We thank the reviewer for their positive feedback and constructive suggestions to improve the manuscript.

The reviewer's main points relate to the need to further explain the advantage of TopREML in computation and conceptual terms, taking Top-Kriging as implemented in the *rtop* package as a baseline comparison. We propose (a) providing a new supplement to the manuscript that details a comparison of the computational algorithms and efficiency for both methods for different gauge densities and different (differentiable and non-differentiable) correlation functions. Secondly, (b) we have revised the text in sections 5.1 to reflect the outcomes of this additional analysis.

While we note the reviewer's comment that the elucidation of the implications of TopREML for the global challenge of PUB could be bolstered by additional analysis, we consider that to provide such analysis would provide a great expansion of the scope of the current paper, and is better suited to a follow-up study.

(a) Proposed new supplementary section:

While we agree with the reviewer that further analysis is justified to explore the computational advantages of TopREML relative to alternative interpolation options, we propose locating this analysis in a supplementary section for two reasons:

(i) to avoid unnecessarily elongating the paper; and

(ii) because we will not undertake a formal optimization of the algorithms used to implement TopREML or Top-Kriging, but simply provide a comparison of the current implementations provided by the authors, we do not want to over-state the finality of a comparison of the numerical implementation. The authors are not computer scientists, and we do not wish to claim that the computational differences between TopREML and Top-Kriging in their current implementations would persist under all possible numerical schema.

Within the supplement, we will provide the following information:

1. A new algorithmic chart that illustrates the computational steps involved in implementing TopREML (Figure S1).

As outlined in this chart, 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 log likelihood, with the option of using a stochastic optimization algorithm (Simulated Annealing, Bellisle 1992) if a non-differentiable (e.g., spherical) covariance function is selected.

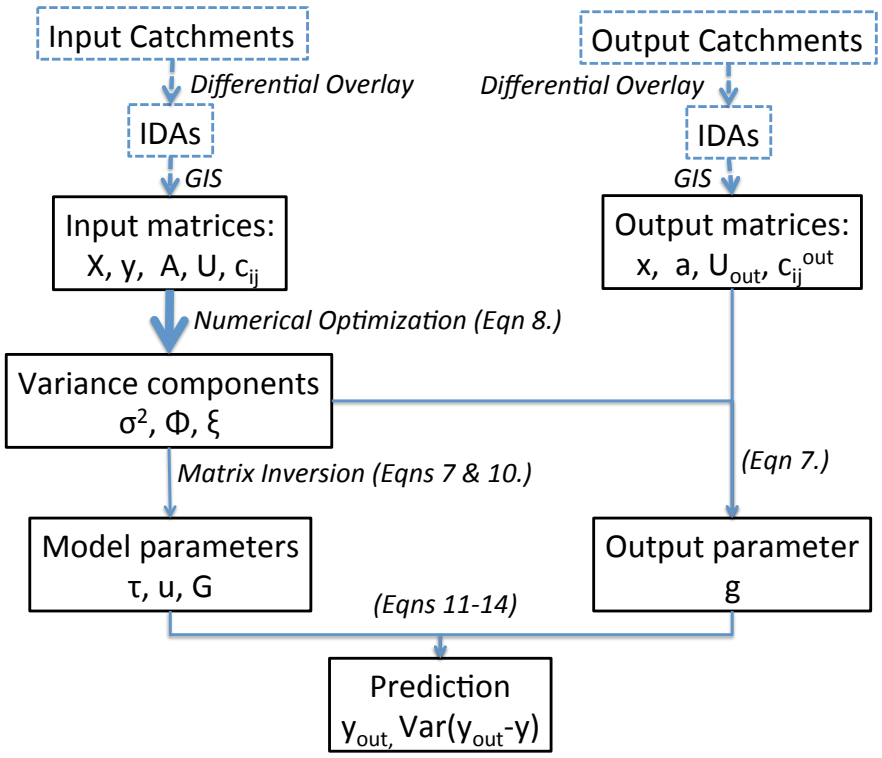

Fig. S1. *Algorithmic chart of the provided TopREML implementation. Dashed frames and arrows represent vector data and operations and the bold arrow represents the step requiring numerical optimization. The complexity of the computational tasks represented by the remaining plain arrows is driven by matrix inversion, which is of polynomial complexity. In the figure, X is a matrix of observed covariate 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, 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 $Var(y_{out}-y)$ are the predicted outcome and the related prediction variance.*

2. A new analysis of the Austrian dataset used to evaluate runtime as a function of the input data complexity, correlation structure and choice of interpolation method (TopREML versus Top-Kriging).

In this analysis, we randomly select one validation gauge, and resample the remaining gauges randomly (no repetition) to generate a given prediction set size. The resampled gauges are used to estimate summer flow at the validation gauge using TopREML and Top-Kriging, and assuming firstly an exponential and secondly a spherical variogram. In each case, relative error and runtime are recorded. This process is repeated 200 times for each size of prediction set. The results are shown in Figure S2:

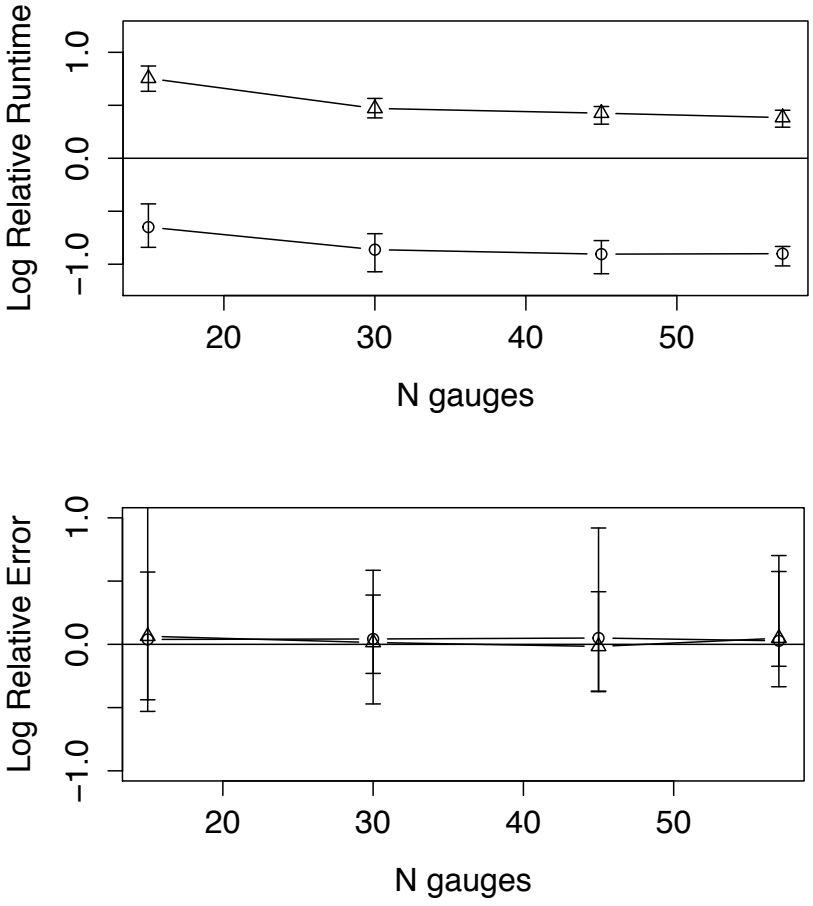

Fig S2. *Leave-one-out cross-validation results for Austrian summer flow when resampling a subset of the training gauges. Computational performances (subfigure a) are represented in terms of log relative runtime [log($RT_{TopREML}/RT_{Top\text{-}Kriging}$)]. Prediction performances are represented as log ratio of relative errors [log($RE_{TopREML}/RE_{Top\text{-}Kriging}$)]. 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.*

The results indicate that the gradient-based optimization algorithm used by TopREML for *differentiable variograms* 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 differentiable variogram function cannot be assumed and stochastic optimization is required. 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.

(b) Amended text in sections 5.1

We propose new text in this section to clarify the computational performance of TopREML.

TopREML has considerably lower computational requirement than Top-Kriging, both in terms of input data and optimization complexity. While vector input data are necessary input data for Top-Kriging, which requires polygons to discretize catchment areas in 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 (Skoien 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 using gradient optimization methods that are much less computationally intensive than the stochastic optimization algorithm required by TopKriging.