



Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach?

S. A. Archfield¹, A. Pugliese², A. Castellarin², J. O. Skøien³, and J. E. Kiang⁴

¹US Geological Survey, Northborough, MA, USA

²Department DICAM, School of Civil Engineering, University of Bologna, Bologna, Italy

³Institute for Environment and Sustainability, Joint Research Centre, European Commission, Italy

⁴US Geological Survey, Reston, VA, USA

Correspondence to: S. A. Archfield (sarch@usgs.gov)

Abstract. In the United States, estimation of flood frequency quantiles at ungauged locations has been largely based on regional regression techniques that relate measurable catchment descriptors to flood quantiles. More recently, spatial interpolation techniques of point data have been shown to be effective for predicting streamflow statistics (i.e., flood flows and low-flow indices) in ungauged catchments. Literature reports successful applications of two techniques, canonical kriging, CK (or physiographical-space-based interpolation, PSBI), and topological kriging, TK (or top-kriging). CK performs the spatial interpolation of the streamflow statistic of interest in the two-dimensional space of catchment descriptors. TK predicts the streamflow statistic along river networks taking both the catchment area and nested nature of catchments into account. It is of interest to understand how these spatial interpolation methods compare with generalized least squares (GLS) regression, one of the most common approaches to estimate flood quantiles at ungauged locations. By means of a leave-one-out cross-validation procedure, the performance of CK and TK was compared to GLS regression equations developed for the prediction of 10, 50, 100 and 500 yr floods for 61 streamgauges in the southeast United States. TK substantially outperforms GLS and CK for the study area, particularly for large catchments. The performance of TK over GLS highlights an important distinction between the treatments of spatial correlation when using regression-based or spatial interpolation methods to estimate flood quantiles at ungauged locations. The analysis also shows that coupling TK with CK slightly improves the performance of TK; however, the improvement is marginal when compared to the improvement in performance over GLS.

1 Introduction

An important application of hydrologic science is to provide an accurate estimate of the design flood (i.e., the flood quantile associated with a given non-exceedance probability, usually expressed in terms of return period) at a site which lacks sufficient measured hydrological information (see Sivapalan et al., 2003). This problem has been addressed by adopting a number of different approaches that are all characterized by the same *fil rouge*: transferring hydrologic information or knowledge from gauged catchments to ungauged or poorly gauged ones (e.g., Blöschl and Sivapalan, 1997; Merz and Blöschl, 2008; Pallard et al., 2009).

One such widely used approach to predict the design flood in ungauged catchments is regional flood frequency analysis (RFFA). RFFA pools flood information across hydrologically homogenous sites and then transfers this information to an ungauged location or a location with data lengths considered too short to provide an estimate of the desired design flood. Although several approaches of RFFA have been proposed (traditional approaches are illustrated for instance in Dalrymple, 1960; Burn, 1990; Gabriele and Arnell, 1991; Stedinger et al., 1993; Hosking and Wallis, 1997; Castellarin et al., 2001; Merz and Blöschl, 2005), standard accepted techniques have been detailed by Hosking and Wallis (1997) and in the Flood Estimation Handbook (FEH, 1999); in the United States, the US Geological Survey utilizes generalized least squares (GLS) regression (RFFA-GLS) as the standard method for the estimation of flood quantiles at ungauged sites.

RFFA is a mature discipline and some aspects are considered to be so well studied that additional investigation of

those aspects would result in limited improvements. Examples include the estimation of the regional parent distribution (how to pool the information found at gauged locations) or statistical homogeneity testing (Viglione et al., 2007). However, some aspects of RFFA have still yet to be resolved and further research could result in substantial improvements to predictions of the design flood in ungauged catchments (Castiglioni et al., 2011). Notably, the determination and use of hydrologically homogenous groups is one such topic (see e.g., McDonnell and Woods, 2004; Di Prinzio et al., 2011; Castiglioni et al., 2011) and is the focus of this paper.

In looking to maximize the information contained within a hydrologically similar region, there has been increased attention to the application of geostatistical techniques to RFFA. Recent studies show that these techniques, which have been originally adopted for the spatial interpolation of point data (see e.g., kriging interpolators, De Marsily, 1986; De Marsily and Ahmed, 1987), can be effectively applied for regionalization of a number of hydrologic indices (Skøien et al., 2006; Skøien and Blöschl, 2007; Chokmani and Ouarda, 2004) and even hydrologic time series (Skøien and Blöschl, 2007). To this end, two geostatistical approaches have been introduced and studied in the literature for the estimation of various properties of streamflow (including design floods) at ungauged locations. The first approach, named topological kriging or top-kriging, exploits the nested structure of the study area for spatially interpolating the streamflow index of interest (e.g., flood quantiles, low-flow indices, etc.) along the stream network (Skøien et al., 2006; Skøien and Blöschl, 2007). The second approach, termed canonical kriging (or physiographical-space-based interpolation, PSBI), interpolates the streamflow index of interest using a two-dimensional spatial representation of the physiographical and climatic descriptors (the physiographic space) of the contributing area to the catchment (Chokmani and Ouarda, 2004), usually through multivariate techniques such as principal components analysis, PCA, or canonical correlation analysis, CCA (Chokmani and Ouarda, 2004; Hundecha et al., 2008).

Topological and canonical kriging (referred to as TK and CK respectively in this study for the sake of brevity) have been shown to be effective methods for the regionalization of flood quantiles (Skøien et al., 2006; Chokmani and Ouarda, 2004) and low flows (Castiglioni et al., 2009, 2011). In particular, CK indirectly takes into account some effects of spatial correlation through catchment descriptors such as latitude and longitude and, to some extent, MAP (mean annual precipitation). Castiglioni et al. (2011) present a comparison of TK and CK for predicting low-flow indices for ungauged sites within a broad geographical region in Central Italy. The study shows that CK slightly outperforms TK and points out the complementarity of the procedures in terms of (i) basic principle of interpolation (i.e., the support for the interpolation is the geographical space for TK and the physiographical space for CK); (ii) data requirements (i.e., topology and

catchment boundaries of the stream network for TK and geomorphologic and climatic descriptors for CK); and (iii) predictive performances (i.e., better results in homogeneous situations along the main rivers for TK, and in heterogeneous situations in headwater catchments for CK).

A comparison between TK and CK that focuses on the prediction of the design flood in ungauged catchments is not available in the literature yet. Furthermore, the possible advantages associated with a combination of CK and TK against traditional RFFA methods were never investigated before, although Castiglioni et al. (2011) indicates the potential for such a combination. These considerations together with the complementarity between TK and CK inspired our study, which primarily addresses three different science questions:

- a. How reliable and accurate are TK and CK predictions of flood quantiles for ungauged sites relative to predictions resulting from traditional RFFA approaches, which for this study is generalized least squares regression?
- b. When should each of the methods (GLS, CK, or TK) be preferred, and what are the strengths and weaknesses of each procedure?
- c. Can we increase the accuracy and reliability of predictions in ungauged catchments by combining TK and CK?

We considered a set of 61 gauged basins located across the southeast US for which several catchment properties and empirical flood quantiles for a number of recurrence intervals are readily available (Gotvald et al., 2009). We compare five design flood quantiles predicted by GLS regression, TK and CK methods, as well as the potential for improvements to TK and CK by blending these two methods. Since GLS is the current method for flood regionalization, we use it as the benchmark approach, while our study focuses on the application of kriging techniques to the Prediction in Ungauged Basins (PUB) problem. We first introduce the study area, data, and flood quantiles in Sect. 2. The GLS regression and kriging procedures considered in this study are briefly illustrated in Sect. 3. Section 4 answers questions (a) and (b) through application and comparison of the kriging methods with the RFFA-GLS method. Section 5 addresses question (c) and describes the utility of combining TK and CK to estimate flood quantiles at ungauged catchments.

2 Study area and data

The study area is located in the southeast United States (Fig. 1a), which is characterized by a temperate climate and annual precipitation between 1000 and 1500 mm per year (Gotvald et al., 2009). The study area was the recent focus of a flood regionalization study where GLS regression had

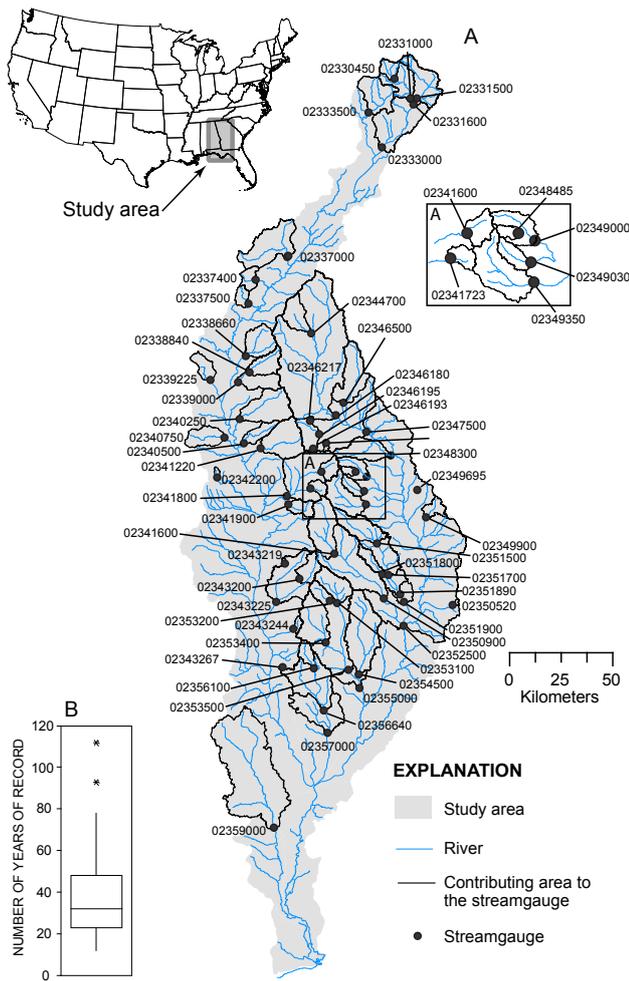


Fig. 1. Map (A) shows the locations of and contributing catchment areas to the 61 study streamgauges located in the southeast United States. The range of record lengths (B) across the study streamgauges is also shown.

been applied to estimate design flood quantiles for rural ungauged catchments in the southeast United States (Gotvald et al., 2009). Study streamgauges are considered to have undeveloped upstream catchments and to have minimal regulation. Additionally, the study streamgauges have at least 10 yr of annual peak-streamflow data (Fig. 1b) and were screened to ensure no significant trends were present in the annual peak streamflows (Gotvald et al., 2009). A number of catchment characteristics describing the morphology, climate, land cover and soil properties of the contributing areas to the study streamgauges are available (see Table 1, Fig. 2, and Gotvald et al., 2009). Where applicable, the catchment characteristics listed in Table 1 were computed as an average over the drainage area.

Empirical quantiles corresponding to the 10, 50, 100, and 500 yr floods (Table 2) were determined as described by the widely used guidelines outlined in Bulletin 17-B of the

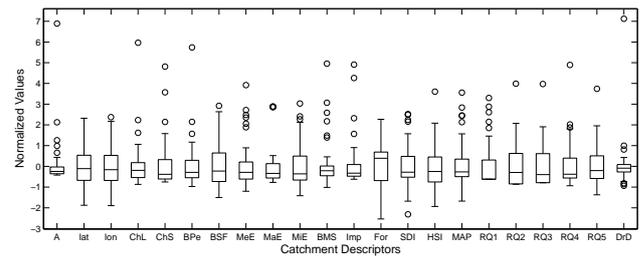


Fig. 2. Normalized values of catchment descriptors (see also Table 1 for label description) used in the study; box plots report minimum and maximum values (whiskers), 25th, 50th and 75th percentiles (box), and outliers (circles).

Hydrology Subcommittee of the Interagency Advisory Committee on Water Data (1982; Gotvald et al., 2009). This interagency document describes how to obtain empirical estimates of flood quantiles by fitting a Pearson Type III distribution to the annual peak-streamflow time series at each of the study streamgauges and pooling at-site and regional information to estimate the parameters of the fitted distribution (Gotvald et al., 2009). Specific details of the procedure as applied to the study streamgauges can be found in Gotvald et al. (2009).

3 Methods to regionalize flood quantiles

3.1 Generalized least squares regression

Regression-based methods to estimate flood quantiles at ungauged locations relate catchment characteristics to the flood quantile of interest and then use this relation to estimate the flood quantile at an ungauged location. For a flood quantile of interest, the regression equation typically has the general form

$$Y = \sum_{i=1}^M a_i X_i + \varepsilon, \quad (1)$$

where Y is a vector of the log-transformed values of the observed floods across the gauged locations, X_i 's are the vectors of the log-transformed values of the observed catchment characteristics, a_i 's are the coefficients estimated by the regression, M is the total number of catchment characteristics, and ε is the vector of the model residuals.

When the true residuals of the regression model have the same variance and are independent, ordinary least squares (OLS) regression can be used to estimate the model coefficients. However, it is unlikely that rivers across a given spatial extent would experience flooding completely independent of one another, and therefore, the assumption of independence of the residuals is likely to be violated (Stedinger et al., 1993). Furthermore, because different record lengths are available to compute the empirical flood quantiles at each

Table 1. Catchment characteristics describing the morphology, climate, land cover and soil properties of the contributing areas to the 61 study streamgauges (Gotvald et al., 2009). Where applicable, catchment characteristics were computed as an average over the drainage area.

Catchment characteristic	Unit of measure	Minimum value	Median value	Maximum value
Morphology				
A – Drainage area	km ²	0.10	13.73	13752.84
DrD – Drainage density	km per km ²	0.05	0.06	0.26
ChL – Main channel length	km	0.14	1.80	17.12
ChS – Main channel slope	m per km	0.03	0.06	0.60
BPe – Catchment perimeter	km	0.26	4.76	1071.68
BSF – Catchment shape factor	dimensionless	0.12	0.23	0.52
MeE – Mean catchment elevation	m	1.98	7.25	31.60
MaE – Maximum catchment elevation	m	4.33	10.51	1351.54
MiE – Minimum catchment elevation	m	0.40	4.61	18.14
BMS – Mean catchment slope	percent	0.07	0.23	1.38
Precipitation				
MAP – Mean annual precipitation	mm	45.0	54.2	79.1
RQ1 – 24 h, 2 yr maximum	mm	4.2	4.2	5.3
RQ2 – 24 h, 10 yr maximum	mm	6.3	6.5	8.1
RQ3 – 24 h, 25 yr maximum	mm	7.4	7.5	9.4
RQ4 – 24 h, 50 yr maximum	mm	8.2	8.5	10.7
RQ5 – 24 h, 100 yr maximum	mm	8.5	9.2	11.8
Land cover and soil properties				
SDI – Soil drainage index	dimensionless	0.10	0.13	0.19
HSI – Hydrologic soil index	dimensionless	0.10	0.09	0.14
Imp – Percent of catchment that is impervious	percent	0.01	0.04	0.36
For – Percent of catchment that is forested	percent	0.16	2.46	3.96

gauged location, the certainty for which the flood quantile is known differs across gauged locations, and the assumption of equal variance is also likely to be violated. To overcome the potential violations of these assumptions, generalized least squares regression is utilized. GLS accounts for the unequal variances and spatial correlation by weighting each flood quantile value by a function of its record length, estimated cross-correlation between floods, and an estimate of the variance determined from OLS regression (Stedinger and Tasker, 1985). Further technical details of GLS as applied to flood quantile estimation at ungauged locations can be found in Tasker and Stedinger (1989). Stedinger and Tasker (1985) found that when the cross-correlation between floods was greater than 0.6, GLS regression was able to estimate regression coefficients that result in more accurate estimates of the regression model coefficients than OLS regression. This is one reason GLS regression is commonly utilized as a method of regionalization, particularly for estimating flood quantiles.

3.2 Geostatistical approaches

In general, smoothing techniques can be applied to interpolate spatially autocorrelated variables, where the spatial coordinates may either identify geographical location (this is the case for TK and all the so-called geostatistical techniques)

or a position in a generic bidimensional space. In particular, kriging is a method for optimizing the estimation of a quantity that is distributed in space and measured at a network of points (see e.g., Journel and Huijbregts, 1978; De Marsily and Ahmed, 1987; Isaaks and Srivastava, 1989; Rossi et al., 1992; Chokmani and Ouarda, 2004). In kriging, the spatial interpolation is obtained by a linear combination of the observed values according to the following equation:

$$\hat{Z}(x_0) = \sum_{i=1}^N \lambda_i Z(x_i), \quad (2)$$

where $\hat{Z}(x_0)$ is the prediction of the variable of interest, Z , at location x_0 , $Z(x_i)$ is the observed value at point (catchment) x_i , with $i = 1, \dots, N$, and λ_i is a weighting coefficient. These weights are estimated by considering spatial correlation and configuration of the observations through variogram models fitted to experimental variograms. An experimental variogram expresses the semivariance between observations as a function of distance and direction of pairs of sampling locations, and describes the spatial correlation structure of the sample data (see e.g., Journel and Huijbregts, 1978; Cressie, 1993).

In practical applications, theoretical variogram models are fitted to experimental variograms to ensure a positive-definite

Table 2. Range of empirical flood quantiles across the 61 study streamgauges (Gotvald et al., 2009).

Flood quantile	Minimum, in $\text{m}^3 \text{s}^{-1}$	Median, in $\text{m}^3 \text{s}^{-1}$	Maximum, in $\text{m}^3 \text{s}^{-1}$
10 yr	3.54	168.77	1673.53
50 yr	7.48	302.99	2489.05
100 yr	9.94	368.12	2860.00
500 yr	13.82	535.19	3822.77

covariance matrix. The literature proposes several theoretical variograms (linear, spherical, gaussian, exponential variogram, etc.; see for instance Cressie, 1993; Journel and Huijbregts, 1978). For this study, statistical interpolation (kriging) is used to regionalize flood quantiles by TK, which is applied over a geographical space, and CK, which is applied over a two-dimensional space of catchment descriptors (physiographical space).

3.2.1 Top-kriging

TK applies kriging methods over a geographical space and combines two groups of forcings for hydrological variability (Skøien et al., 2006; Skøien and Blöschl, 2007). The first group consists of variables that are continuous in space such as rainfall, evapotranspiration and soil characteristics, which are related to local runoff generation. In TK, the variability of these continuous processes in space is represented by the variogram. The second group of forcings are related to aggregation and routing in the stream network. The resulting stream flow variables are only defined for points on the stream network. In TK the aggregation effects that lead to these groups of variables are represented by the catchment boundaries associated with each point on the stream network. Rather than using variograms directly, TK uses point variograms averaged over the catchment areas. These averaged variograms depend on the point variogram as well as the sizes and the relative positions of the two catchments that are compared. In a first step of the analysis, a point variogram model needs to be estimated from the data. This is done by estimating a sample variogram from the data, not only based on the center-to-center distance between catchments, but also on their size. From this, the point variogram model can be back-calculated by fitting aggregated variogram values to the sample variogram, taking into account the nugget effect as proposed in Skøien et al. (2006). Additional details, including a discussion of the fitting method for a point variogram, are also described in Skøien et al. (2006). In a second step, the point variogram is aggregated to the sizes and positions of the catchments for which the hydrometric index of interest is to be estimated and kriging is performed to predict the variables at the ungauged locations (Skøien et al., 2006; Skøien and Blöschl, 2007). TK provides both the estimates as well as the kriging variance (uncertainty).

3.2.2 Canonical kriging

CK applies statistical interpolation techniques, i.e., kriging, to the physiographical space defined by the catchment descriptors of the selected group of catchments (see Chokmani and Ouarda, 2004; Castiglioni et al., 2009). CK is a kernel smoothing technique that uses a covariance-based kernel. CK should not be termed a geostatistical technique in the strict sense as it does not explicitly address autocorrelation of observations or residuals in geographic space (see Pebesma, 2010). The space is defined using two orthogonal coordinates x and y , which can be computed as a function of geomorphoclimatic catchment descriptors; so catchments with similar characteristics have similar coordinates in physiographical space. In particular, any given catchment (gauged or ungauged) can be represented as a point in the x - y space mentioned above. In the same way the set of gauged catchments of the study area can be represented by a cloud of points in this space. The empirical values of the quantity of interest (e.g., empirical flood quantiles associated with a given return period T) can be represented along the third dimension z for each gauged catchment, and can then be interpolated via kriging to estimate it at ungauged sites lying within the same portion of the physiographical space. The term canonical kriging originates from the procedure that is generally adopted to define x and y , that is canonical correlation analysis, or CCA (see e.g., Ouarda et al., 2001; Chokmani and Ouarda, 2004; Di Prinzio et al., 2011).

CCA is an important multivariate statistical tool for reducing the dimensionality of an original data set. CCA is most commonly used in the context where there are two sets of random multidimensional and correlated variables, $N = \{N_1, N_2, \dots, N_n\}$ and $M = \{M_1, M_2, \dots, M_m\}$. For instance, N could be a set of n geomorphologic and climatic catchment descriptors, while M could represent a set of streamflow indices, such as the empirical flood quantiles associated with m different T values. CCA enables one to identify the dominant linear modes of covariability between the sets N and M (e.g., Krzanowski, 1988; Ouarda et al., 2001). In other words, CCA identifies two new groups of artificial variables (canonical variables), $U = \{U_1, U_2, \dots, U_r\}$ and $V = \{V_1, V_2, \dots, V_r\}$, with $r = \min\{n, m\}$, by finding linear combinations of the original N_i , with $i = 1, \dots, n$, and M_j , with $j = 1, \dots, m$, in such a way that the correlation between the canonical variables of a pair (U_i, V_i) is maximized and the correlation between the variables of different pairs is null (Chokmani and Ouarda, 2004; Shu and Ouarda, 2007). If we denote by N and M the independent and dependent variables, respectively, and we consider the linear transformations

$$U = \mathbf{u}_N^T \cdot N \quad \text{and} \quad V = \mathbf{v}_M^T \cdot M \quad (3)$$

characterized by the basis vectors \mathbf{u}_N and \mathbf{v}_M , CCA can be defined as the following optimization problem:

$$\rho = \max_{u_N, v_M} \{ \text{corr}(U, V) \} = \max \frac{\text{cov}(U, V)}{\sqrt{\text{var}(U)}\sqrt{\text{var}(V)}}. \quad (4)$$

Once the linear transformations of Eq. (3) are identified by solving the optimization problem Eq. (4), U_1 and U_2 can be used as x and y coordinates to define the physiographical space. U_1 and U_2 are suitable for defining a Cartesian metric as they are uncorrelated with each other by definition, and therefore orthogonal. Furthermore, U_1 and U_2 are characterized by the maximum linear correlation with the canonical variables V_1 and V_2 (i.e., linear combinations of the streamflow indices, or flood quantiles in our case); hence they are also very effective in explaining the variability of V . In this new physiographic space, U_1 and U_2 are normalized to have zero mean and variance proportional to the explanatory power.

4 Application and comparison of flood regionalization methods

4.1 Implementation of the regionalization methods to the study area

GLS was applied to the study area by closely following the methods described by Gotvald et al. (2009). Log-transformed values of the flood quantiles and basin attributes were used with the procedures described in Tasker and Stedinger (1989) and as implemented by the weighted multiple linear regression (WREG) program (Eng et al., 2009). Gotvald et al. (2009) initially tested all possible subsets of the 22 catchment characteristics shown (Table 1) for inclusion as predictor variables in the final regression equations. Based on an analysis of these regression subsets, Gotvald et al. (2009) determined that drainage area and percent of catchment in each of five hydrologic regions was able to best predict flood quantiles in the GLS regression equations. Hydrologic regions were determined by examining the spatial distribution of residuals obtained from an OLS regression of flood quantiles and catchment area and their correspondence with US Environmental Protection Agency Level III Ecoregions (Gotvald et al., 2009). For this study, we chose not to apply the hydrologic regions because they were defined based on the larger study area in Gotvald et al. (2009).

TK was applied to the study area using the R software package `rtop` (Jon Skøien, personal communication, 2011), which implements the TK procedure as described in Sect. 3.2. The flood quantile values corresponding to the 10, 50, 100 and 500 yr floods were first scaled by the factor $A^{0.65}$, where A is the drainage area of the considered catchment. Considering that the values for a particular flood quantile increase with area, the flood quantiles were first divided by the drainage area to give specific flood quantiles (flood per square kilometer), which is equal to an exponent of one.

However, the specific flood quantile typically decreases with drainage area, as a result of smoothing. After some manual testing, it was found that scaling the flood quantiles by an exponent of 0.65 performed the best. Other studies support the use of a scaling factor in this range for regionalization of flood quantiles (as an example, see Gupta and Dawdy, 1995). The sample variogram was estimated from a cloud variogram (rather than binned), and an exponential theoretical variogram model was then fit to the sample variogram using a neutral weighted least squares method, just as was done in Skøien et al. (2006).

The implementation of CK to the study area was performed using the GLOBEC Kriging Software Package `EasyKrig3.0` for Matlab[®] (Chu and WHOI, 2004). The implementation of CK required some preliminary analyses in cross-validation (see next subsection) to identify the most suitable approach for predicting flood quantiles in ungauged sites. These preliminary investigations show that the best performances of CK are associated with (i) the utilization of the universal kriging (as compared with other types of kriging techniques such as ordinary kriging); (ii) a physiographical space defined by canonical variables U_1 and U_2 derived by applying CCA to the set of 22 available catchment descriptors (N variables; Table 1) and 4 standardized flood quantiles (M variables) obtained by dividing the 10, 50, 100 and 500 yr floods again by the scaling factor $A^{0.65}$, where A is the drainage area of the considered catchment; and using (iii) a gaussian theoretical variogram, which provided the most stable and accurate representation of the empirical variograms amongst all the models we tested (i.e., linear, exponential, spherical, exponential-Bessel, Gaussian-Bessel, etc.). The possibility of anisotropy was investigated but did not improve the fit of the variogram models in the cross-validation. Therefore, isotropy was assumed.

Moreover concerning (ii), it is possible to prove that the first two canonical variables can explain more than 70 % of the original variance of the catchment descriptors; therefore we decided to neglect the residual information of the remaining canonical variables in this study. It is worth noting here that additional canonical variables could in principle be considered when they are informative using higher dimensional kriging procedures.

4.2 Prediction performance in ungauged catchments

To compare the flood quantiles predicted by GLS, TK and CK, a leave-one-out cross-validation approach was applied. In this validation approach, each of the 61 study stream-gauges was subsequently removed from the data set, and the GLS, TK, and CK methods, respectively, were then applied to estimate the flood quantiles at the removed stream-gauge from the remaining 60 stream-gauges. The overall goodness of fit of each method was evaluated by examining the Nash–Sutcliffe efficiency (NSE) values (Nash and Sutcliffe, 1970) computed from the empirical and predicted flood quantiles

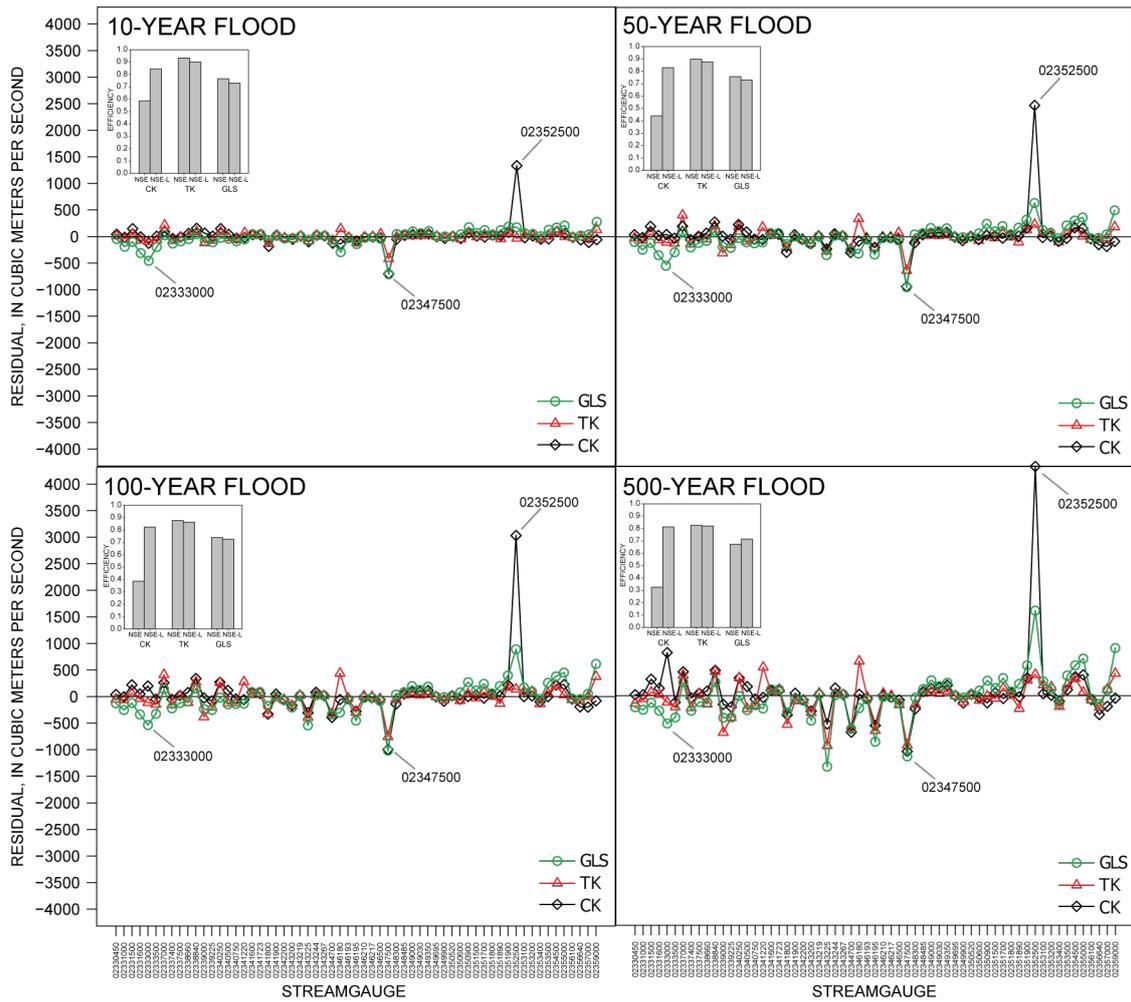


Fig. 3. Error (or residual) between the empirical and predicted flood quantile by study streamgauge and estimation method. Inset in each panel are bars showing the efficiency and log efficiency of the empirical and predicted flood quantiles for the generalized least squares regression, topological kriging, and canonical kriging methods. The log efficiency is the efficiency computed from the natural logarithms of the empirical and predicted flood quantiles.

(Fig. 3). The flood quantile values span 4 orders of magnitude; therefore, the NSE value will be more sensitive to the prediction errors for the largest flood quantiles and the largest catchments. That is, using the NSE index for comparing models, the emphasis is on the magnitude of the deviations between predictions and observations. As floods are to a large degree linearly dependent on the catchment size, the NSE index will then also emphasize on the fit to the largest catchments. To provide a more equally weighted assessment of the fits between the empirical and predicted quantiles across all flood quantiles, the NSE value computed from the natural logarithms of the empirical and predicted flood quantiles (NSE-L) were also compared (Fig. 3). Therefore, the comparison of NSE-L across methods dampens out the effect of the largest catchments and provides an additional and more fair quantification of prediction performance for the whole set of catchments in the study area. Errors between

empirical and predicted flood quantiles for each method were compared by streamgauge to examine which methods performed better than others at the individual streamgauges.

4.3 Discussion of prediction accuracy

Based on these NSE and NSE-L metrics, TK consistently outperformed GLS and CK, with NSE values consistently above 0.8 across all flood quantiles (Fig. 3); GLS and CK were unable to achieve this level of performance for any of the flood quantiles (Fig. 3). GLS, the commonly used method for estimating flood quantiles in ungauged locations in the United States, does not appear to outperform TK or CK (Fig. 3). The difference between NSE and NSE-L values for CK appears to indicate that the lack of agreement between the empirical and CK-predicted flood quantiles is for catchments with the highest values of the flood quantiles. This is

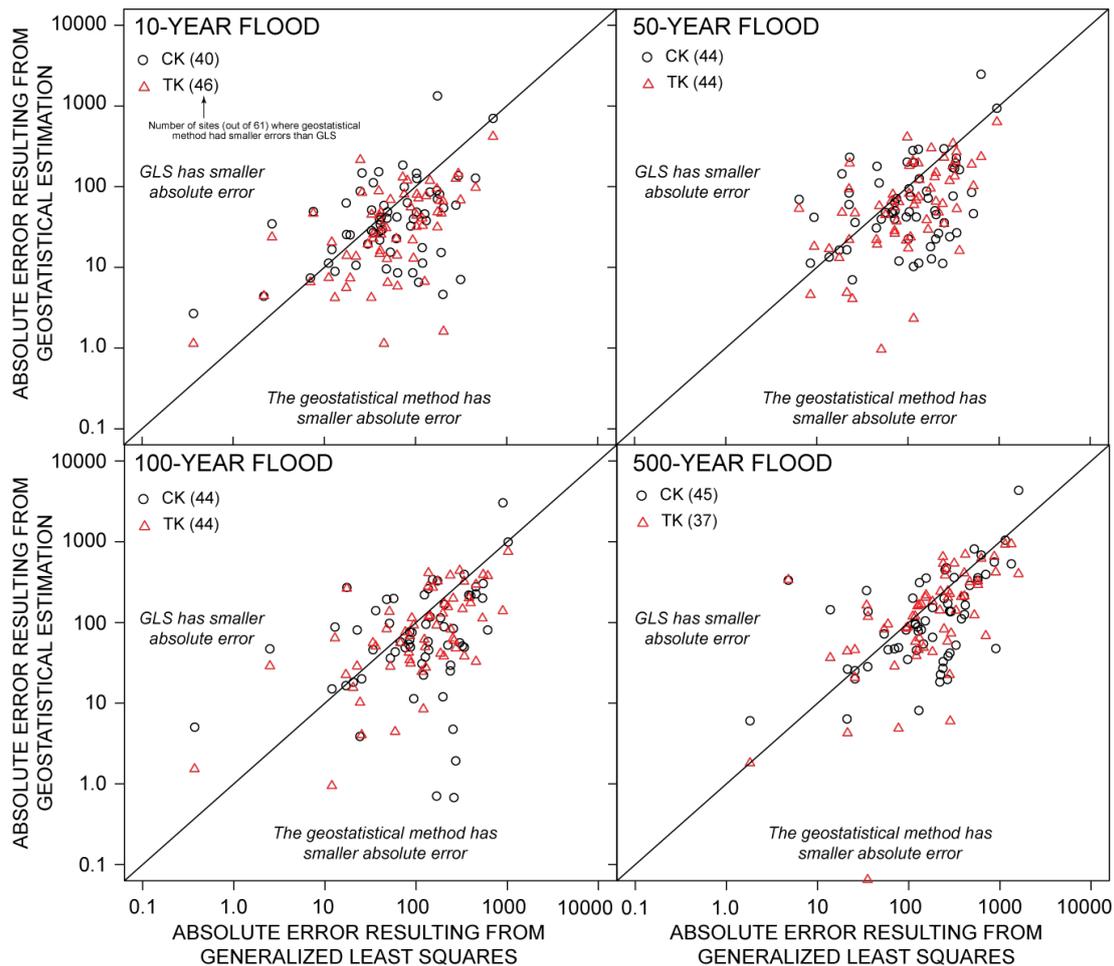


Fig. 4. Comparison of the absolute error between empirical and predicted flood quantiles resulting from topological and canonical kriging to the absolute error resulting from generalized least squares regression.

confirmed in the site-by-site comparison in Fig. 3, showing that catchment 02352500, the largest catchment in the study area, results in a much larger error than the error that results from GLS or TK. The application of CK to the two largest catchments (02352500 and 02347500; Fig. 3) produced two extremely high residuals, one positive ($\sim 3000 \text{ m}^3 \text{ s}^{-1}$) and one negative ($\sim -1000 \text{ m}^3 \text{ s}^{-1}$). The drainage area of these catchments is remarkably larger than the area of all other study catchments (see outliers in Fig. 2). As a result, CK significantly overestimates the former and underestimates the latter, severely impacting the NSE value. This result is partially similar to that observed for low-flow estimation by Castiglioni et al. (2011), which showed that TK performs better in larger catchments than CK. It should be noted that there was no observed relation between length of the stream-gauge record and prediction error across the three methods and four flood quantiles.

The NSE-L values indicate that GLS resulted in the poorest fit between empirical and predicted flood quantiles when compared to CK and TK; however, it should be noted that

all methods have NSE-L values above 0.7. To further understand the differences between the methods, a comparison of the absolute error obtained by GLS was compared to the absolute errors obtained by CK and TK (Fig. 4). In Fig. 4, study streamgauges falling into the lower right portion of the figure indicate that TK or CK resulted in smaller absolute errors than GLS. Both CK and TK generally show smaller absolute errors across sites and across flood quantiles with CK and TK having smaller absolute errors than GLS for more than 40 of the 61 study streamgauges (Fig. 4).

5 Blending flood regionalization methods

5.1 Motivation for blending methods

Castiglioni et al. (2011) show that the nature and data requirements of TK and CK are complementary, implicitly suggesting the possibility of improving the prediction accuracy by blending the two methods. Further analysis of the

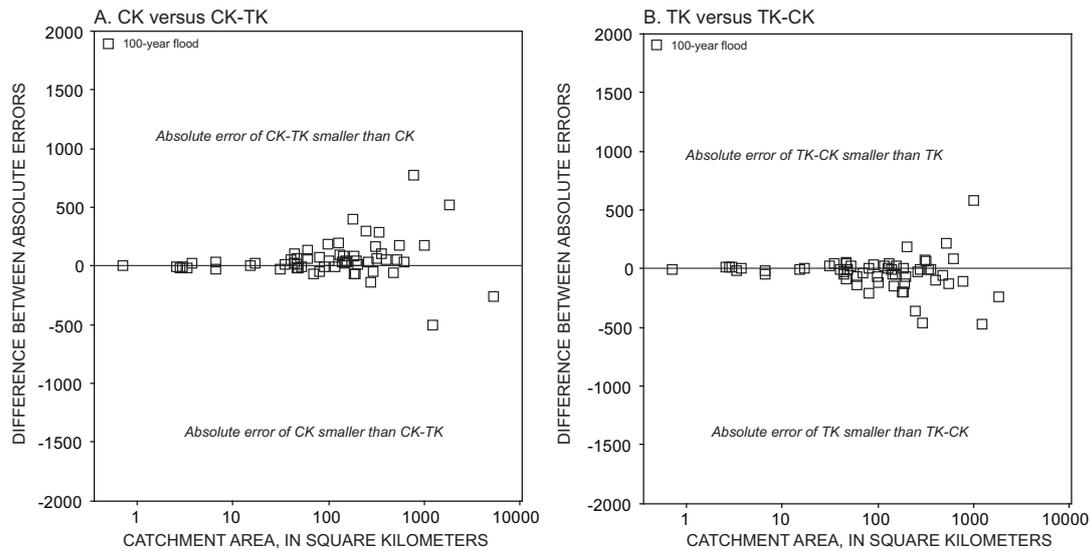


Fig. 5. Differences in absolute error resulting from the application of (A) canonical kriging and the coupling of canonical kriging with topological kriging of the residuals (CK-TK), and (B) topological kriging and the coupling of topological with canonical kriging of the residuals (TK-CK) by catchment area for the 100 yr flood quantile.

results obtained in the previous section may provide some additional insights on possible advantages that result from blending the two kriging procedures. Similarly to the results of this study, Castiglioni et al. (2011) found TK performs better in larger catchments whereas CK performs better in headwater catchments when estimating low-flow statistics at ungauged locations.

5.2 Blending strategies and results

To better assess the utility of blending TK and CK, two strategies were tested for each streamgauge in a leave-one-out cross-validation approach:

1. Use CK to model the flood quantiles, then apply TK to the residuals resulting from cross-validation. Add the TK-predicted residual to the CK-predicted flood quantile to obtain the CK-TK blended estimate.
2. Use TK to model the flood quantiles, then apply CK to the cross-validation residuals. Add the CK-predicted residual to the TK-predicted flood quantile to obtain the TK-CK blended estimate.

The resulting estimates of the flood quantiles were compared by subtracting the absolute error resulting from CK-TK and TK-CK from the absolute errors obtained by TK and CK alone. Negative differences imply that the absolute errors obtained by either CK or TK were smaller than the differences obtained by CK-TK or TK-CK, respectively. This difference was then related to drainage area (Fig. 5). Only the results for the 100 yr flood quantile are shown for clarity; however, this result is similar to that of the other flood quantiles examined

in this study. If the blended methods improve the flood quantile estimates, it is of interest to understand which catchments were improved and if the improvements are related to the size of the catchment. The absolute errors from the blended methods were also compared with the absolute errors resulting from GLS (Fig. 6).

CK-TK generally leads to improved results over the use of CK alone to estimate flood quantiles, particularly for catchments greater than approximately 250 km² (Fig. 5a). In contrast, TK-CK does not seem to improve flood quantile estimates except at a few catchments (Fig. 5b). The blended methods are not effective in improving the flood quantile estimates at the largest catchments. In fact, the TK-CK method substantially degrades the TK-predicted flood quantiles at the largest catchment (Fig. 5b). Despite this, TK-CK-predicted flood quantiles have smaller errors than GLS for a majority of the study streamgauges (Fig. 6). CK-TK was able to improve on predicted flood quantiles when compared to CK-predicted flood quantiles; however, the improvements are not substantial relative to the comparison with GLS-predicted flood quantiles as both CK (Fig. 5) and CK-TK (Fig. 6) have the same number of streamgauges where they have smaller errors than GLS. It should also be noted that for two catchments, CK-TK produced estimates of the flood quantile that were negative.

5.3 Further examination of the blended methods

To have a clearer picture of why blending did not substantially improve the flood quantile predictions in ungauged sites, we investigated the empirical relationship between the residuals resulting from cross-validation and the distance

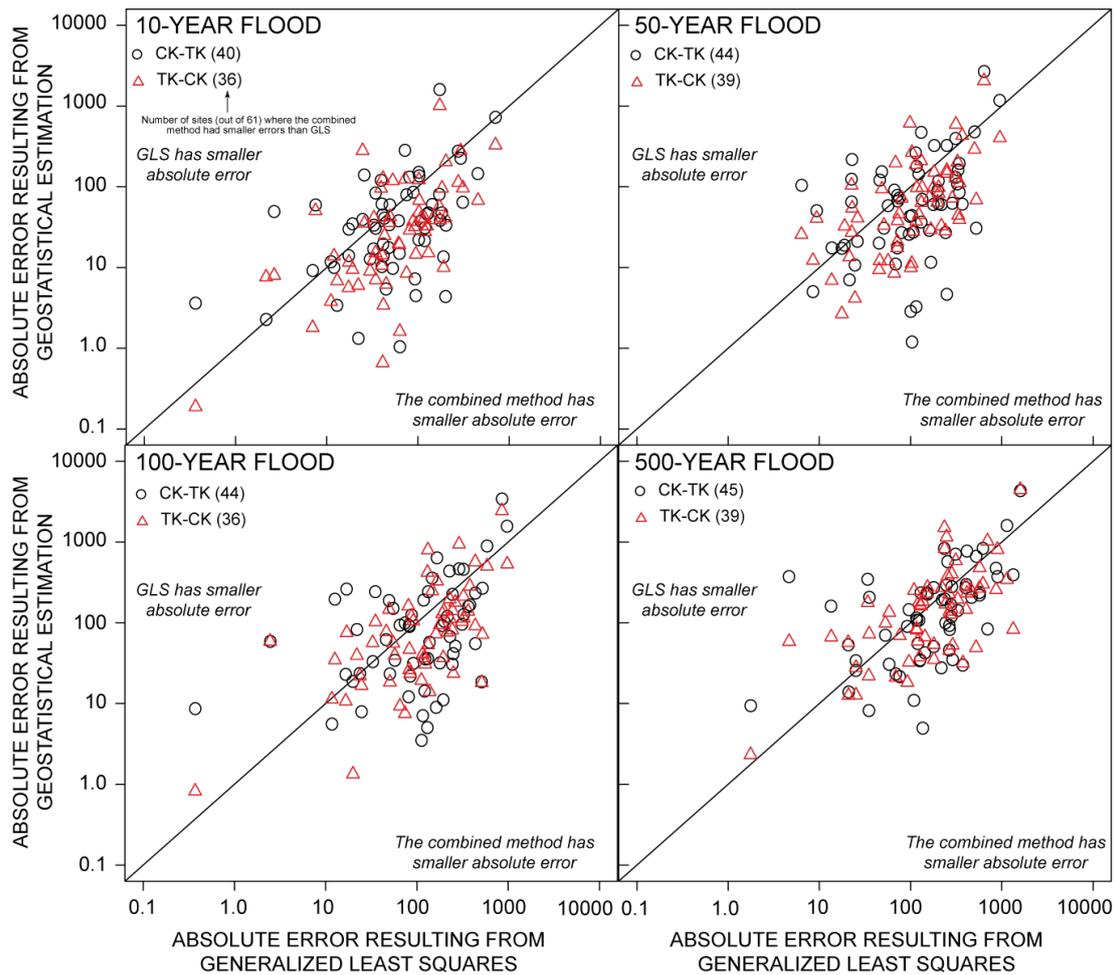


Fig. 6. Comparison of the absolute errors between empirical and predicted flood quantiles resulting from the coupling of topological with canonical kriging of the residuals (TK-CK) and canonical kriging with topological kriging of the residuals (CK-TK) to the absolute error resulting from generalized least squares regression.

between pairs of catchments on the two-dimensional space used for interpolation, that is geographical space for TK and physiographic space for CK. We expressed distances between sites i and j in terms of Euclidean distance,

$$d_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (5)$$

where x and y are the latitude and longitude of catchment centroids if the distance is measured in the geographical space, or the first and second canonical variables (i.e., U_1 and U_2) when distances are measured in the physiographical space. In particular, U_1 and U_2 are obtained in this case by applying CCA to the set of 22 geomorphologic and climatic catchment descriptors (N) and the set of 4 unit residuals (M) obtained in cross-validation from the prediction of the four considered flood quantiles. Catchment descriptors were first normalized over the study area (i.e., coordinates with 0 mean and unit variance) before the distances between the canonical variables were computed.

We then plotted the distance between pairs of catchments against their differences between prediction residuals to evaluate dissimilarity. That is, two catchments are close to each other if they are both associated with large overpredictions of empirical flood quantiles, whereas they are far apart if one catchment is associated with overprediction and the other with underprediction. We defined this distance as follows: first, we considered the 100 yr flood residuals and their relative values (i.e., residuals divided by the empirical quantile) obtained in cross-validation for TK and CK; secondly, for each methodology we normalized residuals and relative residuals to obtain series with 0 mean and unit variance, and we used these series as x and y in Eq. (5) to express the distance between pairs of catchments in terms of prediction errors; finally, we plotted the distance between catchments in the geographical (or physiographical) space against the corresponding distance in terms of residuals for CK (or TK) (see Fig. 7). If pairs of catchments that are far apart in terms of TK (or CK) residuals are also far apart in the physiographical (or

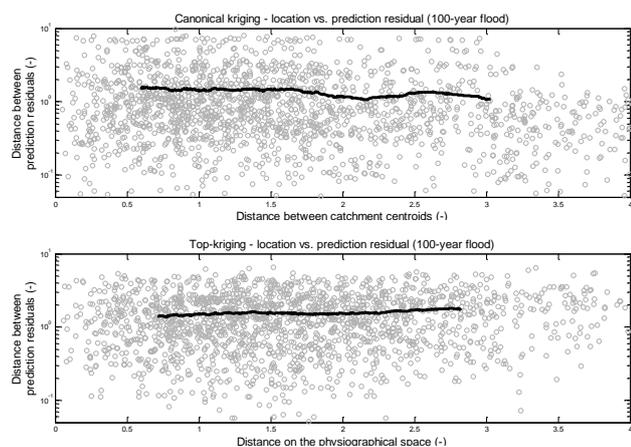


Fig. 7. Dimensionless distances between catchment locations and prediction residuals (gray circles); top panel: distances between catchment centroids in the geographical space (x-axis) and CK residuals (y-axis); bottom panel: distances between catchment pairs in the physiographical space (x-axis) and TK residuals (y-axis) (black thick line: running mean for 250 values).

geographical) space, and vice versa, there are good chances that modeling TK (or CK) residuals with CK (or TK) may improve the overall prediction performance. As illustrated by Fig. 7, distances between catchments in terms of location and residuals do not show an obvious increasing relationship. This empirical evidence suggests why blending TK and CK may have led to limited improvements of the prediction performance over the study area.

6 Discussion

Given the discussion in Sect. 5, the performance of the blended methods was not a surprising result; however, the good performance of CK and TK relative to GLS is cause for further examination of the fundamental objectives of each method. In particular, TK and GLS are fundamentally different in their formulation and in their treatment of spatial correlation in annual flood series. Regression methods use observations as training data for fitting a regression model, where the fitting procedure of GLS reduces the weights of highly correlated observations (> 0.6) as they represent redundant information for the model (Stedinger and Tasker, 1985). Geostatistical methods such as TK will interpolate a surface that passes through the observations, where the value of the surface between the observations is predicted from the observations and the expected local variability.

To explore the extent of spatial correlation that is present amongst the flood quantiles, the cross-correlation between the annual flood time series was computed and compared across the study streamgauges (Fig. 8). Correlations between the annual flood series are relatively low or insignificant for most pairs of streamgauges despite the fact that most

streamgauges had coincident years of record from which to estimate a correlation between annual floods (Fig. 8a); however, some pairs of streamgauges are highly correlated (greater than 0.6) and unsurprisingly these pairs are those which are close together in distance (Fig. 8b). These observations support the suggestion that TK – and to some extent CK – are able to use the correlated information across sites in a way that results in better estimates of flood quantiles at ungauged locations when compared to GLS.

It should also be noted that all of the results in this study are based on a comparison of the ability of TK, CK, and GLS to predict empirical flood quantiles, whereas we are not able to compare their ability to predict (unknown) “true” flood quantiles. When the goal is the prediction of empirical flood quantiles in an ungauged catchment, it is likely that utilizing methods like TK that exploit the cross-correlation between the annual flood series results in better predictive models than GLS. But if the intent is to predict an unknown “true” flood quantile based on a limited set of observations, then it is unclear whether spatial correlation helps to illuminate actual relationships between floods at different locations, or whether the spatial correlation obscures information about the true flood magnitude. Including spatial information in predictive models for flood quantiles may be of benefit if the spatial correlation is caused by real differences in hydrologic processes. If, however, the spatial correlation is merely an artifact of the particular extreme storm events experienced in the region, then inclusion of the spatial correlation may be detrimental to the estimation of flood quantiles.

To illustrate the difference better, we can consider the effect of record-breaking floods in the track of a particular hurricane on predictions of the two different methods. The fact that the hurricane made landfall in one region may or may not mean that areas directly to the north or south are not also susceptible to hurricane flooding. Methods that place too much emphasis on the observed spatial structure may overestimate the flooding potential in the region hit by the hurricane while underestimating the flooding potential in the areas surrounding it. The assumption behind a regression model is that all the regions in the area have the same susceptibility to flooding if long-term records (hundreds to thousands of years) were available and the impact of the hurricane will be smoothed over all regions. If the hurricane made landfall in an area with a high density of streamgauges, the GLS approach will further reduce the impact of the region on the regression model due to the highly correlated samples. Interpolation methods such as TK will treat all observations as correct, although some smoothing between streamgauges can take place if the variogram model has a high nugget effect. On the other hand, if topography or other features may tend to bring storms into the region that experienced the hurricane and this feature is not included in the regressors of the regression model, GLS will underestimate the flood risk in this region and overestimate it elsewhere, whereas TK will predict correctly.

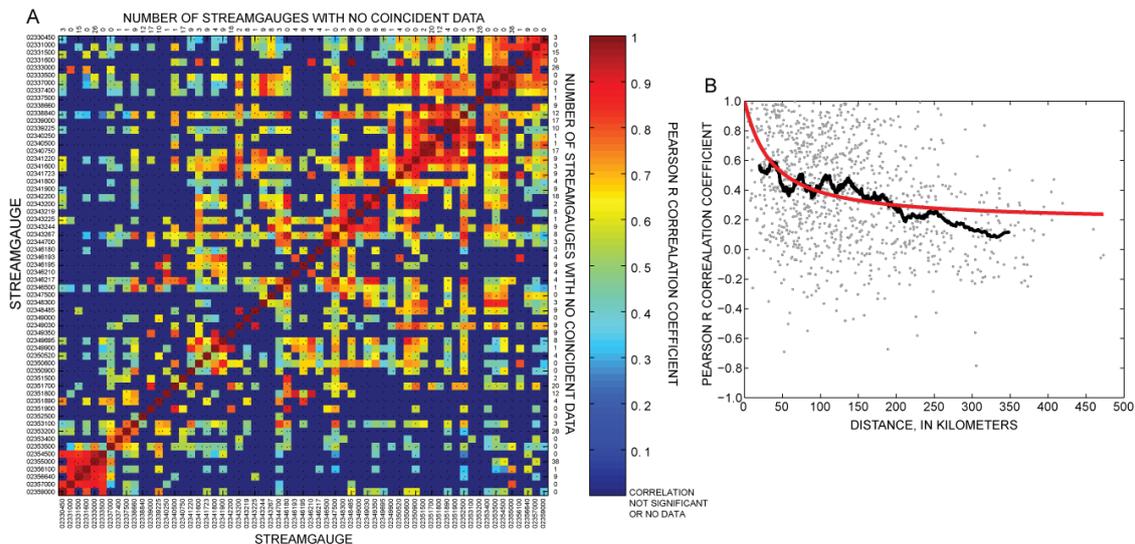


Fig. 8. (A) shows the correlation matrix of the annual flood series between each pair of the 61 study streamgauges. The values on the upper right of the matrix show the number of pairs for which there was no coincident data to compute the correlation between the annual flood series or the correlation was not significant at the $p = 0.05$ level. (B) shows empirical cross-correlation coefficients for couples of catchments (dots), a moving weighted-average curve that weights empirical values proportionally to the record length (black line) and the Tasker and Stedinger (1989) model fitted to data (red line).

It should be noted that GLS requires that the individual model errors are independent. If streamgauges used in development of the GLS regression equations are nested, they are likely to have had the same hydrologic experience and, therefore, could potentially violate this assumption (Veilleux, 2011). In this study, only one of the study catchments (streamgauge 02349695; Fig. 1a) was flagged as a nested basin according to the screening criteria applied by Gotvald et al. (2009). Although nestedness was not likely an issue for the streamgauges used in this study, the effect of nested catchments on GLS remains an open area of research (A. Veilleux, personal communication, 2012).

The results of the comparison between TK and CK study are consistent with those found by Castiglioni et al. (2011), who compared TK and CK for the purpose of estimating low-flow statistics. In both studies, TK outperforms CK for larger catchments. However, in this study, TK and CK performed similarly for smaller catchments whereas, Castiglioni et al. (2011) had found that CK outperformed TK at smaller, head-water catchments. Castiglioni et al. (2009) found that CK outperformed multivariate regression methods to estimate low-flow statistics but did not test TK and GLS regression in their comparison. This paper advances the growing literature which shows that geostatistical methods provide greater prediction accuracy over traditional regression approaches to estimate streamflow at ungauged locations.

7 Conclusions

Recent advances in geostatistical techniques have shown promise as an approach to estimate streamflow at ungauged locations. The performance of two such methods, top-kriging and canonical kriging, were compared to the performance of generalized least squares regression, which is the most common method to estimate design flood values at ungauged locations in the United States. This study represents the first such comparison of these geostatistical methods to GLS regression. The performance of each method was evaluated at 61 streamgauges in the southeast United States using a leave-one-out cross-validation. TK outperformed both CK and GLS regression for estimating the quantiles corresponding to the 10, 50, 100 and 500 yr design floods, particularly for large catchments. Combining methods (adjusting TK-predicted flood quantiles with CK-predicted residuals and vice versa) offered some small improvements; however, this improvement was marginal when compared to the performance of TK and CK over GLS regression. The performance of TK over GLS highlights important differences in geostatistical versus regression-based methods, and the results of this study lend support to other studies which have found that geostatistical methods outperform regression-based methods. As a consequence of the very results of this study, coupling TK and GLS and with consideration of a weighting scheme in blending these methods becomes an interesting research idea. Attention to the role of spatial correlation in regional estimation of flood quantiles is necessary for a more complete understanding of this finding.

Acknowledgements. Tony Gotvald, Toby Feaster, and Curtis Weaver, all from the USGS, provided electronic files of the flood frequency quantiles and upstream basin attributes calculated in their previous study and used for this study. They were gracious in their willingness to answer our questions about the data and their study findings. The authors also thank Edzer Pebesma, of the University of Muenster, Andrea Veilleux, of the US Geological Survey, and one anonymous reviewer for their reviews of the manuscript, which have greatly improved its clarity. Support for Stacey Archfield was provided by the US Geological Survey New England Water Science Center and the Department of the Interior WaterSMART program. The study has been partially supported by the European Cooperation in Science and Technology – COST, through Action ES0901: European procedures for FLOOD FREQUENCY estimation, FLOODFREQ. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

Edited by: F. Laio

References

- Blöschl, G. and Sivapalan, M.: Process controls on regional flood frequency: Coefficient of variation and basin scale, *Water Resour. Res.*, 33, 2967–2980, 1997.
- Burn, D. H.: Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resour. Res.*, 26, 2257–2265, 1990.
- Castellarin, A., Burn, D., and Brath, A.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, *J. Hydrol.*, 241, 270–285, doi:10.1016/S0022-1694(00)00383-8, 2001.
- Castiglioni, S., Castellarin, A., and Montanari, A.: Prediction of low-flow indices in ungauged basins through physiological space-based interpolation, *J. Hydrol.*, 378, 272–280, doi:10.1016/j.jhydrol.2009.09.032, 2009.
- Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., and Blöschl, G.: Smooth regional estimation of low-flow indices: physiological space based interpolation and top-kriging, *Hydrol. Earth Syst. Sci.*, 15, 715–727, doi:10.5194/hess-15-715-2011, 2011.
- Chokmani, K. and Ouarda, T. B. M. J.: Physiographical space-based kriging for regional flood frequency estimation at ungauged sites, *Water Resour. Res.*, 40, W12514, doi:10.1029/2003WR002983, 2004.
- Chu, D. and WHOI: EasyKrig3.0[©]: The GLOBEC Kriging Software Package, copyright (c) 1998, 2001, 2004 property of Dezhang Chu and Woods Hole Oceanographic Institution, All Rights Reserved, 2004.
- Cressie, N.: Statistics for spatial data, Wiley series in probability and mathematical statistics: Applied probability and statistics, J. Wiley, 1993.
- Dalrymple, T.: Flood-frequency analyses, *Manual of Hydrology: Part 3*, 1960.
- De Marsily, G.: Quantitative hydrogeology, Academic Press, London, 1986.
- De Marsily, G. and Ahmed, S.: Application of kriging techniques in groundwater hydrology, *J. Geol. Soc. India*, 29, 57–82, 1987.
- Di Prinzio, M., Castellarin, A., and Toth, E.: Data-driven catchment classification: application to the pub problem, *Hydrol. Earth Syst. Sci.*, 15, 1921–1935, doi:10.5194/hess-15-1921-2011, 2011.
- Eng, K., Chen, Y., and Kiang, J. E.: Users Guide to the Weighted-Multiple-Linear Regression Program (WREG version 1.0), Vol. 4, chap. A8, 21 pp., US Geological Survey Techniques and Methods, 2009.
- FEH: Flood estimation handbook, Insitute of Hydrology, Wallingford, Oxfordshire, 1999.
- Gabriele, S. and Arnell, N.: A hierarchical approach to regional flood frequency analysis, *Water Resour. Res.*, 27, 1281–1289, 1991.
- Gotvald, A. J., Feaster, T. D., and Weaver, J. C.: Magnitude and Frequency of Rural Floods in the Southeastern United States, 2006: Vol. 1, Georgia, 2009-5043, US Geological Survey, Reston, Virginia, USA, 2009.
- Gupta, V. K. and Dawdy, D. R.: Physical interpretations of regional variations in the scaling exponents of flood quantiles, *Hydrol. Process.*, 9, 347–361, doi:10.1002/hyp.3360090309, 1995.
- Hosking, J. and Wallis, J.: Regional Frequency Analysis: An Approach Based on L-Moments, Cambridge University Press, 1997.
- Hundecha, Y., Ouarda, T. B. M. J., and Bárdossy, A.: Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the "spatial" structures of the parameters within a canonical physiographic-climatic space, *Water Resour. Res.*, 44, W01427, doi:10.1029/2006WR005439, 2008.
- Isaaks, E. H. and Srivastava, R.: Applied Geostatistics, Oxford University Press, New York, 1989.
- Journel, A. and Huijbregts, C.: Mining Geostatistics, Academic Press, 1978.
- Krzanowski, W.: Principles of multivariate analysis: a user's perspective, Vol. 3 of Oxford statistical science series, Oxford University Press, 1988.
- McDonnell, J. J. and Woods, R.: On the need for catchment classification, *J. Hydrol.*, 299, 2–3, 2004.
- Merz, R. and Blöschl, G.: Flood frequency regionalisation-spatial proximity vs. catchment attributes, *J. Hydrol.*, 302, 283–306, doi:10.1016/j.jhydrol.2004.07.018, 2005.
- Merz, R. and Blöschl, G.: Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information, *Water Resour. Res.*, 44, W08432, doi:10.1029/2007WR006744, 2008.
- Nash, J. and Sutcliffe, J.: River flow forecasting through conceptual models part I – A discussion of principles, *J. Hydrol.*, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.
- Ouarda, T. B. M. J., Girard, C., Cavadias, G. S., and Bobée, B.: Regional flood frequency estimation with canonical correlation analysis, *J. Hydrol.*, 254, 157–173, doi:10.1016/S0022-1694(01)00488-7, 2001.
- Pallard, B., Castellarin, A., and Montanari, A.: A look at the links between drainage density and flood statistics, *Hydrol. Earth Syst. Sci.*, 13, 1019–1029, doi:10.5194/hess-13-1019-2009, 2009.
- Pebesma, E. J.: Interactive comment on "Geostatistical regionalization of low-flow indices: PSBI and Top-Kriging" by S. Castiglioni et al., *Hydrol. Earth Syst. Sci. Discuss.*, 7, C3351–C3354, 2010.
- Rossi, R., Mulla, D., Journel, A., and Franz, E.: Geostatistical Tools for Modeling and Interpreting Ecological Spatial Dependence, *Ecol. Monogr.*, 62, 277–314, 1992.

- Shu, C. and Ouarda, T. B. M. J.: Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space, *Water Resour. Res.*, 43, W07438, doi:10.1029/2006WR005142, 2007.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., X., L., McDonnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrolog. Sci. J.*, 48, 857–880, doi:10.1623/hysj.48.6.857.51421, 2003.
- Skøien, J. O. and Blöschl, G.: Spatiotemporal topological kriging of runoff time series, *Water Resour. Res.*, 43, W09419, doi:10.1029/2006WR005760, 2007.
- Skøien, J. O., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, *Hydrol. Earth Syst. Sci.*, 10, 277–287, doi:10.5194/hess-10-277-2006, 2006.
- Stedinger, J. R. and Tasker, G. D.: Regional Hydrologic Analysis: 1. Ordinary, Weighted, and Generalized Least Squares Compared, *Water Resour. Res.*, 21, 1421–1432, [with correction, in Stedinger, J. R. and Tasker, G. D., *Water Resour. Res.*, 22, p. 844, 1986], 1985.
- Stedinger, J. R., Vogel, R. M., and Foufula-Georgiou, E.: Frequency analysis of extreme events, in: *Handbook of Hydrology*, edited by: Maidment, R., Chap. 18, 18.11–18.66, McGraw-Hill, New York, 1993.
- Tasker, G. D. and Stedinger, J. R.: An operational GLS model for hydrologic regression, *J. Hydrol.*, 111, 361–375, doi:10.1016/0022-1694(89)90268-0, 1989.
- Veilleux, A. G.: Bayesian GLS Regression, Leverage and Influence for Regionalization of Hydrologic Statistics, Ph.D. dissertation, Cornell University, 2011.
- Viglione, A., Laio, F., and Claps, P.: A comparison of homogeneity tests for regional frequency analysis, *Water Resour. Res.*, 43, W03428, doi:10.1029/2006WR005095, 2007.