JULY 8-18 JUILLET 2019 | MONTRÉAL, CANADA
27th IUGG General Assembly
27e Assemblée Générale de l'UGGI
International Union of Geodesy and Geophysics | Union Géodésique et Géophysique Internationale
IUGG Centennial | 1919-2019 | Centenaire de l'UGGI

IAHS AISH

R in Hydrology Workshop:
GEOSTATISTICAL PREDICTION OF STREAMFLOW INDICES: TOP-KRIGING via the R-package rtop

Attilio Castellarin (attilio.castellarin@unibo.it)
Simone Persiano (simone.persiano@unibo.it)
Alessio Pugliese (alessio.pugliese@unibo.it)
Department of Civil, Chemical, Environmental and Materials Engineering, DICAM University of Bologna, Italy

Overview

WHY?

- Geostatistical interpolation of streamflow indices and its advantages over conventional approaches to statistical regionalization (special focus on Topological Kriging, or Top-kriging)

WHAT?

- Top-kriging interpolation in a nutshell
- Real-world applications (examples)

HOW?

- Tutorial: Top-kriging in R with R-package ‘*rtop*’

Introduction: PUB – Prediction in Ungauged Basins

Decade on PUB, promoted by IAHS (Sivapalan et al., HSJ, 2003)

In operational hydrology, the scientific community has addressed since the '70's the PUB problem by systematically applying hydrological classification as a tool through regionalization studies.

Example: Regional Flood Frequency Analysis

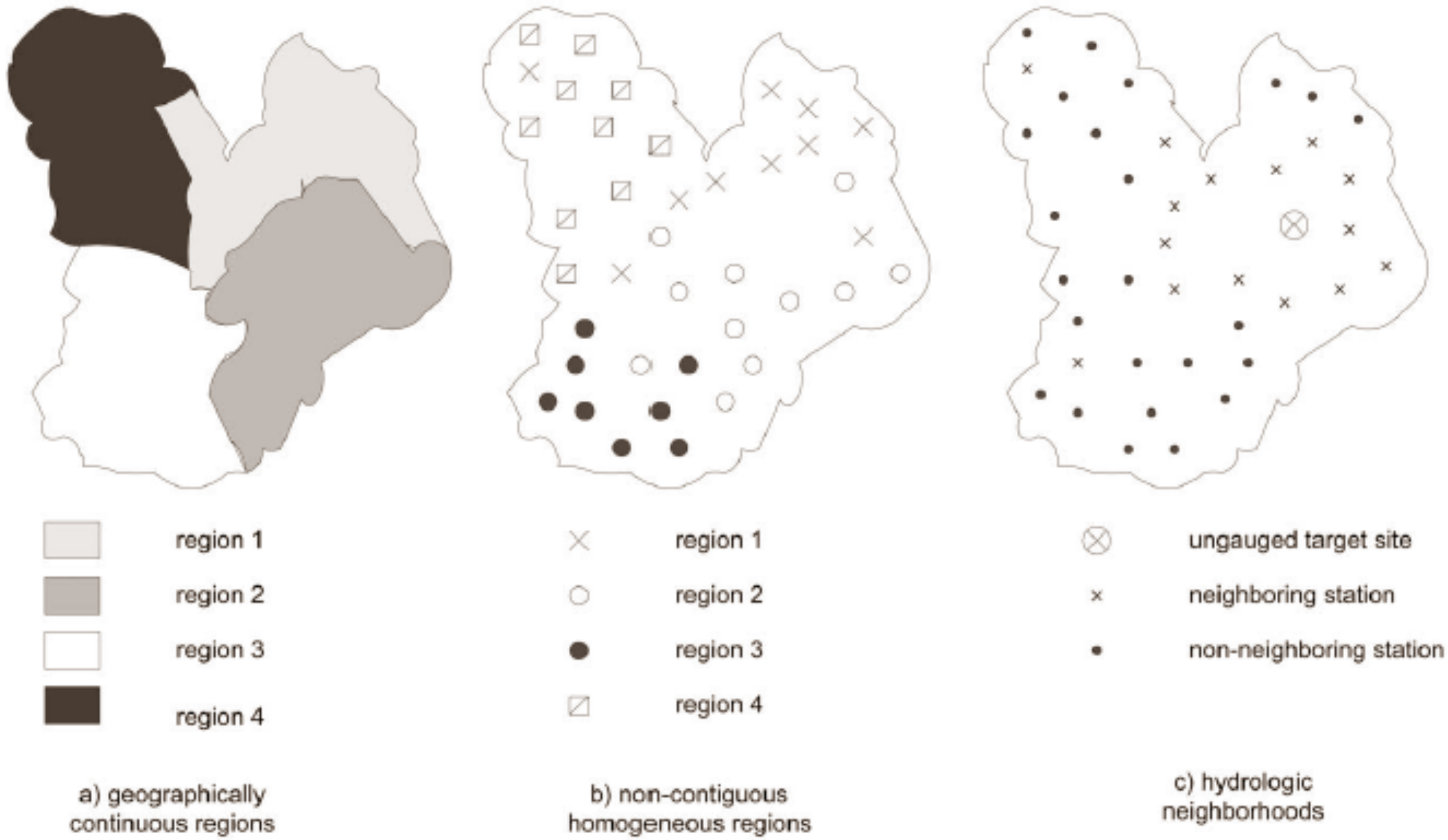
Standard and up-to-date approaches to RFFA are described in detail in the textbook by Hosking and Wallis (1997) and in the Flood Estimation Handbook (FEH, 1999)

Regionalization
(e.g. Hosking & Wallis, 1997)

A. Identification of homogeneous regions
B. Choice of a frequency distribution
C. Estimation of the frequency distribution
D. Validation of the regional model

Introduction: evolution of homogenous regions

Evolution of the concept of hydrologically homogeneous region
(from Ouarda et al., *J. Hydrol.,* 2001)



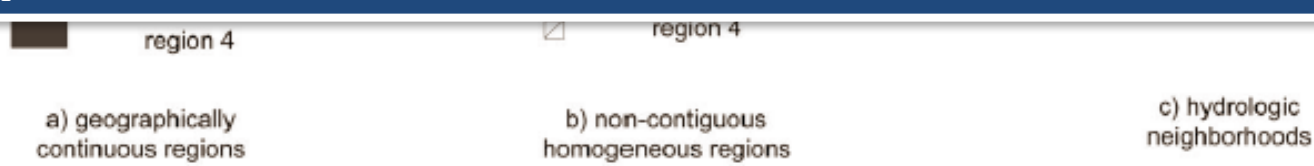
Introduction: evolution of homogenous regions

Evolution of the concept of hydrologically homogeneous region
(from Ouarda et al., *J. Hydrol.,* 2001)

Several studies highlight that geostatistical interpolation can be effectively applied to the problem of regionalization of hydrometric information:

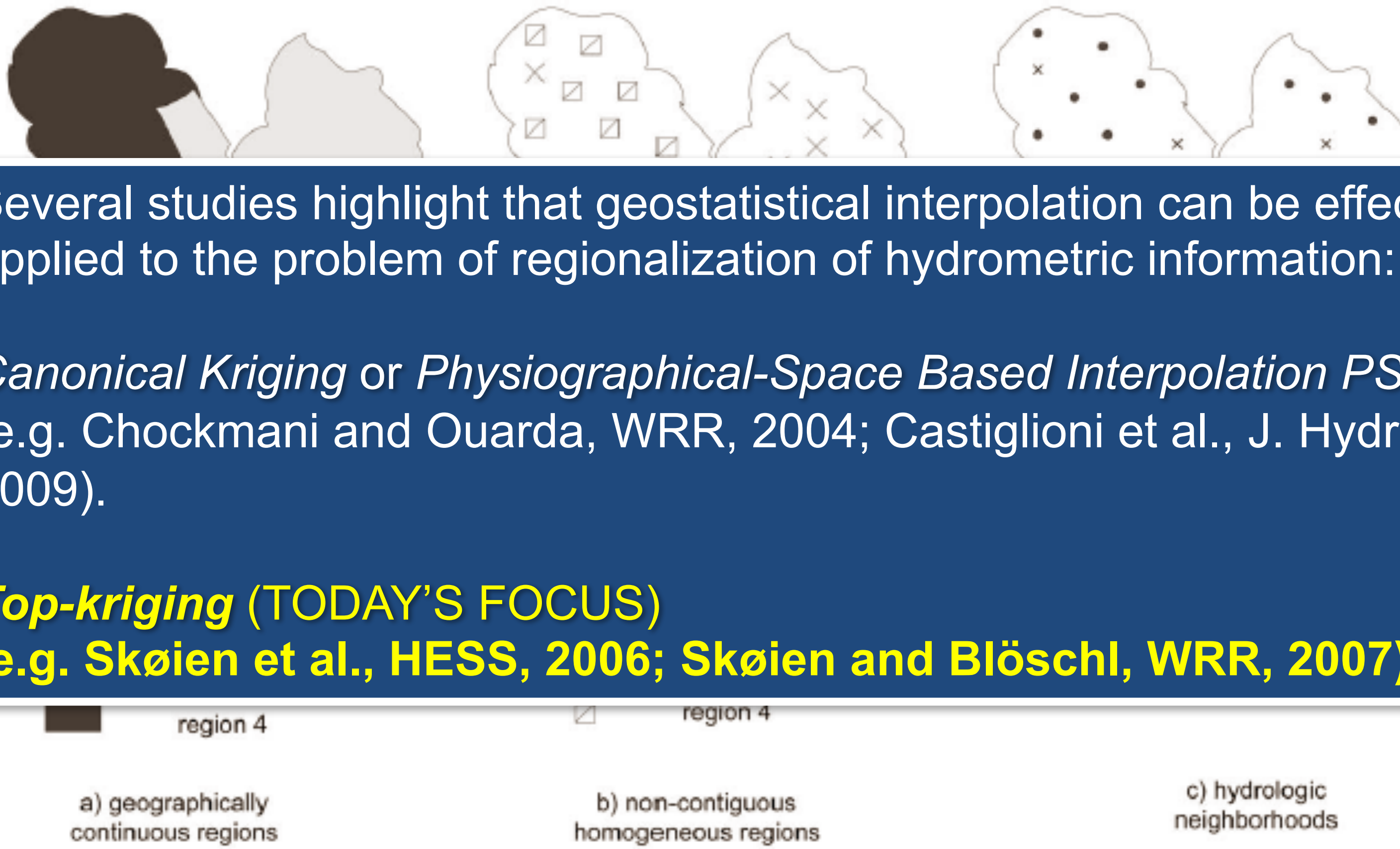
Canonical Kriging or *Physiographical-Space Based Interpolation PSBI*
(e.g. Chockmani and Ouarda, WRR, 2004; Castiglioni et al., J. Hydrol., 2009).

Top-kriging
(e.g. Skøien et al., HESS, 2006; Skøien and Blöschl, WRR, 2007).



Introduction: evolution of homogenous regions

Evolution of the concept of hydrologically homogeneous region
(from Ouarda et al., *J. Hydrol.,* 2001)



Several studies highlight that geostatistical interpolation can be effectively applied to the problem of regionalization of hydrometric information:

Canonical Kriging or *Physiographical-Space Based Interpolation PSBI* (e.g. Chockmani and Ouarda, WRR, 2004; Castiglioni et al., J. Hydrol., 2009).

Top-kriging (TODAY'S FOCUS)
(e.g. Skøien et al., HESS, 2006; Skøien and Blöschl, WRR, 2007).

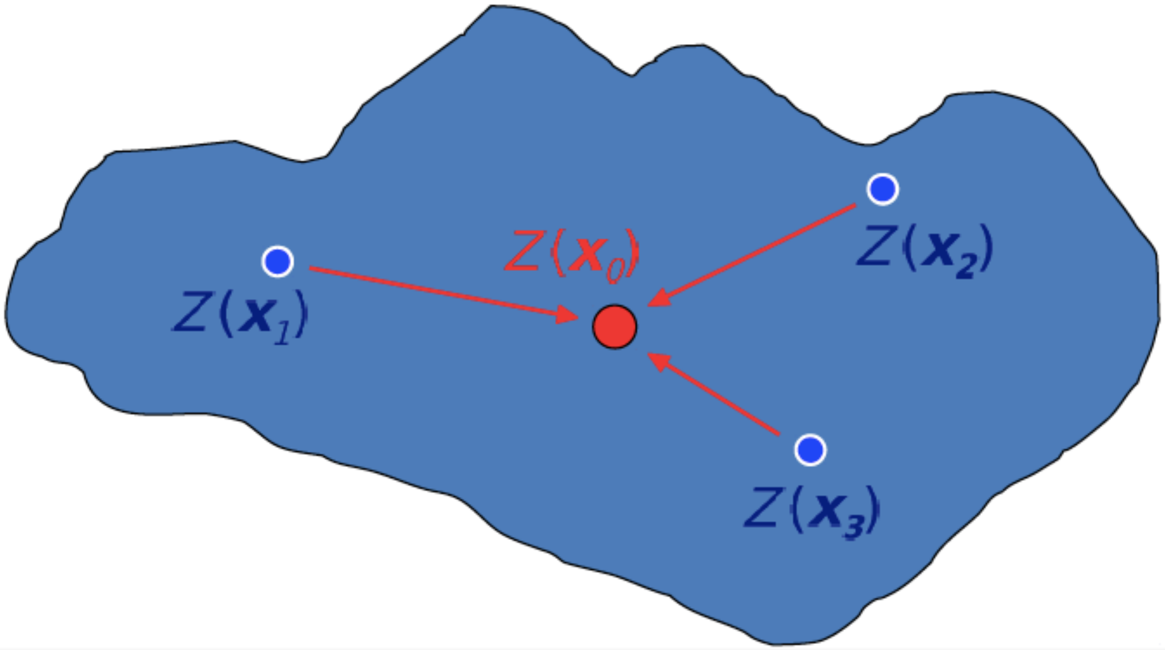
From Kriging to Top-kriging

Kriging as a linear estimator

$$\hat{Z}_0 = f\left(Z_1, Z_2, \ldots, Z_n\right)$$

estimated value for Z_0

function of observed data



From Kriging to Top-kriging

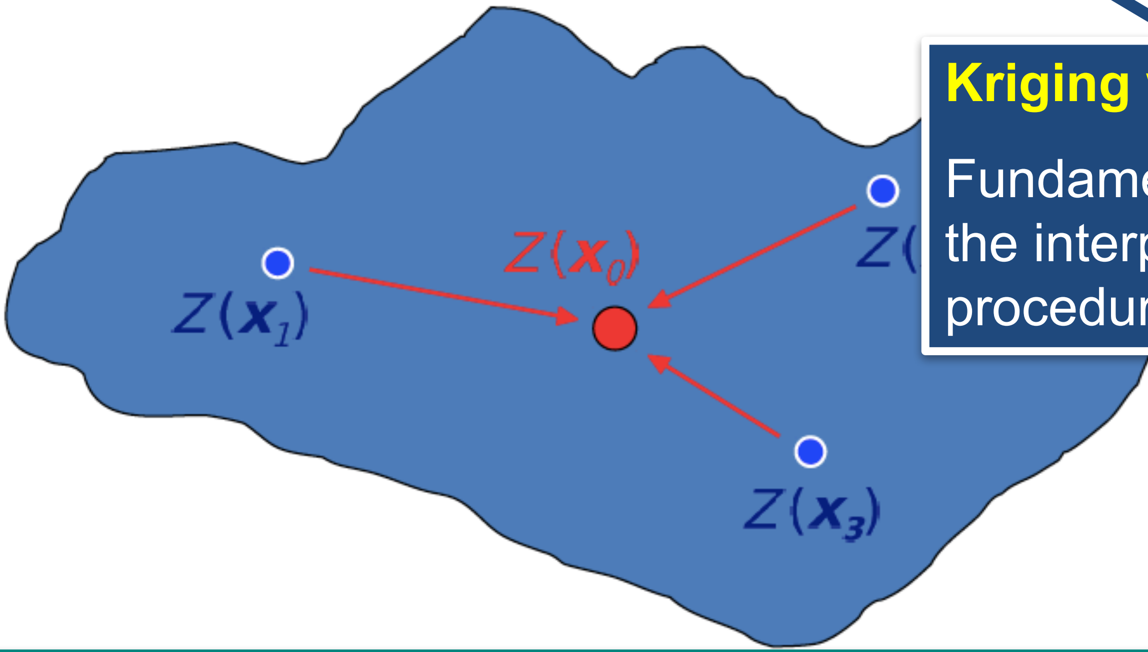
Kriging as a linear estimator

$$\hat{Z}_0 = f\left(Z_1, Z_2, \ldots, Z_n\right)$$

$$\hat{Z}_0 = \lambda_{0,1} \cdot Z_1 + \lambda_{0,2} \cdot Z_2 + \ldots + \lambda_{0,n} \cdot Z_n = \sum_{i=1}^{n} \lambda_{0,i} Z_i$$

Kriging weights

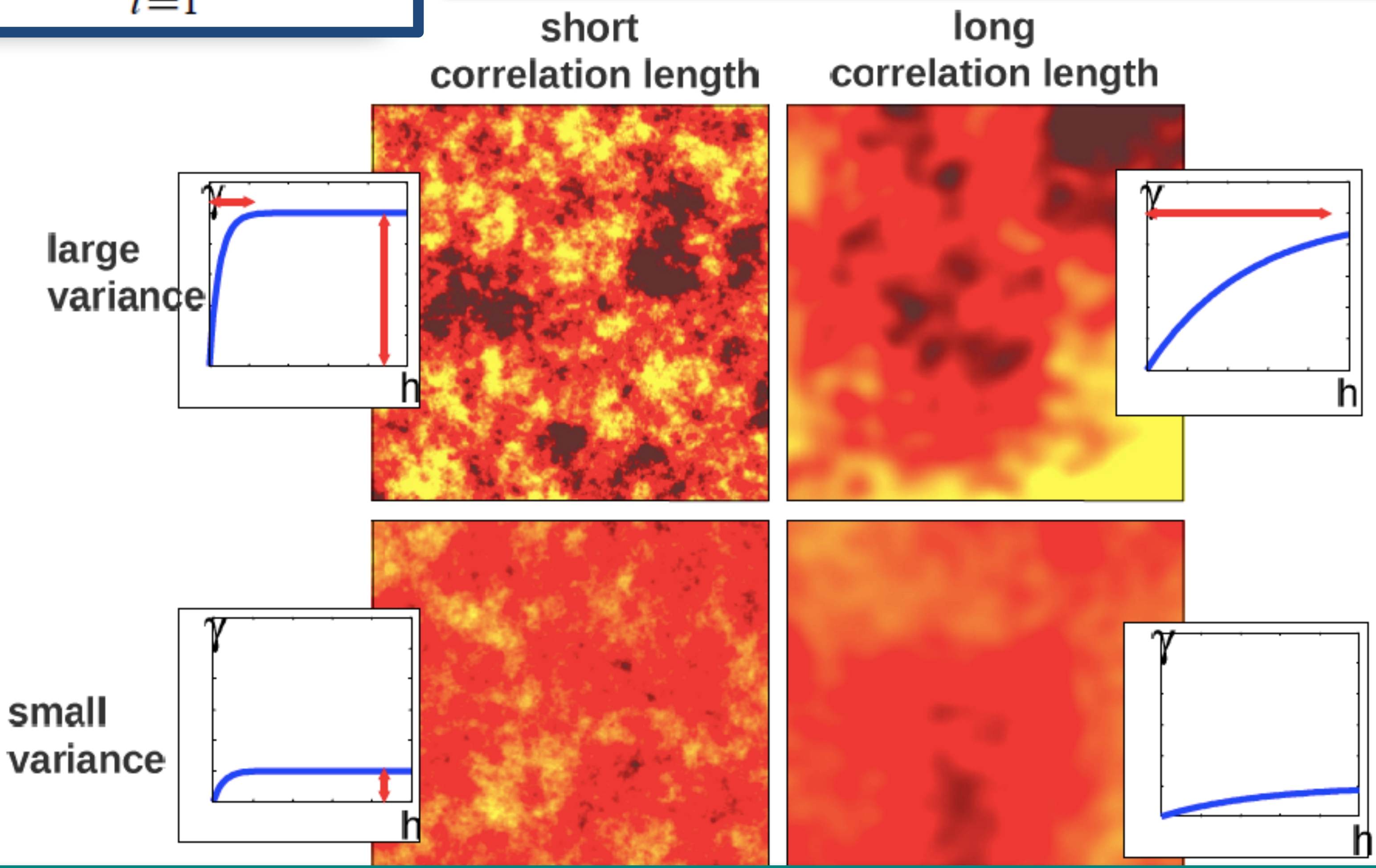
Fundamental role in the interpolation procedure



From Kriging to Top-kriging

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

Weights from the "Variogram" (graphical representation of variance in space)



From Kriging to Top-kriging

Kriging as a BLUe

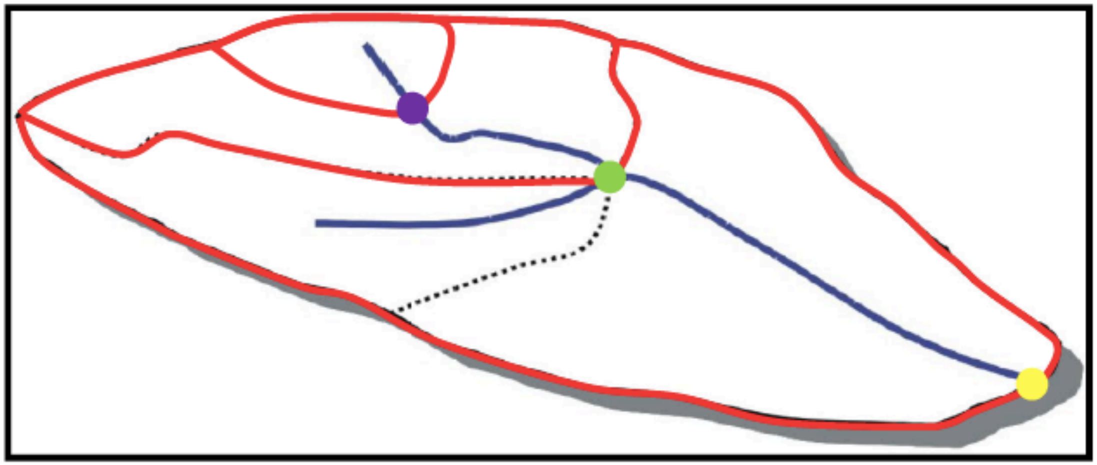
Best $E\left[\left(Z_0 - \hat{Z}_0\right)^2\right] \rightarrow minimize$ — minimum ***square error***

Linear $\hat{Z}_0 = \sum_{i=1}^{n} \lambda_{0,i} \cdot Z_i$ — ***weighted*** average

Unbiased $E\left[Z_0 - \hat{Z}_0\right] = 0$ — no systematic error (***bias***)

From Kriging to Top-kriging

Geostatistics on stream networks: block-kriging using the drainage basin as support area



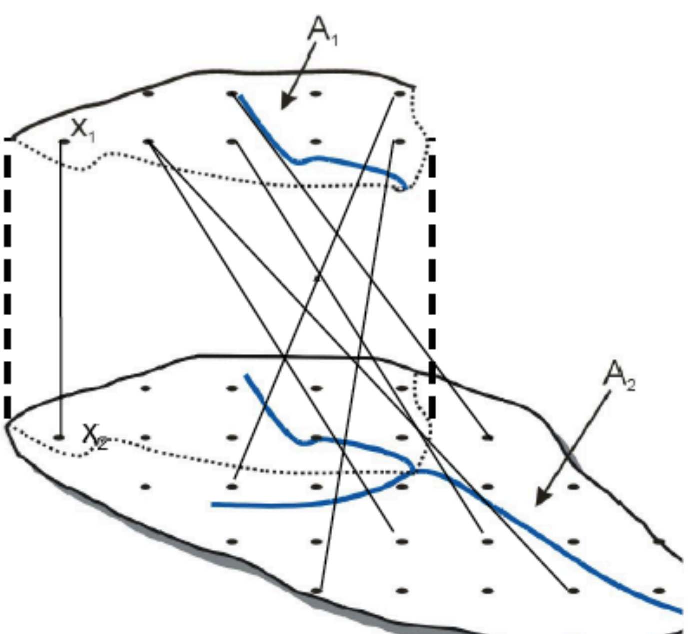
Spatial support variables

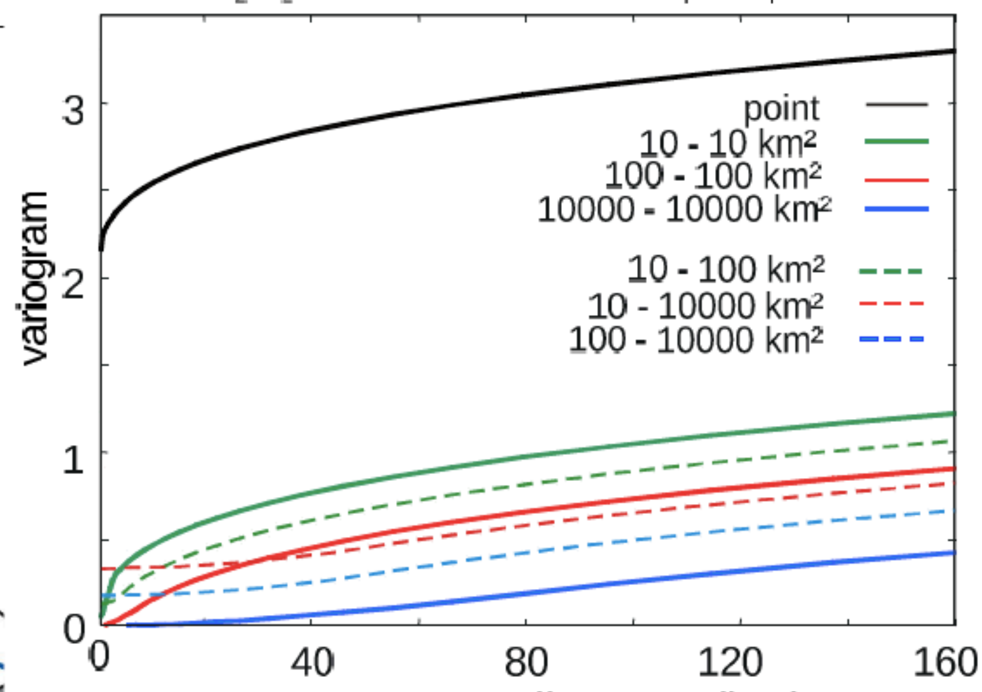
- Standard geostatistics assume random variability
- Topological kriging to account for nested catchment topology

From Kriging to Top-kriging

Geostatistics on stream networks: block-kriging using the drainage basin as support area

$$\gamma_{12}=0.5*Var(z(A_1)-z(A_2))=\frac{1}{A_1 A_2}\int_{A_1}\int_{A_2}\gamma_p(|x_1-x_2|)\,dx_1\,dx_2-$$
$$0.5*\left[\frac{1}{A_1^2}\int_{A_1}\int_{A_1}\gamma_p(|x_1-x_2|)\,dx_1\,dx_2+\frac{1}{A_2^2}\int_{A_2}\int_{A_2}\gamma_p(|x_1-x_2|)\,dx_1\,dx_2\right]$$

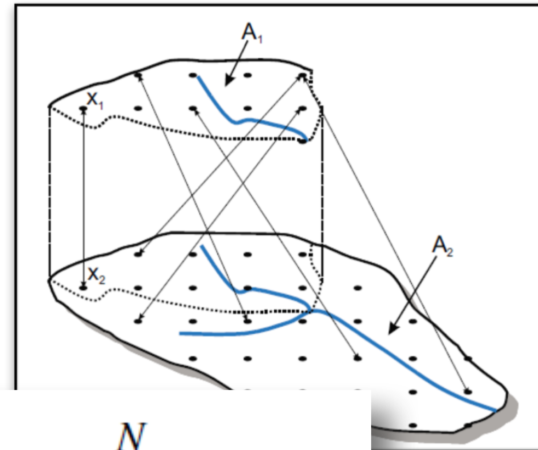




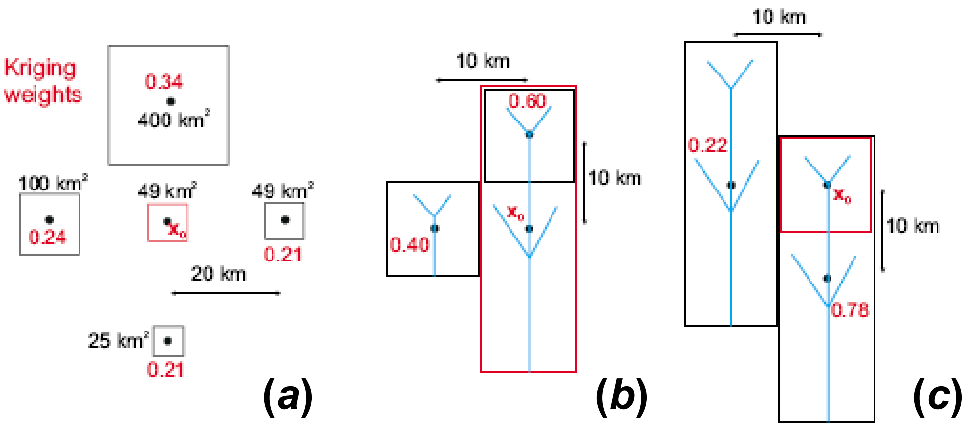
From Jose Salinas' slides: "Prediction of floods in ungauged basins" – PUB Summer school, held in early July in Vienna, Austria

Top-kriging

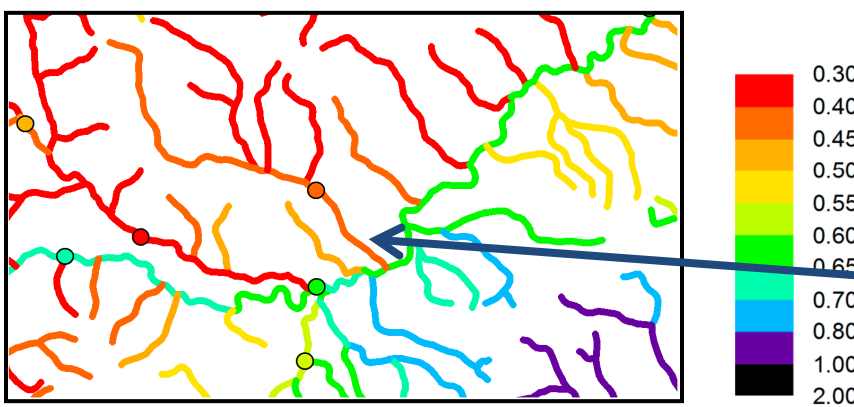
Topological kriging, Top-kriging, predicts the variable of interest along river networks taking both the area and nested nature of catchments into account.



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



Example of catchment size effect (*a*) and the effect of nesting (*b*) and (*c*) in the estimation of λ_i



Applications of Top-kriging:

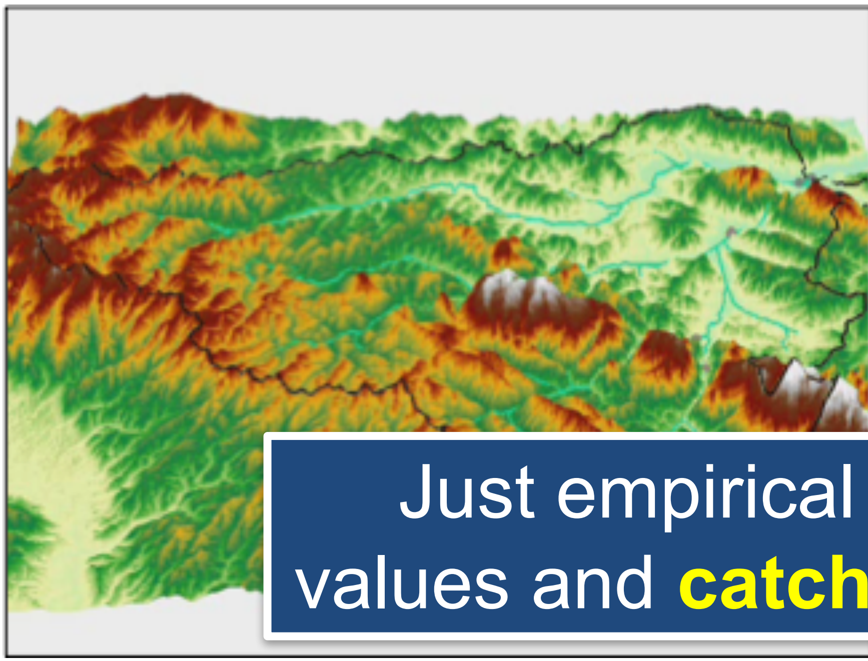
Specific 100-year flood (m³/s/km²)

Example (Skøien et al., HESS, 2006)

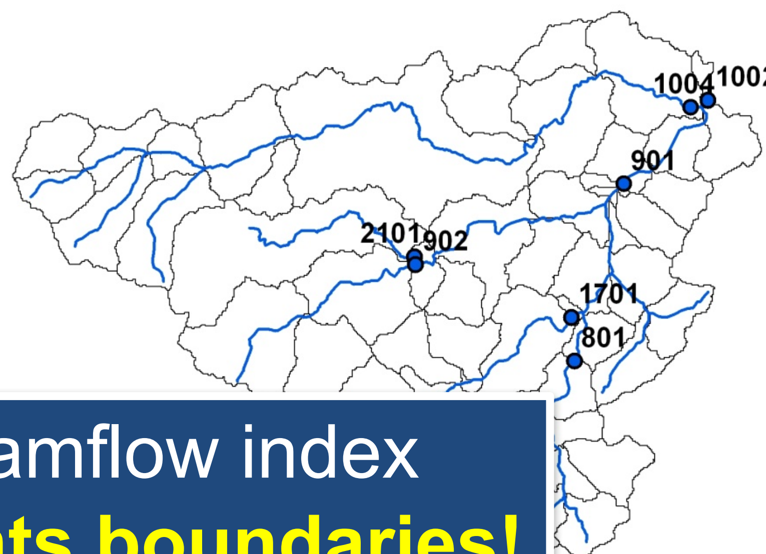
Skøien, J. O., Merz, R., and Blöschl, G. (2006): Top-kriging – geostatistics on stream networks, Hydrol. Earth Syst. Sci., 10, 277–287, doi:10.5194/hess-10-277-2006.

Top-kriging: input data for standard applications

Top-kriging application requires the preliminary identification of the stream network and **watershed delineation** for both **gauged catchments and ungauged catchments** for which the streamflow index is to be predicted



DEM, SRTM 90m Digital Elevation Data (http://csi.cgiar.org/index.asp)



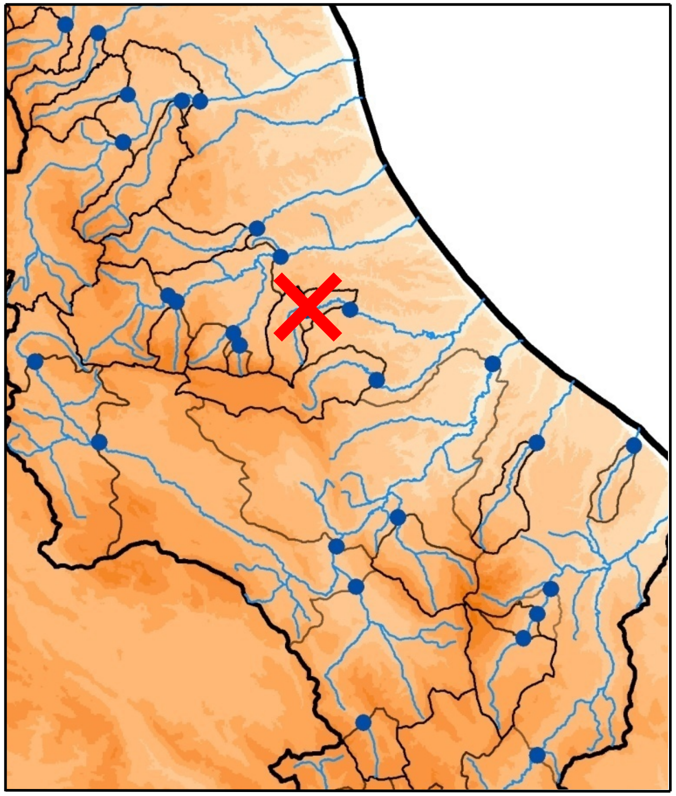
Sub-catchment boundaries (A_{min}>10 Km^2)

Just empirical streamflow index values and **catchments boundaries!**

Leave-One-Out Cross Validation (LOOCV) alternative: Leave-Some-Out Cross Validation

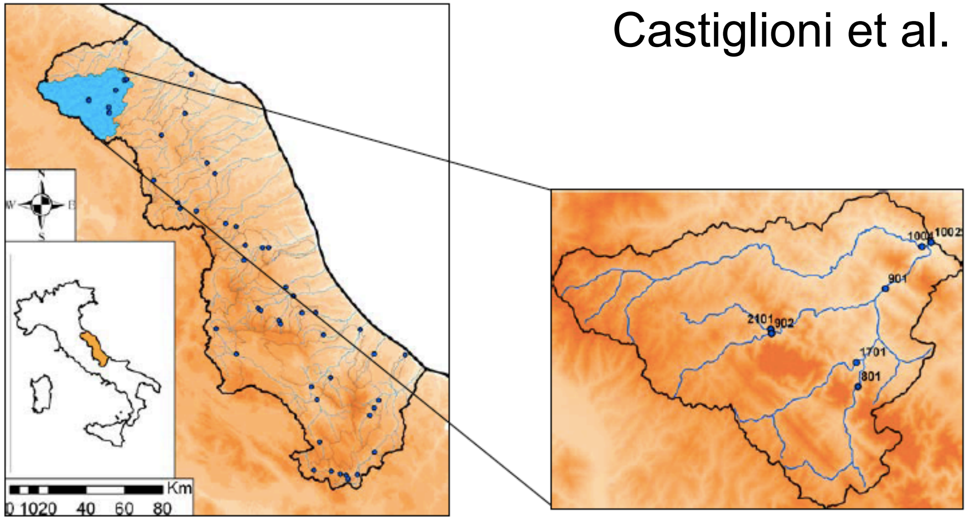
Accuracy of regional predictions in ungauged basins is normally addressed by applying leave-one-out (or leave-some-out) cross-validation procedures (see e.g., Zhang and Kroll, J. Hydrol., 2007):

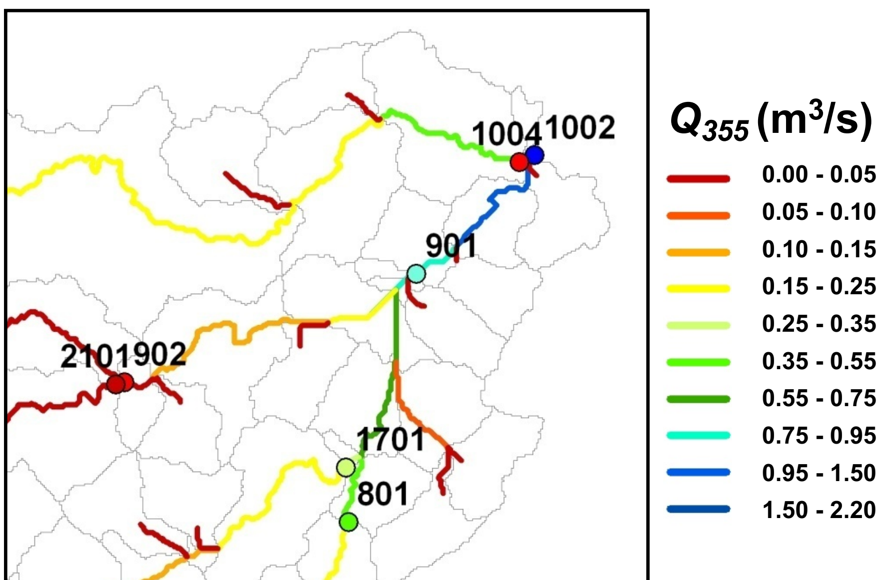
1. assume one of the N basins, let us say ***site i, to be ungauged***;
2. predict the ***Q-index*** value for this site on the basis of the ***remaining N-1 observations***;
3. **repeat this step N times**, considering in turn each of the basins as ungauged, to obtain N cross-validation estimates of ***Q-index*** which can be compared with the corresponding observations.

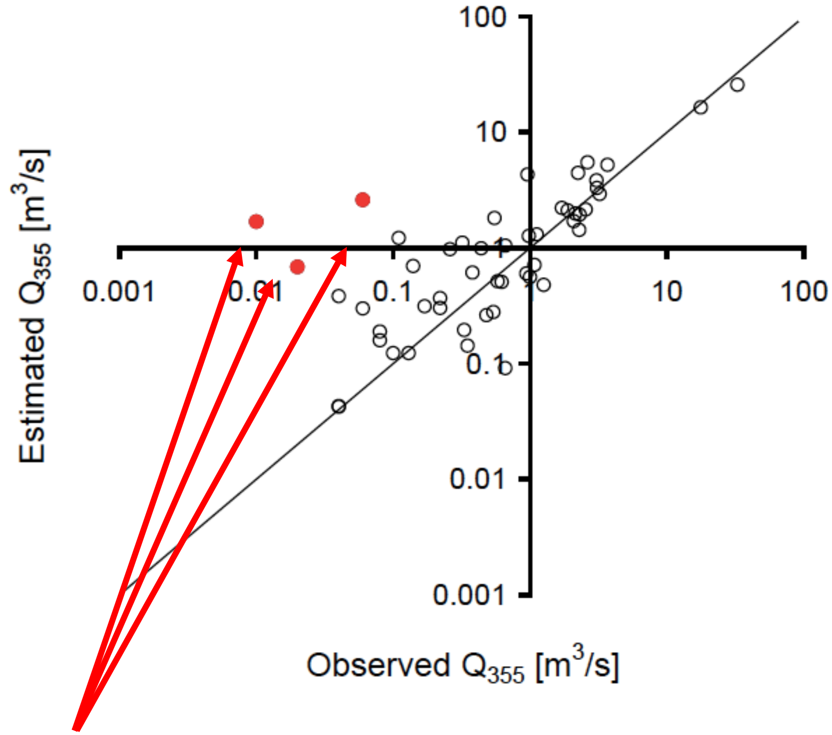


EXAMPLES 1/5: predicting low flows indices

Castiglioni et al. (Hydrol. Earth Syst. Sci., 2011)



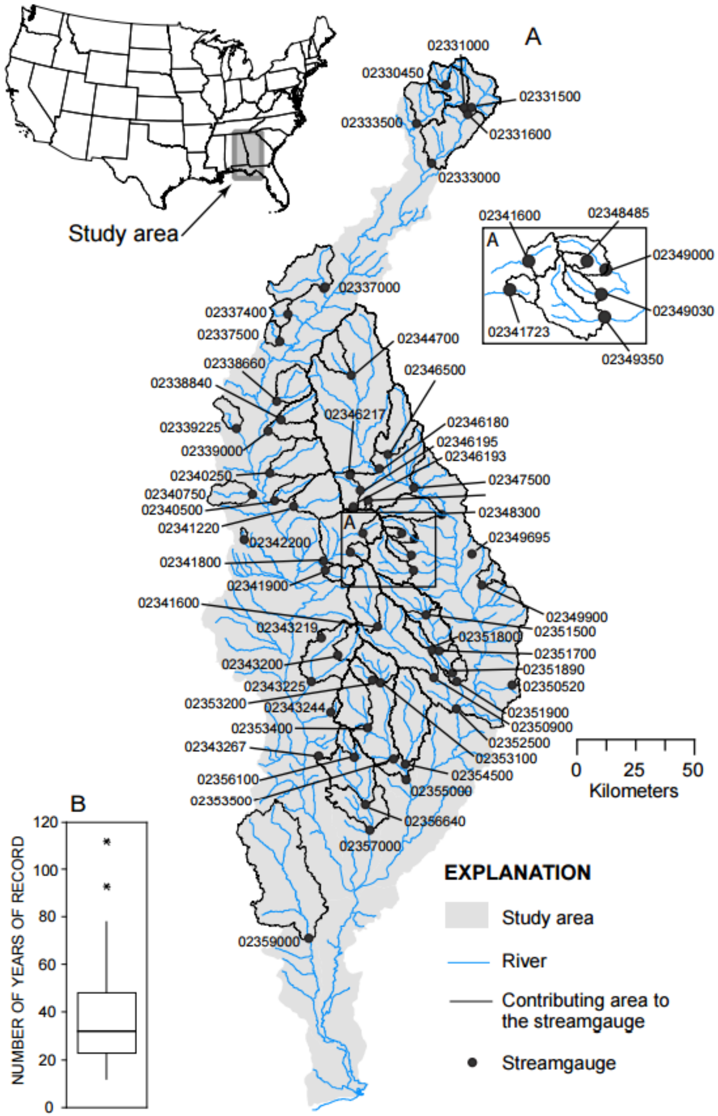




Red-dots: limited accuracy for atypical catchments (low permeability and mean elevations, located in peripheral areas), associated with very low low-flow index

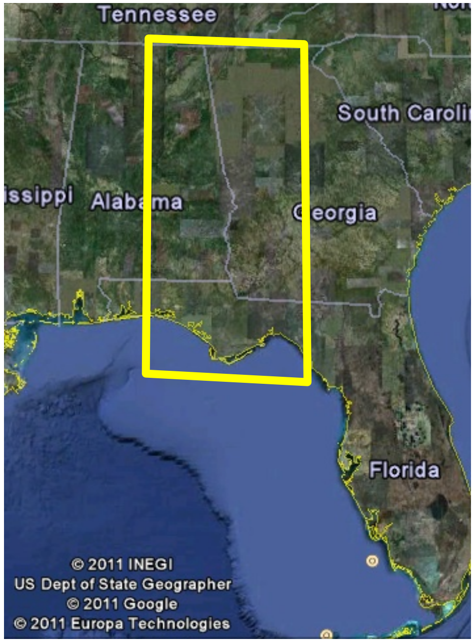
Castiglioni, S., A. Castellarin, A. Montanari, J.O. Skøien, G. Laaha, G. Blöschl (2011): Smooth regional estimation of low-flow indices: physiographical space based interpolation and top-kriging, Hydrol. Earth Syst. Sci., 15, 715-727, 2011.

EXAMPLES 2/5: Predicting flood quantiles



Archfield et al. (HESS, 2013) present a comparison of Top-kriging and PSBI performances, and a possible way to combine the approaches

Study area: 61 gages located across the southeastern US.

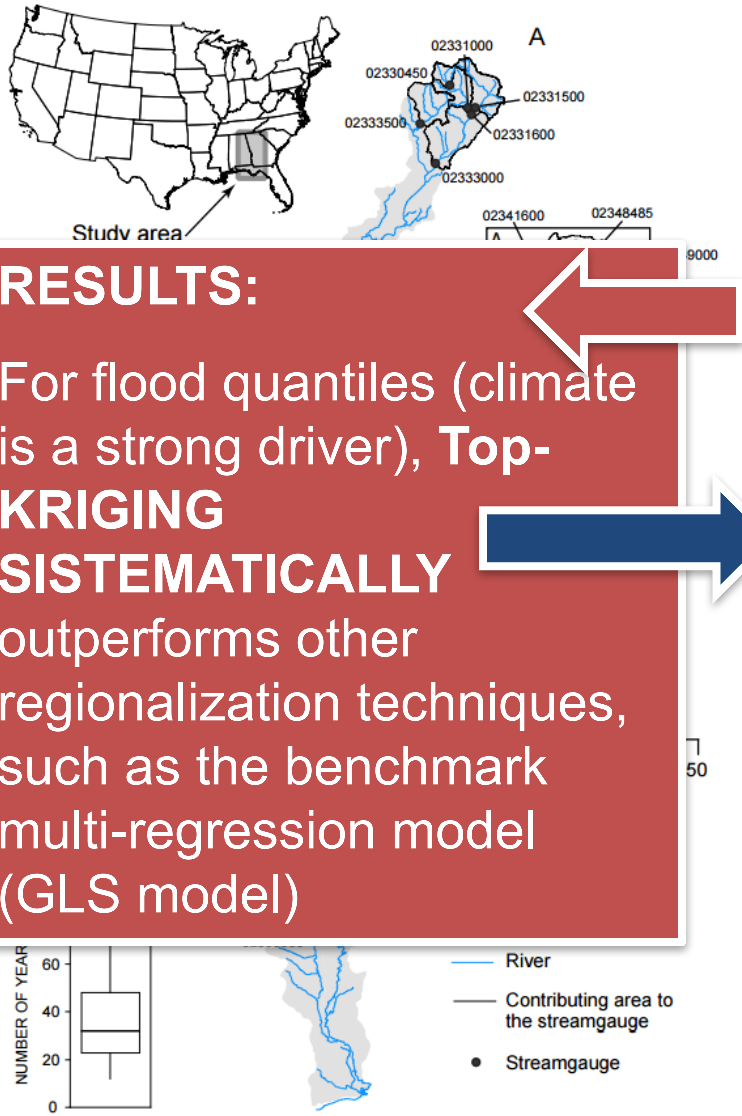


From USGS:

- T-year flood estimates
- Several physiographic and climatic descriptors
- Benchmark prediction procedure (GLS)

Archfield, S. A., Pugliese, A., Castellarin, A., Skøien, J. O., and Kiang, J. E. (2013): Topological and canonical kriging for design flood prediction in ungauged catchments: an improvement over a traditional regional regression approach?, Hydrol. Earth Syst. Sci., 17, 1575-1588

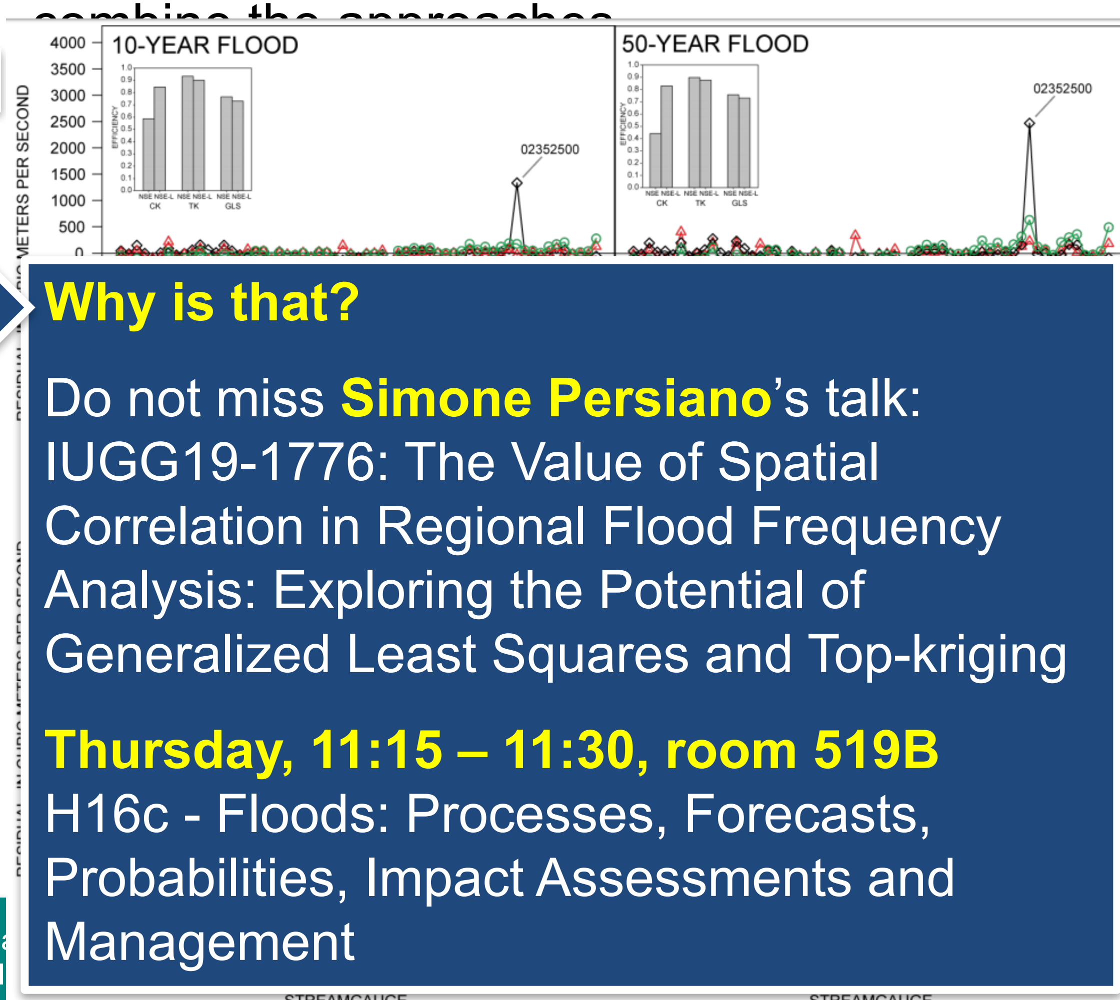
EXAMPLES 2/5: Predicting flood quantiles



RESULTS:

For flood quantiles (climate is a strong driver), **Top-KRIGING SISTEMATICALLY** outperforms other regionalization techniques, such as the benchmark multi-regression model (GLS model)

Archfield et al. (HESS, 2013) present a comparison of Top-kriging and PSBI performances, and a possible way to combine the approaches



Why is that?

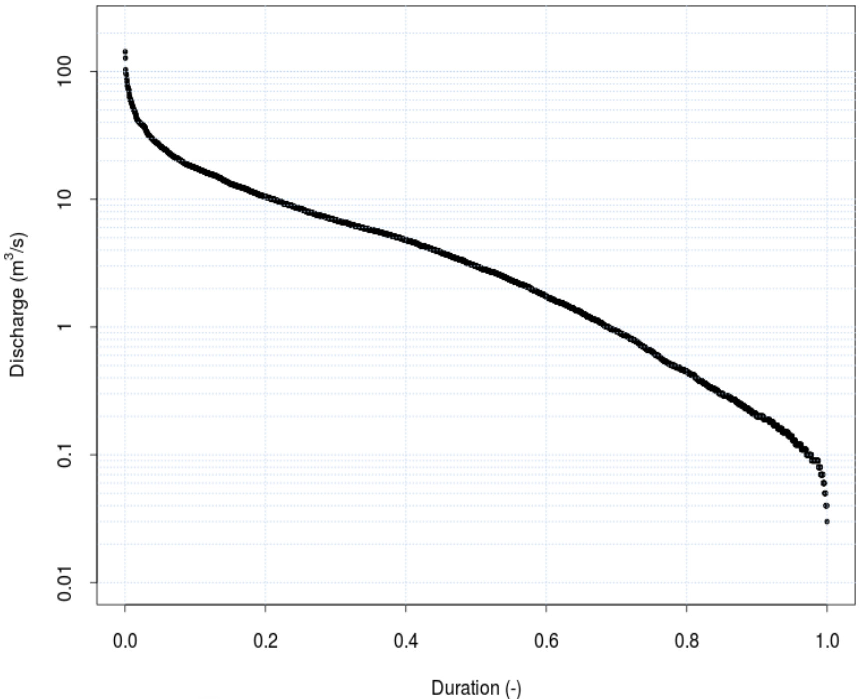
Do not miss **Simone Persiano**'s talk: IUGG19-1776: The Value of Spatial Correlation in Regional Flood Frequency Analysis: Exploring the Potential of Generalized Least Squares and Top-kriging

Thursday, 11:15 – 11:30, room 519B
H16c - Floods: Processes, Forecasts, Probabilities, Impact Assessments and Management

Archfield, S. A., Pugliese, A., Castellarin, A., Skøien, J. O., a
in ungauged catchments: an improvement over a traditional

EXAMPLE 3/5: Predicting flow-duration curves

Flow-Duration Curve (FDC)



Represents the percentage of time (duration) in which the streamflow can be exceeded or equalled

Construction of an empirical period of record FDC

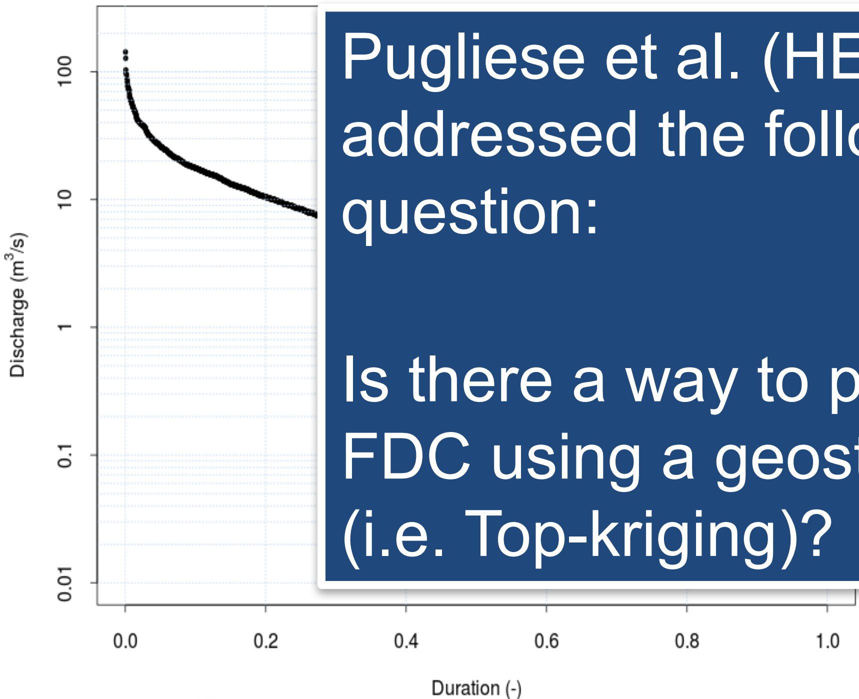
1. Sort all available records in descending order (from high to low flows)
2. Choose an appropriate plotting position (e.g. Weibull)
3. Plot each streamflow value against the corrispondig ranked duration

Provides a comprehensive graphical view of the overall streamflow variability for a specific site of interest.

EXAMPLE 3/5: Predicting flow-duration curves

Flow-Duration Curve (FDC)

Represents the percentage of time (duration) in which the streamflow can be exceeded or equalled



Pugliese et al. (HESS, 2013) addressed the following science question:

Is there a way to predict the entire FDC using a geostatistical technique (i.e. Top-kriging)?

ical period

cords in
rom high to

iate
g. Weibull)

w value
against the correspondig
ranked duration

Provides a comprehensive graphical view of the overall streamflow variability for a specific site of interest.

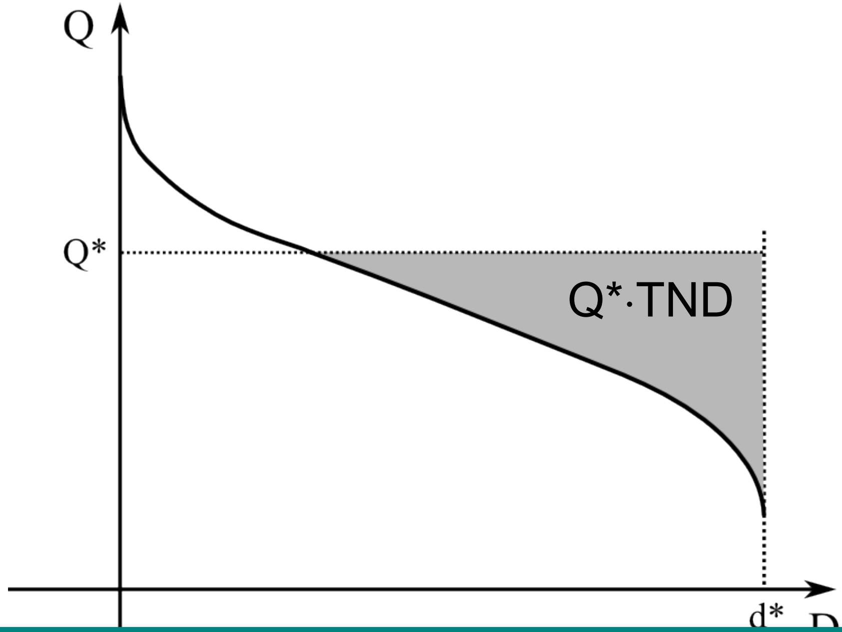
Pugliese, A., Castellarin, A., and Brath, A. (2013): Geostatistical prediction of flow-duration curves, Hydrol. Earth Syst. Sci. Discuss., 10, 13053-13091, doi:10.5194/hessd-10-13053-2013.

EXAMPLE 3/5: Predicting flow-duration curves

$$FDC = \Psi(x, D)$$

→ Multidimensional problem

The idea is to summarize the information from the entire curve through a single index (a metric) which identifies its main characteristics (slope, extremes values, mean value, etc…).



TND: Total Negative Deviation

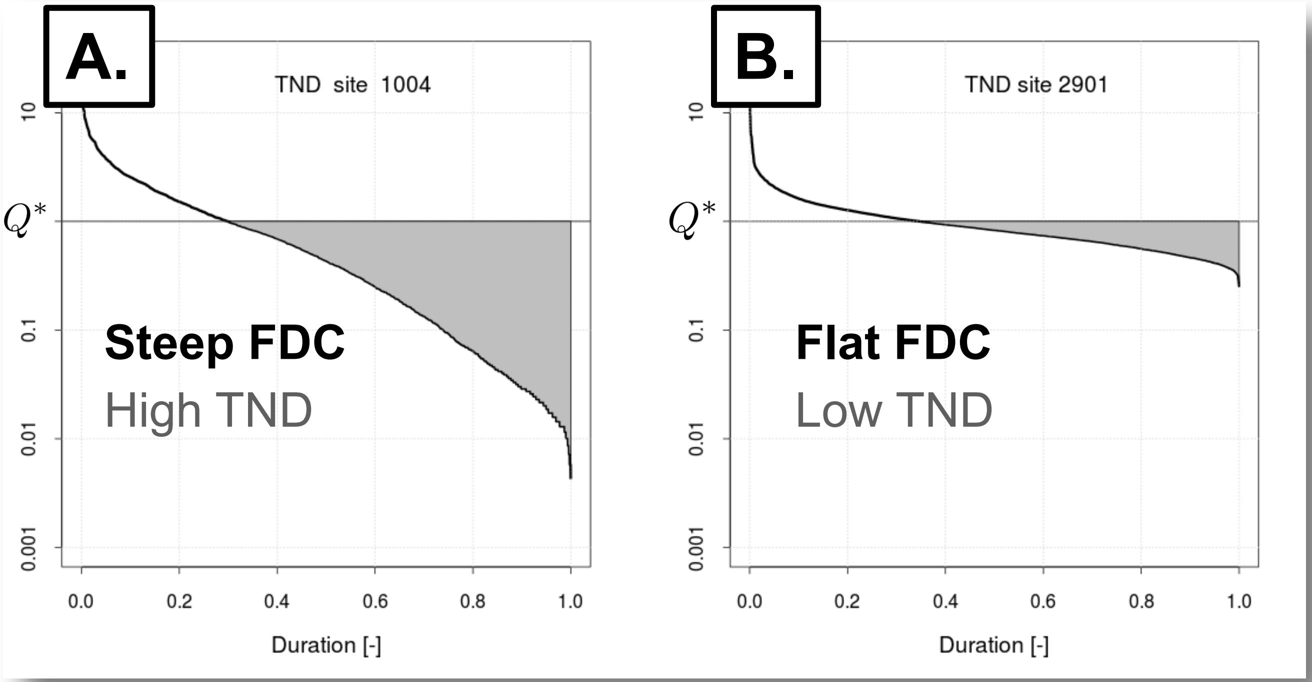
$$TND = \sum_{i=1}^{m} \left| \frac{Q_i - Q^*}{Q^*} \right| \cdot \Delta D_i = \sum_{i=1}^{m} \left| q_i - 1 \right| \cdot \Delta D_i$$

where ***Q**** is a reference streamflow (e.g. Mean Annual Flow).

Pugliese, A., Castellarin, A., and Brath, A. (2013): Geostatistical prediction of flow-duration curves, Hydrol. Earth Syst. Sci. Discuss., 10, 13053-13091, doi:10.5194/hessd-10-13053-2013.

EXAMPLE 3/5: Predicting flow-duration curves

Why TND? 1) objective; 2) easy to compute; 3) meaningful



A. Catchment with rapidly responding runoff

B. Catchment with higher storage capacity

Pugliese, A., Castellarin, A., and Brath, A. (2013): Geostatistical prediction of flow-duration curves, Hydrol. Earth Syst. Sci. Discuss., 10, 13053-13091, doi:10.5194/hessd-10-13053-2013.

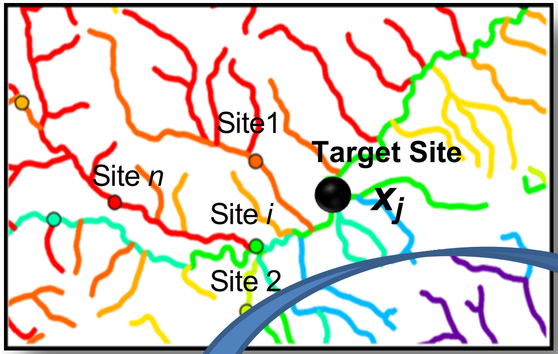
EXAMPLE 3/5: Predicting flow-duration curves

Using ***Q****** as the standardization value, we can combine a traditional **"index-flow" regional method**, with a **geostatistical weighting scheme**:

1. Compute empirical TND values for all N_{staz} gauged sites

$$TND_k = \sum_{i=1}^{m} |q_{i,k} - 1|\Delta_{i,k} \qquad k = 1, \ldots, N_{\text{staz}}$$

2. Predict TND in the ungauged site x_j by using Top-kriging (TK).



3. Obtain a weighting array

$$\{\lambda_1, \ldots, \lambda_n\}$$

GEOSTATISITCAL WEIGHTING SCHEME

$$\hat{\psi}(x_j, D) = \sum_{i=1}^{n} \lambda_i \psi(x_i, D)$$

INDEX-FLOW

Reliable model for predicting ***Q****

$$\Psi(x_j, D) = \hat{Q}_j^* \hat{\psi}(x_j, D)$$

Pugliese, A., Castellarin, A., and Brath, A. (2013): Geostatistical prediction of flow-duration curves, Hydrol. Earth Syst. Sci. Discuss., 10, 13053-13091, doi:10.5194/hessd-10-13053-2013.

EXAMPLE 3/5: Predicting flow-duration curves

Extensive application: USGS-SEMC Dataset

Dataset

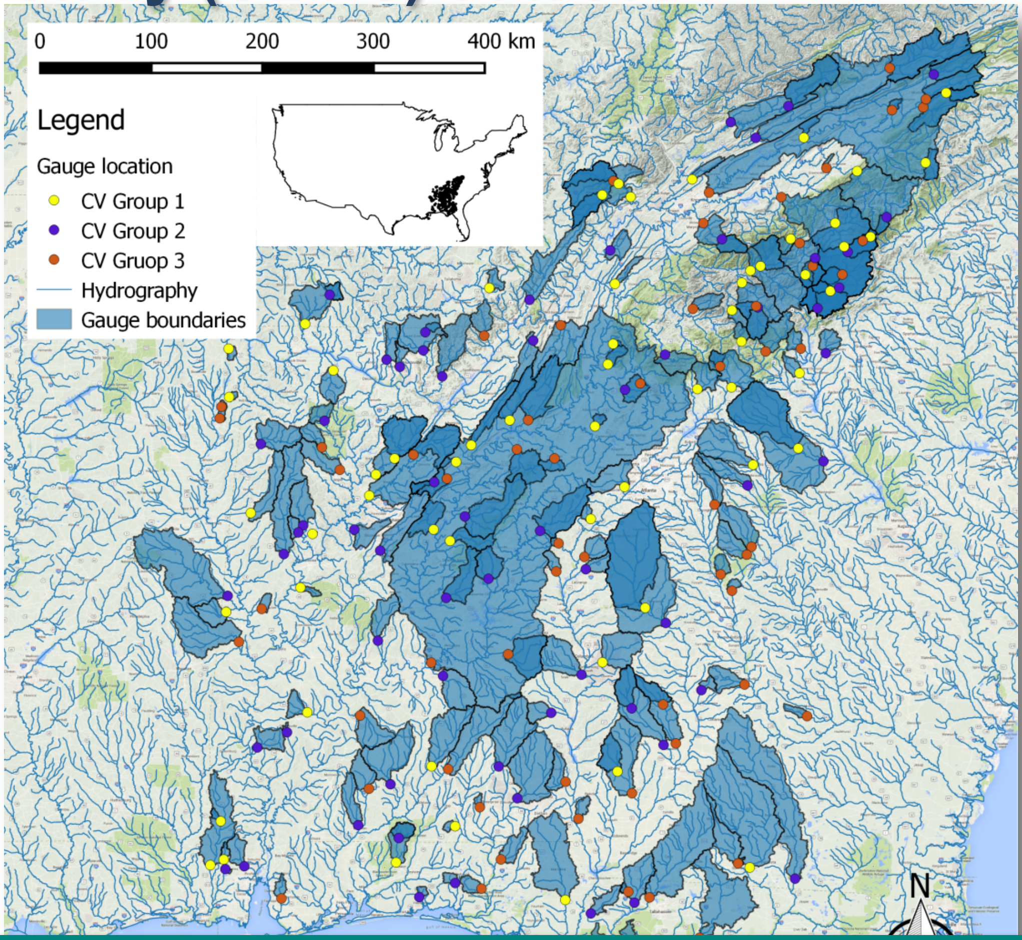
- 182 river catchments
- 30 yrs. observed streamflows
- Several geomorphological and climatic catchment descriptors
- USGS-SEMC FDCs prediction for benchmarking (quantile regression method, 27 quantiles resampling, 3-fold cross-validation **3FCV**)

Information used by TNDTK

- Watershed boundaries (SHP files)
- Streamflow data

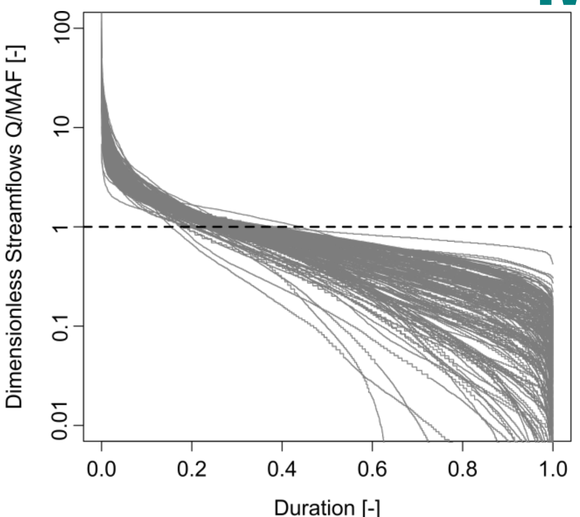
(a **3FCV** may have a significant impact on a geostatistical procedure…)

Southeast Model-Comparison study (SEMC) dataset

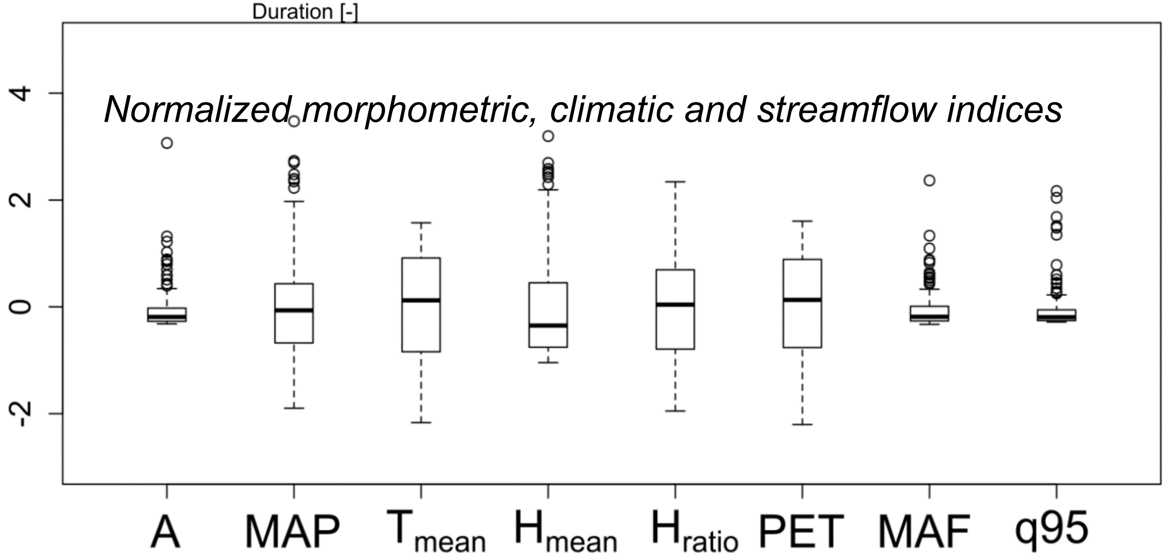


Pugliese, A., W.H. Farmer, A. Castellarin, S.A. Archfield, R.M. Vogel (2016): Regional flow duration curves: geostatistical techniques versus multivariate regression, Advances in Water Resources, Advances in Water Resources, 96, 11-22, doi: 10.1016/j.advwatres.2016.06.008.

USGS-SEMC Dataset: Main characteristics



	A (km^2)	MAP (mm)	T_{mean} (°C)	H_{mean} (m)	H_{ratio} (-)	PET (mm/d)	MAF (m^3/s)	q95 (m^3/s)
Min.	15.3	1146	9.5	17.7	0.2	576.5	0.3	0
1st Qu.	225.9	1358	13.1	116.0	0.3	753.5	3.3	0.05
Median	588.4	1462	15.7	252.1	0.4	862.2	6.9	0.12
Mean	1427.2	1473	15.4	370.8	0.4	846.1	15.4	0.13
3rd Qu.	1317.2	1547	17.8	520.7	0.5	954.7	15.9	0.21
Max.	56609.5	2072	19.6	1452.4	0.6	1042.3	598.1	0.59



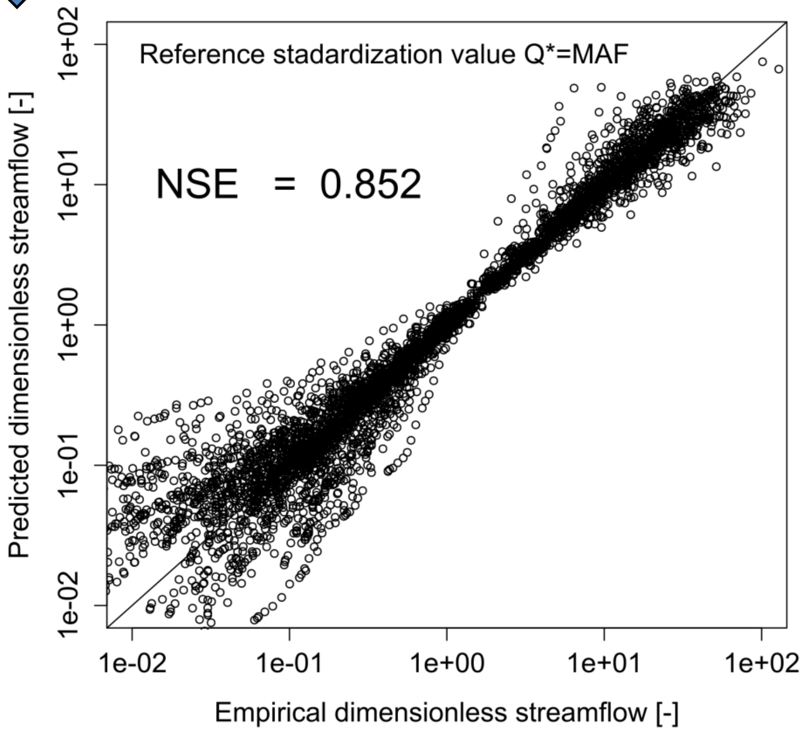
A:	Drainage Area
MAP:	Mean Annual Precipitation
T_{mean}:	Mean Temperature
H_{mean}:	Mean Elevation
H_{ratio}:	Dimensionless elevation relief ratio (difference between mean and minimum elevation divided by the full range of elevations)
PET:	Potential Evapotranspiration
MAF:	Mean Annual Flow
q95:	daily-streamflow exceeded 95% of the time divided by MAF

Pugliese, A., W.H. Farmer, A. Castellarin, S.A. Archfield, R.M. Vogel (2016): Regional flow duration curves: geostatistical techniques versus multivariate regression, Advances in Water Resources, Advances in Water Resources, 96, 11-22, doi: 10.1016/j.advwatres.2016.06.008.

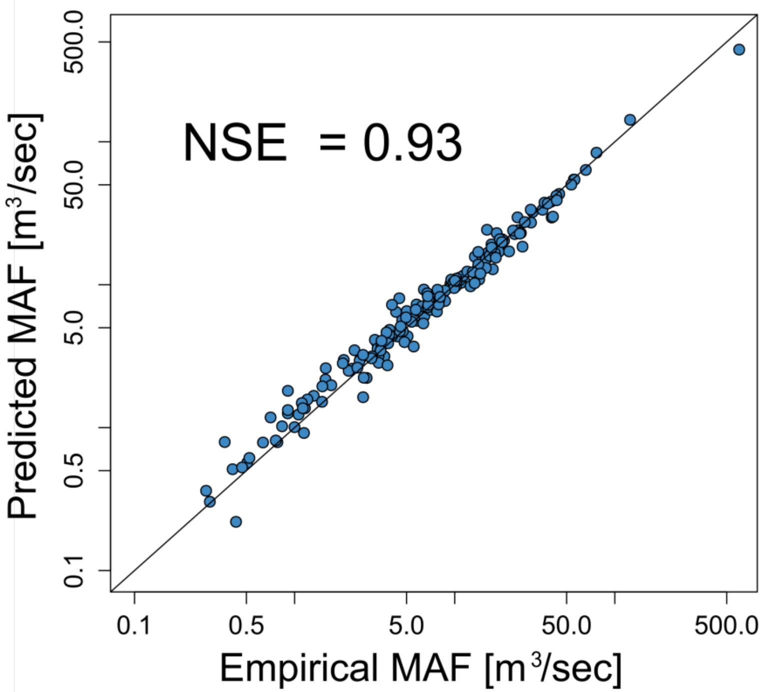
Results: prediction of standardized FDCs and MAF

Settings: (*i*) *Q** = MAF; (*ii*) 3-fold cross-validation (3FCV); **(*iii*) 6 donor sites**

(1) Prediction of dimensionless FDCs



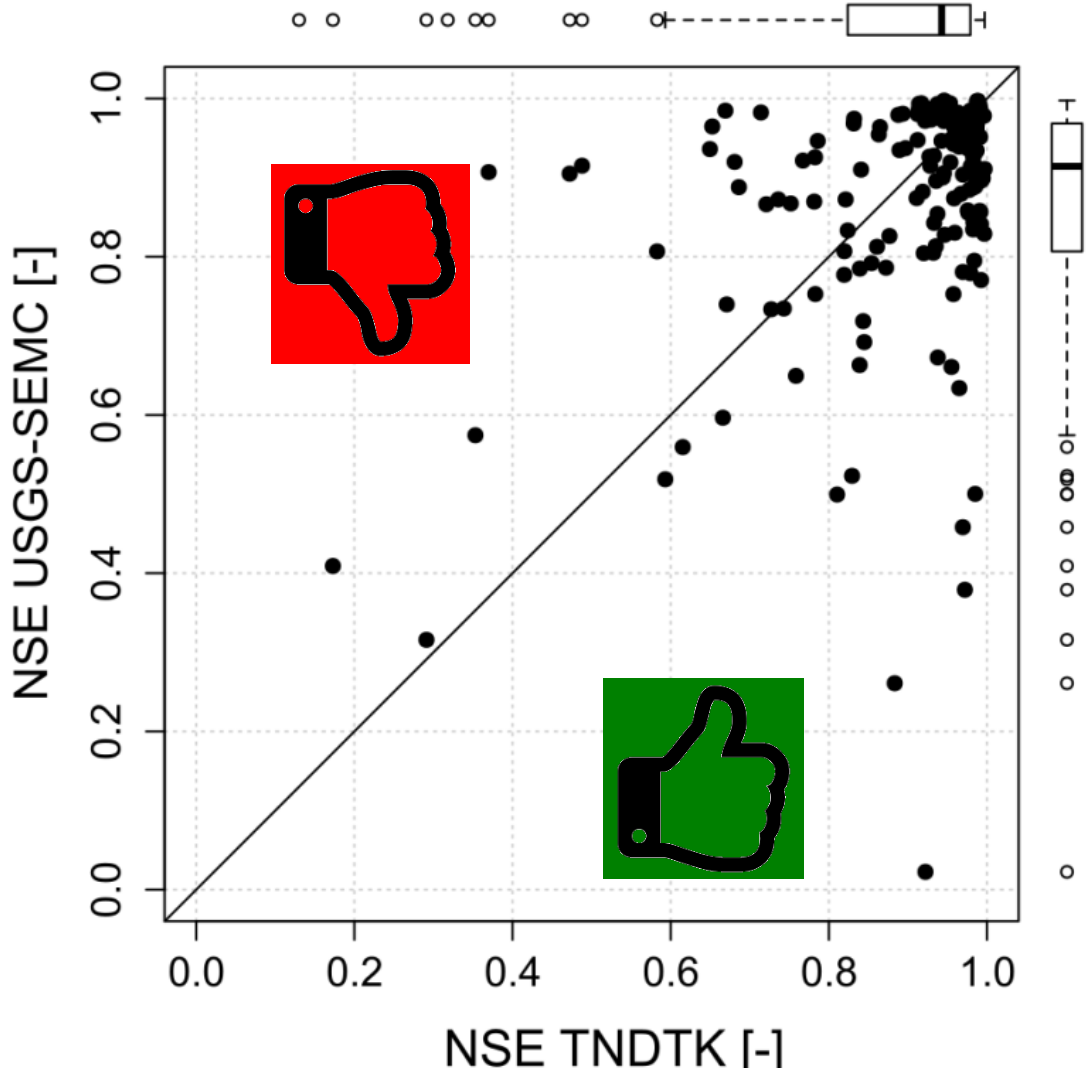
(2) Prediction of index-flow value (Q*=MAF)



Pugliese, A., W.H. Farmer, A. Castellarin, S.A. Archfield, R.M. Vogel (2016): Regional flow duration curves: geostatistical techniques versus multivariate regression, Advances in Water Resources, Advances in Water Resources, 96, 11-22, doi: 10.1016/j.advwatres.2016.06.008.

Results: comparison between Top-kriging and USGS-SEMC (site-wise for *n*=6)

Top-kriging outperforms USGS-SEMC for **102/171** (~60%) sites



Pugliese, A., W.H. Farmer, A. Castellarin, S.A. Archfield, R.M. Vogel (2016): Regional flow duration curves: geostatistical techniques versus multivariate regression, Advances in Water Resources, Advances in Water Resources, 96, 11-22, doi: 10.1016/j.advwatres.2016.06.008.

HYDROLOGICAL SCIENCES JOURNAL
2018, VOL. 63, NO. 6, 845–861
https://doi.org/10.1080/02626667.2018.1445855

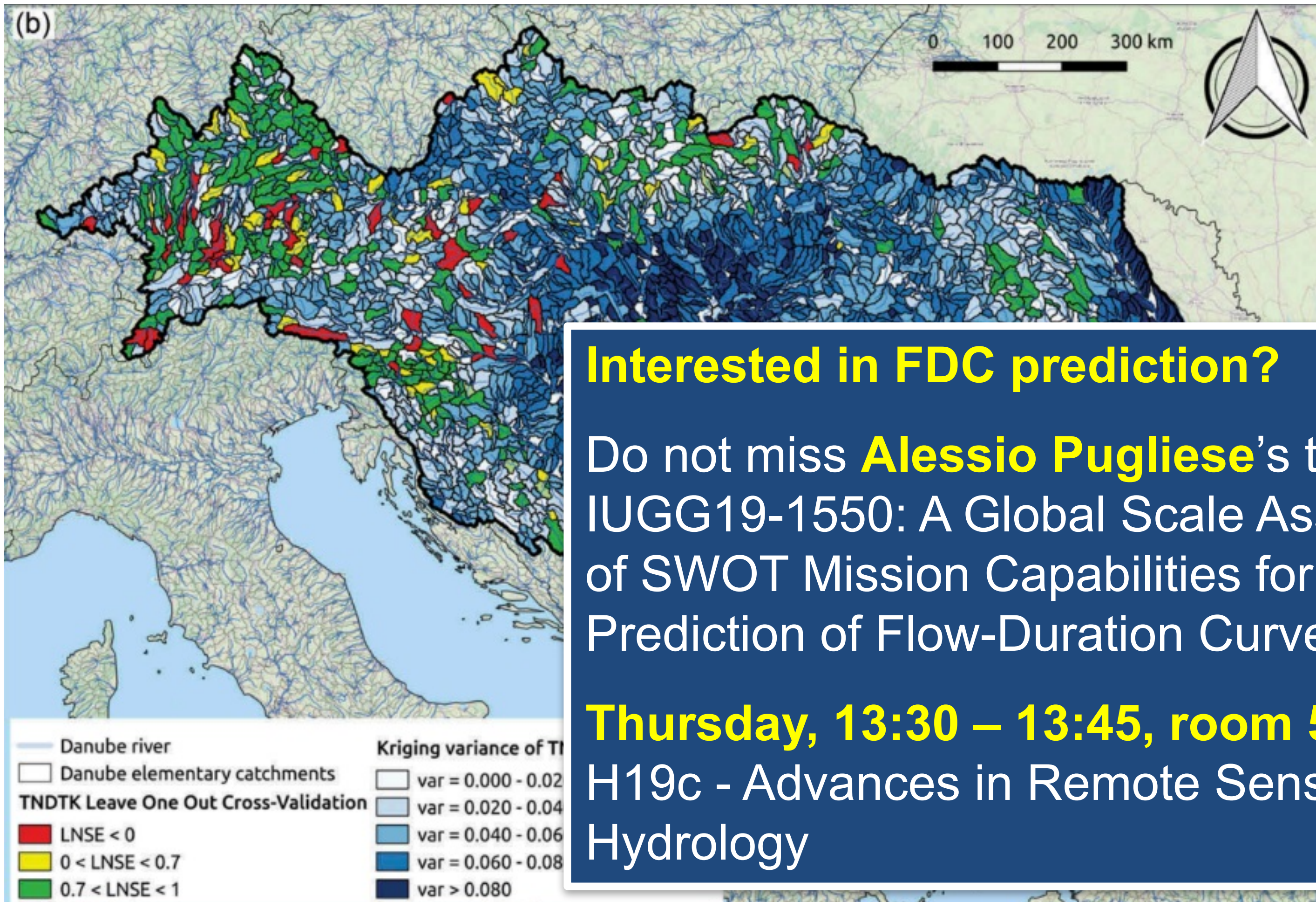
Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Prediction of streamflow regimes over large geographical areas: interpolated flow–duration curves for the Danube region

A. Castellarin [a], S. Persiano [a], A. Pugliese [a], A. Aloe[b], J. O. Skøien [b] and A. Pistocchi [b]

[a]Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), School of Civil Engineering, University of Bologna, Bologna, Italy; [b]European Commission, DG Joint Research Centre (JRC), Ispra, Italy



ARTICLE HISTORY
Received 5 May 2017
Accepted 5 February 2018

EDITOR
D. Koutsoyiannis

ASSOCIATE EDITOR
not assigned

KEYWORDS
flow–duration curve; total negative deviation top-kriging; prediction in

Interested in FDC prediction?

Do not miss **Alessio Pugliese**'s talk:
IUGG19-1550: A Global Scale Assessment of SWOT Mission Capabilities for the Prediction of Flow-Duration Curves

Thursday, 13:30 – 13:45, room 516D
H19c - Advances in Remote Sensing Hydrology

EXAMPLE 4/5: predicting daily streamflow series

Parajka et al. (J. Hydrol. Hydromech., 2015) investigated the role of station density for predicting daily streamflow series by top-kriging interpolation in Austria (more than 550 stations; i.e. >6.5 stations/1000km^2)

Results of the cross-validation:

In Austria, top-kriging interpolation is superior to hydrological model regionalisation if station density exceeds approximately 2 stations per 1000 km^2 (175 stations in Austria). The average median of Nash-Sutcliffe cross-validation efficiency is larger than 0.7 for densities above 2.4 stations/1000 km^2. For such densities, the variability of runoff efficiency is very small over ten random samples. Lower runoff efficiency is found for low station densities (less than 1 station/1000 km^2) and in some smaller headwater basins.

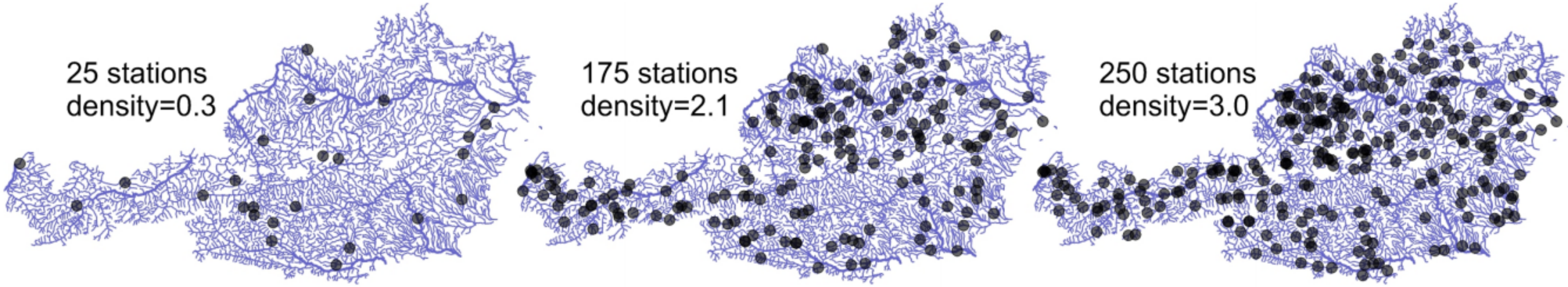


Fig. 2. Example of 3 different station densities (25, 175 and 250 stations).

Parajka, J., Merz, R., Skøien, J.O., Viglione, A. (2015). The role of station density for predicting daily runoff by top-kriging interpolation in Austria. J. Hydrol. Hydromech. 63. doi:10.1515/johh-2015-0024

EXAMPLE: interpolation of stream-temperature

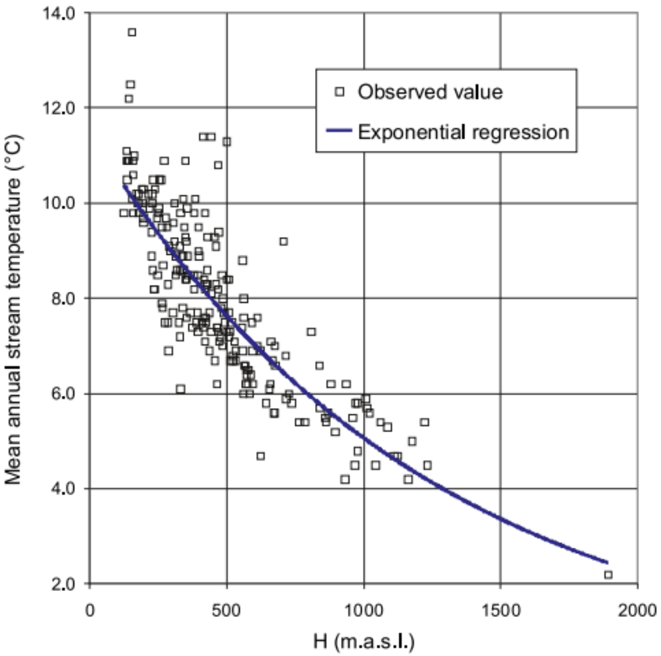
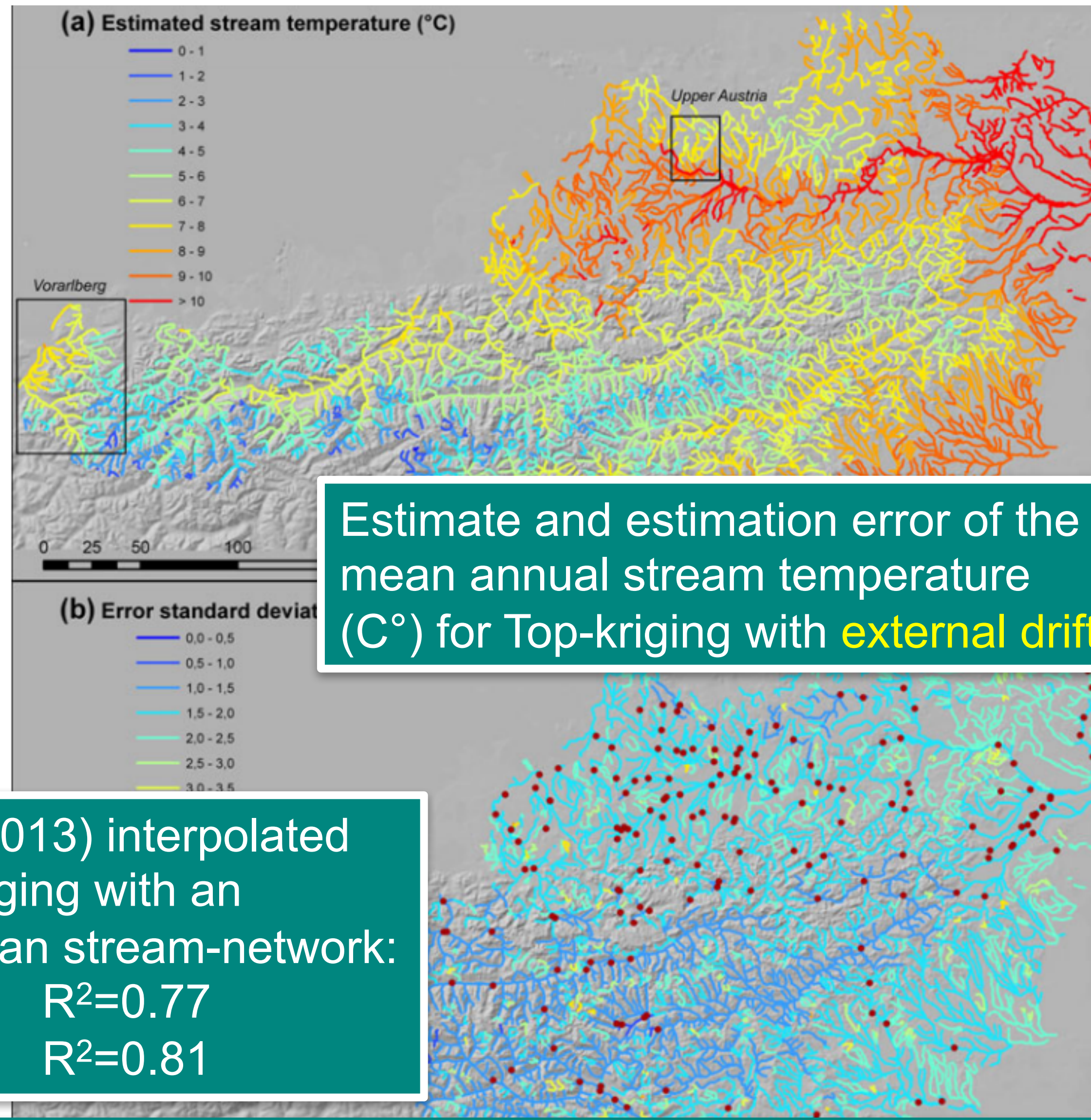


Fig. 4 Exponential regression model between the mean annual stream temperature (in degree Celsius) and altitude of the gauge (in metres above sea level) used as external drift function, fitted to 214 catchments



Estimate and estimation error of the mean annual stream temperature (C°) for Top-kriging with external drift

Laaha et al. (Env. Mod. Assess., 2013) interpolated stream temperatures using top-kriging with an external drift over the whole Austrian stream-network:

Exponential regression:	R^2=0.77
Top-kriging with external drift:	R^2=0.81

Laaha, Gregor; Skoien, Jon O.; Nobilis, Franz; et al. (2013) Spatial Prediction of Stream Temperatures Using Top-Kriging with an External Drift Environmental Modeling & Assessment, 18(6), 671-683

Top-kriging applications (incomplete list)

Literature reports successful applications for predicting:

- ✓ Flood quantiles (e.g. Skøien et al., HESS, 2006; Archfield et al., HESS, 2013)
- ✓ Low-flows (Castiglioni et al., HESS, 2011; Laaha et al., HP, 2014)
- ✓ Flow-duration curves (Ad. Water. Resour., 2016; Castellarin et al., HSJ, 2018)
- ✓ Streamflow series (Skøien & Bloeschl, WRR, 2007; Parajka et al., J. Hydrol. Hydromech., 2015)
- ✓ Streamflow temperature (Laaha et al., Env. Mod. Assess., 2013)
- ✓ Fish-habitat suitability indices (Ceola et al., Adv. Wat. Resourc., 2018)
- ✓ ...

Next application could be yours!
Have fun with the "rtop" R-package
(Skøien et al., Comp. & Geosci., 2014)

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Attilio Castellarin

DICAM - Dipartimento di Ingegneria Civile, Chimica, Ambientale e dei Materiali

attilio.castellarin@unibo.it

www.dicam.unibo.it