EGU 2010

Spatially smooth regional estimation of the flood frequency curve (with uncertainty)

D. Ganora, P. Claps, F.Laio

Politecnico di Torino, Italy

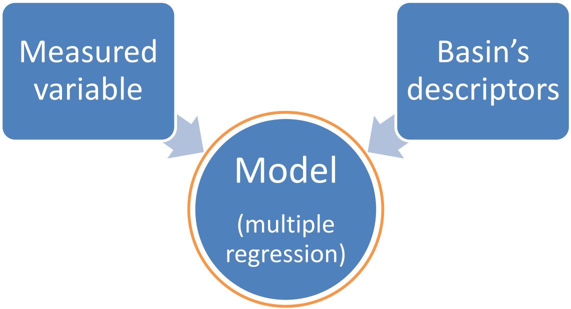
Introduction

Flood frequency curve estimation in ungauged basins

- Regional approach
- Ease of use
- Handle short records

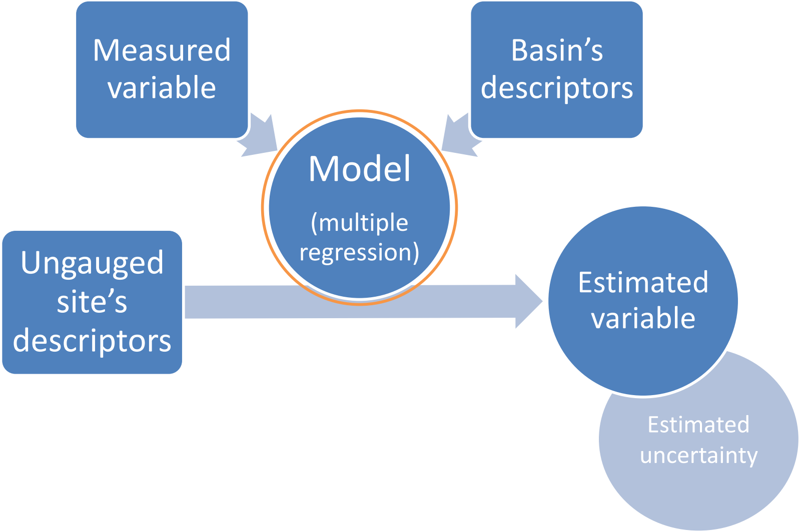
Regionalization approach

Statistical regionalization through morpho-climatic basin-scale descriptors



Regionalization approach

Statistical regionalization through morpho-climatic basin-scale descriptors



Approach characteristics

- Challenging some standard approach assumptions
  - choice of a priori probability distribution
  - regions definition

Approach characteristics

- Challenging some standard approach assumptions
  - choice of a priori probability distribution
  - regions definition
- Uncertainty (variance) of final estimates

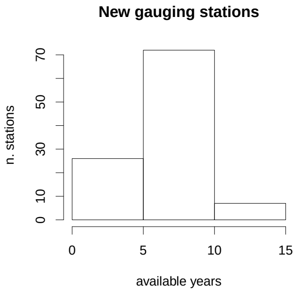
Approach characteristics

- Challenging some standard approach assumptions
  - choice of a priori probability distribution
  - regions definition
- Uncertainty (variance) of final estimates
- Maximization of available at-site information:
  - non-systematic data
  - "poorly" gauged sites (short records)

Approach characteristics

- Challenging some standard approach assumptions
  - choice of a priori probability distribution
  - regions definition
- Uncertainty (variance) of final estimates
- Maximization of available at-site information:
  - non-systematic data
  - "poorly" gauged sites (short records)

Suitable method for Northwestern Italy
(about 100 new stations)



Data handling

Use of *L*-moments statistics

- Possibility to incorporate non-systematic data:
  - historical floods
  - occasional extreme events during gap in time series
- *L*-moments variance via simplified formulae

Data handling

Use of L-moments statistics

- Possibility to incorporate non-systematic data:
  - historical floods
  - occasional extreme events during gap in time series
- L-moments variance via simplified formulae

Flood frequency curve

$$Q_T = Q_{ind} \cdot \mathcal{P}(T, L_{CV}, L_{CA})$$

$$\begin{cases} Q_{ind} \longrightarrow \text{scale (index-flood)} \\ L_{CV} \longrightarrow \text{dispersion} \\ L_{CA} \longrightarrow \text{skewness} \end{cases}$$

Which variable to regionalize?

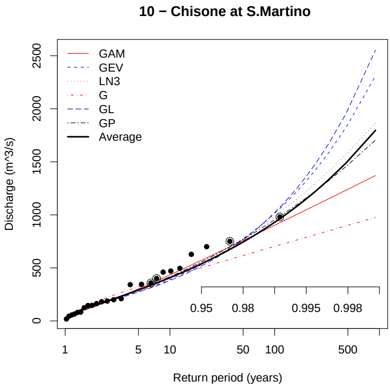
Usually: a priori choice of distribution

- Quantile regionalization
- Regionalization of the parameters of the distribution

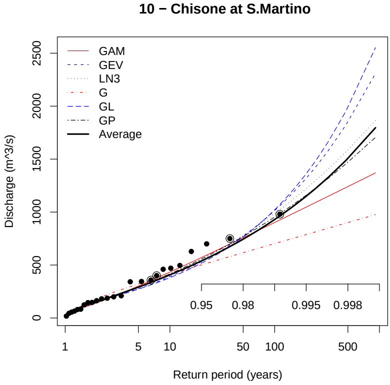
Which variable to regionalize?

Usually: a priori choice of distribution

- Quantile regionalization
- Regionalization of the parameters of the distribution



Which variable to regionalize?



Usually: a priori choice of distribution

- Quantile regionalization
- Regionalization of the parameters of the distribution

Distribution-free statistics

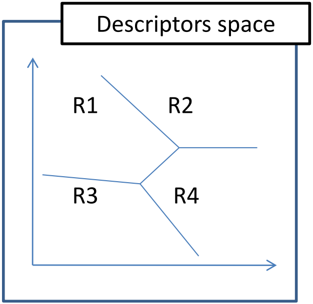
- *L*-moments regionalization
- A posteriori reconstruction of frequency distribution

How to regionalize?

Usually: homogeneous regions

Difficult uncertainty estimation due to:

- regions creation
- regions border effects

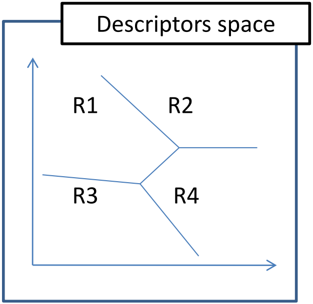


How to regionalize?

Usually: homogeneous regions

Difficult uncertainty estimation due to:

- regions creation
- regions border effects



Smooth variability of L-moments in the descriptors space

- No homogeneity required
- Easier uncertainty evaluation

Regional model definition

- Multiple linear regression
- Error structure as in *Stedinger & Tasker (1985)*
  - Model error $\mathbf{Y}_T = \mathbf{X}\,\beta + \boldsymbol{\delta}$
  - Sampling error $\mathbf{Y} = \mathbf{Y}_T + \boldsymbol{\eta}$

Regional model definition

- Multiple linear regression
- Error structure as in *Stedinger & Tasker (1985)*
  - Model error $\mathbf{Y}_T = \mathbf{X}\,\beta\; + \boldsymbol{\delta}$
  - Sampling error $\mathbf{Y} = \mathbf{Y}_T + \boldsymbol{\eta}$

GLS concurrent estimation of regression coefficients and model variance

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$$

$$\boldsymbol{\Lambda} = \sigma_\delta^2 \mathbf{I} + \boldsymbol{\Sigma}$$

Regional model definition

- Multiple linear regression
- Error structure as in *Stedinger & Tasker (1985)*
  - Model error $\mathbf{Y}_T = \mathbf{X}\ \boldsymbol{\beta}\ + \boldsymbol{\delta}$
  - Sampling error $\mathbf{Y} = \mathbf{Y}_T + \boldsymbol{\eta}$

GLS concurrent estimation of regression coefficients and model variance

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \varepsilon$$

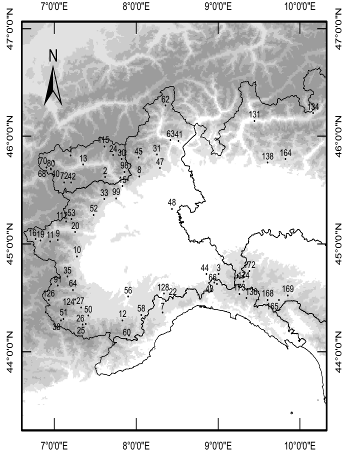
$$\boldsymbol{\Lambda} = \sigma_\delta^2 \mathbf{I} + \boldsymbol{\Sigma}$$

Variance of prediction

$$VP = \sigma_Y^2 = \sigma_\delta^2 + \mathbf{x}\left(\mathbf{X}^T \boldsymbol{\Lambda}^{-1} \mathbf{X}\right)^{-1} \mathbf{x}^T$$

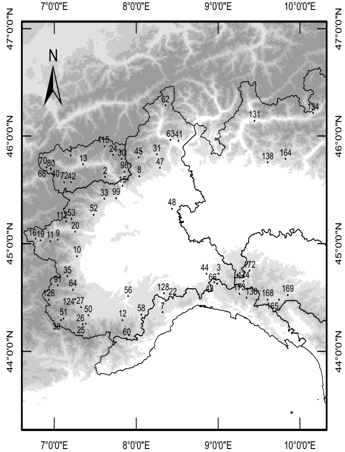
Case study: 70 basins - 35 descriptors

Application to a large descriptors set: all the combination with 1 to 4 descriptors (+ intercept) are calculated



Case study: 70 basins - 35 descriptors

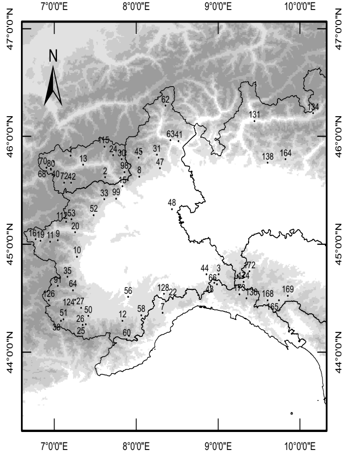
Application to a large descriptors set: all the combination with 1 to 4 descriptors (+ intercept) are calculated



- Model validation
  - Multicollinearity test
  - Student's test
  - Residuals check

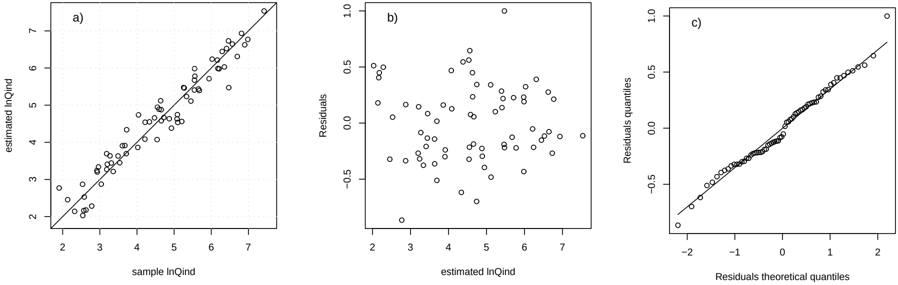
Case study: 70 basins - 35 descriptors

Application to a large descriptors set: all the combination with 1 to 4 descriptors (+ intercept) are calculated



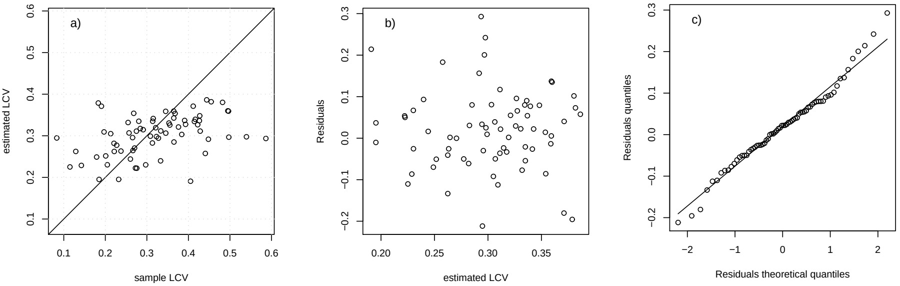
- Model validation
  - Multicollinearity test
  - Student's test
  - Residuals check
- Model selection (among all combinations)
  - Model variance σ_δ^2
  - Average variance of predictions (*AVP*)
  - Simplest model

Example: index-flood estimation



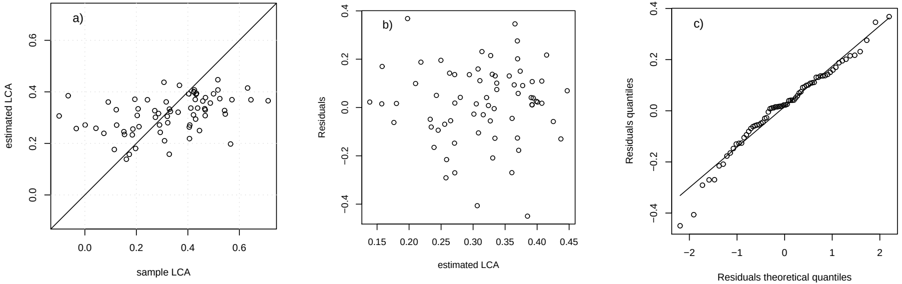
$$\log\left(Q_{ind}\right) = -8.76 + 7.99_{\text{E-01}} \cdot A + 1.09 \cdot IDFa$$
$$+ 9.53_{\text{E-01}} \cdot MAP + 7.85_{\text{E-01}} \cdot perm$$

Example: L_{CV}



$$L_{CV} = 6.44_{\text{E-01}} - 4.28_{\text{E-07}} \cdot X_c - 5.00_{\text{E-04}} \cdot P - 1.44_{\text{E-04}} \cdot H_{min}$$

Example: L_{CA}



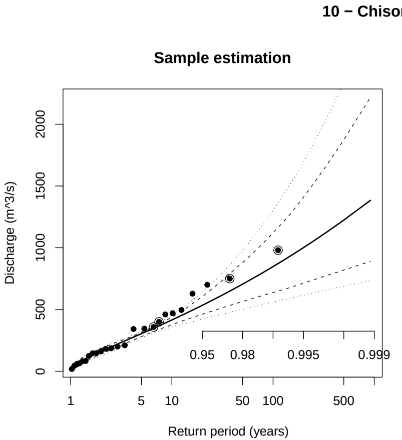
$$L_{CA} = 9.38\text{E-01} - 1.40\text{E-02} \cdot IDFa - 1.39\text{E-01} \cdot P - 2.65\text{E-04} \cdot H_{min}$$

Quantile estimation

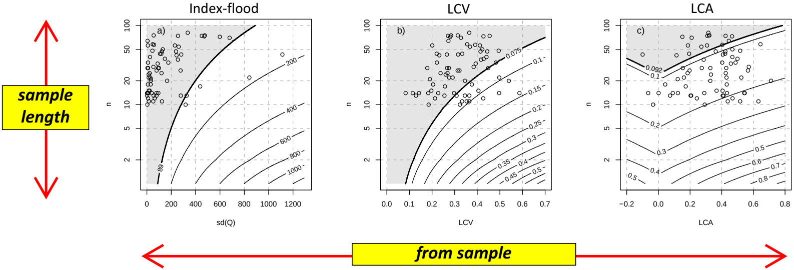
- Q_{ind}, L_{CV} and L_{CA} $\longrightarrow$ different suitable distributions
- Flood frequency curve as the average
- Confidence bands through Monte Carlo simulations (using also $\sigma^2_{Q_{ind}}$, $\sigma^2_{L_{CV}}$ and $\sigma^2_{L_{CA}}$)

Quantile estimation

- Q_{ind}, L_{CV} and L_{CA} $\longrightarrow$ different suitable distributions
- Flood frequency curve as the average
- Confidence bands through Monte Carlo simulations (using also $\sigma^2_{Q_{ind}}$, $\sigma^2_{L_{CV}}$ and $\sigma^2_{L_{CA}}$)



Final remarks: a practical tool



Shaded area = sample estimate with variance lower than regional one