Manuscript prepared for Hydrol. Earth Syst. Sci.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 15 June 2015

A topological restricted maximum likelihood (TopREML) approach to regionalize trended runoff signatures in stream networks

Marc F. Müller[1] and Sally E. Thompson[1]

[1]Department of Civil and Environmental Engineering, Davis Hall, University of California, Berkeley CA, USA

Correspondence to: Marc F. Müller (marc.f.muller@gmail.com)

Abstract. We introduce TopREML as a method to predict runoff signatures in ungauged basins. The approach is based on the use of linear mixed models with spatially correlated random effects. The nested nature of streamflow networks is taken into account by using water balance considerations to constrain the covariance structure of runoff and to account for the stronger spatial correlation between flow-connected basins. The restricted maximum likelihood (REML) framework generates the best linear unbiased predictor (BLUP) of both the predicted variable and the associated prediction uncertainty, even when incorporating observable covariates into the model. The method was successfully tested in cross validation analyses on mean streamflow and runoff frequency in Nepal (sparsely gauged) and Austria (densely gauged), where it matched the performance of comparable methods in the prediction of the considered runoff signature, while significantly outperforming them in the prediction of the associated modeling uncertainty. TopREML's ability to combine deterministic and stochastic information to generate BLUPs of the prediction variable and its uncertainty makes it a particularly versatile method that can readily be applied in both densely gauged basins, where it takes advantage of spatial covariance information, and data-scarce regions, where it can rely on covariates, which are increasingly observable via remote sensing technology.

1 Introduction

Regionalizing runoff and streamflow for the purposes of making Predictions in Ungauged Basins (PUB) continues to be one of the major contemporary challenges in hydrology. At global, regional and local scales only a small fraction of catchments are monitored for streamflow (Blöschl et al., 2013), and this fraction is at risk of decreasing given the ongoing challenge of maintaining existing gauging stations (Stokstad, 1999). Reliable information about local streamflows is essential for

the management of water resources, especially in the context of changing climate, ecosystem and demography; and flow prediction uncertainties are bound to propagate and lead to significantly sub-optimal design and management decisions (e.g., Sivapalan et al., 2014; **?**). Techniques are needed to maximize the use of available data in data scarce regions to accurately predict streamflow, while providing a reliable estimate of the related modeling uncertainty.

1.1 Linear Models

There are a number of approaches to predicting runoff in ungauged catchments, including process-based modeling (e.g., Müller et al., 2014), graphical methods based on the construction of isolines (e.g., Bishop and Church, 1992), and statistical approaches. Statistical approaches are often implemented via linear regression, wherein the runoff signature of interest is considered to be an unobservable random variable correlated with observable features of both gauged and ungauged basins (e.g. rainfall, topography). Such linear models are well understood and widely implemented, not only for PUB (see review in Blöschl et al., 2013, p.83) but also across a wide variety of fields in the physical and social sciences (e.g., Myers, 1990).

Spatial correlation is generally problematic for linear model predictions, including the multiple regression approaches commonly used for regionalization. For example if these models predict a hydrologic outcome y using a matrix X of observed features then the linear model has the form:

$$y = X\tau + \eta \tag{1}$$

Here τ is an *a priori* unknown set of weights that represent the influence of each external trend on the hydrological outcome being modeled. The residuals, η, are the observed variation of y that cannot be explained by a linear relation with X. If the residuals are *independent and identically distributed* (iid), the *best linear unbiased predictions* (BLUP) of both y and its uncertainty (i.e. $\mathsf{Var}(y)$) can readily be obtained using ordinary least squares (OLS) regression. Unfortunately, residuals are rarely *iid* in hydrological applications due to the spatial organization of hydrological processes around the topology of river channel networks. This organization has the potential to introduce non-random spatial correlations with a structure imposed by the river network. To recover a suitable model in which residuals remain independent requires that the model structure be altered to explicitly account for the spatial and topological correlation in the residuals.

1.2 Spatial Correlation Models

There are several techniques available to address spatially correlated data. Within PUB, kriging (Cressie, 1993) based geostatistical methods have been widely used (e.g, Huang and Yang, 1998; Gottschalk et al., 2006; Sauquet, 2006; Sauquet et al., 2000; Skøien et al., 2006). In a geostatistical framework, a parametric function is used to model the relationship between distance and covariance in observations. The ensuing semi-variogram is assumed to be homogenous in space, and predictions

at a point are computed as a weighted sum of the available observations. The weights are chosen to minimize the variance while meeting a given constraint on the expected value of the prediction. In ordinary kriging for PUB applications, that constraint is simply the average of the streamflow signature as observed in gauged catchments. Ordinary kriging can also be extended as 'universal kriging' to include a linear combination of observable features (Olea, 1974). Kriging approaches are widely used to predict spatially-distributed point-scale processes like soil properties (e.g., Goovaerts, 1999) and climatic features (e.g., Goovaerts, 2000). Although ordinary kriging has also been used to interpolate runoff (e.g., Huang and Yang, 1998), the theoretical justification for this approach is less robust than for point-scale processes. Runoff is organized around a topological network of stream-channels, and the covariance structure implied for prediction should reflect the higher correlation between streamflow at watersheds that are 'flow connected' (i.e. share one or more subcatchments), compared to unconnected but spatially proximate catchments. Currently, two broad classes of geostatistical methods accommodate this network-aligned correlation structure.

The first suite of methods posits the existence of an underlying point-scale process, which is assumed to have a spatial autocorrelation structure that allows kriging to be applied. Because the runoff point-scale process is only observed as a spatially integrated measure made at specific gauged locations along an organized network of streams, the spatial autocorrelation structure of the point-scale process cannot itself be observed. Block-kriging approaches (Gottschalk et al., 2006; Sauquet, 2006; Sauquet et al., 2000) infer the semi-variogram of the (unobserved) point-scale so as to best reproduce the observed spatial correlation of the area-integrated runoff at the gauges – a procedure known as regularization. The topology of the network is implicitly accounted for by the fact that nested catchments have overlapping areas, which affects the relation between observed (area integrated) and modeled (point scale) covariances. Yet, complex catchment shapes complicate the regularization of semi-variograms, meaning that the estimation of the point-scale process becomes analytically intractable and requires a trial-and-error approach in most practical applications (e.g., Top-kriging (Skøien et al., 2006)). Top-kriging is an extension of the block-kriging approach that accommodates non-stationary variables and short observation records. Top-kriging provides an improved prediction method for hydrological variables when compared to ordinary kriging or linear regression techniques (Laaha et al., 2014; Viglione et al., 2013; Castiglioni et al., 2011) and was recently extended to account for deterministic trends (Laaha et al., 2013). Top-kriging represents an important advance for PUB, but it does have a few drawbacks: (i) The regularization process is unintuitive, and requires extensive trial-and-error to determine both the form of a suitable point-scale variogram, and its parameters; (ii) This trial-and-error process is likely to be computationally expensive; (iii) Like all kriging techniques, the estimation of the variogram is challenging when accounting for observable features: the presence of an unknown trend coefficient and variogram leads to an under-determined problem, making consistent estimates for both challenging. Cressie (1993) (p. 166) showed that the presence of a trend tends to impose a spatially inhomogeneous, negative bias on the estimated semi-

variogram. The bias increases quadratically with distance, meaning that estimates of the long-range variance (the *sill*) are strongly impacted by the presence of the trend, leading to an underestimation of the prediction uncertainty. This bias, however, only marginally affects the prediction itself.

Geomorphological considerations of the topology of a river network generally focus on the channels, and lead to an intuitive conceptualization that topological interpolation should focus on runoff correlations along flow paths. The second type of approach embraces this topological structure. It does not consider a point-scale runoff generation process, but instead models the hillslope-scale runoff delivery process to the channel network as a unidimensional directed tree (Cressie et al., 2006; Ver Hoef et al., 2006). Runoff correlation is expected to decrease with the distance along the stream following a known parametric function. However, unlike Euclidian distances, the stream-wise distance does not have the necessary properties to provide a solvable kriging system. This issue is addressed in Cressie et al. (2006) and Ver Hoef and Peterson (2010), where streamflow is modeled as a random process represented by a Brownian motion that starts at the trunk of the tree (i.e. the river mouth) moves upstream, bifurcates and evolves independently on each branch. The resulting model only allows spatial dependence with points that are upstream on the river network and provides a positive definite covariance matrix that is estimated through restricted maximum likelihood (REML). Models of this nature have been successfully tested on stream chemistry data (Ver Hoef et al., 2006) and further developed to also allow spatial autocorrelation among random variables on stream segments that do not share flow, with potential applications to the modeling of the concentration of upstream moving species (e.g., fishes or insects)(Ver Hoef and Peterson, 2010). While these methods do not account for the streamflow generation process, they avoid the conceptual and prediction uncertainty challenges confronted by kriging techniques.

1.3 The TopREML Approach

Inspired by both types of approaches, here we present a method based on the use of linear mixed models to generate a BLUP for hydrological variables on a flow network. Rather than using a kriging estimator, we adopt a Restricted Maximum Likelihood (REML) framework (Gilmour et al., 2004; Patterson and Thompson, 1971; Lark et al., 2006) to estimate variance parameters. This reduces the bias on the semivariogram by allowing the variance to be estimated independently from the trend coefficients (Cressie, 1993; Lark et al., 2006). This use of a REML framework to estimate a linear mixed effect model on a topological support is termed TopREML. The approach is based on the following conceptual assumptions:

Flow generation and propagation: Similar to Top-Kriging, runoff is assumed to be generated at a point scale on the landscape, from where it is routed to a channel and measured at a gauge (Figure 1 (i)). Runoff observations made at any individual gauge (Figure 1 (ii)) can be broken up into a *local* contribution, derived from a never-previously-gauged catchment area, and an *upstream* contribution that was previously observed at upstream gauge(s) along the channel (Figure 1 (iii)).

TopREML disaggregates all flow contributions into a cascade of local components, as observed at each successive gauge, and uses these characteristics to constrain the covariance structure of runoff and to account for the stronger spatial correlations between flow-connected basins.

Treatment of time: For the local effects to form a suitable basis for spatial interpolation, variations associated with temporal correlation (e.g. travel time effects) need to be removed. This is achieved by considering time-averaged streamflow data, with the proviso that the time averaging window is much greater than the characteristic catchment and channel response timescales. This treatment of time has several specific consequences. First, TopREML is only suitable for the regionalization of time-averaged and statistically stationary runoff properties (i.e. *runoff signatures*). Stationarity is necessary to ensure that the water balance assumption used to separate local from upstream runoff contributions is valid. However, as a consequence, TopREML cannot be used to interpolate transient signatures, such as those associated with real-time forecasting. Nor can it be used to describe runoff properties that are correlated over time scales larger than the time averaging window. Because of the stationarity assumption applied, all correlation arguments described in this manuscript refer to the *spatial*, and not *temporal*, correlation of the runoff signatures.

Network topology: Network topology in TopREML also follows a conceptual model that is similar to the model posited by Top-kriging. Topology is conceptualized by area connectivity. That is, flow-connected gauges are characterized by overlapping drainage areas. Unlike Top-kriging, TopREML does not require information about a spatially random point process, but solely relies on information measured at the gauges. It uses the inter-centroidal Euclidian distance between drainage areas of the local flow contributions at each gauge – the *isolated drainage areas* (IDA) – as a distance metric to compute streamflow correlation. The underlying assumption is that runoff signatures of local flow generation regions that are close to each other (in Euclidian space) are more likely to be identical. Although TopREML doesn't require that the characteristics of a point-scale runoff generation process are known in order to support interpolation (a necessary requirement for Top-kriging), the existence of such a point process is consistent with the treatment of spatial correlation in TopREML. To illustrate this consistency, a stylized example relating point-scale runoff generation to the existence of a covariance-structure that relates flow-connected gauges is outlined as an Appendix (Appendix A).

1.4 Paper Outline

We first derive the TopREML estimator and its variance for mass conserving (i.e. linearly aggregated) variables, with extensions to some non-conservative variables (Section 2). We then apply the approach in two case studies to evaluate its ability to predict mean runoff and runoff frequency by comparison to other available interpolation techniques: Sections 3.1 and 4.1 present leave-one-out cross-validations in Nepal (sparse gauges, significant trends) and Austria (dense gauge network, no observed trends). In both cases, TopREML performed similarly to the best alternative geostatistical

method. We then use numerical simulations to illustrate the effect of the two distinguishing features of TopREML: its ability to properly predict runoff using highly nested networks of stream gauges and its ability to properly estimate the prediction variance when accounting for observable features (Sections 3.2 and 4.2). Finally, we discuss the limits and delineate the context in which TopREML – and geostatistical methods in general – can successfully be applied to predict streamflow signatures in ungauged basins (Section 5).

2 Theory

2.1 Accounting for spatially correlated residuals

Linear models can be used to make predictions about hydrological variables along a network, provided that the models explicitly address the effects of network structure. A mixed linear model approach provides a suitable framework for this accounting. In this framework, the effects of observable features on the hydrological outcome are assumed to be independent of the network, and retain their influence independently, as so-called 'fixed effects'. The role of spatial structure is assumed to lead to correlation specifically in the residuals η. The residuals are split into two parts: (i) one containing 'random effects' u that exhibit spatial correlation along the flow network and (ii) a remaining, spatially independent, white noise term ϵ, which does not have any spatial structure. With these assumptions, the mixed linear model is written as:

$$y = \underbrace{X}_{\substack{\text{Trends: Explanatory variables} \\ (\text{Nxk})}} \underbrace{\tau}_{\substack{\text{Coefficients} \\ (\text{kx1})}} + \underbrace{I_N}_{\substack{\text{Identity Matrix} \\ (\text{NxN})}} \underbrace{u}_{\substack{\text{Correlated random effects} \\ (\text{Nx1})}} + \underbrace{\epsilon}_{\substack{\text{Residuals, uncorrelated errors} \\ (\text{Nx1})}} \tag{2}$$

To proceed, we assume that u and ϵ (and therefore y) are normally distributed with zero mean and are independent from each other. The variance associated with ϵ is denoted σ^2, the variance of u is assumed to be proportional to σ^2 according to some ratio, ξ, and finally, u is assumed to have a spatial dependence captured by a correlation structure G, which is related to the spatial layout of gauges along the river network and a distance parameter ϕ (the correlation range). Thus, the random effects can be specified as:

$$\begin{bmatrix} u \\ \epsilon \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \xi G(\phi) & 0 \\ 0 & I_N \end{bmatrix} \right) \tag{3}$$

To solve this mixed model, five unknowns must be found: σ^2 , ξ, ϕ, the fixed (τ) and random (u) effects. Once τ and u are known, the empirical best linear unbiased prediction (E-BLUP) of y can be made at ungauged locations (Lark et al., 2006). The solution strategy adopted here is to prescribe a parametric form for $G(\phi)$, allowing the covariance structure along the network to be specified,

and the likelihood function for the model to be written in terms of *all* five unknowns. Identifying the parameter values that optimize this model thus simultaneously solves for the correlation structure, covariance parameters, fixed and random effects. To proceed with the specification of $G(\phi)$, however, the form of the covariance structure that arises along the network needs to be addressed.

2.2 Covariance structure of mass conserving variables

In the linear mixed model framework, the propagation of hydrological variables through the flow network introduces topological effects into the covariance structure of that variable. Firstly, linearly propagated variables, such as annual specific runoff, are discussed. Nonlinearly-propagating variables can in some cases be transformed to allow the linear solutions to be used (as outlined in Section 2.5). Consider a set of streamflow gauges monitoring a watershed as illustrated in Figure 1 (ii). Because of the nested nature of the river network, the catchment area related to any upstream gauge is entirely included within the area drained by all downstream gauges. To account for the network structure, the catchment at any location along a stream can be subdivided into the *isolated drainage areas* (IDA) that are *monitored for the first time* by an upstream gauge. This is illustrated in Figure 1 (iii), and leads to a subdivision into non-overlapping areas, each associated with the most upstream gauge that monitors them. In making this subdivision, it is implicitly assumed that the timescales at which a hydrological variable is propagated in the channel are negligible compared with the timescales on which hillslope effects operate (a generally valid assumption for small to moderately sized watersheds (see D'Odorico and Rigon, 2003)). IDA's can be associated with both gauged locations and ungauged locations. In what follows, indices i, j, k, and m are used to refer to gauged sites, while index n refers to ungauged sites where a prediction is to be made.

With these assumptions, observations of y_i made at gauge i can be expressed as a linear combination of contributions from the upstream IDAs:

$$y_i = \sum_{k=i}^{k\in \mathsf{UP}_i} a_k y'_k \tag{4}$$

where y'_k is the contribution of the IDA related to gauge k (that is, y_i is equivalent to y'_i only if there are no gauges upstream of gauge i); UP is the set of isolated drainage areas monitored by gauges that are located upstream of i; $a_k = A_k / \sum_{m=i}^{\mathsf{UP}} A_m \leq 1$ is the surface area of the drainage area k normalized by the total watershed area upstream of gauge i. The covariance between observations of y made at different gauges can then be expressed as

$$\begin{aligned}\mathsf{Cov}\,(y_i, y_j) &= \mathsf{E}\,[y_i y_j] - \mathsf{E}\,[y_i]\,\mathsf{E}\,[y_j] \\ &= \sum_{k=i}^{k\in \mathsf{UP}_i} \sum_{m=j}^{m\in \mathsf{UP}_j} a_k a_m \mathsf{E}\,[y'_k y'_m] - \left(\sum_{k=i}^{k\in \mathsf{UP}_i} a_k \mathsf{E}\,[y'_k] \right) \left(\sum_{m=j}^{m\in \mathsf{UP}_j} a_m \mathsf{E}\,[y'_m] \right)\end{aligned}$$

With $\mathsf{E}[y'_k y'_m] = \mathsf{Cov}(y'_k y'_m) + \mathsf{E}[y'_k]\,\mathsf{E}[y'_m]$, we have

$$\mathsf{Cov}(y_i, y_j) = \sum_{k=i}^{k\in \mathsf{UP}_i} \sum_{m=j}^{m\in \mathsf{UP}_j} a_k a_m \mathsf{Cov}(y'_k, y'_m) \tag{5}$$

where $\mathsf{Cov}(y'_k, y'_m)$ is the covariance between the contributions of sub-catchments k and m. By summing over UP in Equation 5 (rather then the complete set of available gauges), the model assumes no correlation between runoff observed at flow-unconnected gauges.

Here we assume that the area-averaged process y' is drawn from a second order stationary random process, and that the covariance between y'_k and y'_m will depend only on the relative position of sub-catchments m and k, given some specified correlation function $\rho(\cdot)$ of the distance c_{km} between the centroids of the two sub catchments (Cressie, 1993). We assume that this function is well approximated by an exponential function $\rho(c_{km}, \phi) = exp(-c/\phi)$. A justification for this assumption, which reproduces the streamflow variances observed in our case studies well (Figure 8), is derived for strongly idealized conditions in Appendix A. Finally, because the observations made at the gauges represent an area-averaged process, the averaging generates a nugget variance σ^2 that is homogenous across observations. The nugget consists of the variance of processes that are spatially correlated over scales smaller than the sub-catchments (see Appendix A) and of measurement errors at the gauges.

With this background, the covariance matrix of y can be expressed as

$$\mathsf{Cov}(y_i, y_j) = \xi\sigma^2 \sum_{k=i}^{k\in \mathsf{UP}_i} \sum_{m=j}^{m\in \mathsf{UP}_j} a_k a_m \rho(c_{km}, \phi) + \sigma^2 = \sigma^2 \cdot (\xi U[A \diamond R]U^T + I_N) \tag{6}$$

where $\sigma^2 = \mathsf{Var}(y'_k, y'_k)$, $U_{i,j} = \mathbf{1}\{j \in \mathsf{UP}_i\}$, $A = aa^T$, and $R_{i,j} = \rho(c_{i,j}, \phi)$. $[\cdot\diamond\cdot]$ denotes the element-by-element matrix multiplication. The matrix G describing the correlation between the random effects in Equation 3 is finally

$$G(\phi) = U[A \diamond R(\phi)]U^T \tag{7}$$

The topology of the network is described by the matrix U, which ensures that only those catchments that are on the same sub-network (upstream or downstream) of the considered gauge are utilized in the determination of the covariance of y. This spatial constraint comes at the expense of neglecting potential correlations with neighboring catchments that are not flow-connected, and the effects of this tradeoff are investigated in the Monte Carlo experiment described in Section 3.2. The effect of spatial proximity is addressed by use of the Euclidian distance between catchment centroids (matrix R), and the effect of scale is accounted for by weighting by the catchment area of the IDAs (matrix A).

2.3 REML estimation

The restricted maximum likelihood approach partitions the likelihood of $y \sim \mathcal{N}\left(X\tau, \sigma^2(\xi G + I_N)\right)$ into two parts, one of which is independent of τ (Corbeil and Searle, 1976). This allows the determination of fixed effects and the variance parameters of the model (here σ^2, ϕ and ξ) to be undertaken separately. The variance parameters are then estimated by maximizing the restricted log likelihood expression (Gilmour et al., 1995)

$$\lambda_R(\sigma^2, \phi, \xi) = -\frac{1}{2}\left(\log\det\left(X^T H^{-1} X\right) + \log\det\left(H\right)) + \nu\log\sigma^2 + \frac{1}{\sigma^2} y^T P y\right)$$

where $\det(\cdot)$ is the matrix determinant operator, $\nu = N - k$, $H = I_N + \xi G$, and $P = I_N - W K^{-1} W^T$, $W = [X : I_N]$ and R is the correlation matrix in Equation 7, and K is the block matrix:

$$K = \begin{bmatrix} X^T X & X^T \\ X & I_N + \xi^{-1} G^{-1} \end{bmatrix}$$

The REML estimators $\hat{\sigma}^2$ and $\hat{\phi}$ that maximize λ_R can be obtained through numerical optimization.

2.4 E-BLUP and prediction variance at ungauged catchments

Once the variance components $\hat{\phi}$ and $\hat{\xi}$ are estimated, the fixed effect coefficients $\hat{\tau}$ and the random effects $\tilde{u}$ can be obtained by solving the linear system (Henderson, 1975):

$$K(\hat{\phi}, \hat{\xi}) \begin{bmatrix} \hat{\tau} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} Xy \\ y \end{bmatrix} \tag{8}$$

The empirical best linear unbiased prediction of $\tilde{y}_n$ at an ungauged site n can be computed by summing the fixed and random effect predictions (Lark et al., 2006)

$$\tilde{y}_n = x_n^T \hat{\tau} + \tilde{u}_n = x_n \hat{\tau} + g_n^T G^{-1} \tilde{u} \tag{9}$$

where x_n is the vector of fixed covariates at ungauged site n, g_n a correlation vector between site n and each gauge; given $\hat{\phi}$, g_n can be readily obtained from the relative position of site n and the gauges in the river network.

The variance of the TopREML prediction error can be expressed as

$$\begin{aligned} \mathsf{Var}\left(\tilde{y}_n - y_n\right) &= \mathsf{Var}\left(x_n^T(\hat{\tau} - \tau) + g_n^T G^{-1}(\tilde{u} - u)\right) \\ &= x_n^T \mathsf{Var}\left(\hat{\tau} - \tau\right) x_n + g_n^T G^{-1} \mathsf{Var}\left(\tilde{u} - u\right) G^{-1} g_n + 2 x_n^T \mathsf{Cov}\left(\tilde{u} - u, \hat{\tau} - \tau\right) G^{-1} g_n \end{aligned} \tag{10}$$

The covariance matrix of the error on τ and u in Equation 10 can be expressed as a function of the inverted model matrix K (Lark et al., 2006):

$$\mathsf{Cov}\begin{pmatrix} \hat{\tau} - \tau \\ \tilde{u} - u \end{pmatrix} = \sigma^2 K^{-1} \tag{11}$$

This provides:

$$\mathsf{Var}\left(\tilde{y}_n - y_n\right)_- = \sigma^2 \left(x_n^T K_{11}^{-1} x_n + g_n^T G^{-1} K_{22}^{-1} G^{-1} g_n + 2 x_n^T K_{12}^{-1} G^{-1} g_n\right) \quad (12)$$

Where K_{11}^{-1}, K_{22}^{-1}, K_{12}^{-1} are $k \times k$, $N \times N$ and $k \times N$ partitions of the inverted K matrix. If ϵ is an error that is truly *iid* and does not affect the true value of y_n (e.g., measurement errors), then Equation 12 corresponds to the mean square error of the TopREML prediction of y_n. If, by contrast, ϵ represents random variations of the true value of y_n that are correlated over short distances (and so do not appear correlated in our data), then ϵ should be included in Equation 10 and the prediction variance becomes

$$\mathsf{Var}\left(\tilde{y}_n - y_n\right)_+ = \mathsf{Var}\left(\tilde{y}_n - y_n\right)_- + \sigma^2, \quad (13)$$

because ϵ and u are independent. In reality ϵ is likely composed of both spatially correlated and *iid* error components and the true variance will be somewhere between these two bounds (Lark et al., 2006).

2.5 Application to non-conservative variables

Unlike mean specific runoff, numerous streamflow signatures (e.g., runoff frequency or descriptors of the recession behavior) are non-conservative and cannot be expressed as linear combinations of their values in upstream sub-catchments. In such conditions the derivations in section 2.2 cannot be applied and the correlation structure in Equation 7 will lead to biased REML predictions. The effect of the network structure on streamflow can nonetheless be accounted if the non-linearities can be neglected or eliminated through algebraic transformations.

For instance, runoff frequency λ is defined as the probability, on daily timescales, that a gauge will record a positive increment in streamflow (Botter et al., 2007; Müller et al., 2014). Provided all sub basins are large enough to significantly contribute to streamflow, a runoff pulse at *any* of the upstream sub-basins causes a streamflow increase at the gauge. Therefore runoff frequency does not scale linearly through the river network. It can nonetheless be shown (see Appendix B) that if runoff pulses occur independently for each sub-basin, the logarithm of the complement to runoff probability (i.e. $ln(1-\lambda)$) propagates linearly throughout the network enabling the application of TopREML to predict runoff probability at ungauged catchments.

A similar reasoning can be applied to predict recession parameters. For example, the exponential function $Q(t) = Q_0 exp(-k_r t)$ is a widely used approach to model base flow recession, where $Q(t)$ is the discharge at time t, Q_0 the peak discharge, and k_r the recession constant which can be considered to represent the inverse of the average response time in storage (Wittenberg and Sivapalan, 1999). Because expected values scale linearly, the average response time at a gauge can be modeled as a linear combination of the mean response times of the upstream IDAs. Therefore, although recession constants themselves do not propagate linearly, their value in ungauged basins can be estimated by taking the inverse of TopREML predictions of average response times.

2.6 Implementation

The computational implementation of TopREML in R (R Core Team, 2008) is described in Appendix C and an operational script is provided as a supplement to this manuscript. To run the script, two vector datasets (e.g., ESRI Shapefiles) are needed as inputs – one containing the catchments where runoff is available and another containing the basins where predictions are to be made. Catchment polygons and explanatory and predicted variables must be provided as attributes of the vector polygons. The way in which the catchment polygons are nested provides the topology of the stream network. TopREML uses the BFGS algorithm (Wright and Nocedal, 1999) to maximize the restricted log likelihood, though stochastic algorithms are required if a non-differentiable (e.g., spherical) covariance function is selected. The selection of initial values for σ^2, ϕ and ξ is a key user input that may affect the performance of optimization algorithms by causing them to converge to a local extrema. We found that initial values of $[\sigma_0^2, \phi_0, \xi_0] = [\sigma_{LM}^2, \mathsf{E}[c_{km}], 1]$ worked well in our case studies, with σ_{LM}^2 the variance of the OLS residuals of the linear model and $\mathsf{E}[c_{km}]$ the average distance between IDA centroids.

3 Methods

3.1 Case studies

Observed streamflow data are used to evaluate the ability of TopREML to predict streamflow signatures in ungauged basins. The assessment is based on leave-one-out cross validations, where the tested model is applied to predict runoff at one basin based on observations from all the other basins. After predicting runoff at all available basins in that manner, the model is evaluated based on its mean absolute prediction error. Streamflow variables from 57 catchments in Upper Austria (Skøien et al., 2014) and 52 catchments in Nepal (Department of Hydrology and Meteorology, 2011) are used in two separate leave-one-out analyses. The location of the gauges is shown in Figure 3, and Table 2 provides a summary of relevant catchment characteristics. Further details on the datasets are provided in Skøien et al. (2014) for Austria and Müller et al. (2014) in Nepal. The two regions differ significantly with respect to gauge density (high in Austria and low in Nepal) and in the nature of the runoff signature and observable features. The Nepalese datasets provides specific runoff and wet season runoff frequency, as well as gauge elevation and bias-adjusted annual rainfall derived from the Tropical Rainfall Measurement Mission 3B42v7 dataset (Müller and Thompson, 2013). Gauge elevation and annual rainfall are used as observable features for specific runoff (Chalise et al., 2003). The Austrian dataset was directly taken from the *rtop* package (Skøien et al., 2014), where mean summer runoff observations are provided to demonstrate Top-kriging. The Austrian dataset did not contain additional observable features and previous studies have found spatial proximity to be a significantly better predictor of runoff than catchment attributes in Austria (Merz and Blöschl, 2005).

The predictive ability of TopREML was evaluated on (a) specific annual runoff in Nepal, (b) wet season runoff frequency in Nepal and (c) average summer streamflow in Austria. The performance of TopREML (TR) was compared to five other widely used regionalization methods: sample mean (LM_0), linear regression (LM), universal kriging (UK) and Top-kriging (TK). As shown in Table 1, these methods cover a wide spectrum of incrementally specific assumptions and comparing them provides an assessment of the value added by increased model complexity for regionalization of these streamflow parameters. Code to implement all four methods is readily available in R, with dedicated packages available for Top-kriging – *rtop* – and universal kriging – *gstat* (Pebesma, 2004).

3.2 Numerical Simulations

3.2.1 Network Effects

Conventional geostatistical methods predict runoff by weighing observations from surrounding basins based on their geographic distance. TopREML also incorporates the topology of the stream network by including or excluding basins based on their flow-connectedness. This adds topological information to the determination of the covariance structure of runoff, at the expense of discarding information that could be derived from correlations between spatially proximate regions that are not connected to the gauge of interest by a flow path. Assessing the net benefits of accounting for network effects requires being able to control the topology of the network, and thus requires numerical simulations. A series of Monte Carlo experiments as described in Figure 4 were run to simulate network complexity by varying the number of flow-connected basins that are within (N_{inner}) and beyond (N_{outer}) the predefined spatial auto-correlation range of the randomly generated runoff. A non-topological geostatistical method like universal kriging would include all basins within and exclude all basins beyond the spatial auto-correlation range. We expect TopREML to outperform universal kriging when the number of flow-connected basins beyond the autocorrelation range increases and the number of connected basins within the autocorrelation range decreases.

3.2.2 Variance Estimation and Observable Features.

A key advantage of the Reduced Maximum Likelihood framework is its ability to avoid the downward bias in the covariance function that affects kriging-based methods (including Top-kriging) when external trend coefficients are simultaneously estimated. This bias particularly affects the prediction of the variance. Again, empirical cross validation analysis does not allow an assessment of this bias, because the observation datasets used contained only one observation per location. Numerical simulations, however, allow many realizations of the underlying stochastic process to be made at each location, and thus allow the prediction variance to be compared with the numerical variance. We evaluate TopREML's ability to predict variances (and therefore evaluate prediction uncertainties) at ungauged locations using the Monte Carlo procedure on the synthetic catchments described

in Figure 4. We construct the observed prediction uncertainty by taking the standard deviation of the prediction errors across all 1000 Monte Carlo runs and compare it to the square root of the median predicted variance. The external trend is omitted from the model specification (i.e. it is *not observed*) in a first experiment, and explicitly included in the model in the second experiment. We compare TopREML and Top-kriging based on their ability to model prediction variance. We expect TopREML to provide a better estimate of the variance than Top-kriging when accounting for observable features. Because the trend is spatially correlated, omitting it in the model specification adds a significant spatially correlated component to the error and Equation 13 should be used to predict the variance. Conversely, including a trend in the model will cause the remaining error to mostly consists of (spatially uncorrelated) residuals so in this case Equation 12 is used.

4 Results

4.1 Case Studies

Basin-level predictions of the considered signatures are presented in Figure 5 for the three cross validation analyses described in Section 3.1. Figure 5 also provides box plots summarizing the distribution of the ensuing cross validation errors. In the three analyses, the prediction errors related to TopREML were comparable to the best alternate method: a linear model for annual specific runoff (Nepal) and Top-kriging for runoff frequency (Nepal)) and summer runoff (Austria).

Figure 5 (a) presents results for annual specific runoff in Nepal and shows that observable features play a significant role in the prediction of runoff. The linear model showed a highly significant effect of annual precipitation ($\hat{\tau}^{(LM)}_{\mathsf{yearlyPrecip}} = 0.99$, t-stat: 9.1) a moderately significant effect of altitude ($\hat{\tau}^{(LM)}_{\mathsf{meanElev}} = 0.39$, t-stat: 2.5) and an overall fit of $R^2 = 0.63$. The positive sign of the altitude coefficient can be attributed to the effects of glacial melt on runoff, which are more significant at higher altitudes, while the average effect of evapotranspiration explains the negative and noisy intercept of -313 mm/y. While including rainfall and altitude in the model decreased the median absolute error by 43% (LM to LM_0), further increasing the complexity of the model by allowing for spatial (UK) and topological effects (TK and TR) did not improve the predictive performance: residuals from the linear regression appeared to be correlated at a range shorter than the distance between the gauges in Nepal. Indeed, fitting the empirical semivariograms with exponential functions revealed spatial correlation ranges that were on the order of the mean distance between IDA centroids for annual streamflow (21.6 km), and significantly below that distance (7.0 km) for the regression residuals. Nonetheless, the lack of parsimony of TopREML did not appear to affect its predictive performance, which almost perfectly reproduced the performance of the linear model – the most parsimonious method.

In contrast, the analysis revealed significant spatial effects for both runoff frequency in Nepal, which has a much larger spatial correlation range than annual streamflow (426 km – presumably

set by meteorology and the correlation range of storm events (Müller and Thompson, 2013)), and summer average streamflow in Austria, which has a range of 19.1km but is sampled by a much higher density of streamflow gauges than in Nepal. Allowing for spatial correlation in the residuals (UK) decreased the median absolute error by 11% compared to the linear model (LM) for runoff frequency in Nepal and 31% for summer runoff in Austria. Accounting for topological effects further reduced errors by 33% (runoff frequency) and 40% (summer runoff) for both TopREML and Top-kriging methods.

4.2 Numerical Simulation

Results from the Monte Carlo analysis are presented in Figure 6, showing the outcomes of the two numerical experiments described in section 3.2.

Figure 6 (a) and (b) shows the effect of network complexity on the performance of TopREML relative to the baseline performance of universal kriging. This effect is measured as the difference in the relative errors of the two methods as a function of N_{outer}, the ratio of basins *beyond* the spatial correlation range of runoff that *are* flow-connected, and N_{inner}, the ratio of basins *within* range that are *not* flow-connected. The effect is expected to increase with N_{outer} and decrease with N_{inner}, reaching zero when 100% of observed basins lie within the spatial correlation range and 0% of the basins beyond the range are flow-connected. In that case (not shown in the figure), TopREML and universal kriging perform similarly and the mean difference in the relative error of the two methods is zero. Figure 6 (a) shows that the relative performance of TopREML improves with the number of flow-connected catchments that are located beyond the spatial correlation range, and which are therefore not properly accounted for by universal kriging. Conversely, Figure 6 (b) shows that the relative performance of TopREML decreases with decreasing network effects within the spatial correlation range. A linear regression of the relative performance of TopREML against N_{outer} and N_{inner} showed that both trends are significant and in the expected direction. However, the positive coefficient associated to N_{outer} (9.1, t-stat: 11.9) is larger in absolute value and more statistically significant than the negative coefficient associated to N_{inner} (-2.6, t-stat: -2.6), which suggests that the benefits of including distant flow-connected basins outweigh the costs of discarding nearby (but unconnected) IDAs.

In Figure 6 (c), the Monte Carlo analysis showed that model uncertainty is well predicted by TopREML and strongly underestimated by Top-kriging, both with and without considering an external trend. Including a trend in the model reduces the prediction variance of TopREML – this effect is expected because the variance explained by the trend is no longer included in the modeling error ϵ. The decrease in the prediction variance is well modeled by TopREML, which predicts the observe model uncertainty almost exactly.

5 Discussion

5.1 Performance of TopREML

Cross validation outcomes suggest that TopREML is an attractive operational tool for predicting streamflow in ungauged basins. The method performs as well as the best alternative approach in the prediction of the considered runoff signatures in Nepal and Austria, and significantly outperforms Top-kriging in the prediction of modeling uncertainties in the numerical analysis.Two distinguishing features of TopREML are responsible for these encouraging results. First, TopREML incorporates the topology of the stream network by restricting correlations to runoff observed at flow-connected catchments. This allows TopREML to explicitly model the higher correlation in streamflow anticipated along channels, but comes at the expense of discarding correlations with neighboring, but not flow-connected catchments. Such correlations can, for instance, be driven by large scale weather patterns. This tradeoff was investigated in a Monte Carlo analysis showing that modeling performance increases more rapidly when including distant flow-connected basins (slope in Figure 6 (a)), than it decreases when discarding nearby unconnected basins (slope in Figure 6 (b)). Further, empirical correlograms of Austrian summer runoff (Figure 2) reveal significantly lower and shorter-ranged spatial correlations when basins are not flow-connected. Both results suggest that the benefit of accounting for network effects on correlations outweighs the cost of losing some information on the correlation between unconnected basins. Second, the Restricted Maximum Likelihood framework provides an unbiased estimation of variance parameters, even when accounting for observable features. This allows TopREML to accurately predict modeling uncertainties even for highly trended and autocorrelated runoff signatures, as visible in the Monte Carlo analysis presented on Figure 6 (c). By contrast, the expected downward bias in the kriging estimation of partial sills (Cressie, 1993) is clearly visible in the underestimation of prediction uncertainties by the Top-kriging method.

TopREML also has considerably lower computational requirements than Top-Kriging, both in terms of input data and optimization complexity. Unlike Top-kriging, where watershed polygons are necessary inputs for the regularization procedure, vectors are not fundamentally indispensable for TopREML. Indeed, TopREML does not rely on a distributed point process but assumes homogenous IDAs. It follows that its only fundamental data requirement is a table (i.e. a data.frame) of IDAs displaying the observed regionalization variable and the area, centroid coordinates and network position (i.e. own ID and downstream ID) of the IDA. When considering runtime, both methods rely on numerical optimization, but Top-Kriging uses it to back-calculate the point semi-variogram in its regularization procedure. This may substantially increase the dimensionality of the optimization task, depending on the grid resolution chosen for the discretization of the catchment areas, which in turn has a highly significant effect on prediction performances (Skøien et al., 2006). By contrast, the dimensionality of the optimization in TopREML is driven by the number of catchments, not an arbitrary grid. More importantly, TopREML admits a well- defined objective function, the

restricted likelihood, that is differentiable if the selected variogram function is differentiable. This allows gradient optimization methods to be used, which are much less computationally intensive than the stochastic algorithm required by Top-kriging. The resampling analysis shown in Appendix C suggests that TopREML reduces the computation runtime by an order of magnitude, relative to the implementation of Top-kriging in the *rtop* package, for comparable prediction performances.

Despite these encouraging results, TopREML is subject to stringent linearity assumptions on the nature of the regionalized runoff signature. The predicted variable should aggregate linearly both on hillslope surfaces and at channel junctions that are subject to mass conservation. This limitation also affects block-kriging aproaches, as pointed out by Skøien et al. (2006), who suggest that Top-kriging can still be applied, *in an approximate way* on non-conservative variables. Here we assert that hydrologic arguments can be used to convert some non-conservative variables into linearly aggregating processes using simple algebraic transformations. This theoretically more robust approach was here successfully tested in a cross-validation analysis of runoff frequency in Nepal.

5.2 Model selection

The regionalization methods assessed in the cross validation analysis range from simple linear regressions with strong independence assumptions, to complex geostatistical methods that allow for both spatial and topological correlations. Results indicate that while complex methods perform best in general, there seems to be a threshold, beyond which increasing the complexity of the statistical method does not significantly improve the prediction performance: while a linear model is better than a simple average for the prediction of annual streamflow in Nepal (Figure 5 (a)), accounting for spatial (UK) and topological (TR) correlation does not further improve predictions. In that situation, parsimony prescribes selecting the least complex of the best performing methods.

Under these conditions, the selection of the optimal method is driven by the interplay between the layout of the gauges and the spatial correlation range of the considered runoff signature. A dense network of flow gauges is necessary for geostatistical methods to properly estimate the semivariogram and improve on predictions from linear regressions – the case studies suggest that the mean distance between the gauges must be on the order of half the spatial correlation range of the runoff signature. Sparser gauge densities do not allow geostatistical methods to capture spatial correlations and their prediction is effectively driven by the deterministic components of the model, i.e. the intercept and (when available) observable features.

An interesting tradeoff arises if observable features are themselves spatially correlated and explain a significant part of the spatial correlation of the predicted variable. Including these observable features in the model reduces the correlation scale of the residuals, possibly crossing the threshold below which geostatistics are not the most parsimonious approach. In Nepal, controlling for rainfall reduced the spatial correlation range of annual streamflow from 21.6 km to 7 km – well below the mean distance between the gauges (13.9 km). In that case there is a tradeoff between relying on

observable features or variance information to make a prediction, and parsimony and stationarity considerations come into play when selecting the regionalization model. For instance, while parsimony generally prescribes the use of observable features, a climate may be less stationary – and therefore a less reliable external trend – than embedded geology or geomorphology.

In general, geostatistical approaches improve on the prediction of ungauged basins if the distance between the stream gauges is significantly smaller than the spatial correlation scale of runoff. Favorable areas are characterized by high drainage densities or localized rainfall, in addition to a high density of streamflow gauges. All three variables are highly heterogeneously distributed at a global scale, as seen on Figure 7. The multiplicity of local settings likely explains the large diversity of existing regionalization methods and suggests that the selection of the optimal regionalization approach has to be made locally.

Lastly, the decreasing returns to improvements in the complexity of the model also suggest that the performance of statistical methods for PUB is ultimately bounded by the spatial heterogeneity of runoff generating processes. Statistical methods resolve parts of that heterogeneity using the spatial distribution of observable features (linear regressions) and/or based on the analysis of the variance of a sample of the predicted variable (geostatistics). Yet very important parts of the hydrological activity related to storage and flow path characteristics take place underground: they cannot be observed and included in the statistical models (Gupta et al., 2013). This residual spatial heterogeneity can utlimately only be resolved through a better understanding of the particular catchment processes governing runoff in the considered region. Approaches coupling statistical regionalization with process based models that assimilate both a conceptual understanding of catchment scale processes and the random nature of runoff (e.g., Botter et al. (2007); Schaefli et al. (2013); Müller et al. (2014)) are particularly promising.

6 Conclusions

We introduced TopREML as a method to predict runoff signatures in ungauged basins. The approach takes into account the spatially correlated nature of runoff and the nested character of streamflow networks. Unlike kriging approaches, the restricted maximum likelihood (REML) estimators provide the best linear unbiased predictor (BLUP) of both the predicted variable and the associated prediction uncertainty, even when incorporating observable features in the model.

The method was successfully tested in cross validation analyses on mass conserving (mean streamflow) and non-conservative (runoff frequency) runoff signatures in Nepal (sparsely gauged) and Austria (densely gauged), where it matched the performance of the best alternative method: Top-kriging in Austria and linear regression in Nepal. TopREML outperformed Top-kriging in the prediction of uncertainty in Monte Carlo simulations and its performance is robust to the inclusion of observable features.

TopREML's ability to combine deterministic (observable features) and stochastic (covariance) information to generate a BLUP makes it a particularly versatile method that can readily be applied in densely gauged basins, where it takes advantage of spatial covariance information, as well as data-scarce regions, where it can rely on covariates with spatial distributions that are increasingly observable thanks to remote sensing technology. This flexibility, along with its ability to provide a reliable estimate of the prediction uncertainty, offer considerable scope to use this computationally inexpensive method for practical PUB applications.

Appendix A: Covariance of a spatially averaged process

The aim of this analysis is to explore the likely forms of a correlation structure between spatially aggregated processes, given that the underlying point-scale processes are also spatially correlated. In order to maintain tractability, the analysis will consider a strongly idealized case. While we anticipate deviations from the results in non-ideal situations, we nontheless interpret this idealized analysis as offering insight that constrains the choice of correlation function in the TopREML analysis.

Assuming that the underlying point-scale process Y is conservative, the aggregated process y'_k related to the subcatchment S_k of gauge k can be expressed as:

$$y'_k = \frac{1}{A_k} \int_{S_k} Y(x) dx$$

where A_k is the area of S_k.

To proceed, we make the assumption that the area of the drainage areas S_k are approximately equal. While this is a strong constraint, under situations where gauges are placed near confluences and where subcatchments for a given stream ratio are adequately monitored by the gauge network, Horton Scaling ensures that the drainage areas are of a similar order of magnitude. Thus, we will take ($A_k = A \forall k$). The subcatchments are further assumed to have similar shapes and (by definition) do not overlap.

Following Cressie (1993) (p. 68), the covariance between two aggregated random variables y'_k and y'_m is expressed as a function of the covariogram $C_P(\cdot)$ of the underlying point-scale process:

$$\mathsf{Cov}\,(y'_k, y'_m) = \frac{1}{A^2} \int_{S_k} \int_{S_m} C_P(\mid x_2 - x_1 \mid) dx_1 dx_2 = \int_0^\infty \nu(D) C_P(D) dD \tag{A1}$$

where S_k and S_m are the surfaces of subcatchments k and m, and $\nu(D)$ is the probability density function of the distance between randomly chosen points within S_k and S_m – two identical and non-overlapping shapes. Analytical expressions for $\nu(D)$ can be derived for simple geometries (e.g. Mathai, 1999), although complex algebraic expressions typically result. For analytical tractability

we adopt a simplified expression:

$$\nu(D) = \begin{cases} a_0 exp(-a_D D + a_c c) & \text{if } c - D_1 \leq D \leq c + D_2 \\ 0 & \text{otherwise} \end{cases} \tag{A2}$$

which approximates distance frequency function of adjacent elliptical subcatchments, as shown in Figure 8. In equation A2 the parameters a_0, $a_D > a_c$, D_1 and D_2 are positive functions of A, and c is the distance between the centroids of the subcatchments.

We also assume that the underlying point-scale process is second-order stationary and follows an exponential correlation function:

$$C_P(D) = \sigma_p^2 exp(-a_p D) \tag{A3}$$

where σ_p^2 and a_p are respectively the point variance and spatial range of the process.

Inserting Equations A2 and A3 into Equation A1 allows the covariance of the two spatially aggregated random variables to also be expressed as an exponential function of the distance c between their supports

$$C_A(c) = \xi\sigma^2 exp(-\phi c)$$

where $\xi\sigma^2 = \frac{\sigma_p^2 a_0}{a_p + a_D}\left[exp(a_p D_2 + a_D D_2) - exp(-a_p D_1 - a_D D_1)\right] > 0$ and $\phi = a_p + a_D - a_c > 0$. This exponential form was adopted in the covariance derivation in the main text.

We note that within this analysis, the spatial aggregation of the point-scale process creates a nugget variance arising from spatial correlation scales smaller than the subcatchments. The nugget variance can be derived (for this idealized case) by considering the average covariance of points within the catchments:

$$\mathsf{Cov}\,(y'_k, y'_k) = \frac{1}{A^2}\int_{S_k}\int_{S_k} C_P(\mid x_2 - x_1 \mid) dx_1 dx_2 = \int_0^\infty \nu_0(D) C_P(D) dD \tag{A4}$$

where $\nu_0(D)$ now represents the pdf of the distance between two randomly selected points within S_k:

$$\nu(D) = \begin{cases} a_0 exp(-a_D D) & \text{if } 0 \leq D \leq D_0 \\ 0 & \text{otherwise} \end{cases} \tag{A5}$$

where D_0 is the maximum distance between two points within S_k. Again, inserting Equations A5 and A3 into Equation A4, we get the nugget variance resulting from spatial aggregation:

$$C_{A,0} = \frac{\sigma_p^2 a_0}{a_p + a_D}\left[1 - exp(-a_p D_0 - a_D D_0)\right]$$

Therefore, under the aforementioned assumptions, catchment scale variance parameters σ^2 and ξ in Equation 6 can be expressed in terms of point scale parameters:

$$\sigma^2 = \frac{\sigma_p^2 a_0}{a_p + a_D}\left[1 - exp(-a_p D_0 - a_D D_0)\right]$$
$$\xi = \frac{exp(a_p D_2 + a_D D_2) - exp(-a_p D_1 - a_D D_1)}{1 - exp(-a_p D_0 - a_D D_0)}$$

Appendix B: Propagation of runoff frequency in a stream network

We describe runoff occurrence as a binary random variable taking the value of 1 if an increase in daily streamflow occurs and 0 otherwise. If runoff events are uncorrelated in time, the random variable follows a Bernouilli distribution with frequency λ. At a given gauge on a given day, the random variable takes a value of 0 if *all* of the upstream gauges take a value of 0.

In a simple situation with two upstream sub-basins described by the random variables X and Y, the frequency P_Z of the random variable $Z = max(X, Y)$ can be described as:

$$1 - P_Z = P_{!X,!Y} = P_{!X} P_{!Y|!X} = P_{!X}(1 - P_{Y|!X}) = (1 - P_X)(1 - P_{Y|!X})$$

where $!X$ stands for the event $X = 0$. Applying the law of total probabilities to substitute $P_{Y|!X}$ gives:

$$1 - P_Z = (1 - P_X)\left(1 - \frac{P_Y - P_X P_{Y|X}}{1 - P_X}\right)$$

The covariance of X and Y can be derived as:

$$\mathsf{Cov}\,(X, Y) = \mathsf{E}\,[XY] - \mathsf{E}\,[X]\,\mathsf{E}\,[Y] = P_X P_{Y|X} - P_X P_Y$$

with $\mathsf{E}\,[XY] = 0 \cdot P_{!X,!Y} + 0 \cdot P_{!X,Y} + 0 \cdot P_{X,!Y} + 1 \cdot P_{X,Y} = P_X P_{Y|X}$. Finally, substituting $P_X P_{Y|X}$ for the covariance expression, yields:

$$1 - P_Z = (1 - P_X)\left(1 - \frac{P_Y - [\mathsf{Cov}\,(X, Y) + P_X P_Y]}{1 - P_X}\right) = (1 - P_X)(1 - P_Y) + \mathsf{Cov}\,(X, Y)$$

Extending the above derivation to multiple sub-basins and neglecting the covariance term leads to a linear relation between runoff frequencies at gauge i and at upstream gauges in the following form:

$$ln(1 - \lambda_i) \approx \sum_{k=i}^{k \in \mathsf{UP}_i} ln(1 - \lambda_k)$$

Thus, if runoff pulses occur independently for each sub-basin, TopREML can be applied to $ln(1 - \lambda)$ (setting $a_k = 1$), to estimate runoff frequency at ungauged sites.

Appendix C: Computational Performance of TopREML

An algorithmic chart of TopREML, as implemented in the provided script, is presented in Figure 9. IDAs and the topology of the stream network are extracted from the nested catchment using differential overlay. TopREML uses the BFGS algorithm (Wright and Nocedal, 1999) to maximize the restricted likelihood, with the option of using a stochastic optimization algorithm (Simulated Annealing, (Belisle, 1992)) if a non-differentiable (e.g., spherical) covariance function is selected.

A resampling analysis was performed on Austrian dataset to evaluate the runtime and predictive performance of each method as a function of the topological complexity of the considered region (as proxied by the size of the considered sample of gauges) and the considered semi-variogram model. We randomly selected one validation gauge, and resampled the remaining gauges randomly (no repetition) to obtain the chosen sample size. The resampled gauges were used to estimate summer flow at the validation gauges using TopREML and Top-kriging, and successively assuming an exponential (differentiable) and a spherical (non-differentiable) variogram. In each case, relative error and runtime were recorded. This process was repeated 200 times for each sample size. Results (shown in Figure 10) indicate that the gradient-based optimization algorithm used by TopREML for the differentiable (i.e. exponential) variogram reduces the computation runtime by an order of magnitude, relative to the implementation of Top-Kriging in the rtop package. This computational advantage vanishes if a non-differentiable (i.e. spherical) variogram must be used, which requires stochastic optimization. The results also indicate that the relative computational performance of TopREML improves with the number of gauges, while its predictive performance remains constant and approximately equivalent to Top-kriging.

Acknowledgements. The authors would like to thank Michèle Müller, Morgan Levy, David Dralle, Gabrielle Boisramé and two anonymous reviewers for their helpful review and comments. Data have been graciously provided by the Department of Hydrology and Meteorology of Nepal, the HKH-FRIEND project and the *rtop* package. The Swiss National Science Foundation are gratefully acknowledged for funding (M.F.M). Publication made possible in part by support from the Berkeley Research Impact Initative (BRII) sponsored by the UC Berkeley Library.

References

Anders, A. M., Roe, G. H., Hallet, B., Montgomery, D. R., Finnegan, N. J., and Putkonen, J.: Spatial patterns of precipitation and topography in the Himalaya, Geological Society of America Special Papers, 398, 39–53, 2006.

Bishop, G. D. and Church, M. R.: Automated approaches for regional runoff mapping in the northeastern United States, Journal of Hydrology, 138, 361–383, 1992.

Belisle, C.J.P.: Convergence theorems for a class of simulated annealing algorithms on Rd, Journal of Applied Probability, 29, 885–895, 1992.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. and Savenije, H.: Runoff prediction in ungauged basins: Synthesis across processes, places and scales, Cambridge University Press, 2013.

Bosch, D., Sheridan, J., and Davis, F.: Rainfall characteristics and spatial correlation for the Georgia Coastal Plain, Transactions of the ASAE, 42, 1637–1644, 1999.

Botter, G., Porporato, A., Rodriguez-Iturbe, I., and Rinaldo, A.: Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response, Water Resources Research, 43, 2417, 2007.

Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J., Laaha, G., and Blöschl, G.: Smooth regional estimation of low-flow indices: physiographical space based interpolation and Top-kriging, Hydrology and Earth System Sciences, 15, 715–727, 2011.

Chalise, S., Kansakar, S., Rees, G., Croker, K., and Zaidman, M.: Management of water resources and low flow estimation for the Himalayan basins of Nepal, Journal of Hydrology, 282, 25–35, 2003.

Corbeil, R. R. and Searle, S. R.: Restricted maximum likelihood (REML) estimation of variance components in the mixed model, Technometrics, 18, 31–38, 1976.

Cressie, N.: Statistics for Spatial Data, Wiley, New York, NY, 1993.

Cressie, N., Frey, J., Harch, B., and Smith, M.: Spatial prediction on a river network, Journal of Agricultural, Biological, and Environmental Statistics, 11, 127–150, 2006.

Department of Hydrology and Meteorology: Daily Streamflow and Precipitation Data,, Kathmandu, 2011.

D'Odorico, P. and Rigon, R.: Hillslope and channel contributions to the hydrologic response, Water Resources Research, 39, 2003.

Gilmour, A., Thompson, R., and Cullis, B.: Average Information REML: an efficient algorithm for variance parameter estimation in linear mixed models, Biometrics, 51, 1440–1450, 1995.

Gilmour, A., Cullis, B., Welham, S., Gogel, B., and Thompson, R.: An efficient computing strategy for prediction in mixed linear models, Computational Statistics & Data Analysis, 44, 571–586, 2004.

Global Runoff Data Center: Global Runoff Data Base, Global Runoff Data Centre. Koblenz, Federal Institute of Hydrology (BfG), 2014.

Goovaerts, P.: Geostatistics in soil science: state-of-the-art and perspectives, Geoderma, 89, 1–45, 1999.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, Journal of Hydrology, 228, 113–129, 2000.

Gottschalk, L.: Interpolation of runoff applying objective methods, Stochastic Hydrology and Hydraulics, 7, 269–281, 1993.

Gottschalk, L., Krasovskaia, I., Leblois, E., Sauquet, E.: Mapping mean and variance of runoff in a river basin, Hydrology and Earth System Sciences, 10, 469–484, 2006.

Gupta, H., Blöschl, G., McDonnell, J., Savenije, H., Sivapalan, M., Viglione, A., and Wagener, T.: Outcomes of Synthesis, in: Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales, ed.: Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. and Savenije, H., Cambridge University Press, 2013.

Henderson, C. R.: Best linear unbiased estimation and prediction under a selection model, Biometrics, pp. 423–447, 1975.

Huang, W.-C. and Yang, F.-T.: Streamflow estimation using Kriging, Water Resources Research, 34, 1599–1608, 1998.

HYDRO1k Elevation Derivative Database, U.S. Geological Survey Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, 2004.

Laaha, G., Skøien, J. O., Nobilis, F., and Blöschl, G.: Spatial prediction of stream temperatures using Top-kriging with an external drift, Environmental Modeling & Assessment, 18, 671–683, 2013.

Laaha, G., Skøien, J., and Blöschl, G.: Spatial prediction on river networks: comparison of Top-kriging with regional regression, Hydrological Processes, 28, 315–324, 2014.

Lark, R., Cullis, B., and Welham, S.: On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML, European Journal of Soil Science, 57, 787–799, 2006.

LIS Global lightning Image, NASA EOSDIS Global Hydrology Resource Center (GHRC) DAAC, Huntsville, AL, 2011.

Mathai, A. M.: An introduction to geometrical probability: distributional aspects with applications, vol. 1, CRC Press, Boca Raton, FL, 1999.

Merz, R. and Blöschl, G.: Flood frequency regionalisation - spatial proximity vs. catchment attributes, Journal of Hydrology, 302, 283–306, 2005.

Müller, M. F. and Thompson, S. E.: Bias adjustment of satellite rainfall data through stochastic modeling: Methods development and application to Nepal, Advances in Water Resources, 60, 121–134, 2013.

Müller, M. F., Dralle, D. N., and Thompson, S. E.: Analytical model for flow duration curves in seasonally dry climates, Water Resources Research, 50, 5510–5531, 2014.

Myers, R.: Classical and modern regression with applications, PWS-KENT Publishing Company, 1990.

Olea, R.: Optimal contour mapping using universal kriging, Journal of Geophysical Research, 79, 695–702, 1974.

Patterson, H. D. and Thompson, R.: Recovery of inter-block information when block sizes are unequal, Biometrika, 58, 545–554, 1971.

Pebesma, E. J.: Multivariable geostatistics in S: the gstat package, Computers & Geosciences, 30, 683–691, 2004.

R Core Team: R: A language and environment for statistical computing, R Foundation Statistical Computing, 2008.

Sauquet, E.: Mapping mean annual river discharges: geostatistical developments for incorporating river network dependencies, Journal of Hydrology, 331, 300–314, 2006.

Sauquet, E., Gottschalk, L., and Leblois, E.: Mapping average annual runoff: a hierarchical approach applying a stochastic interpolation scheme, Hydrological Sciences Journal, 45, 799–815, 2000.

Schaefli, B., Rinaldo, A., and Botter, G.: Analytic probability distributions for snow-dominated streamflow, Water Resources Research, 49, 2701–2713, 2013.

Sivapalan, M., Konar, M., Srinivasan, V., Chhatre, A., Wutich, A., Scott, C., Wescoat, J., and Rodríguez-Iturbe, I.: Socio-hydrology: Use-inspired water sustainability science for the Anthropocene, Earth's Future, 2, 225–230, 2014.

Skøien, J., Merz, R., and Blöschl, G.: Top-kriging – geostatistics on stream networks, Hydrology and Earth System Sciences, 10, 277–286, 2006.

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

Smith, D. F., Gasiewski, A. J., Jackson, D. L., and Wick, G. A.: Spatial scales of tropical precipitation inferred from TRMM microwave imager data, Geoscience and Remote Sensing, IEEE Transactions on, 43, 1542–1551, 2005.

Srinivasan, V. , Thompson, S.E., Madhyastha, K, Penny, G. Jeremiah, K. and Lele, S.: Why is the Arkavathy River drying? A multiple hypothesis approach in a data scarce region, Hydrology and Earth System Sciences, 19, 1905–1917, 2015.

Stokstad, E.: Scarcity of rain, stream gages threatens forecasts, Science, 285, 1199–1200, 1999.

Ver Hoef, J. M. and Peterson, E. E.: A moving average approach for spatial statistical models of stream networks, Journal of the American Statistical Association, 105, 2010.

Ver Hoef, J. M., Peterson, E., and Theobald, D.: Spatial statistical models that use flow and stream distance, Environmental and Ecological statistics, 13, 449–464, 2006.

Viglione, A., Parajka, J., Rogger, M., Salinas, J., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins–Part 3: Runoff signatures in Austria, Hydrology and Earth System Sciences, 17, 2263–2279, 2013.

Watersheds of the World, IUCN, IWMI, Rasmar Convention Bureau and WRI World Resources Institute, Washington, DC, 2003.

Wittenberg, H. and Sivapalan, M.: Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation, Journal of Hydrology, 219, 20–33, 1999.

World Conservation Monitoring Centre (UNEP-WCMC): Mountains and Forests in Mountains, UNEP-WCMC, Cambridge, 2000.

Wright, S. and Nocedal, J.: Numerical optimization, vol. 2, Springer New York, NY, 1999.

Xu, T., Croke, B., and Hutchinson, M. F.: Identification of spatial and temporal patterns of Australian daily rainfall under a changing climate, in: Proceedings of the 7th International Congress on Environmental Modelling and Software, ed.: Ames, D.P, Quinn, N.W.T. and Rizzoli, A.E. , 2014.

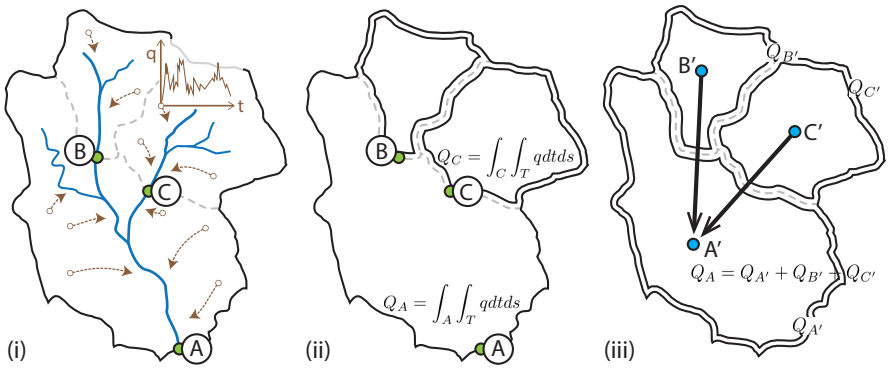

Figure 1. Conceptual flow propagation model. (i) Runoff is generated continuously by a spatially distributed point process and drained to the stream network. (ii) When monitored by stream gauges, runoff is spatially integrated over the corresponding catchment and temporally averaged at the chosen observation frequency (e.g., daily streamflow). (iii) The model conceptualizes the catchments as isolated drainage areas (A', B', and C') representing the *local* runoff contribution to each gauge. The flow actually measured at each gauge is the sum of the upstream isolated drainage areas.

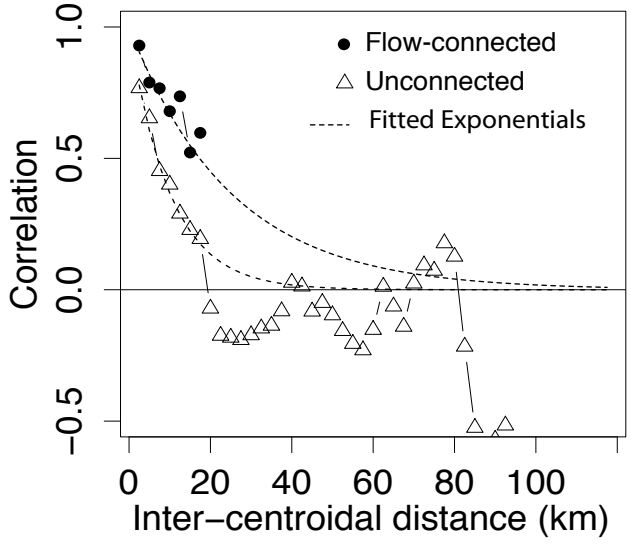

Figure 2. Empirical correlograms of the mean specific summer flow recorded at the 57 gauges of the Austrian dataset. Distance has a different effect on the correlation between flow-connected (black circles) and flow-unconnected (white triangles) gauges. Both correlograms are well fitted by an exponential function but the spatial correlation range doubles when gauges are flow connected. Both empirical correlograms are constructed using 5km bins.

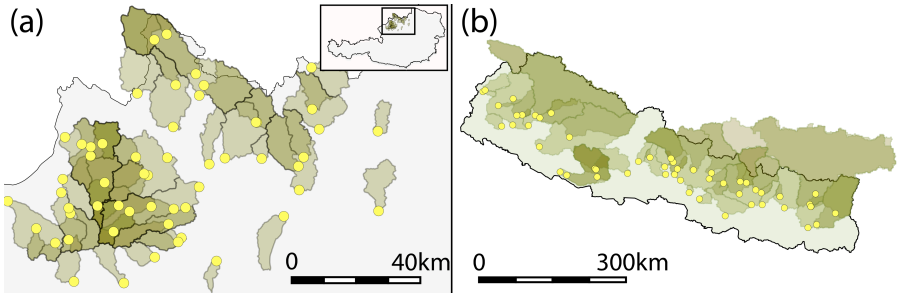

Figure 3. Location of the gauges and related catchments included in the cross validation analyses in Upper Austria (a) and Nepal (b). Coloring is semi-transparent to emphasize overlapping catchment areas. Dark colors represent upstream catchments, which runoff is monitored by many gauges. downstream. Light colors represent downstream catchments with only few downstream gauges to monitor runoff.

Table 1. Taxonomy of the compared regionalization approaches.

	Explanatory Variables	Spatial Covariance	Network Topology	Unbiased Variance
Sample mean				
Linear regression	X			
Universal kriging	X	X		
Top-kriging	X	X	X	
TopREML	X	X	X	X

Table 2. Catchment characteristics of the case studies.

	N	Q	λ	A	c	Dpt	P_y	z_g
Nepal	52	1660	0.42	2121	13.9	10	1683	320
		(1062, 2228)	(0.40, 0.46)	(513, 5267)	(9.2, 25.2)		(1482, 1909)	(507, 750)
Austria	57	0.68		68	4.5	8		
		(0.42,1.43)		(44, 136)	(3.9, 6.3)			

N is the number of catchments; Q the specific runoff $[mm/y]$ in Nepal and the mean summer streamflow $[m^3/s]$ in Austria; λ is the rainy season runoff frequency (d^{-1}) in Nepal; A the catchment area in km^2; c the distance in km between the centroids of isolated drainage areas; Dpt the depth of the stream network graph (i.e. the maximum number of flow-connected gauges); P_y the annual rainfall in mm given by TRMM over Nepal and adjusted according to (Müller and Thompson, 2013); z_g is the gauge elevation in meters above sea level. Median values are provided with 25^{th} and $75th$ quantiles in parenthesis.

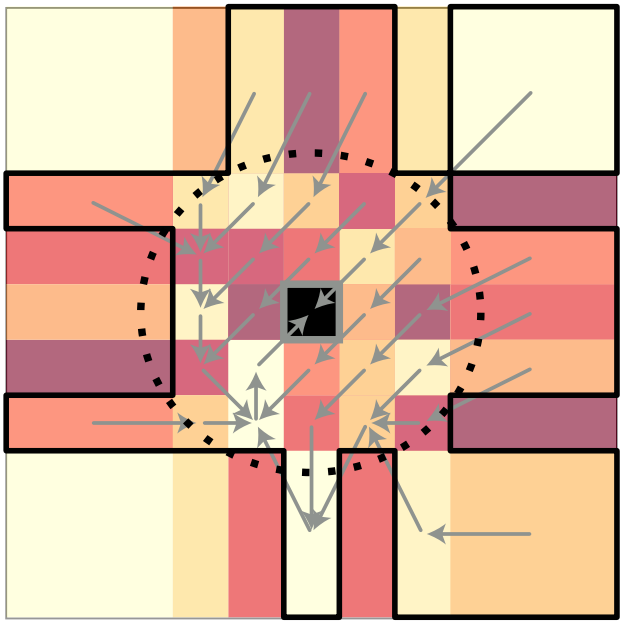

Figure 4. Monte Carlo generation procedure: (i) a spatially correlated gaussian field with an exponential covariance function (mean=30 , partial sill=8, nugget=2, range=3) is generated along a 7×7 irregular grid. The central pixel (in black) represents the downstream-most catchment, where runoff is to be predicted. Among the remaining pixels, 24 *inner* isolated drainage areas (IDA) are within a radius of one spatial correlation range (dashed circle) of the central pixel, and 24 *outer* pixels are beyond that radius. (ii) A predefined number of *inner* and *outer* pixels are randomly selected as part of the set of catchments that are flow-connected to the central pixel. In the figure, all 24 *inner* pixels and 12 *outer* pixels are selected and form the flow catchment outlined with a thick black line. (iii) A tree graph is randomly generated (grey arrows) with its trunk at the prediction pixel and branches passing through all the flow connected pixels. The random field generated in step one is aggregated along the tree by summing the value of all lower order branches at each confluence. (iv) A new spatially correlated field (mean=1, partial sill=0.15, nugget=0, range=0.5) is generated at each pixel – that is the observed trend. The trend is multiplied by a predefined trend coefficient (τ=10) and added to the aggregated runoff at each pixel – that is the observed runoff. (v) Based on the observed runoff and (if applicable) trend at the 48 non-central pixels, TopREML and the compared baseline method (Top-kriging or universal kriging) are used to predict runoff at the central pixel. Prediction errors are recorded and the procedure repeated 1000 times to get the mean and variance of the errors.

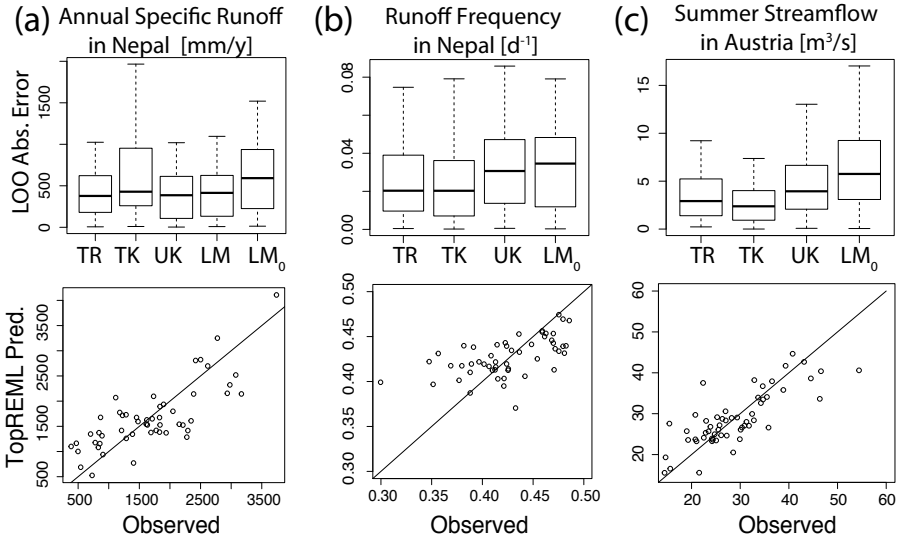

Figure 5. Results of the comparative cross validation analyses of (a) specific runoff and (b) wet season runoff frequency in Nepal, and (c) mean summer streamflow in Austria. *First row:* Box plots with the quartiles and 95% confidence intervals around the median of leave-one-out (LOO) absolute prediction errors. Compared models are TopREML (TR), Top-kriging (TK), universal kriging (UK), linear regression models (LM) and the sample mean (LM_0). Note that without observable trends ((b) and (c)), LM and LM_0 are equivalent. *Second row*: Catchment level performance of TopREML. Signatures predicted by TopREML for each catchment in the leave-one-out crossvalidation analysis are plotted against the corresponding observed signature. Diagonal lines (x=y) representing perfect fit are also displayed for indicative purposes.

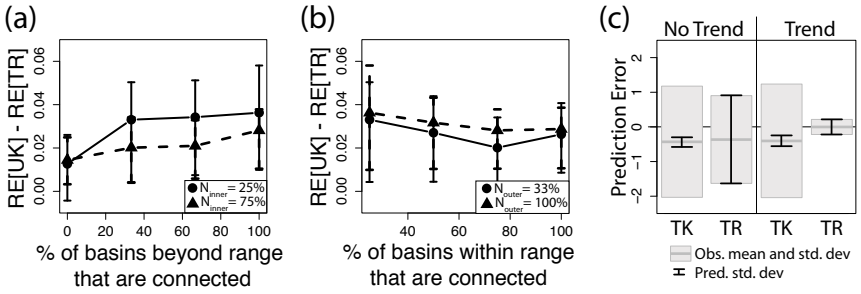

Figure 6. Results of the Monte Carlo experiments. Subfigures (a) and (b) display the effect of network complexity on the performance of TopREML relative to universal kriging. Network complexity is given as the ratio of basins *beyond* (N_{outer}) and *within* (N_{inner}) the spatial correlation range that are flow-connected – minimum network complexity is modeled when *no* basins beyond and *all* basins within the range are flow-connected. Relative performance is computed at each Monte Carlo run as the difference in relative prediction errors between universal kriging and TopREML (i.e. $\mathsf{RE}[\mathsf{UK}] - \mathsf{RE}[\mathsf{TR}]$ on subfigures (a) and (b)). The graphs display the expectation and standard deviation of that difference over the 1000 Monte Carlo runs. Subfigure (c) presents the observed (grey boxes) and predicted (black error bars) standard deviation on the prediction errors for Top-kriging (TK) and TopREML (TR). Note that the slight downward biases that appear on the graph remain below 1% of the expected value of the predicted outcome.

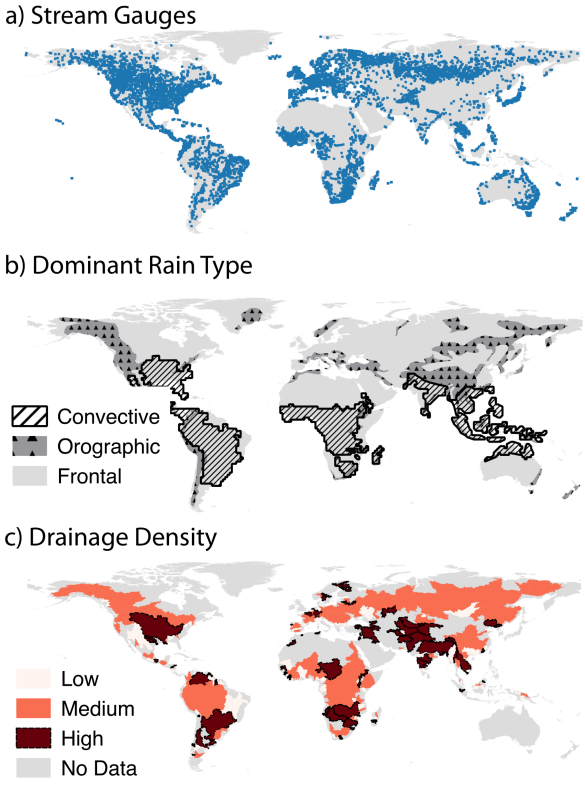

Figure 7. Global distribution of factors affecting model selection. (a) Spatial repartition of the 8540 stream gauges indexed by the Global Runoff Data Center (Global Runoff Data Center, 2014). (b) Dominant rainfall type: orographic rainfall are assumed to occur in mountains, as defined by the United Nations Environment Programme (WCM, 2000), and have a typical range of of 1-10km (Anders et al., 2006). Convective rainfall are assumed dominant in region with a high frequency of lighting strikes ($\geq 10/km^2yr^{-1}$) as recorded by the TRMM satellite (LIS, 2011) and have a typical scale of 10-100km (Bosch et al., 1999; Smith et al., 2005). Frontal precipitations are assumed dominant in the remaining regions and have a typical scale in excess of 100km (Bosch et al., 1999; Xu et al., 2014). (c) Drainage density is estimated based on the DEM-based Hydro1k dataset (Hyd, 2004), using 154 large basins (Wot, 2003) as units of analysis. Drainage densities are displayed in three classes: low ($0.01 - 0.025km^{-1}$), medium ($0.025 - 0.027km^{-1}$) and high ($> 0.027km^{-1}$).

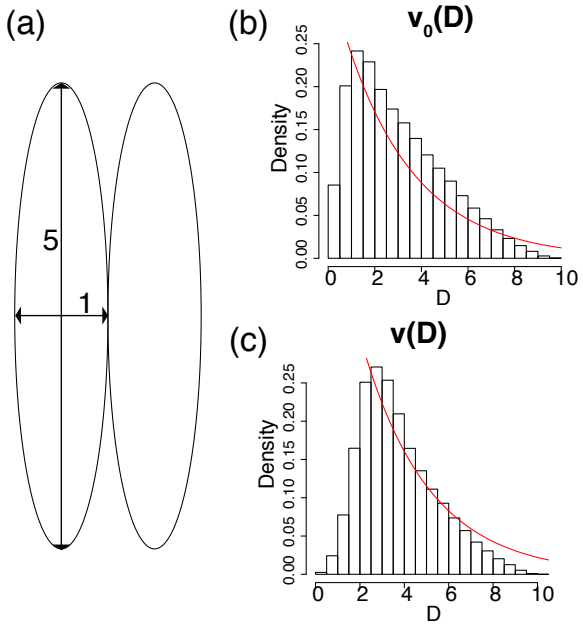

Figure 8. The pdfs assumed in Equations A5 and A2 represent well the case of adjacent ellipsoidal watersheds illustrated in subfigure (a). Subfigure (b) displays the histogram of distance between two random points *within* a watershed, overlaid by a plot of Equation A5 with $a_0 = 3$ and $a_D = 1/3$. Subfigure (c) displays the histogram of distance between one random point *on each* watershed, overlaid by a plot of Equation A2 with $a_c = 1/3$.

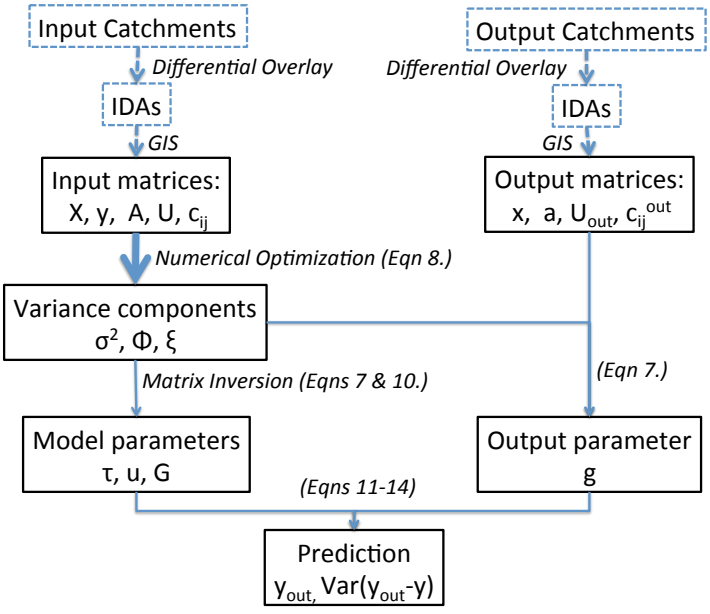

Figure 9. Algorithmic chart of the provided TopREML implementation. Dashed frames and arrows represent vector data and operations and the bold arrow represents the step requiring numerical optimization. The complexity of the computational tasks represented by the remaining plain arrows is driven by matrix inversion, which is of polynomial complexity. In the figure, X is a matrix of observed covariates and y a vector of outcomes measured at the available gauges, as defined in Eqn. 1; x is a vector of identical covariates observed at the prediction location. A, U and c_{ij} are matrices of relative catchment areas, network topology and inter-centroidal distances of the available gauges, as defined in Eqn 6; a, U_{out}, and c_{ij}^{out} are equivalent matrices for the prediction location. σ^2, ϕ and ξ are estimated variance parameters as defined in Eqn 3; τ, u and G are the estimated fixed and random effects (Eqn 10) and variance-covariance matrix (Eqn 7); g is the estimated covariance at the prediction location (used in Eqn 11) . Finally, y_{out} and $\mathsf{Var}\left(y_{out} - y\right)$ are the predicted outcome and the related prediction variance.

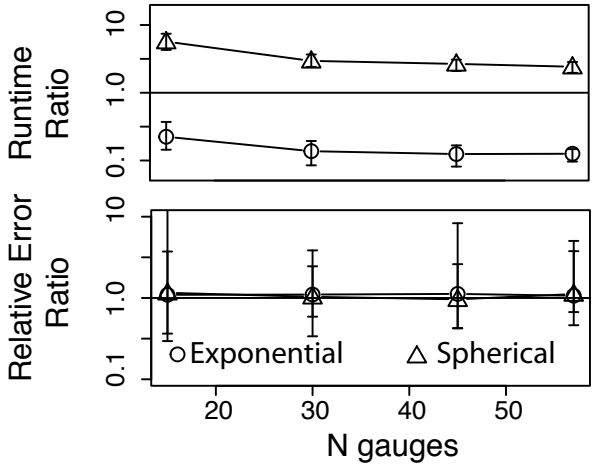

Figure 10. Leave-one-out cross-validation results for Austrian summer flow when resampling a subset of the training gauges. Computational performances are represented as the ratio of runtimes for TopREML against Top-Kriging. Prediction performances are represented as the ratio of relative errors. TopREML performances when using gradient based and stochastic optimization algorithms are represented as circles and triangles respectively. Points represent the median value and error bars represent 90% confidence intervals over 200 repetitions.