



TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

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

M. F. Müller and S. E. Thompson

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

Received: 12 December 2014 – Accepted: 9 January 2015 – Published: 29 January 2015

Correspondence to: M. F. Müller (marc.muller@berkeley.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



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 thanks to 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

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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.

5 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.

25 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

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

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 semivariogram. 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.

streams but does not explicitly model the distribution of an underlying runoff-generating point-scale process. Similar to (Huang and Yang, 1998), we allocate the representative value of the runoff to the centre of gravity of the considered support units, with the important distinction that we consider isolated drainage areas, *not* nested catchments, as area-averaging units (Fig. 1a). Finally, we refrain from using a kriging estimator and 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.

We first derive the TopREML estimator and its variance for mass conserving (i.e. linearly aggregated) variables, with extensions to some non-conservative variables (Sect. 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: Sects. 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 (Sects. 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 (Sect. 5).

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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:

$$y = \underbrace{X}_{\substack{\text{Trends:} \\ \text{Explanatory} \\ \text{variables} \\ (N \times k)}} \underbrace{\tau}_{\substack{\text{Coefficients} \\ (k \times 1)}} + \underbrace{I_N}_{\substack{\text{Identity} \\ \text{Matrix} \\ (N \times N)}} \underbrace{u}_{\substack{\text{Correlated} \\ \text{random} \\ \text{effects} \\ (N \times 1)}} + \underbrace{\epsilon}_{\substack{\text{Residuals,} \\ \text{uncorrelated} \\ \text{errors} \\ (N \times 1)}} \quad (2)$$

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 \mathbf{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

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



specified as:

$$\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \xi \mathbf{G}(\phi) & 0 \\ 0 & \mathbf{I}_N \end{bmatrix} \right) \quad (3)$$

To solve this mixed model, five unknowns must be found: σ^2 , ξ , ϕ , the fixed ($\boldsymbol{\tau}$) and random (\mathbf{u}) effects. Once $\boldsymbol{\tau}$ and \mathbf{u} are known, the empirical best linear unbiased prediction (E-BLUP) of \mathbf{y} can be made at ungauged locations (Lark et al., 2006). The solution strategy adopted here is to prescribe a parametric form for $\mathbf{G}(\phi)$, 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 $\mathbf{G}(\phi)$, 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 Sect. 2.5). Consider a set of stream-flow gauges monitoring a watershed as illustrated in Fig. 1a. 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 Fig. 1a, and leads to a subdivision into non-overlapping areas, each associated with the most upstream gauge that monitors them. In making

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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:

$$y_i = \sum_{k=i}^{k \in \text{UP}_i} a_k y'_k \quad (4)$$

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 ; $a_k = A_k / \sum_{m=i}^{\text{UP}} A_m \leq 1$ 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

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

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

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

where $\text{Cov}(y'_k, y'_m)$ is the covariance between the contributions of sub-catchments k and m . By summing over UP in Eq. (5) (rather than 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 \mathbf{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 (Fig. 7), 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 \mathbf{y} can be expressed as

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

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

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

The topology of the network is described by the matrix \mathbf{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 \mathbf{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 Sect. 3.2. The effect of spatial proximity is addressed by use of the Euclidian distance between catchment centroids (matrix \mathbf{R}), and the effect of scale is accounted for by weighting by the catchment area of the IDAs (matrix \mathbf{A}).

2.3 REML estimation

The restricted maximum likelihood approach partitions the likelihood of $y \sim \mathcal{N}(\mathbf{X}\boldsymbol{\tau}, \sigma^2(\xi\mathbf{G} + \mathbf{I}_N))$ into two parts, one of which is independent of $\boldsymbol{\tau}$ (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)

$$\lambda_R(\sigma^2, \phi, \xi) = -\frac{1}{2} \left(\log \det(\mathbf{X}^T \mathbf{H}^{-1} \mathbf{X}) + \log \det(\mathbf{H}) + \nu \log \sigma^2 + \frac{1}{\sigma^2} \mathbf{y}^T \mathbf{P} \mathbf{y} \right), \quad (8)$$

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

$$\mathbf{K} = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \\ \mathbf{X} & \mathbf{I}_N + \xi^{-1} \mathbf{G}^{-1} \end{bmatrix} \quad (9)$$

The REML estimators $\hat{\sigma}^2$ and $\hat{\phi}$ that maximize λ_R (Eq. 8) can be obtained through numerical optimization.

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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 $\tilde{\mathbf{u}}$ can be obtained by solving the linear system (Henderson, 1975):

$$\mathbf{K}(\hat{\phi}, \hat{\xi}) \begin{bmatrix} \hat{\tau} \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}\mathbf{y} \\ \mathbf{y} \end{bmatrix} \quad (10)$$

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)

$$\tilde{y}_n = \mathbf{x}_n^T \hat{\tau} + \tilde{u}_n = \mathbf{x}_n^T \hat{\tau} + \mathbf{g}_n^T \mathbf{G}^{-1} \tilde{\mathbf{u}} \quad (11)$$

where \mathbf{x}_n is the vector of fixed covariates at ungauged site n , \mathbf{g}_n a correlation vector between site n and each gauge; given $\hat{\phi}$, \mathbf{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

$$\begin{aligned} \text{Var}(\tilde{y}_n - y_n) &= \text{Var}(\mathbf{x}_n^T (\hat{\tau} - \boldsymbol{\tau}) + \mathbf{g}_n^T \mathbf{G}^{-1} (\tilde{\mathbf{u}} - \mathbf{u})) \\ &= \mathbf{x}_n^T \text{Var}(\hat{\tau} - \boldsymbol{\tau}) \mathbf{x}_n + \mathbf{g}_n^T \mathbf{G}^{-1} \text{Var}(\tilde{\mathbf{u}} - \mathbf{u}) \mathbf{G}^{-1} \mathbf{g}_n + 2\mathbf{x}_n^T \text{Cov}(\tilde{\mathbf{u}} - \mathbf{u}, \hat{\tau} - \boldsymbol{\tau}) \mathbf{G}^{-1} \mathbf{g}_n \end{aligned} \quad (12)$$

The covariance matrix of the error on $\boldsymbol{\tau}$ and \mathbf{u} in Eq. (12) can be expressed as a function of the inverse of the model matrix \mathbf{K} defined in Eq. (9) (Lark et al., 2006):

$$\text{Cov} \begin{pmatrix} \hat{\tau} - \boldsymbol{\tau} \\ \tilde{\mathbf{u}} - \mathbf{u} \end{pmatrix} = \sigma^2 \mathbf{K}^{-1} \quad (13)$$

This provides:

$$\text{Var}(\tilde{y}_n - y_n) = \sigma^2 \left(\mathbf{x}_n^T \mathbf{K}_{11}^{-1} \mathbf{x}_n + \mathbf{g}_n^T \mathbf{G}^{-1} \mathbf{K}_{22}^{-1} \mathbf{G}^{-1} \mathbf{g}_n + 2\mathbf{x}_n^T \mathbf{K}_{12}^{-1} \mathbf{G}^{-1} \mathbf{g}_n \right) \quad (14)$$

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Where \mathbf{K}_{11}^{-1} , \mathbf{K}_{22}^{-1} , \mathbf{K}_{12}^{-1} are $k \times k$, $N \times N$ and $k \times N$ partitions of the inverted \mathbf{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 Eq. (14) 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 Eq. (12) and the prediction variance becomes

$$\text{Var}(\tilde{y}_n - y_n)_+ = \text{Var}(\tilde{y}_n - y_n)_- + \sigma^2, \quad (15)$$

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 Sect. 2.2 cannot be applied and the correlation structure in Eq. (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.

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



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

TopREML is implemented in R (R Core Team, 2008). The script is provided as a supplement to this manuscript. To run the script, two vector datasets (e.g., ESRI Shapefile) 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 $[\sigma_0^2, \phi_0, \xi_0] = [\sigma_{LM}^2, E[c_{km}], 1]$ worked well in our case studies, with σ_{LM}^2 the variance of the OLS residuals of the linear model and $E[c_{km}]$ the average distance between IDA centroids.

crementally 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 Fig. 3 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.

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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.

5 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
10 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 Fig. 3. We construct the observed prediction uncertainty by taking the SD 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
15 specification adds a significant spatially correlated component to the error and Eq. (15) 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 Eq. (14) is used.

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



ables are highly heterogeneously distributed at a global scale, as seen on Fig. 6. 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.

5 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
10 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 ultimately only be resolved through a better understanding of the particular catchment processes governing runoff in the considered region. Approaches coupling
15 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.
20

25 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 perfor-

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



mance 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 (co-
variance) 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 pre-
diction 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 be-
tween 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 nonetheless interpret this idealized analysis as offering insight that con-
strains 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

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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.

5 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:

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

10 where S_k and S_m are the surfaces of subcatchments k and m , and $v(D)$ 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 $v(D)$ can be derived for simple geometries (e.g. Mathai, 1999), although complex algebraic expressions typically result. For analytical tractability we adopt a simplified expression:

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

15 which approximates distance frequency function of adjacent elliptical subcatchments, as shown in Fig. 7. In Eq. (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:

$$20 \quad C_P(D) = \sigma_p^2 \exp(-a_p D) \quad (\text{A3})$$

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

Inserting Eqs. (A2) and (A3) into Eq. (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

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

5 where $\xi \sigma^2 = \frac{\sigma_p^2 a_0}{a_p + a_D} [\exp(a_p D_2 + a_D D_2) - \exp(-a_p D_1 - a_D D_1)] > 0$ 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 sub-catchments. The nugget variance can be derived (for this idealized case) by considering the average covariance of points within the catchments:

$$10 \text{Cov}(y'_k, y'_k) = \frac{1}{A^2} \int \int_{S_k S_k} C_P(|x_2 - x_1|) dx_1 dx_2 = \int_0^\infty v_0(D) C_P(D) dD \quad (\text{A4})$$

where $v_0(D)$ now represents the pdf of the distance between two randomly selected points within S_k :

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

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

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

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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

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

$$\xi = \frac{\exp(a_p D_2 + a_D D_2) - \exp(-a_p D_1 - a_D D_1)}{1 - \exp(-a_p D_0 - a_D D_0)}$$

5 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 Bernoulli 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:

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

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



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

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

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:

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

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.

The Supplement related to this article is available online at doi:10.5194/hessd-12-1355-2015-supplement.

Acknowledgements. The authors would like to thank Michèle Müller, Morgan Levy, David Dralle and Gabrielle Boisramé for their helpful assistance, 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 is gratefully acknowledged for funding (M.F.M). Publication made possible in part by support from the Berkeley Research Impact Initiative (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, Geol. Soc. Am. Spec. Pap., 398, 39–53, 2006. 1395

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

- Bishop, G. D. and Church, M. R.: Automated approaches for regional runoff mapping in the northeastern United States, *J. Hydrol.*, 138, 361–383, 1992. 1357
- Blöschl, G., Sivapalan, M., and Wagener, T.: *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, Cambridge University Press, Cambridge, 2013. 1356, 1357
- Bosch, D., Sheridan, J., and Davis, F.: Rainfall characteristics and spatial correlation for the Georgia Coastal Plain, *T. ASAE*, 42, 1637–1644, 1999. 1395
- 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 Resour. Res.*, 43, 2417, doi:10.1029/2006WR005043, 2007. 1368, 1378
- Castiglioni, S., Castellarin, A., Montanari, A., Skøien, J. O., Laaha, G., and Blöschl, G.: Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging, *Hydrol. Earth Syst. Sci.*, 15, 715–727, doi:10.5194/hess-15-715-2011, 2011. 1359
- 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, *J. Hydrol.*, 282, 25–35, 2003. 1370
- Corbeil, R. R. and Searle, S. R.: Restricted maximum likelihood (REML) estimation of variance components in the mixed model, *Technometrics*, 18, 31–38, 1976. 1366
- Cressie, N.: *Statistics for Spatial Data*, Wiley, New York, NY, 1993. 1358, 1359, 1361, 1365, 1376, 1380
- Cressie, N., Frey, J., Harch, B., and Smith, M.: Spatial prediction on a river network, *J. Agric. Biol. Envir. S.*, 11, 127–150, 2006. 1360
- Daily Streamflow and Precipitation Data: Department of Hydrology and Meteorology, Government of Nepal, Kathmandu, 2011. 1370
- D’Odorico, P. and Rigon, R.: Hillslope and channel contributions to the hydrologic response, *Water Resour. Res.*, 39, 1113, doi:10.1029/2002WR001708, 2003. 1364
- 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. 1366
- Gilmour, A., Cullis, B., Welham, S., Gogel, B., and Thompson, R.: An efficient computing strategy for prediction in mixed linear models, *Comput. Stat. Data An.*, 44, 571–586, 2004. 1361

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- 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, *Comput. Geosci.*, 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, *IEEE T. Geosci. Remote*, 43, 1542–1551, 2005. 1395
- Srinivasan, V., Thompson, S., Madhyastha, K., Penny, G., Jeremiah, K., and Lele, S.: Why is the Arkavathy River drying? A multiple hypothesis approach in a data scarce region, *Hydrol. Earth Syst. Sci. Discuss.*, 12, 25–66, doi:10.5194/hessd-12-25-2015, 2015. 1357
- Stokstad, E.: Scarcity of rain, stream gages threatens forecasts, *Science*, 285, 1199–1200, 1999. 1356
- Ver Hoef, J. M. and Peterson, E. E.: A moving average approach for spatial statistical models of stream networks, *J. Am. Stat. Assoc.*, 105, 6–18, doi:10.1198/jasa.2009.ap08248, 2010. 1360
- Ver Hoef, J. M., Peterson, E., and Theobald, D.: Spatial statistical models that use flow and stream distance, *Environ. Ecol. Statist.*, 13, 449–464, 2006. 1360
- Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci.*, 17, 2263–2279, doi:10.5194/hess-17-2263-2013, 2013. 1359
- Watersheds of the World: IUCN, IWMI, Rasmussen Convention Bureau and WRI World Resources Institute, Washington, DC, 2003. 1395
- Wittenberg, H. and Sivapalan, M.: Watershed groundwater balance estimation using streamflow recession analysis and baseflow separation, *J. Hydrol.*, 219, 20–33, 1999. 1369
- World Conservation Monitoring Centre (UNEP-WCMC): Mountains and Forests in Mountains, raster dataset, UNEP-WCMC, Cambridge, 2000. 1395
- Wright, S. and Nocedal, J.: Numerical Optimization, vol. 2, Springer, New York, NY, 1999. 1369
- 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, edited by: Ames, D. P., Quinn, N. W. T., and Rizzoli, A. E., San Diego, California, USA, 15–19 June, 2014. 1395

HESSD

12, 1355–1396, 2015

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)**Table 1.** Taxonomy of the compared regionalization approaches.

	Explanatory variables	Spatial covariance	Network topology	Unbiased variance
Sample mean				
Linear regression	×			
Universal kriging	×	×		
Top-kriging	×	×	×	
TopREML	×	×	×	×

TopREML – runoff regionalization on stream networks

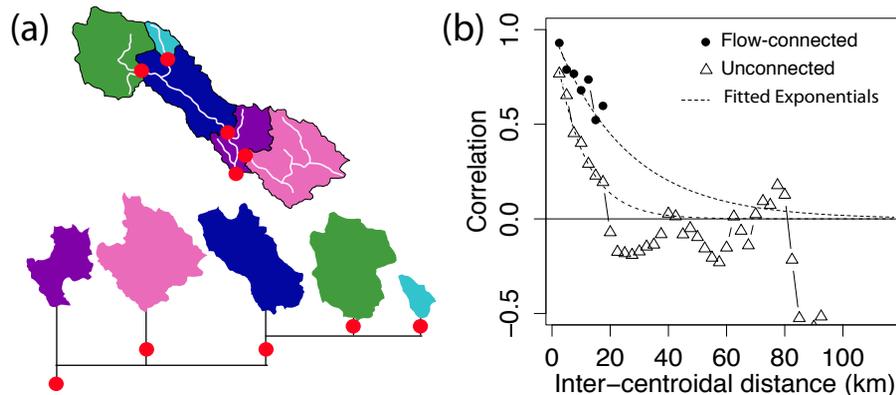
M. F. Müller and
S. E. Thompson

Figure 1. (a) Example of isolated drainage area (IDA) representation: physical layout of the gauges in the watershed (above) and IDA representation of the stream network (below). (b) 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 5 km bins.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

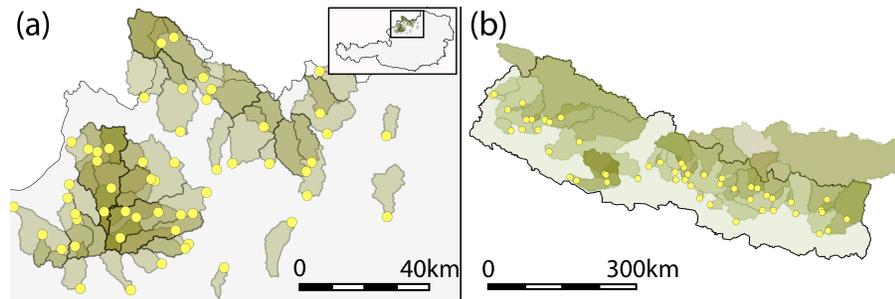
TopREML – runoff regionalization on stream networksM. F. Müller and
S. E. Thompson

Figure 2. Location of the gauges and related catchments included in the cross validation analyses in Upper Austria **(a)** and Nepal **(b)**.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

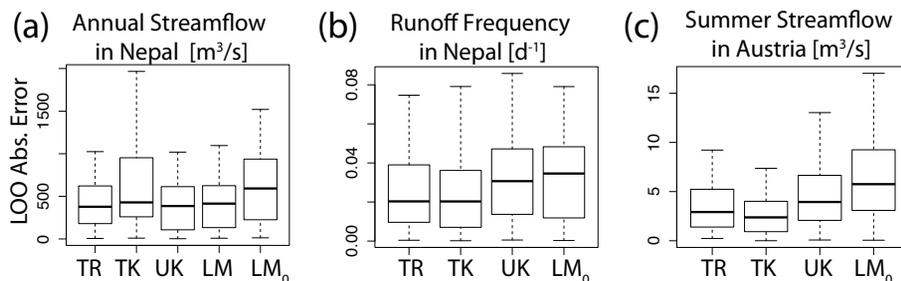


Figure 4. Results of the comparative cross validation analyses of **(a)** annual streamflow and **(b)** wet season runoff frequency in Nepal, and **(c)** mean summer streamflow in Austria. Box plots with the quartiles and 95% confidence intervals around the median of leave-one-out prediction errors are given. Compared models are TopREML (TR), Top-kriging (TK), universal kriging (UK), linear regression models (LM) and the sample mean (LM₀). Note that without observable trends **(b)** and **(c)**, LM and LM₀ are equivalent.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[⏪](#)
[⏩](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

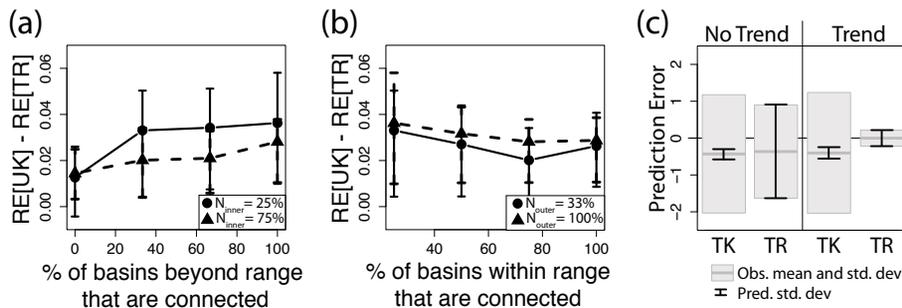


Figure 5. Results of the Monte Carlo experiments. Subfigures (a) and (b) display the effect of network complexity on the performance of TopREML relative to universal kriging. Network complexity is given as the ratio of basins *beyond* (N_{outer}) and *within* (N_{inner}) the spatial correlation range 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 – the graphs display the expectation and SD of that difference over the 1000 Monte Carlo runs. Subfigure (c) presents the observed (grey boxes) and predicted (black error bars) SD 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.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

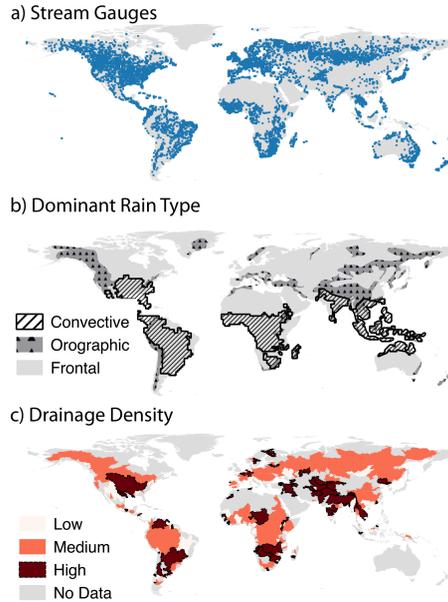


Figure 6. Global distribution of factors affecting model selection. **(a)** Spatial repartition of the 8540 stream gauges indexed by the Global Runoff Data Center (Global Runoff Data Center, 2014). **(b)** Dominant rainfall type: orographic rainfall are assumed to occur in mountains, as defined by the United Nations Environment Programme (WCM, 2000), and have a typical range of of 1–10 km (Anders et al., 2006). Convective rainfall are assumed dominant in region with a high frequency of lightning strikes ($\geq 10 (\text{km}^2 \text{yr}^{-1})^{-1}$) as recorded by the TRMM satellite (LIS, 2011) and have a typical scale of 10–100 km (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 100 km (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 ($0.01\text{--}0.025 \text{ km}^{-1}$), medium ($0.025\text{--}0.027 \text{ km}^{-1}$) and high ($> 0.027 \text{ km}^{-1}$).

TopREML – runoff regionalization on stream networks

M. F. Müller and
S. E. Thompson

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



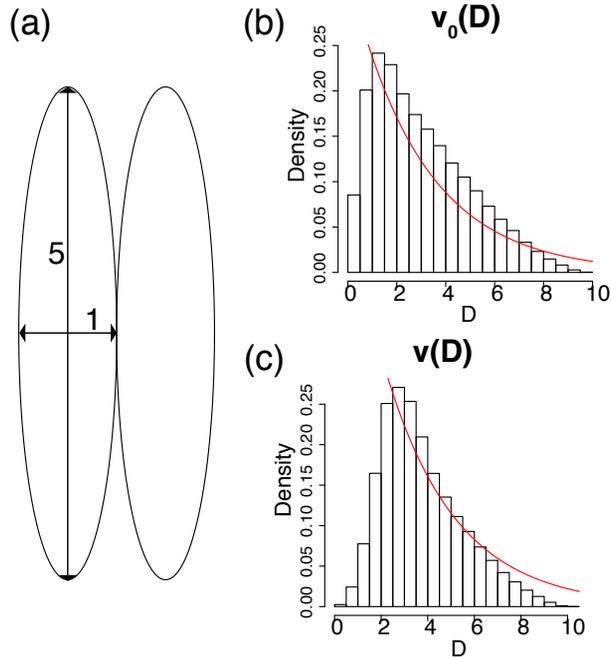


Figure 7. The pdfs assumed in Eqs. (A5) and (A2) represent well the case of adjacent ellipsoidal watersheds illustrated in subfigure (a). Subfigure (b) displays the histogram of distance between two random points *within* a watershed, overlaid by a plot of Eq. (A5) with $a_0 = 3$ and $a_D = 1/3$. Subfigure (c) displays the histogram of distance between one random point *on each* watershed, overlaid by a plot of Eq. (A2) with $a_c = 1/3$.