g. Gruber and Stedinger 2008, Veilleux *et al.* 2011). For each realization of the region and level of cross-correlation, we apply GLS and TK to obtain predictions of at-site flood quantiles for return periods T equal to 10, 30, 50 and 100 years in a leave-one-out cross-validation (LOOCV) scheme, consistently with several studies on regionalization, including Archfield *et al.* (2013). Differently from Archfield *et al.* (2013), which considered a simple-regression application (i.e. based on Gotvald *et al.* 2009, drainage area was the only significant physiographic explanatory variable considered for their study area) of the two methods, we apply GLS and TK in both simple (i.e. univariate) and multiple (i.e. multivariate: more than one significant physiographic explanatory variable) versions. Finally, we compare the cross-validated GLS and TK regional flood quantiles against (a) the empirical estimators of flood quantiles and (b) their known theoretical values at each site in the realizations of the region.

As the basis for our simulations experiments, we consider a dataset of annual maximum series (AMS) of peak flow discharges collected at 20 nested catchments in Triveneto (north-eastern Italy; see Persiano *et al.* 2016, and see the Supplementary material of the present article for a more detailed illustration), which may be regarded as possibly homogeneous in terms of flood frequency regime according to the test proposed by Hosking and Wallis (1997) (see also Castellarin *et al.* 2008). In the Supplementary material we also show that performing the exercise described in Archfield *et al.* (2013) for the Triveneto case study (i.e. application of GLS and TK in an LOOCV scheme for predicting at-site empirical flood quantiles in a nested region, see S.3 in the Supplementary material) leads to exactly the same results and conclusions as those obtained by Archfield *et al.* (2013) for the study area located in the south-east United States.

2 Structure of the analysis

2.1 Monte Carlo simulation framework

To assess the behaviour of GLS and TK under different cross-correlation scenarios and their accuracy in predicting the theoretical unknown flood quantiles at ungauged sites, we implement a Monte Carlo simulation experiment for generating realizations of a given cross-correlated homogeneous region, which can be summarized as follows (see also Fig. 1):

Step 1. We focus on a real-world regional dataset (see the Supplementary material) to (a) mimic the regional flood frequency regime and controls of relevant catchment descriptors, as well as evaluate spatial correlation structure of flood flows, and (b) define theoretical flood quantiles at each and every site in the region referring to a unique regional parent distribution.

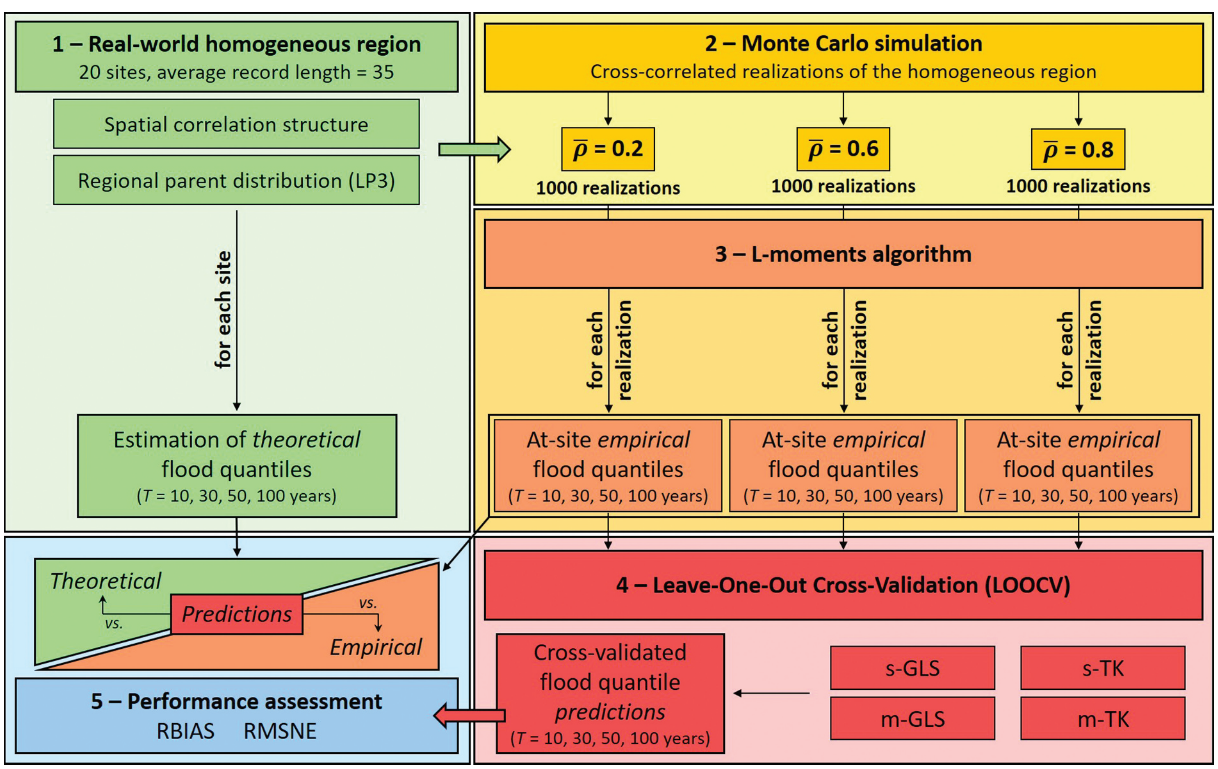

Figure 1. Flowchart describing the main steps of the adopted procedure. s-GLS, m-GLS, s-TK, m-TK identify the univariate (s-) and multivariate (m-) applications of GLS and TK. RBIAS (relative bias) and RMSNE (root mean square normalized error) are the metrics used for assessing the prediction performance for the considered methods.

Step 2. We generate 1000 realizations of the homogeneous region, with three different degrees of regional cross-correlation; each realization consists of N = 20 concurrent sequences (i.e. sites) of 35 annual floods (see Section 2.2 for the Monte Carlo simulation algorithm).

Step 3. We apply the L-moments algorithm (Hosking and Wallis 1993, 1997) for predicting at-site flood quantiles associated with return periods T = 10, 30, 50, 100 years at each and every site in each realization (see Section 2.3).

Step 4. We then refer to at-site flood quantiles for predicting flood quantiles in ungauged locations by means of simple and multiple (i.e. univariate and multivariate) applications of GLS (i.e. s-GLS and m-GLS) and TK (i.e. s-TK and m-TK) within an LOOCV procedure for each realization (see Section 2.4);

Step 5. Finally, we compare GLS and TK cross-validated flood quantile predictions with their empirical and theoretical counterparts in terms of relative bias (RBIAS) and root mean square normalized error (RMSNE).

The main steps of the analysis and performance metrics are synthetically reported in Fig. 1. A more detailed description is reported in the following subsections.

2.2 Step 2. Realizations of cross-correlated regions

To investigate the impact of spatial correlation in flood data on the prediction accuracy of both GLS and TK, we refer to realizations of cross-correlated homogeneous regions generated in a Monte Carlo framework for three different degrees of intersite (or spatial, or cross-) correlation. We generate numerous realizations of cross-correlated hypothetical regions (each one composed of cross-correlated annual sequences of flood flows) that mimic the main characteristics of the real-world homogeneous reference region in terms of its flood frequency regime, these being: (1) the regional parent distribution of annual flood flows (we refer to the LP3 distribution, for which the Stedinger-Tasker GLS regression is adapted; see S.1); (2) the geomorphological and climatic controls on annual flood; and (3) the cross-correlation structure (see S.1 and S.2 in the Supplementary material).

We adopt as parameters of the theoretical parent distribution for each realization the LP3 regional L-moments, μ^R, σ^R and χ^R, presented in S.1, hence we consider as theoretical dimensionless quantiles the regional LP3 quantiles q_T^R with return periods T = 10, 30, 50 and 100 years. It is worth noting here that we adopt as theoretical dimensional flood quantiles at each and every site i, Q_{Ti}^R, the product between q_T^R and local empirical mean annual flood, MAF_i (i.e. the mean value of the observed AMS at site i).

The cross-correlation structure of the study region is modelled in the standard-normal space through the nonlinear model of Tasker and Stedinger (1989) (see Equation (A6) in the Appendix) and without distinguishing between cases where basins are nested and where they are not, for the reasons described above and according to the empirical evidence illustrated in the Supplementary material (see Fig. S4). We consider a scenario with $\bar{\rho}$ equal to 0.6 (values of the parameters of the model by Tasker and Stedinger 1989, equal to θ~0.95 and α~0.07; see Equation (A6) in the Appendix) and two alternative models that adopt similar laws of decay between correlation and distance, describing a lower and higher degree of cross-correlation (or cross-correlation scenario) in the standard-normal space, and resulting in $\bar{\rho}$ equal to 0.2 (θ~0.91, α~0.03) and 0.8 (θ~0.96, α~0.18), respectively.

For each cross-correlation scenario (i.e. $\bar{\rho}$ = 0.2, 0.6, 0.8), 1000 realizations of the hypothetical homogeneous region are compiled by generating cross-correlated random annual flood sequences. Each realization of the region consists of 20 overlapping annual sequences with record length equal to 35 years (average record length in the real-world dataset). Flood sequences are simulated by first generating cross-correlated annual series from a multivariate standard-normal distribution with the selected record length (i.e. 35) and cross-correlation structure (i.e. a 20 × 20 correlation matrix resulting from the nonlinear model for the specific cross-correlation scenario). The generated standard-normal variates are then back-transformed to dimensionless LP3 flows (with parameters μ^R, σ^R and χ^R) through a quantile-quantile transformation. Note that the quantile-quantile back-transformation from the standard normal variate to the dimensionless LP3 flows introduces a slight distortion in terms of average cross-correlation: a slight negative bias is observed ($\bar{\rho}$ = 0.2, 0.6, 0.8 in the standard-normal space correspond to $\bar{\rho}$ = 0.17, 0.55, 0.76 in the real space, after back-transformation), yet the variation range does not change noticeably. For this reason, the distortion is found to be not relevant for the aims of our study, and in this manuscript we always refer to $\bar{\rho}$ computed in the standard-normal space.

Finally, to obtain the synthetic dimensional annual sequence of flood flows at site i, the i-th synthetic dimensionless sequence is multiplied by the local empirical mean annual flood, MAF_i. Note that, due to sampling variability (i.e. limited record length), the empirical mean value of each generated dimensionless series is not exactly equal to 1; for this reason, the resulting dimensional series for site MAF_i shows an arithmetic mean value that can differ from the local theoretical mean annual flood (MAF_i, that is the empirical estimate from the real-world annual sequence for site i), and varies from realization to realization across the region according to the regional cross-correlation structure used in the random generation.

Moreover, it is worth noting here that, consistent with the hypothesis of homogeneity, the Monte Carlo generation of the synthetic dimensionless series considers no spatial pattern in the flood frequency distribution of the regional dimensionless samples, only different values of cross-correlation. The regional spatial pattern in the theoretical dimensional series is introduced by the local empirical mean annual flood, MAF_i, which is always the same for the different realizations of the region, i.e. we mimic the same regional spatial pattern in terms of flood frequency distribution in all the Monte Carlo realizations.

Figure 2 illustrates the cross-correlation structure in the standard-normal space as results from an empirical analysis of 1000 realizations for each cross-correlation scenario: (a) $\bar{\rho}$ = 0.2, (b) $\bar{\rho}$ = 0.6, (c) $\bar{\rho}$ = 0.8. Figure 2(d) shows the overall distributions of empirical cross-correlation coefficients for the 1000 realizations

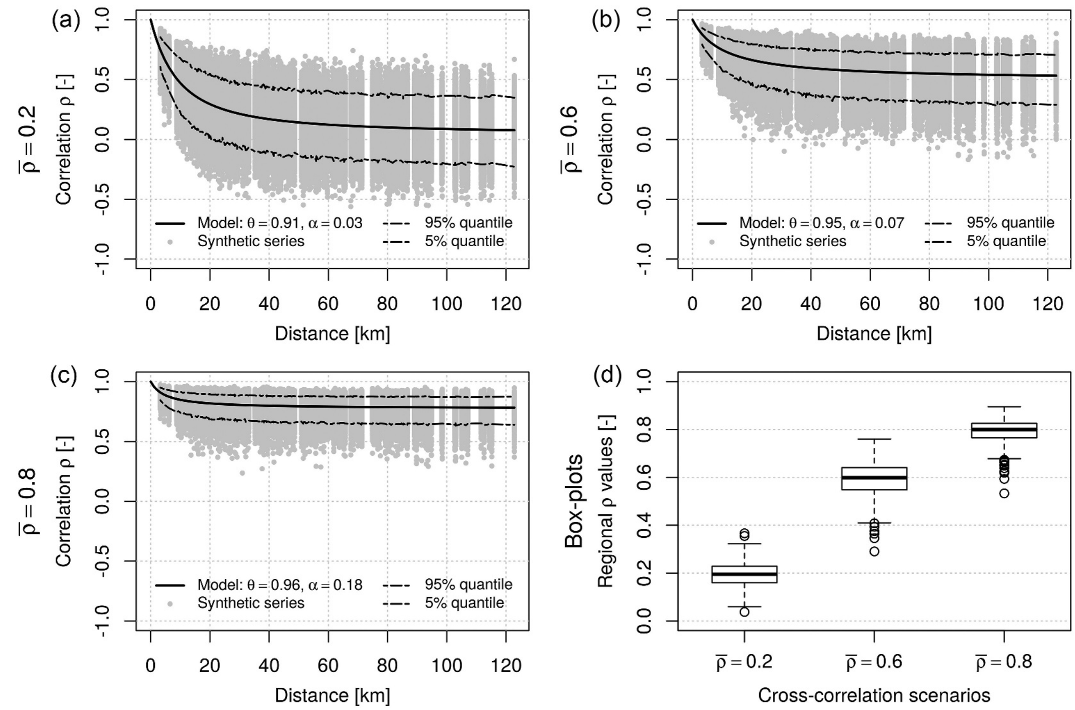

Figure 2. One thousand realizations of the hypothetical cross-correlated region for each cross-correlation scenario. Grey dots (panels (a), (b) and (c)) show the distribution of empirical cross-correlation coefficients around the corresponding cross-correlation model (black solid line) for $\bar{\rho}$ = 0.2, 0.6 and 0.8, respectively; the dashed lines indicate the 90% confidence band (upper and lower lines represent 95th and 5th quantiles, respectively). Box plots in panel (d) show the overall distribution of average correlation values for the 1000 realizations of the three cross-correlation scenarios ($\bar{\rho}$ = 0.2, 0.6 and 0.8, respectively).

of the three correlation scenarios, which are centred around the corresponding average cross-correlation (i.e. the median values of each box plot are consistent with the imposed $\bar{\rho}$ values).

2.3 Step 3. Empirical flood quantiles

For each realization and each site in the region, empirical flood quantiles are then estimated for arbitrarily selected return periods T (i.e. $T = 10, 30, 50, 100$ years). The quantile estimator combines at-site and regional information. In particular, we refer to the LP3 distribution (i.e. no uncertainty on the model selection) and estimate the LP3 parameters of location and scale, μ and σ, respectively, locally by using the L-moments method (Hosking and Wallis 1997) (i.e. uncertainty from sampling error only on the moments of order 1 and 2); while the shape parameter χ is set equal to χ^R (theoretical shape parameter, i.e. no uncertainty on the third-order moment).

2.4 Step 4. Prediction of flood quantiles in ungauged sites: application of GLS and TK in cross-validation

The Appendix recalls the theoretical bases of GLS and TK. Both methods are applied to the locally estimated (dimensional) empirical flood quantiles recalled in Section 2.3 for any realization that we generated as described above in order to predict flood quantiles at ungauged locations. Ungauged conditions at each and every site are simulated through an LOOCV scheme. For each realization in each cross-correlation scenario, we apply a simple and a multiple implementation of GLS and TK, as described in the following sections.

2.4.1 Application of GLS

The GLS analysis is carried out using the R-package WREG (i.e. Farmer 2017) (see also Eng *et al.* 2009) in the statistical environment R (R Core Team 2016). Applying the GLS procedure to a given site i requires estimates of standard deviation s_i, regional skew $G_{R,i}$, weighted skew $\tilde{G}_{w,i}$ and LP3 distribution standard deviate K_i (see also Griffis and Stedinger 2007b). These variables are computed for the log-transformed samples of each synthetic region, as follows:

- standard deviation s_i is evaluated as the at-site empirical standard deviation;
- we use the same value of regional skew G_R for each and every catchment in the synthetic region (i.e. assumption of homogeneity); this value is computed as the empirical skew of the log-transformed regional dimensionless sample;
- weighted skew $\tilde{G}_{w,i}$ for each site is computed as indicated in Equation (A4) in the Appendix, by combining the at-site empirical skew g_i with the regional skew G_R; weights are computed as indicated in Equation (A5), evaluating the estimated mean square errors $MSE(g_i)$ and $MSE(G_R)$ according to Bulletin 17B of the Interagency Advisory Committee on Water Data (1982) (see also Eng *et al.* 2009);
- the standard deviates K_i are computed as a function of non-exceedance probability (i.e. return period T) and weighted skew $\tilde{G}_{w,i}$, as indicated in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982) and in the most recent Bulletin 17C (England *et al.* 2018).

The application of the GLS procedure requires an estimate of the cross-correlation structure, in terms of parameters θ and α in Equation (A6) (see Tasker and Stedinger 1989). We estimate the cross-correlation model relative to the distances between catchment centroids by optimizing model parameters θ and α through the ordinary least squares (OLS) procedure.

GLS regressions may use several catchment descriptors, e.g. drainage area, precipitation, elevation, etc. In the present study, we implement a GLS regional model in two ways:

- simple-GLS (hereinafter referred to as s-GLS): the GLS quantile regression analysis is performed by considering drainage area A alone (as did in Archfield *et al.* 2013):

$$Q_{T_i} = a_T {A_i}^{b_T} \quad (1)$$

  where Q_{T_i} is the T-year flood for site i, A_i is the drainage area for site i, and a_T and b_T are the GLS parameters;

- multiple-GLS (hereinafter referred to as m-GLS): the GLS quantile regression analysis considers more catchment descriptors. In particular, we refer to drainage area A, mean annual precipitation MAP, latitude of catchment centroid Y_g, and mean elevation H_{mean}, which preliminary OLS stepwise log-linear regression analyses (see Draper and Smith 1981, Weisberg 1985, Chambers 1992) indicated are the most significant subset of descriptors (i.e. adjusted $R^2{}_{adj} \approx 0.87$ and normally distributed standardized residuals) for predicting the mean annual flood (MAF) for the reference real-world dataset described in the Supplementary material. The resulting multiple-GLS model can be described as:

$$Q_{T_i} = a_T {A_i}^{b_T} {MAP_i}^{c_T} {Y_{g_i}}^{d_T} {H_{mean_i}}^{e_T} \quad (2)$$

where a_T, b_T, c_T, d_T and e_T are the GLS parameters.

Equations (1 and 2) are then reduced to linear additive forms by means of log-transformation (see e.g. Thomas and Benson 1970, Pandey and Nguyen 1999, Griffis and Stedinger 2007a, Laio *et al.* 2011).

Both the s- and m-GLS applications are explored in an LOOCV scheme, by removing in turn one site from the dataset and by referring to the remaining N–1=19 sites while estimating (1) the cross-correlation structure (i.e. fitting of the model of Tasker and Stedinger 1989), (2) the GLS parameters, and finally (3) the flood quantiles Q_T at the discarded site. Consistently with the Monte Carlo generation, where nestedness was not modelled, GLS is applied without removing eventual redundant nested sites.

2.4.2 Application of TK

The first step of the ungauged application of TK is the use of OLS to identify a regional power-law model between flood quantile Q_T and basin area (see e.g. Pugliese *et al.* 2014, 2016). The OLS estimates are then used to standardize LP3 quantiles (i.e. Q_T with T = 10, 30, 50, 100 years) at all sites: this fundamental step is necessary as TK directly handles drainage area as a key variable of the model itself.

We opt for two types of OLS regional power-law to keep the applications of GLS and TK consistent (see Section 2.4.1):

- simple-OLS (s-OLS; i.e. considering only drainage area), resulting in a standard application of TK (hereinafter referred to as s-TK);
- multiple-OLS (m-OLS; i.e. considering the five significant descriptors identified via stepwise regression analysis; see e.g. Equation (2)), resulting in a TK with external drift (hereinafter referred to as m-TK; see e.g. Laaha *et al.* 2013, for the use of m-TK for predicting stream temperatures).

Both s-OLS and m-OLS are applied in linear additive forms by means of log-transformation, and their natural estimates (i.e. back-transformed from the logarithmic to the natural space) are then used to standardize LP3 quantiles. TK interpolation is then applied using the R-package rtop (i.e. Skøien et al., 2014; Skøien, 2014) by fitting the sample variogram of the standardized quantiles with the five-parameter fractal-exponential model suggested in Skøien *et al.* (2006) through a modified version of weighted least squares (WLS) regression (see Cressie 1993). The fitted variogram model is then used to compute the kriging weights, referring to the six closest neighbouring stations (in line with recent sensitivity analyses, see e.g. Pugliese *et al.* 2014, 2016). The standardized quantiles are then predicted site by site using Equation (A9) in the Appendix. Finally, the TK estimates of the standardized quantiles are combined with the regression estimates resulting from s-OLS and m-OLS to obtain the s-TK and m-TK estimated T-year value at each site.

Both the s-TK and m-TK analyses are performed in an LOOCV scheme, by removing in turn one site from the dataset and referring to the remaining N–1=19 sites while estimating (1) the OLS regional power-law useful for standardizing flood quantiles, (2) the variogram (i.e. five-parameter fractal-exponential model suggested in Skøien *et al.* 2006), and (3) the flood quantiles Q_T at the discarded site.

2.5 Performance metrics

We use two metrics for assessing the prediction performance for the considered versions of GLS (i.e. s-GLS and m-GLS) and TK (i.e. s-TK and m-TK). In particular, to assess the overall performance of GLS and TK in the entire region, we consider two error measures, RBIAS and RMSNE:

$$\mathrm{RBIAS} = \frac{1}{N}\sum_{j=1}^{N}\left(\frac{\hat{x}_i - x_i}{x_i}\right) \quad (3)$$

$$\mathrm{RMNSE} = \left[\frac{1}{N}\sum_{j=1}^{N}\left(\frac{\hat{x}_i - x_i}{x_i}\right)^2\right]^{1/2} \quad (4)$$

where N is the number of sites, $\hat{x}_i$ the estimated variable at site i and x_i the observed value of the variable at site i. The choice of relative (i.e. RBIAS) or normalized (i.e. RMSNE) measures is made to quantify errors and performance, regardless of the size of the drainage area of each catchment.

3 Results

Box plots in Figs 3 and 4 depict the efficiency of GLS and TK in estimating empirical and theoretical flood

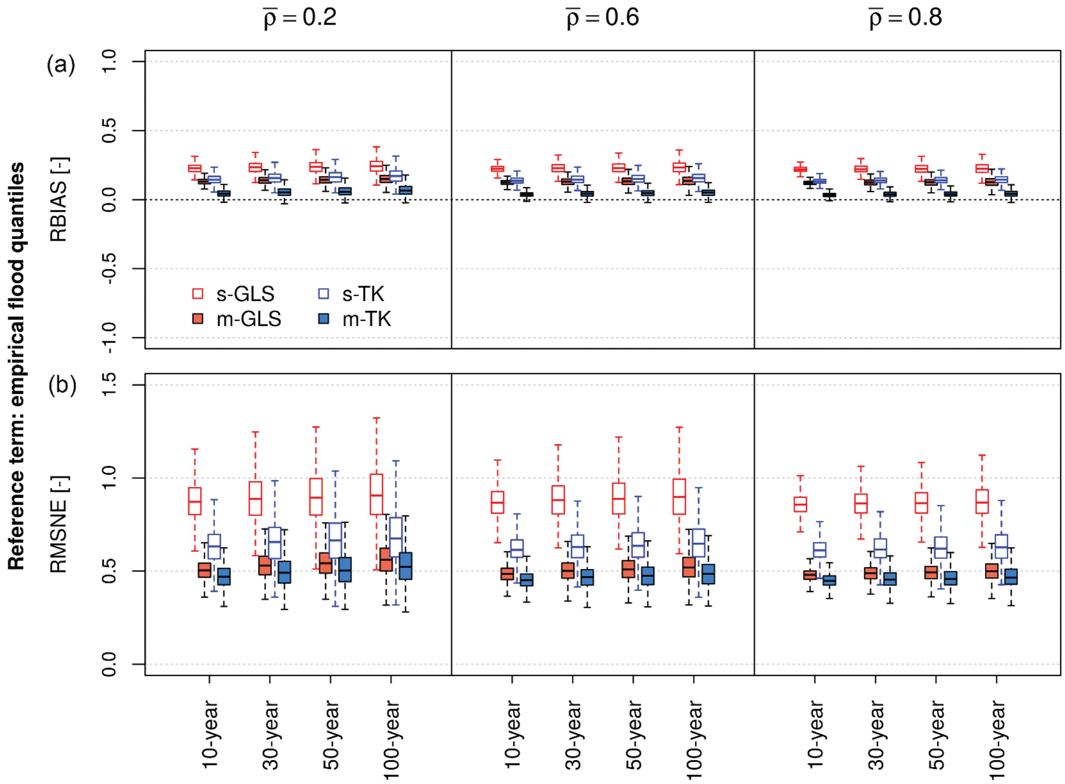

Figure 3. Prediction performance of s-GLS (red-bordered box plots), m-GLS (red-filled box plots), s-TK (blue-bordered box plots) and m-TK (blue-filled box plots) in predicting empirical flood quantiles for the three degrees of cross-correlation (i.e. $\bar{\rho}$ = 0.2, 0.6, 0.8) in ungauged sites (LOOCV scheme): box plots represent the distribution of 1000 values for each metric (i.e. RBIAS and RMSNE) computed for each realization and for selected return periods (i.e. T = 10, 30, 50, 100 years).

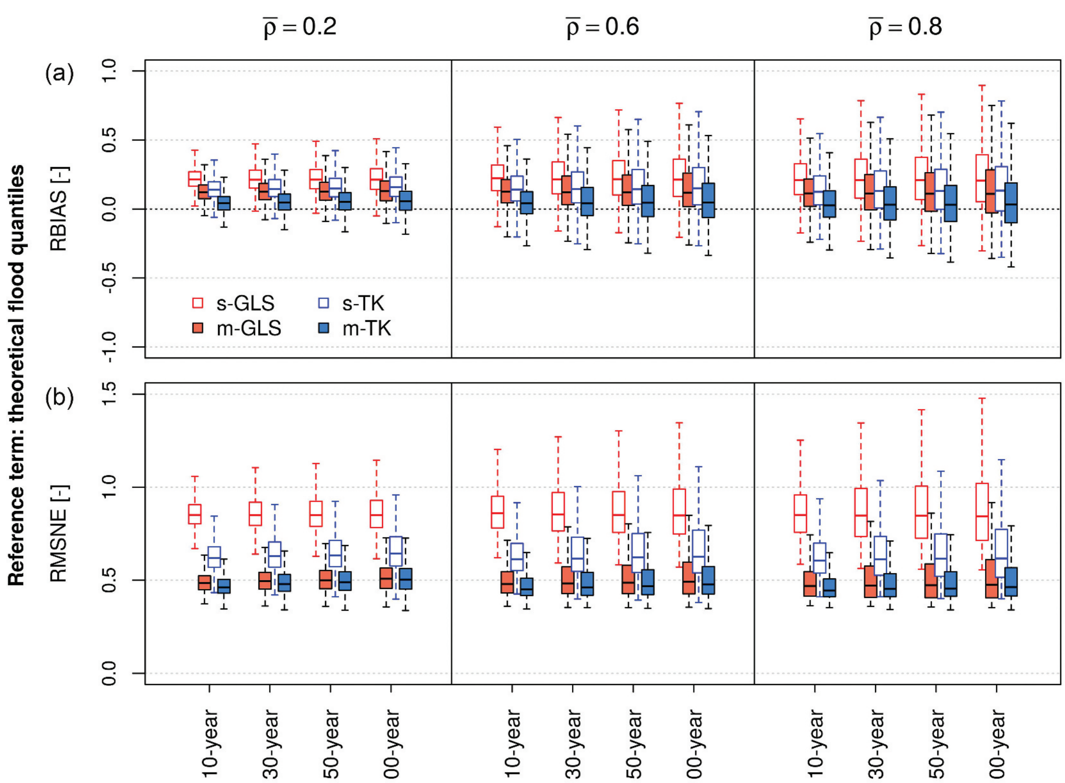

Figure 4. Prediction performance of s-GLS (red-bordered box plots), m-GLS (red-filled box plots), s-TK (blue-bordered box plots) and m-TK (blue-filled box plots) in predicting theoretical flood quantiles for the three degrees of cross-correlation (i.e. $\bar{\rho}$ = 0.2, 0.6, 0.8) in ungauged sites (LOOCV scheme): box plots represent the distribution of 1000 values for each metric (i.e. RBIAS and RMSNE) computed for each realization and for selected return periods (i.e. T = 10, 30, 50, 100 years).

quantiles, respectively, in ungauged sites (note that "ungauged" refers here to the application of the LOOCV scheme). Each box plot depicts the distribution of 1000 values (i.e. one value for each realization of the region) of the selected performance measure for the selected method, return period and cross-correlation scenario. To enable one to visually compare the results across different return periods, methods and degree of cross-correlation, the y-axes in Figs 3 and 4 are fixed for each row.

Concerning the efficiencies in estimating the empirical estimates of flood quantiles, TK results in generally better predictions of empirical quantiles (i.e. lower RBIAS and lower RMSNE) than the corresponding version of GLS (see Fig. 3): s-TK outperforms s-GLS, and m-TK shows similar median RMNSE, but smaller RBIAS than m-GLS. Moreover, the comparison of s-GLS with m-GLS, and of s-TK with m-TK, shows that the inclusion of more catchment descriptors in the regression analysis (i.e. m-GLS; m-OLS for m-TK) leads to substantially improved performances. In particular, the weak performances of s-GLS can be explained by the fact that drainage area alone is not enough for fully describing *MAF* (and therefore dimensional flood quantiles) in the study region. Moreover, the similar behaviour of m-GLS and m-TK can be attributed to the fact that they both use the important physiographic information needed to explain variations in the mean flood. In this regard, as observed for the real-world study area (see S.3 in the Supplementary material), all the considered procedures are slightly positively biased, with median RBIAS values that are similar to the corresponding values obtained for the real-world application (see Table S1 for the 100-year flood quantiles), and RBIAS substantially decreasing from s-GLS to m-TK.

Another aspect shown in Fig. 3 is the dependence on T: for each $\bar{\rho}$ and considered method, we observe that the higher the T value, the lower the performance (i.e. the higher the bias or uncertainty). This behaviour is expected and confirms that estimates of flood quantiles associated with lower probability of occurrence (i.e. higher return period T) are affected by higher uncertainties.

Finally, we observe a dependence of the performance in estimating empirical flood quantiles on the average regional cross-correlation $\bar{\rho}$: for a given method and return period T, the higher the $\bar{\rho}$ the better the median of the performance and the lower the uncertainty, with much less dispersed performance. This behaviour is expected, especially for TK, which explicitly exploits cross-correlation in predicting flood quantiles.

With regards to the efficiencies of GLS and TK in estimating the theoretical flood quantiles, Fig. 4 confirms the trend already observed for empirical flood quantiles: s-TK, m-GLS and m-TK outperform s-GLS in predicting theoretical flood quantiles (i.e. lower RBIAS and lower RMSNE), and, in particular, m-GLS and m-TK perform similarly, with slightly less biased predictions for m-TK. This confirms that incorporating multiple regression in m-GLS and m-TK improves substantially the prediction performances. Moreover, the weak dependence on T is confirmed: the higher the T value, the higher the uncertainty.

4 Discussion

Notwithstanding the above-mentioned similarities in the relative behaviour of the methods, a significant difference is present in the extent of the performances: all the considered methods show generally higher uncertainty in estimating theoretical flood quantiles than empirical quantiles. This can be explained by the fact that the presence of cross-correlation hampers the identification of the theoretical flooding potential in the region, and none of the considered methods is able to effectively overcome this effect, unless the regional average cross-correlation is very limited (i.e. see results for $\bar{\rho}$ = 0.2). An exception is observed for $\bar{\rho}$ = 0.2, for which RBIAS values show a larger dispersion in predicting theoretical flood quantiles (compare panels (a) in Figs 3 and 4), yet RMNSE indicates higher accuracies for all the considered methods in predicting theoretical flood quantiles than their empirical estimates. In this regard, other important indications come from the dependence on $\bar{\rho}$: contrary to what is observed for empirical flood quantiles, for estimating theoretical flood quantiles, the higher the $\bar{\rho}$ the lower the performance for both GLS and TK. As the spatial correlation has the effect of hampering the identification of the flood magnitude for a limited set of flood-flow observations, and TK exploits cross-correlation to perform its estimates, the decreasing performances and increasing uncertainty of s-TK and m-TK with increasing $\bar{\rho}$ are expected. On the other hand, the increasing uncertainty of GLS with increasing $\bar{\rho}$ shows that, even if GLS is able to reduce the hampering effect due to cross-correlation thanks to its efficient estimation of the covariance matrix (Kroll and Stedinger 1998), a remaining uncertainty, increasing with increasing levels of spatial correlation, is still present.

In summary, the Monte Carlo experiment performed in this study enables us to address the unsolved issue raised by Archfield *et al.* (2013) regarding the ability of GLS and TK in predicting the theoretical unknown flood quantiles in ungauged sites when the flood sequences observed at nested gauged catchments are affected by cross-correlation. The main outcome of our study is that, despite the different nature of GLS and TK and their different ways to treat spatial correlation, cross-correlation controls GLS and TK accuracy of prediction of empirical and theoretical flood quantiles in a very similar way. In particular, TK outperforms GLS when catchment area is the only catchment descriptor used for predicting theoretical flood quantiles, regardless of the level of cross-correlation. These results are valid under the simplifying assumptions adopted in the analysis; a fair interpretation of the outcomes of our study, which is the first to address this very interesting issue, needs to acknowledge the simplifying hypotheses we adopted, which are recalled and discussed below.

We refer to synthetic regions that mimic the real-world reference dataset (i.e. an acceptably homogeneous region, with partially nested catchments, and no noticeable influence of nestedness on the cross-correlation structure; see the Supplementary material). Indeed, given the partial nestedness of catchments in the study area, the LOOCV scheme performed in our analyses refers to ungauged sites that are likely to have other gauges upstream or downstream within the same catchment, and therefore are not fully ungauged catchments in the strict sense. Although the presence of nested catchments represents a common situation in hydrological applications, it is important to recall that Stedinger and Tasker (1985) considered GLS very explicitly for fully ungauged catchments (no nested sites), and Veilleux *et al*. (2011) suggested to remove (redundant) nested gauged sites when they share a large portion of drainage area. For this reason, future studies could explore the more general case where cross-correlation is modelled by considering nestedness and GLS is applied by removing redundant sites (as suggested by e.g. Veilleux *et al*. 2011).

Another important aspect is related to the homogeneity of our study region: we refer to the straightforward case of a homogeneous region, yet the homogeneity hypothesis (i.e. the same growth factor but index flood varying in space) is not a fundamental prerequisite for the application of GLS and TK, which can indeed be applied to heterogeneous areas. For this reason, future studies could consider the application of a Monte Carlo experiment similar to the one considered here, but referring to a heterogeneous region (i.e. theoretical flood quantiles for each site referring to different values for the parameters of the chosen distribution). Also, our experiment (see Section 2.2) mimics the same regional spatial pattern (in terms of regional flood frequency distribution) in all the Monte Carlo realizations of the region: while no spatial pattern is introduced in the regional dimensionless samples (i.e. hypothesis of homogeneity), the generated dimensional samples depend on the local empirical mean annual flood, MAF_i, which is always the same for the different realizations of the region. Therefore, our Monte Carlo generation does not model the dependence of the generated series from the considered catchment descriptors (i.e. covariates). In this context, future studies could consider synthetic series generated as a function of the covariates (as done e.g. by Stedinger and Tasker 1986).

Moreover, we considered overlapping annual sequences with record length equal to 35 years for every site, without investigating the sensitivity of the methods to different record lengths. Finally, we referred to the L-moments approach (Hosking and Wallis 1993, 1997) to fit the LP3 distribution, but we should consider that the GLS procedure implements another approach for fitting LP3 (see Section 2.4.1), and this difference could in part affect the results.

Relaxing the above-mentioned simplifying assumptions opens up interesting future research avenues that can deepen our current understanding on how best to handle spatial correlation of flood sequences for predicting flood quantiles in partially gauged or fully ungauged catchments.

5 Conclusions

The present study addresses an important research question raised by Archfield *et al*. (2013), namely understanding which of two techniques, Tasker-Stedinger generalized least squares (GLS) regression or top-kriging (TK), is better suited for predicting the theoretical unknown flood quantiles in ungauged sites when the observed flood sequences are affected by cross-correlation and nestedness does not play a significant role on the cross-correlation structure of the study region. The performance of GLS and TK is evaluated relative to the prediction in ungauged conditions of the local empirical quantile estimates (i.e. those that would have been estimated if data were available at that location) and the theoretical flood quantiles.

The preliminary LOOCV analysis performed over a real-world study area (i.e. a spatially limited, hydrologically homogeneous region in Triveneto consisting of 20 partially nested catchments; see the Supplementary material) highlights that the behaviour of the simple (i.e. function of the drainage area alone) versions of GLS (i.e. s-GLS) and TK (i.e. s-TK) applied for predicting the 10-year flood is consistent with the results reported by Archfield *et al*. (2013). The better performances of s-TK compared to s-GLS are expected: referring to *n* neighbouring sites, TK is implicitly able to take some climate and geomorphological similarities between catchments into account. However, the inclusion of more catchment descriptors in the analysis (i.e. m-GLS, m-TK) can lead to substantially improved performances, especially for GLS. Although informative, our preliminary analysis cannot address the unsolved problem raised by Archfield *et al*. (2013), as the theoretical flood quantiles are unknown in real-world study areas.

To shed some light on the above-mentioned research question, we resort to the Monte Carlo simulation experiment described in Section 2, generating a total of 3000 realizations of the homogeneous region with different values of average regional cross-correlation (i.e. equal to 0.2, 0.6, 0.8), where the cross-correlation structure does not distinguish between nested and non-nested catchments. For each realization, GLS and TK are applied to obtain predictions of at-site flood quantiles (with return periods equal to 10, 30, 50, 100 years) in an LOOCV scheme. This application provides us with some significant information about the ability of GLS and TK to predict empirical estimates of flood quantiles and theoretical flood quantiles at nested ungauged catchments, when GLS is applied in its traditional formulation (i.e. non-Bayesian and without removing redundant nested gauged sites), consistent with Archfield *et al*. (2013). First, in agreement with what is observed for the real-world application, the multiple versions of both GLS and TK are better suited than the corresponding simple versions for predicting both empirical and theoretical flood quantiles. Moreover, the analyses highlight an analogous dependence of GLS and TK performance on the degree of cross-correlation: the larger the regional average cross-correlation, the better the empirical estimators of flood quantiles, and the poorer the estimators of the theoretical flood quantiles. These findings prove that, under the adopted simplifying assumptions, the presence of cross-correlation in the annual flood sequences hampers the identification of the theoretical flood magnitude for both GLS and TK. This behaviour is totally expected for TK (which explicitly exploits spatial correlation in performing its estimates), whereas what was observed for GLS may be explained by acknowledging the very specific and simplified situation considered in our study (i.e. homogeneous region with nested gauged sites). The investigation of a more general case where nested catchments may impact the cross-correlation structure, which should therefore be modelled in GLS by removing redundant sites, is left for future studies.

In particular, the multiple versions of GLS and TK show very similar performances: the application of m-GLS or m-TK is almost equivalent when descriptive geomorphological and climatic catchment attributes are available for representing mean annual flood; for practical estimation in cases like this, one could consider the application of a model-averaging approach between the two candidate models. On the contrary, when only a simple-regression analysis with drainage area is performed, the application of TK is recommended, especially in the presence of high degrees of spatial correlation. In this context, an interesting issue to be investigated in future studies could be the use of GLS (instead of OLS) for identifying the regional power-law model between flood quantiles and catchment descriptors when applying TK.

The findings outlined are valid for the simplified situation that is investigated here and suggest that, for a homogeneous region, the benefits of applying traditional GLS over TK are rather limited even in estimating theoretical flood quantiles. Therefore, in such a situation, the application of TK is suggested over traditional GLS. Further analyses regarding heterogeneous regions (i.e. the presence of a regional spatial signal), flood series generated as a function of catchment descriptors (see e.g. Stedinger and Tasker 1986), fully ungauged conditions (no information is available upstream or downstream within a given catchment, i.e. the absence of nested catchments) and/or the application of more recent versions of GLS (e.g. with removal of redundant nested gauged catchments, B-GLS), as well as sensitivity analyses of the results to the presence of different record lengths in the AMS series are suggested for future studies.

Acknowledgements

The dataset used for the analyses of this study can be accessed at the following repository: https://zenodo.org/badge/latestdoi/194627072. All figures included in the study were produced with the use of free and open-source software (i.e. Quantum GIS Geographic Information System – Open Source Geospatial Foundation Project, http://qgis.osgeo.org; and the R Project for Statistical Computing, https://www.R-project.org/).

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Simone Persiano http://orcid.org/0000-0002-9857-738X
Jose Luis Salinas http://orcid.org/0000-0002-3045-9811
Jery Russell Stedinger http://orcid.org/0000-0002-7081-729X
William H. Farmer http://orcid.org/0000-0002-2865-2196
David Lun http://orcid.org/0000-0003-0382-3371
Alberto Viglione http://orcid.org/0000-0002-7587-4832
Günter Blöschl http://orcid.org/0000-0003-2227-8225
Attilio Castellarin http://orcid.org/0000-0002-6111-0612

References

Acreman, M.C. and Wiltshire, S.E., 1989. The regions are dead: long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. *In*: L. Roald, K. Nordseth, and K.A. Hassel, eds. *FRIENDS in Hydrology*. Wallingford, UK: IAHS Press, 187.

Archfield, S.A., *et al.*, 2013. Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach? *Hydrology and Earth System Sciences*, 17 (4), 1575–1588. doi:10.5194/hess-17-1575-2013

Archfield, S.A. and Vogel, R.M., 2010. Map correlation method: selection of a reference streamgage to estimate daily streamflow at ungaged catchments. *Water Resources Research*, 46 (10). doi:10.1029/2009WR008481

Blöschl, G., *et al.*, 2013. *Runoff prediction in ungauged basins: synthesis across processes, places and scales*. Cambridge, UK: Cambridge University Press.

Burn, D.H., 1990. Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26 (10), 2257–2265. doi:10.1029/WR026i010p02257

Castellarin, A., 2007. Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resources Research*, 43 (4), 4. doi:10.1029/2005WR004384

Castellarin, A., *et al.*, 2018. Prediction of streamflow regimes over large geographical areas: interpolated flow–duration curves for the Danube region. *Hydrological Sciences Journal*, 63 (6), 845–861. doi:10.1080/02626667.2018.1445855

Castellarin, A., Burn, D.H., and Brath, A., 2008. Homogeneity testing: how homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology*, 360 (1–4), 67–76. doi:10.1016/j.jhydrol.2008.07.014

Castiglioni, S., *et al.*, 2011. Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging. *Hydrology and Earth System Sciences*, 15 (3), 715–727. doi:10.5194/hess-15-715-2011

Castiglioni, S., Castellarin, A., and Montanari, A., 2009. Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. *Journal of Hydrology*, 378 (3–4), 272–280. doi:10.1016/j.jhydrol.2009.09.032

Ceola, S., *et al.*, 2018. Hydro-power production and fish habitat suitability: assessing impact and effectiveness of ecological flows at regional scale. *Advances in Water Resources*, 116, 29–39. doi:10.1016/j.advwatres.2018.04.002

Chambers, J.M., 1992. Linear models. *In*: J.M. Chambers and T.J. Hastie, eds. *Statistical models* (pp. 95–144). Pacific Grove, CA: S. Wadsworth Brooks/Cole.

Chokmani, K. and Ouarda, T.B.M.J., 2004. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resources Research*, 40 (12), W12514. doi:10.1029/2003WR002983

Cressie, N.A.C., 1993. *Statistics for spatial data Wiley series in probability and mathematical statistics: applied probability and statistics*. New York, NY: J. Wiley.

Dalrymple, T., 1960. Flood frequency analysis. US geological survey, water supply paper, 1543 A.

de Lavenne, A., *et al.*, 2016. Transferring measured discharge time series: large-scale comparison of top-kriging to geomorphology-based inverse modeling. *Water Resources Research*, 52 (7), 5555–5576. doi:10.1002/2016WR018716

De Marsily, G., 1986. *Quantitative hydrogeology: groundwater hydrology for engineers*. Orlando, FL: Academic Press.

Draper, N.R. and Smith, H., 1981. *Applied regression analysis*. 2nd. New York, USA: JohnWiley & Sons, Inc.

Durocher, M., *et al.*, 2019. Estimating flood quantiles at ungauged sites using nonparametric regression methods with spatial components. *Hydrological Sciences Journal*, 64 (9), 1056–1070. doi:10.1080/02626667.2019.1620952

Eng, K., Chen, Y., and Kiang, J., 2009. *User's guide to the weighted-multiple-linear regression program (WREG Version 1.0)*. Techniques and methods. U.S. geological survey.

England, J.F., *et al.* 2018. Guidelines for determining flood flow frequency - bulletin 17C. U.S. Geological survey techniques and methods. book 4, chap. B5.

Farmer, W.H., 2016. Ordinary kriging as a tool to estimate historical daily streamflow records. *Hydrology and Earth System Sciences*, 20 (7), 2721–2735. doi:10.5194/hess-20-2721-2016

Farmer, W.H., 2017. *WREG: USGS WREG v. 2.02. R package version 2.02*. GitHub repository. Available from: https://github.com/USGS-R/WREG (accessed 23 February 2021).

Feaster, T.D. and Tasker, G.D., 2002. *Techniques for estimating the magnitude and frequency of floods in rural basins of south carolina, 1999*. Technical report, US department of the interior, US geological survey.

Gotvald, A.J., Feaster, T.D., and Weaver, J.C., 2009. *Magnitude and frequency of rural floods in the Southeastern United States, 2006: vol. 1, Georgia, 2009–5043, US Geological Survey*. Reston, Virginia, USA: U. S. Geological Survey.

Greene, W.H., 2002. *Econometric analysis*. 5th ed. Upper Saddle River, N. J: Prentice Hall.

Grehys, G.D.R.E.S., 1996. Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology (Amsterdam)*, 186 (1–4), 63–84.

Griffis, V.W. and Stedinger, J.R., 2007a. Log-Pearson type 3 distribution and its application in flood frequency analysis. I: distribution characteristics. *Journal of Hydrologic Engineering*, 12 (5), 482–491. doi:10.1061/(ASCE)1084-0699(2007)12:5(482)

Griffis, V.W. and Stedinger, J.R., 2007b. The use of GLS regression in regional hydrologic analyses. *Journal of Hydrology*, 344 (1–2), 82–95. doi:10.1016/j.jhydrol.2007.06.023

Griffis, V.W. and Stedinger, J.R., 2009. Log-Pearson type 3 distribution and its application in flood frequency analysis. III: sample skew and weighted skew estimators. *Journal of Hydrologic Engineering*, 14 (2), 121–130. doi:10.1061/(ASCE)1084-0699(2009)14:2(121)

Gruber, A.M., Reis, D.S., and Stedinger, J.R., 2007. Models of regional skew based on Bayesian GLS regression. *In*: *World environmental & water resources conference*, Tampa, FL: ASCE EWRI.

Gruber, A.M. and Stedinger, J.R., 2008. Models of LP3 regional skew, data selection, and Bayesian GLS regression. *In*: *World environmental & water resources conference 2008 AHUPUA'A*. Honolulu, Hawaii.

Grygier, J.C., Stedinger, J.R., and Yin, H.-B., 1989. A generalized maintenance of variance extension procedure for extending correlated series. *Water Resources Research*, 25 (3), 345–349. doi:10.1029/WR025i003p00345

Hardison, C.H., 1971. Prediction error of regression estimates of streamflow characteristics at ungaged sites, Chapter C. US geological survey professional paper. 750–C.

Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982. Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18 (1), 107–121. doi:10.1029/WR018i001p00107

Hosking, J.R.M. and Wallis, J.R., 1988. The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research*, 24 (4), 588–600. doi:10.1029/WR024i004p00588

Hosking, J.R.M. and Wallis, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resources Research*, 29 (2), 271–281.

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

Interagency Advisory Committee on Water Data, 1982. *Guidelines for determining flood-flow frequency. Bulletin 17B of the hydrology subcommittee*. Office of water data coordination: U.S. geological survey, Reston, VA.

Isaaks, E.H. and Srivastava, R.M., 1990. *Applied geostatistics*. USA: OUP.

Johnston, J., 1972. *Econometric methods*. New York, USA: McGraw-Hill.

Journel, A.G. and Huijbregts, C.J., 1978. *Mining geostatistics*. London, UK: Academic Press.

Kite, G.W., 1975. Confidence limits for design events. *Water Resources Research*, 11 (1), 48–53. doi:10.1029/WR011i001p00048

Kite, G.W., 1976. Reply to comment on confidence limits for design events. *Water Resources Research*, 12 (4), 826. doi:10.1029/WR012i004p00826

Kroll, C.N. and Stedinger, J.R., 1998. Regional hydrologic analysis: ordinary and generalized least squares revisited. *Water Resources Research*, 34 (1), 121–128. doi:10.1029/97WR02685

Laaha, G., *et al.*, 2013. Spatial prediction of stream temperatures using top-kriging with an external drift. *Environmental Modeling and Assessment*, 18, 671–683. doi:10.1007/s10666-013-9373-3.

Laaha, G., Skøien, J.O., and Blöschl, G., 2014. Spatial prediction on river networks: comparison of top-kriging with regional regression. *Hydrological Processes*, 28 (2), 315–324. doi:10.1002/hyp.9578

Laio, F., *et al.*, 2011. Spatially smooth regional estimation of the flood frequency curve (with uncertainty). *Journal of Hydrology*, 408 (1–2), 67–77. doi:10.1016/j.jhydrol.2011.07.022

Lettenmaier, D.P., Wallis, J.R., and Wood, E.F., 1987. Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, 23 (2), 313–323. doi:10.1029/WR023i002p00313

Madsen, H., *et al.*, 2002. Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics. *Water Resources Research*, 38 (11), 21. doi:10.1029/2001WR001125

Madsen, H. and Rosbjerg, D., 1997. Generalized least squares and empirical bayes estimation in regional partial duration series index-flood modeling. *Water Resources Research*, 33 (4), 771–781. doi:10.1029/96WR03850

Martins, E.S. and Stedinger, J.R., 2002. Cross correlations among estimators of shape. *Water Resources Research*, 38 (11), 34. doi:10.1029/2002WR001589

Matalas, N.C. and Langbein, W.B., 1962. Information content of the mean. *Journal of Geophysical Research*, 67 (9), 3441–3448. doi:10.1029/JZ067i009p03441

Matheron, G., 1971. The theory of regionalized variables and its applications. Cahiers du centre de morphologie mathematique no.5. Fontainebleau, France.

Merz, R., Blöschl, G., and Humer, G., 2008. National flood discharge mapping in Austria. *Natural Hazards*, 1 (1), 53–72. doi:10.1007/s11069-007-9181-7

Nash, J. and Sutcliffe, J., 1970. River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, 10 (3), 282–290. doi:10.1016/0022-1694(70)90255-6

NERC, 1975. *Flood studies report*. London, UK: NERC (Natural Environment Research Council).

Ouarda, T.B., *et al.*, 2001. Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, 254 (1), 157–173. doi:10.1016/S0022-1694(01)00488-7

Pandey, G. and Nguyen, -V.-T.-V., 1999. A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, 225 (1–2), 92–101. doi:10.1016/S0022-1694(99)00135-3

Parajka, J., *et al.*, 2015. The role of station density for predicting daily runoff by top-kriging interpolation in Austria. *Journal of Hydrology and Hydromechanics*, 63 (3), 228–234. doi:10.1515/johh-2015-0024

Patil, S. and Stieglitz, M., 2012. Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrology and Earth System Sciences*, 16 (2), 551–562. doi:10.5194/hess-16-551-2012

Persiano, S., *et al.*, 2016. Climate, orography and scale controls on flood frequency in triveneto (Italy). *In*: *IAHS-AISH proceedings and reports, volume 373*. Bochum, Germany.

Pugliese, A., *et al.*, 2016. Regional flow duration curves: geostatistical techniques versus multivariate regression. *Advances in Water Resources*, 96, 11–22. doi:10.1016/j.advwatres.2016.06.008

Pugliese, A., *et al.*, 2018. A geostatistical data-assimilation technique for enhancing macro-scale rainfall–runoff simulations. *Hydrology and Earth System Sciences*, 22 (9), 4633–4648. doi:10.5194/hess-22-4633-2018.

Pugliese, A., Castellarin, A., and Brath, A., 2014. Geostatistical prediction of flow–duration curves in an index-flow framework. *Hydrology and Earth System Sciences*, 18 (9), 3801–3816. doi:10.5194/hess-18-3801-2014

R Core Team, 2016. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Reis, D.S.J., *et al.*, 2020. Operational Bayesian GLS regression for regional hydrologic analyses. *Water Resources Research*, 56 (8), 8. doi:10.1029/2019WR026940

Reis, D.S.J., Stedinger, J.R., and Martins, E.S., 2005. Bayesian generalized least squares regression with application to log pearson type 3 regional skew estimation. *Water Resources Research*, 41 (10). doi:10.1029/2004WR003445

Riggs, H.C., 1973. Regional analysis of stramflow characteristics. US geological survey techniques of water resources investigation.

Rosbjerg, D., 2007. Regional flood frequency analysis. *In*: O.F. Vasiliev, *et al*., eds. *Extreme hydrological events: new concepts for security. NATO Science Series*, Vol. 78, 151–171. Springer, Dordrecht. doi:10.1007/978-1-4020-5741-0_12

Serinaldi, F., 2015. Dismissing return periods! *Stochastic Environmental Research and Risk Assessment*, 29 (4), 1179–1189. doi:10.1007/s00477-014-0916-1

Sivapalan, M., *et al*. 2003. IAHS decade on predictions in ungauged basins (pub), 2003–2012: shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48 (6), 857–880. doi:10.1623/hysj.48.6.857.51421

Skøien, J.O., 2014. *rtop: interpolation of data with variable spatial support*. R package version, 3–45.

Skøien, J.O. and Blöschl, G., 2007. Spatiotemporal topological kriging of runoff time series. *Water Resources Research*, 43 (9). doi:10.1029/2006WR005760

Skøien J.O., Blöschl G., Laaha G., Pebesma E., Parajka J., and Viglione A., 2014. rtop: An R package for interpolation of data with a variable spatial support, with an example from river networks. *Computers & Geosciences*, 67, 180–190. doi:10.1016/j.cageo.2014.02.009

Skøien, J.O., Merz, R., and Blöschl, G., 2006. Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences*, 10 (2), 277–287. doi:10.5194/hess-10-277-2006

Stedinger, J.R., 1983. Estimating a regional flood frequency distribution. *Water Resources Research*, 19 (2), 503–510. doi:10.1029/WR019i002p00503

Stedinger, J.R. and Lu, L.H., 1995. Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics*, 9 (1), 49–75. doi:10.1007/BF01581758

Stedinger, J.R. and Tasker, G.D., 1985. Regional hydrologic analysis: 1. Ordinary,weighted, and generalized least squares compared. *Water Resources Research*, 21 (9), 1421–1432. doi:10.1029/WR021i009p01421

Stedinger, J.R. and Tasker, G.D., 1986. Regional hydrologic analysis, 2, model-error estimators, estimation of Sigma and Log-Pearson Type 3 distributions. *Water Resources Research*, 22 (10), 1487–1499. doi:10.1029/WR022i010p01487

Tasker, G.D., 1980. Hydrologic regression with weighted least squares. *Water Resources Research*, 16 (6), 1107–1113. doi:10.1029/WR016i006p01107

Tasker, G.D., Eychaner, J.H., and Stedinger, J., 1986. Application of generalized least squares in regional hydrologic regression analysis. US geological survey. *Water Supply Paper*, 2310 (1), 107–115.

Tasker, G.D. and Stedinger, J.R., 1986. Regional skew with weighted LS regression. *Journal of Water Resources Planning and Management*, 112 (2), 225–236.

Tasker, G.D. and Stedinger, J.R., 1989. An operational GLS model for hydrologic regression. *Journal of Hydrology*, 111 (1–4), 361–375.

Thomas, D.M. and Benson, M.A., 1970. Generalization of streamflow characteristics from drainage basin characteristics. US geological survey. Water supply paper 1975.

Veilleux, A.G., Stedinger, J.R., and Lamontagne, J.R., 2011. Bayesian WLS/GLS regression for regional skewness analysis for regions with large cross-correlations among flood flows. *In*: *World environmental and water resources congress 2011 bearing knowledge for sustainability*. Palm Springs, California: ASCE

Vogel, R.M. and Stedinger, J.R., 1985. Minimum variance streamflow record augmentation procedures. *Water Resources Research*, 21 (5), 715–723. doi:10.1029/WR021i005p00715

Vormoor, K., *et al*., 2011. Geostatistical regionalization of daily runoff forecasts in Norway. *International Journal of River Basin Management*, 9 (1), 3–15. doi:10.1080/15715124.2010.543905

Weisberg, S., 1985. *Applied linear regression*. 2nd. New York, USA: John Wiley & Sons, Inc.

Appendix Methods.

A.1 Generalized least squares (GLS)

Multi-linear regressions are one of the most commonly used approaches for estimating the *T*-year flood, or design flood (i.e. flood quantile associated with a given return period *T*), for ungauged catchments. Such models relate observable catchment characteristics (e.g. drainage area, mean annual precipitation, catchment slope, etc.) to streamflow characteristics (e.g. flood quantiles), as follows:

$$\hat{\boldsymbol{Y}} = \boldsymbol{X}\beta + \varepsilon \tag{A1}$$

where $\hat{\boldsymbol{Y}}$ is a $(n \times 1)$ vector of streamflow characteristics at *n* sites, $\boldsymbol{X}$ is the $(n \times k)$ matrix of $(k-1)$ catchment characteristics augmented with a column of 1, β is the $(k \times 1)$ vector of regression parameters, and ε is the $(n \times 1)$ vector of total errors. In particular, each component of ε is envisioned as a random variable with zero mean and variance equal to σ_ϵ^2.

Traditionally, the regression parameters are estimated using least squares procedures. The general expression that indicates the estimation of regression parameters is given by:

$$\hat{\beta} = \left(\boldsymbol{X}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{Y}} \tag{A2}$$

where $\boldsymbol{X}^T$ indicates the transposition of matrix $\boldsymbol{X}$, and $\boldsymbol{\Lambda}^{-1}$ is the inverse of the weighting matrix $\boldsymbol{\Lambda}$. Once estimated, $\hat{\beta}$ can be used to compute the regression estimates $\hat{\boldsymbol{Y}}$ (i.e. $\hat{y}_i$ at the *i*th gauge).

Ordinary least squares (OLS) regression is traditionally used for estimating the regression parameters β, by setting $\boldsymbol{\Lambda} = \mathbf{I}$ in Equation (A2), where $\mathbf{I}$ indicates the identity matrix (Thomas and Benson 1970, Hardison 1971, Riggs 1973). To estimate $\hat{\beta}$, OLS considers the total errors of the model to be homoscedastic and independently distributed (Riggs 1973); however, these assumptions are often violated in hydrological applications, where different streamgauges usually have different record lengths, and concurrent streamflows observed at different gauges in a region are often cross-correlated (Tasker and Stedinger 1989). If not properly represented in a regional analysis, cross-correlation affects the precision of regression parameters, and the estimators of flood characteristics are inefficient in that they are not as accurate as they could be (i.e. the higher the cross-correlation, the higher the model errors; see Stedinger and Tasker 1985).

To deal with situations where regression model residuals are heteroscedastic and cross-correlated, the regression parameters can be estimated by means of a generalized least squares (GLS) technique, which is the best linear unbiased estimator (BLUE) when the true residual error covariance matrix $\boldsymbol{\Lambda}$ is known (Johnston 1972). Unfortunately, in general $\boldsymbol{\Lambda}$ is unknown and an estimator must be employed. Stedinger and Tasker (1985, 1986) and Tasker and Stedinger (1989) addressed the issue of the loss of efficiency from using an approximation of the covariance matrix when estimating streamflow characteristics in non-nested ungauged (i.e. fully ungauged) basins: they developed a procedure for estimating $\boldsymbol{\Lambda}$ accounting for both correlated streamflows and time-sampling errors. The Stedinger-Tasker GLS improves the representation of the overall regression error ε, by assuming that it equals the sum of the sampling error η for the estimates of the streamflow statistics (e.g. flood statistics), and the modelling error δ in modelling the theoretical streamflows (e.g. flood quantiles) across catchments. Kroll and Stedinger (1998) demonstrated that when estimating flood quantiles with the Stedinger-Tasker GLS $\boldsymbol{\Lambda}$-estimator, estimation of $\boldsymbol{\Lambda}_{GLS}$ results in little loss of efficiency. More recently, Reis *et al.* (2005) considered Bayesian estimators of model error and regression parameters. Note that, unlike feasible GLS (FGLS) estimation (which can be applied to any frequency distribution; see e.g. Greene 2002), the Stedinger-Tasker GLS regression is applied for a log-Pearson type III (LP3) frequency analysis (see Bulletin 17C by England *et al.* 2018), which is the reference procedure for estimating streamflow characteristics in ungauged catchments in the US (Eng *et al.* 2009).

In particular, the Stedinger-Tasker GLS (hereinafter referred to as GLS, for the sake of brevity) computes the regression parameters by setting $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_{GLS}$ in Equation (A2), where $\boldsymbol{\Lambda}_{GLS}$ contains the estimates of the covariances of ε_i among gauged sites. The main diagonal elements of $\boldsymbol{\Lambda}_{GLS}$ include a part associated with the model error δ_i and all elements include the effect of the time-sampling errors η_i. For streamflow characteristics that are computed from a log-Pearson type III (LP3) frequency analysis (see Bulletin 17B of the Interagency Advisory Committee on Water Data 1982, and the most recent Bulletin 17C by, England *et al.* 2018), Tasker and Stedinger (1989) propose to estimate $\boldsymbol{\Lambda}_{GLS}$ as follows:

$$\hat{\Lambda}_{GLS,ij} = \begin{cases} \sigma_{\delta i}^2 + \frac{\sigma_i^2}{m_i}\left[1 + K_i G_i + 0.5K_i^2\left(1 + 0.75G_i^2\right)\right], \; if \; i = j \\ \frac{\hat{\rho}_{ij}\sigma_i\sigma_j m_{ij}}{m_i m_j}\left[1 + 0.5K_i G_i + 0.5\,K_j G_j + 0.5K_i K_j\left(\hat{\rho}_{ij} + 0.75\,G_i G_j\right)\right], \; if \; i \neq j \end{cases} \tag{A3}$$

where i and j are indices of the gauged sites in the region of interest, $\sigma_{\delta i}^2$ is the model-error variance at site i, K_i and K_j are the LP3 standard deviates (function of exceedance probability, i.e. return period T, and at-site empirical skew g; see Bulletin 17B of the Interagency Advisory Committee on Water Data 1982, and Bulletin 17C by England *et al.* 2018), G_i and G_j are the corresponding skew values (i.e. equal to either at-site empirical skews g, or weighted skew G_w), m_i and m_j are the corresponding record lengths, m_{ij} is the concurrent record length, and $\hat{\rho}_{ij}$ is an estimated value for the cross-correlation of time series of streamflow values used to calculate the streamflow characteristics at gauges i and j.

As reported in Bulletin 17B of the Interagency Advisory Committee on Water Data (1982), and consistent with the more recent Bulletin 17C (England *et al.* 2018), the weighted skew G_{wi} for the i-th gauged site can be computed as:

$$G_{w,i} = \omega_i\, g_i + (1 - \omega_i)\, G_{R,i} \tag{A4}$$

where $G_{R,i}$ is the regional skew estimate for the i-th gauged site, and

$$\omega_i = \frac{MSE(G_R)}{MSE(g_i) + MSE(G_R)} \tag{A5}$$

where $MSE(g_i)$ is the estimated mean square error of the skew value at the i-th gauged site, and $MSE(G_R)$ is the estimated mean square error of the regional skew values. Several methods are available for the determination of G_R values: Tasker and Stedinger (1986) developed a weighted least squares procedure for estimating regional skewness coefficients based on sample skewness coefficients for the logarithms of annual peak-discharge data; more recently, a Bayesian-GLS (B-GLS) regression model for regional skewness analyses was introduced (see e.g. Reis *et al.* 2005, Gruber *et al.* 2007, Gruber and Stedinger 2008, Veilleux *et al.* 2011; Bulletin 17C by England *et al.* 2018).

Empirical estimates of ρ_{ij} are imprecise due to the short record lengths of observed flows. To overcome this problem, values of the cross-correlation are usually approximated by referring to the nonlinear relationship introduced by Tasker and Stedinger (1989), which is useful for smoothing the sample correlations as a function of distance between gauges:

$$\hat{\rho}_{ij} = \theta^{\frac{d_{ij}}{1+\alpha\, d_{ij}}} = e^{\frac{\ln(\theta)\, d_{ij}}{1+\alpha\, d_{ij}}} \tag{A6}$$

where d_{ij} is the distance between gauges i and j, and θ and α are the dimensionless model parameters estimated from data. In particular, $\hat{\rho}_{ij}$ is a convex, monotonically decreasing function of d_{ij} when $0 < \theta < 1$ and $\alpha > 0$.

As reported by Stedinger and Tasker (1985), the $\hat{\beta}$ values of Equation (A2) for GLS and the $\sigma_{\delta i}^2$ values in Equation (A3) are jointly determined by recursively searching for a non-negative solution to the following equation:

$$\left(\hat{Y} - X\hat{\beta}\right)^T \Lambda_{GLS}^{-1}\left(\hat{Y} - X\hat{\beta}\right) = n - (k+1) \tag{A7}$$

As Equation (A3) does not account for the error associated with estimating G, Griffis and Stedinger (2007b) introduced a modified version of $\mathbf{\Lambda}_{GLS}$, named $\mathbf{\Lambda}_{GLS,skew}$, which accounts for the uncertainty in the skew estimates. For the estimation of $\mathbf{\Lambda}_{GLS,skew}$ values, Griffis and Stedinger (2007b) consider additional terms, such as the partial derivatives for the gauges (calculated from the approximation for K given by Kite 1975, 1976), the covariance between the skew values at the different gauged sites, which, in turn, depends on the correlation $\rho_{g_i g_j}$ between skew values (estimated by Martins and Stedinger 2002) and on the variances of the skew values at gauges i and j (estimated by Griffis and Stedinger 2009). Further details concerning the computation of $\mathbf{\Lambda}_{GLS,skew}$ can be found in Griffis and Stedinger (2007b) and in the User's Guide to the Weighted-Multiple Linear Regression Program (WREG) (Eng *et al.* 2009), which is the software used by the US Geological Survey for estimating streamflow characteristics at ungauged basins in the United States.

GLS also provides estimates of the model error variance, σ_δ^2. In particular, the average variance of prediction, AVP, can be computed as follows (Tasker and Stedinger 1986):

$$AVP = \sigma_\delta^2 + \frac{1}{n}\sum_{p=1}^{n} \mathrm{x_p}\left(\boldsymbol{X}^T\boldsymbol{\Lambda}_{GLS}^{-1}\boldsymbol{X}\right)^{-1}\mathrm{x_p}^T \tag{A8}$$

where x_p is a vector containing the values of the independent variables of the p-th gauge augmented by a value of 1 (see also Eng *et al.* 2009).

A.2 Top-kriging

Topological kriging (or top-kriging) is a geostatistical procedure developed by Skøien *et al.* (2006) for the prediction of hydrological variables, accounting for catchment area and catchment nestedness (which are not taken into account by ordinary kriging). Top-kriging produces predictions of hydrological variables at ungauged sites with a linear combination of the empirical information collected at neighbouring gauging stations. Using this method, the unknown value of the streamflow index of interest at prediction location x_0, $Z(x_0)$, can be estimated as a weighted average of the variable measured at other sites within a neighbourhood:

$$Z(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i) \tag{A9}$$

where λ_i is the kriging weight for the empirical value $Z(x_i)$ at location x_i, and n is the number of neighbouring stations used for interpolation. Kriging weights λ_i can be found by solving the typical ordinary kriging linear system with the constraint of unbiased estimation:

$$\begin{cases} \sum_{j=1}^{n} \gamma_{i,j}\lambda_j + \phi = \gamma_{0,i} \qquad i = 1,\ldots,n \\ \sum_{j=1}^{n} \lambda_j = 1 \end{cases} \tag{A10}$$

where ϕ is the Lagrange parameter and $\gamma_{i,j}$ is the semi-variance between catchments i and j (Isaaks and Srivastava 1990). The relationship between semi-variance and distance is represented by the semi-variogram (hereinafter referred to as variogram), which represents the spatial variability of the regionalized variable Z.

Top-kriging is based on block kriging (see e.g. Journel and Huijbregts 1978), which considers the variable defined over a non-zero support S; in top-kriging, S coincides with the catchment drainage area A (Cressie 1993, Skøien *et al.* 2006). In particular, the point variable $Z(x)$ is averaged over the drainge area A to obtain the spatially averaged variable $Z_r(x)$:

$$Z_r(x) = \frac{1}{A}\int_A Z(x)dx \tag{A11}$$

In this way, the kriging system of Equation (A10) remains the same, but the semi-variances between the measurements need to be obtained by regularization, i.e. smoothing the point variogram over the support area. In particular, considering two measurements with catchment area A_i and A_j, respectively, the regularization consists of assuming the existence of a point variogram $\gamma_p(h)$, where $h = |x_i - x_j|$ represents the Euclidean distance (evaluated on a horizontal plane) between two generic position vectors x_i and x_j within the corresponding catchments, and evaluating the semi-variance $\gamma_{i,j}^r$ between the two measurements as:

$$\gamma^r_{i,j} = \frac{1}{A_iA_j}\int_{A_i}\int_{A_j}\gamma_p(|x_i - x_j|)dx_idx_j - \frac{1}{2}\left\{\frac{1}{A_i^2}\int_{A_i}\int_{A_i}\gamma_p(|x_i - x_j|)dx_idx_j + \frac{1}{A_j^2}\int_{A_j}\int_{A_j}\gamma_p(|x_i - x_j|)dx_idx_j\right\} \quad \text{(A12)}$$

where the first part integrates all the variance between the two catchments, while the second one (in curly brackets) subtracts the averaged variance within the catchments (i.e. representing the smoothing effect of the support, which indicates that the variance of the averaged variable decreases as the support area increases; Skøien *et al.* 2006). In this way, Equation (A12) can be used to evaluate the variogram of the averaged variable from the point variogram. Then, $\gamma^r_{i,j}$ can be inserted into the kriging matrix of Equation (A10) and the kriging system can be solved to compute the weights λ_j. The integration shown in Equation (A12) is performed over the catchment area that drains to a particular location on the stream network (i.e. the outlet of the target catchment). Computationally, the catchment area is discretized by a grid and the integrals in Equation (A12) are replaced by sums (Skøien 2014).

Top-kriging also accounts for the possible presence of a nugget effect in the point variogram, i.e. a discontinuity close to the origin of the variogram (at a distance infinitesimally larger than zero), caused by measurement errors and variability at scales much smaller than the distance between measurements. Many variables of interest in hydrological applications, such as streamflow data, are likely to have a nugget effect. As the direct regularization with Equation (A12) would make the nugget vanish (even for small catchments), Skøien *et al.* (2006) suggest regularizing the nugget separately by considering the nugget variance as the variance of a spatially independent random variable and adding the regularized nugget effect to the regularized variogram of Equation (A12). In particular, the regularized nugget variance for two catchments of different size, $C_0(A_i, A_j)$, is computed as follows:

$$C_0(A_i, A_j) = \frac{1}{2}\left(\frac{C_{0p}}{A_i} + \frac{C_{0p}}{A_j} - \frac{2\,C_{0p}\,Meas(A_i \cap A_j)}{A_iA_j}\right) \quad \text{(A13)}$$

where C_{0p} is the nugget variance of the point variogram, and $Meas(A_i \cap A_j)$, representing the area shared by the two catchments with areas A_i and A_j, accounts for the nestedness of the two: if the catchments are nested then $Meas(A_i \cap A_j) = min(A_i, A_j)$, otherwise $Meas(A_i \cap A_j) = 0$. In other words, the effect of the point nugget variance C_{0p} is dependent on the catchment size and degree of overlapping (i.e. nestedness). More details about the regularization of the nugget effect can be found in Skøien *et al.* (2006).

In practice, the application of top-kriging requires the preliminary estimation of a point variogram $\gamma_p(h)$. To this aim, the sample point variogram is computed by using the binned variogram technique (for details, see Skøien 2014), which aggregates sample points in distance and area classes or bins under the hypothesis of isotropy (i.e. the variogram does not vary with direction). The sample point variogram can then be modelled through a suitable theoretical model (e.g. exponential, fractal, Gaussian, spherical, etc.).

Like any other kriging method, top-kriging provides an estimate of the kriging variance σ^2_R (see e.g. Skøien *et al.* 2006), which represents the uncertainty of the estimates at any location:

$$\sigma^2_R = \sum_{j=1}^{n}\gamma_{j,0}\lambda_j + \phi \quad \text{(A14)}$$

where $\gamma_{j,0}$ is the gamma value between the target catchment and the neighbouring catchments, and ϕ is the Lagrange parameter.

Thanks to its structure, top-kriging has the merit (over traditional kriging methods, i.e. ordinary kriging) to account for the influence of catchment area and the nested structure of catchments on the evaluation of the weights λ_j. Indeed, while ordinary kriging would assign the same weights λ_j for all the neighbouring catchments having the same centre-to-centre distance to the target one, top-kriging weighs them differently. In particular, top-kriging assigns larger weights to larger catchments (which are regarded as the most certain, or as having the least uncertain measurements in comparison to the mean). Moreover, for identical catchment areas and centre-to-centre distances, a sub-catchment of the target catchment is given a larger weight than a non-nested catchment; on the other hand, given two neighbours with same area and same centre-to-centre distances to the target catchment, more weight is attached to the catchment into which the target catchment drains (Skøien *et al.* 2006).