Streamflow prediction in ungauged catchments using a copula-based similarity measure

Luis Samaniego and Rohini Kumar

Helmholtz - Centre for Environmental Research - UFZ, Germany (luis.samaniego@ufz.de)

1. Introduction

In this study we propose a method to estimate daily streamflow time series in an ungauged basin using a process-based distributed hydrologic model (HBV-UFZ) and sets of model parameters obtained from donor basins. Suitable donor basins are those whose distance to the ungauged catchment in a m-dimension embedding space of catchment predictors is the least. The embedding space is found as a linear or nonlinear transformation of the n-dimension space ($n > m$) of the catchment descriptors.

2. Research questions

1. How to find an appropriate similarity measure suitable for finding basins with an homogeneous hydrologic response?
2. How to use this similarity measure to estimate daily streamflow in an ungauged basin?

3. Traditional catchment characterization

Donor basins $\longrightarrow \mathcal{D} = \{(\mathbf{x}_i, y_i) \quad i = 1, \dots n\}$

$\mathcal{Q}_i = \{q_i^t \quad t = 1, \dots \mathcal{T}\}$

Ungauged basin $\longrightarrow \mathcal{D}_u = \{\mathbf{x}_j \quad j \neq i \quad \forall i\}$

Similarity measure $\longrightarrow$ Small Euclidian distance in $\mathbf{x}$

- Similarity measure selected *a priori*, e.g. Euclidian distance.
- Employ unsupervised approaches that are highly uncertain.
- Characterization is not unique.

$\Rightarrow$ Lack of transferability.

4. Notation

i, j	Indexes for basins.	q_i^t		Discharge for basin i at day t.
t	Time index.	$F(\bullet)$		Hydrologic model HBV-UFZ.
k	Index for a realization.	$\mathbf{M}_j^t$		Meteorologic inputs.
$\mathbf{x}$	m-dimensional vector of inputs.	$\mathbf{U}_j^t$		Land cover inputs.
$\mathbf{u}$	n-dimensional vector of inputs in the embedding space.	$\mathbf{G}_j$		Morphologic inputs.
n	The sample size of the data set $\mathcal{D}$.	$\hat{\beta}_{ik}$		k-realization of the parameters from donor catchment i.
B	Embedding transformation (possibly nonlinear).	x_1	$[m^2]$	Area
λ_{ij}	Similarity measure based on a density copula $c(\cdot)$.	x_2	[1]	Trimmed mean slope $P_{15} - P_{85}$
y	Validation runoff characteristic (e.g. $\hat{q}$).	x_3	[1]	Fraction of north facing slopes
$\mathcal{T}, \mathcal{D}$	Training sets.	x_4	$[m]$	$h_{max} - h_{min}$ (elevation)
$\mathcal{V}$	Validation set.	x_5	[1]	Fraction of karstic formation
N	Number of pairs separated with a distance less than $D(N)$.	x_6	$[mm]$	Mean Maximum water capacity
		x_7	[1]	Mean share of impervious areas
		x_8	$[mm]$	30y-mean annual precipitation
		x_9	$[K]$	30y-mean annual temperature January
		p		Moment threshold

5. Copula based technique

Training $\longrightarrow \mathcal{T} = \{\mathbf{x}_i, [\lambda_{ij}], y_i \quad \forall\, i, j \in n\}$

$\mathcal{Q}_i = \{q_i^t \quad t = 1, \dots \mathcal{T}\}$

Embedding $\longrightarrow \mathbf{u} = B[\mathbf{x}]$

Similarity Measure $\longrightarrow$ Small $d(\mathbf{u}_i, \mathbf{u}_j) \rightarrow$ Small λ_{ij}

Metric $\longrightarrow d_{ij}^2 = (\mathbf{u}_i - \mathbf{u}_j)\mathbf{g}(\mathbf{u}_i - \mathbf{u}_j)^{\mathrm{T}}$

Crossvalidation $\longrightarrow \mathcal{V} = \{(\mathbf{x}, y)\}$

$$y = \frac{1}{N} \sum_{d_B(\mathbf{u}, \mathbf{u}_i) < D(N)} y_i$$

Streamflow Prediction $\longrightarrow q_j^t = \frac{1}{N} \sum_{k \in \mathrm{NN}_j} \sum_{d_B(\mathbf{u}, \mathbf{u}_i) < D(N)} \hat{q}_{jik}^t$

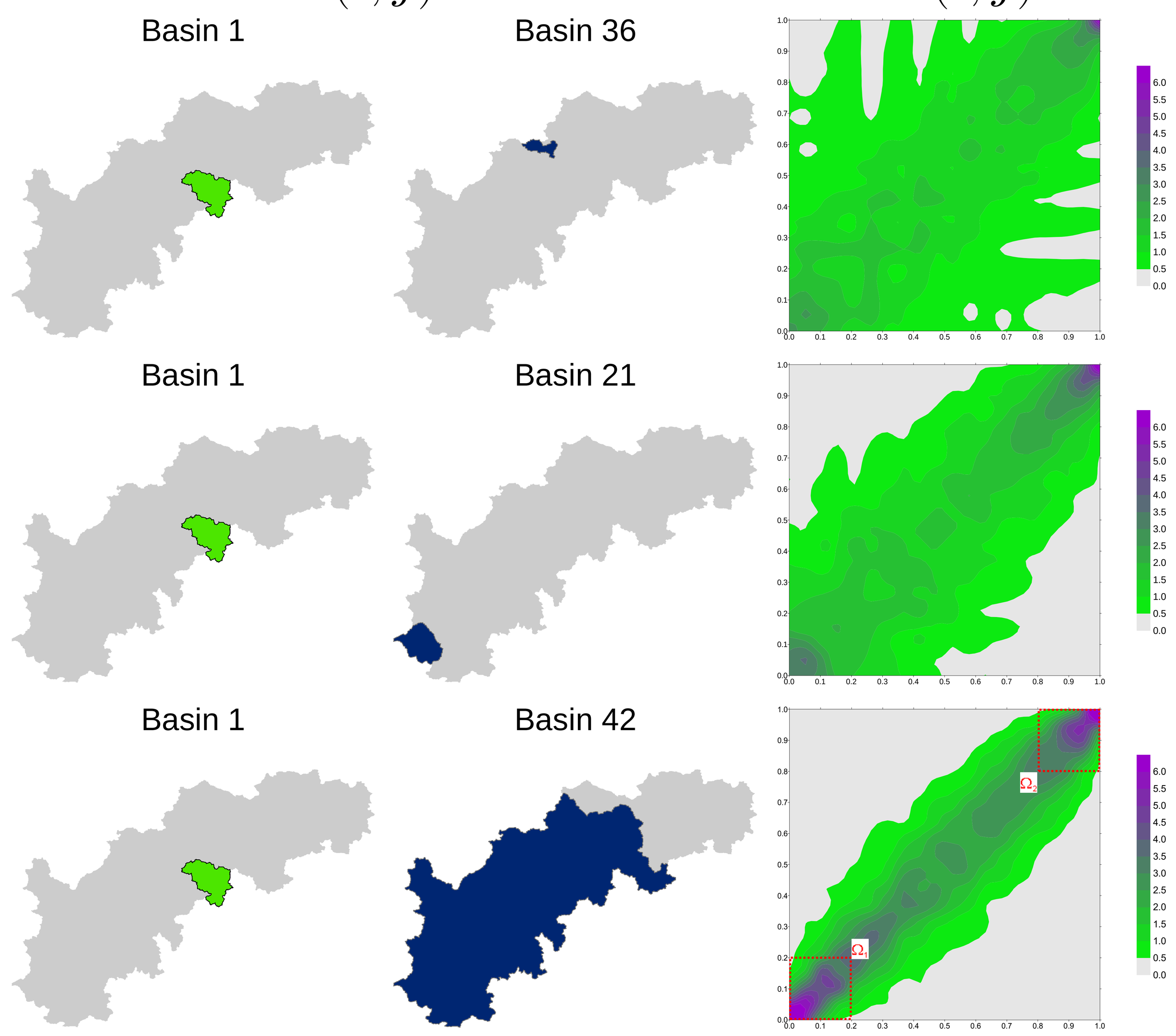
$$\hat{q}_{jik}^t = F(\mathbf{M}_j^t, \mathbf{G}_j, \mathbf{U}_j^t, \hat{\beta}_{ik})$$

- The streamflow prediction for an ungauged basin is the ensemble mean of the streamflows generated by the hydrological model $F(\bullet)$ using parameters obtained from its k-nearest neighboring donor basins.

$\Rightarrow$ Suitable for transferability

6. Similarity measure based on runoff density copulas

Basin pair (i, j)	Density copula $c(i, j)$	Sim. Measure λ_{ij}
Basin 1, Basin 36		57.7
Basin 1, Basin 21		47.1
Basin 1, Basin 42		22.4



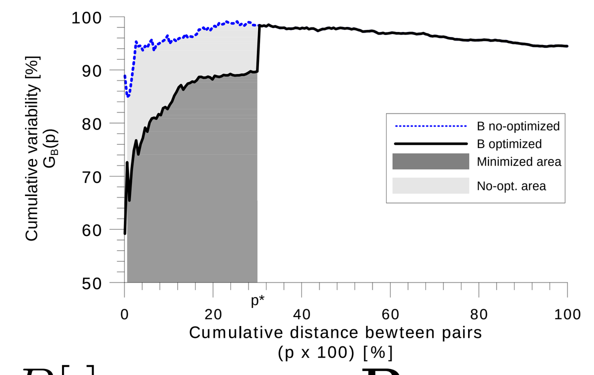
Moments $\longrightarrow L_{ij} = \int \int_{\Omega_1} c(v_i, v_j) d\Omega$

$U_{ij} = \int \int_{\Omega_2} c(v_i, v_j) d\Omega$

Similarity Measure $\longrightarrow \lambda_{ij} = \frac{|U_{ij} - L_{ij}|}{U_{ij} + L_{ij}} + (p - L_{ij})$

7. Local variance minimization

Define a variance function [1]: $G_B(p) = \frac{1}{N(D_B(p))} \Sigma_{d_B(i,j) < D_B(p)} \left(\lambda_{ij}\right)^2$



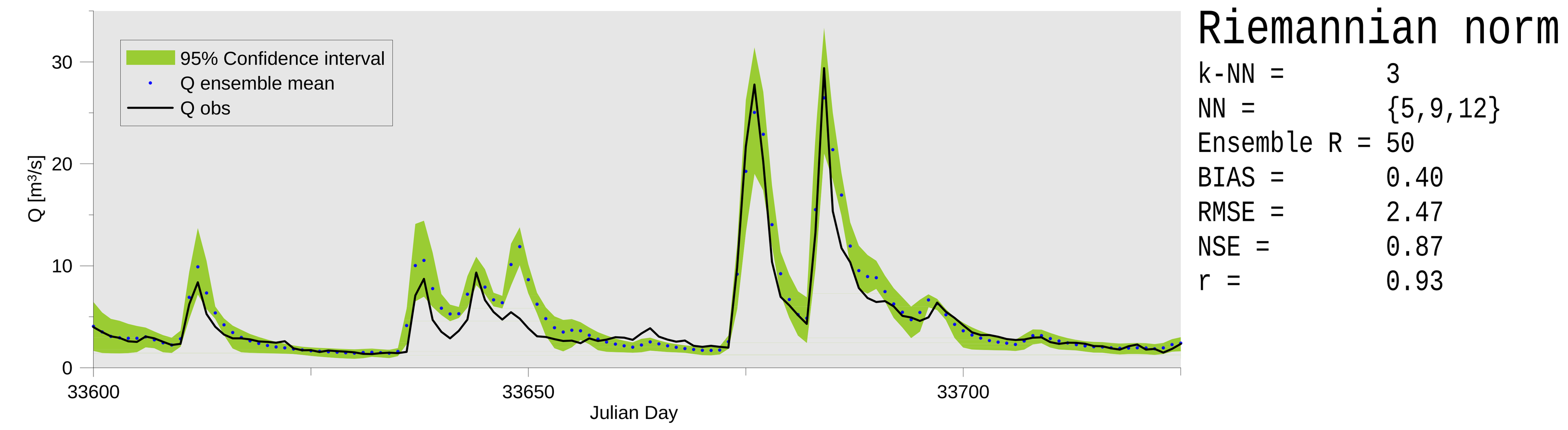
Find a transformation and a metric: $B[\cdot]$, e.g. $\mathbf{u} = \mathbf{B}\mathbf{x}$

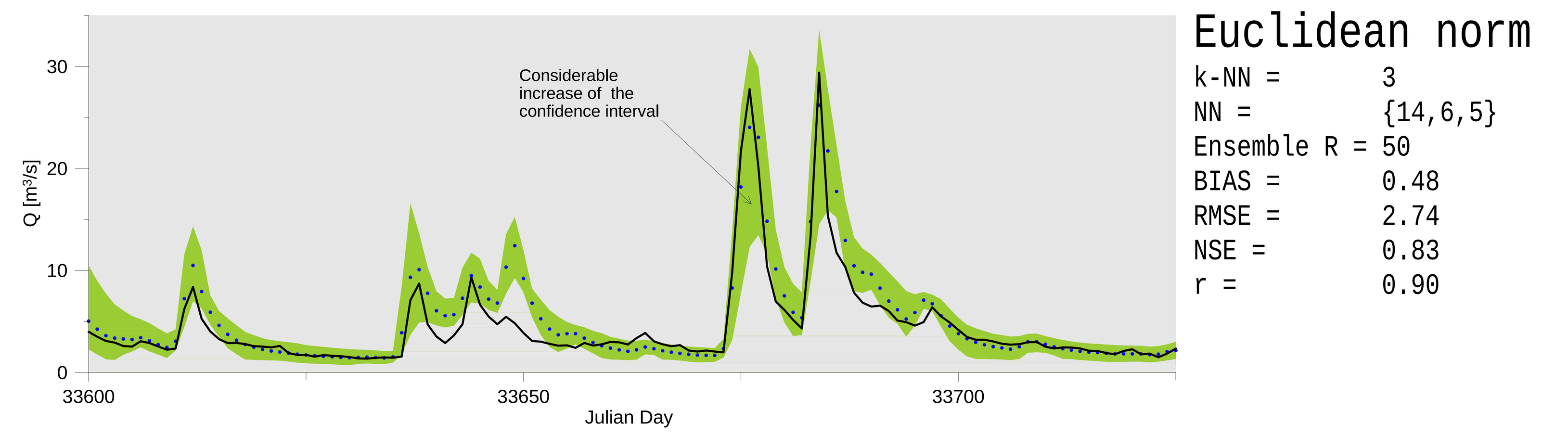
$\mathbf{g} = [g_{ij}]$, positive definite

so that $\int_0^{p^*} G_B(p) dp \rightarrow \min$

8. Results

Crossvalidation results for basin 7, using k=3 Nearest Neighbors (NN) and R =50 ensemble predictions using parameters sets from donor k-NN catchments.





9. Conclusions

The crossvalidation results of the study indicate that:

- The Euclidian distance is not appropriate to find NN in the space of the catchment descriptors.
- The copula based method reduced considerably the confidence intervals, BIAS and RMSE of the predictions. NSE and r were, in turn, increased.

References

[1] A. Bárdossy, G. S. Pegram, and L. Samaniego, "Modeling data relationships with a local variance reducing technique: Applications in hydrology," *Water Resour. Res.*, vol. 41, 2005.

HELMHOLTZ CENTRE FOR ENVIRONMENTAL RESEARCH – UFZ