UNIVERSITAT POLITÈCNICA DE VALÈNCIA
Instituto de Ingeniería del Agua y Medio Ambiente

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
DIPARTIMENTO di SCIENZE della TE[RRA] e GEOLOGICO-AMBIENTALI

Predictive uncertainty estimation at ungauged basins in a Bayesian framework

Juan Camilo Múnera[(1)], Félix Francés[(1)], Ezio Todini[(2)]

*1 - Research Institute of Water and Environmental Engineering, **Universitat Politècnica de València**, Spain*
*2 – Earth Science and Geology Department, **Università di Bologna**, Italy*

IAHS AISH

PUB SYMPOSIUM 2012 – Delft – 24 October 2012

Introduction

- Flow simulations and forecasts from models **are not error free**
- Implications in **decision making** process

=> Growing interest in assessing **uncertainty** of predictions

- Several developed methods for gauging stations:
  - Hydrologic Uncertainty Processor (Krzysztofowicz, 1999)
  - Bayesian Model Averaging (Raftery, 1993; Raftery et al, 2003; 2005)
  - Meta-Gaussian error model (Montanari and Brath, 2004)
  - Model Conditional Processor (Todini, 2008)
  - Some others statistical approaches…

- How to estimate PU at **ungauged sites**?

Introduction

- Two different approaches for Predictive Uncertainty estimation:

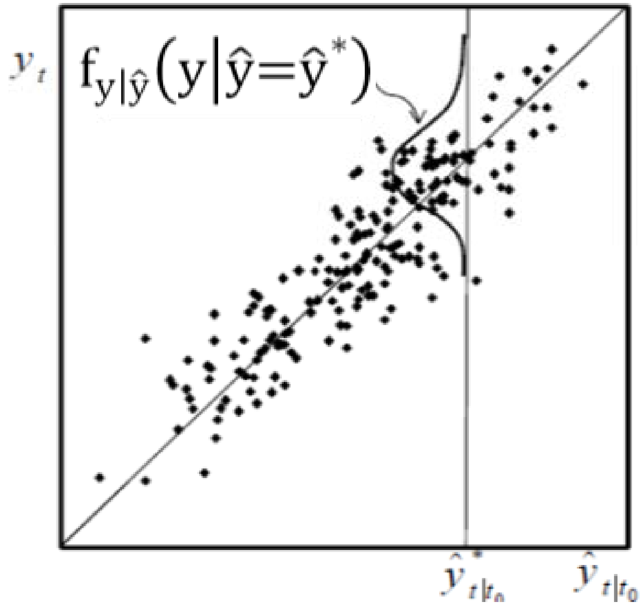
1st proposal: One hydrological model and multi-site data assimilation using a Kalman filter based algorithm **MISP** (*Mutually Interactive State-Parameter Estimation)*

Presented at "Floods in 3D" Leonardo EGU Conference. Bratislava,November 2011

2nd proposal: Multi-site and multi-model approach using the bayesian post-processor **MCP** (*Model Conditional Processor*) by combining historical simulations from various hydrological models at several sites on a region, in a regionalization framework

Predictive Uncertainty definition

- Probability density function (pdf) of the predictand given our model prediction (simulation or forecast).



From Todini (2011)

- NQT transformation (Krzysztofowicz,1999):
  - Observations: y
  - Models predictions: $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N$

$$\xrightarrow{\text{NQT}} \begin{cases} \eta \\ \hat{\eta}_1, \ldots, \hat{\eta}_N \end{cases}$$

MCP with one prediction model

- The predictive distribution of an event conditional on the prediction of a model in the Gaussian space:

$$f(\eta/\hat{\eta}) = f(\eta,\hat{\eta}) / f(\hat{\eta}) \qquad (1)$$

- The joint distribution in the Normal space has moments:

  - Mean: $\mu_{\eta,\hat{\eta}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

  - Variance: $\sum_{\eta,\hat{\eta}} = \begin{bmatrix} 1 & \sigma_{\eta,\hat{\eta}} \\ \sigma_{\eta,\hat{\eta}} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \rho_{\eta,\hat{\eta}} \\ \rho_{\eta,\hat{\eta}} & 1 \end{bmatrix}$ Marginals of η and $\hat{\eta} \approx \mathrm{N}(0,1)$

- The predictive conditional pdf resulting from (1) is Normal with moments:

$$\mu_{\eta/\hat{\eta}} = \rho_{\eta\hat{\eta}} \cdot \hat{\eta}$$

$$\sigma^2_{\eta/\hat{\eta}} = 1 - \rho^2_{\eta\hat{\eta}}$$

MCP with multiple prediction models

- The predictive distribution of an event conditional on the prediction of more than one model in the Gaussian field:

$$f\left(\eta / \hat{\eta}_1,\ldots,\hat{\eta}_N\right) = f\left(\eta,\hat{\eta}_1,\ldots,\hat{\eta}_N\right) / \, f\left(\hat{\eta}_1,\ldots,\hat{\eta}_N\right) \qquad (2)$$

- The predictive conditional pdf from eq. (2) is Normal and it has moments:

  - Mean: $\mu_{\eta/\hat{\eta}} = \sum_{\eta\hat{\eta}} \sum_{\hat{\eta}\hat{\eta}}^{-1} \hat{\eta}$

  - Variance: $\sigma^2_{\eta/\hat{\eta}} = \sum_{\eta\eta} - \sum_{\eta\hat{\eta}} \sum_{\hat{\eta}\hat{\eta}}^{-1} \sum_{\eta\hat{\eta}}^{T}$

$$= 1 - \sum_{\eta\hat{\eta}} \sum_{\hat{\eta}\hat{\eta}}^{-1} \sum_{\eta\hat{\eta}}^{T}$$

where:

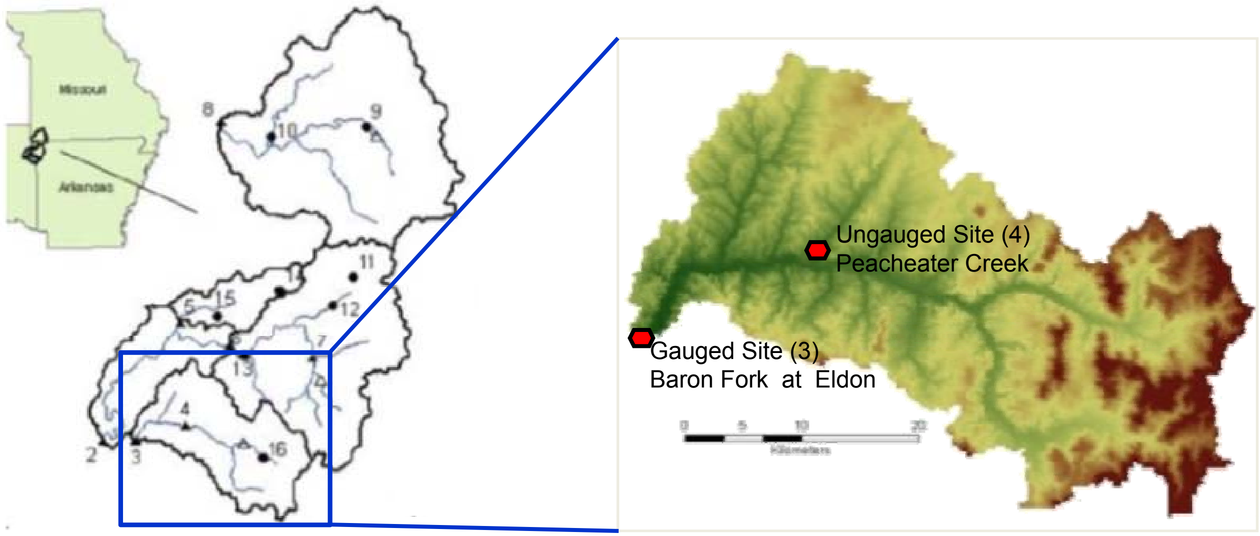
$$\sum_{\eta\eta} = 1$$

$$\sum_{\eta\hat{\eta}} = \left[\rho_{\eta\hat{\eta}_1} \quad \rho_{\eta\hat{\eta}_2} \quad \cdots \quad \rho_{\eta\hat{\eta}_N}\right]$$

$$\sum_{\hat{\eta}\hat{\eta}} = \begin{bmatrix} 1 & \rho_{\hat{\eta}_1\hat{\eta}_2} & \cdots & \rho_{\hat{\eta}_1\hat{\eta}_N} \\ \rho_{\hat{\eta}_2\hat{\eta}_1} & 1 & \cdot & \rho_{\hat{\eta}_2\hat{\eta}_N} \\ \vdots & \cdot & \ddots & \vdots \\ \rho_{\hat{\eta}_N\hat{\eta}_1} & \cdots & \rho_{\hat{\eta}_N\hat{\eta}_{N-1}} & 1 \end{bmatrix}$$

Case study

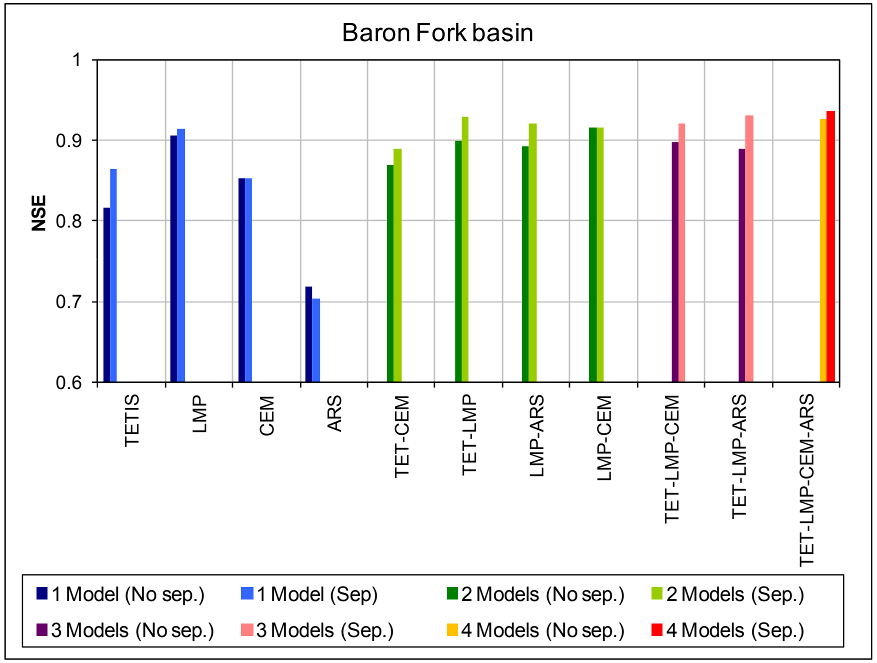
- Distributed Model Intercomparison Project (DMIP2) by NOAA/NWS.
- Four Hydrological models selected from DMIP2 participants: TETIS (UPV), LMP SAC (NWS), SWAT (ARS), GR4J (CEM)
- We are using 13 gauge stations:
  - Calibration at Baron Fork (3) 795 km^2
  - Ungauged basin: Peacheater Creek (4) 65 km^2

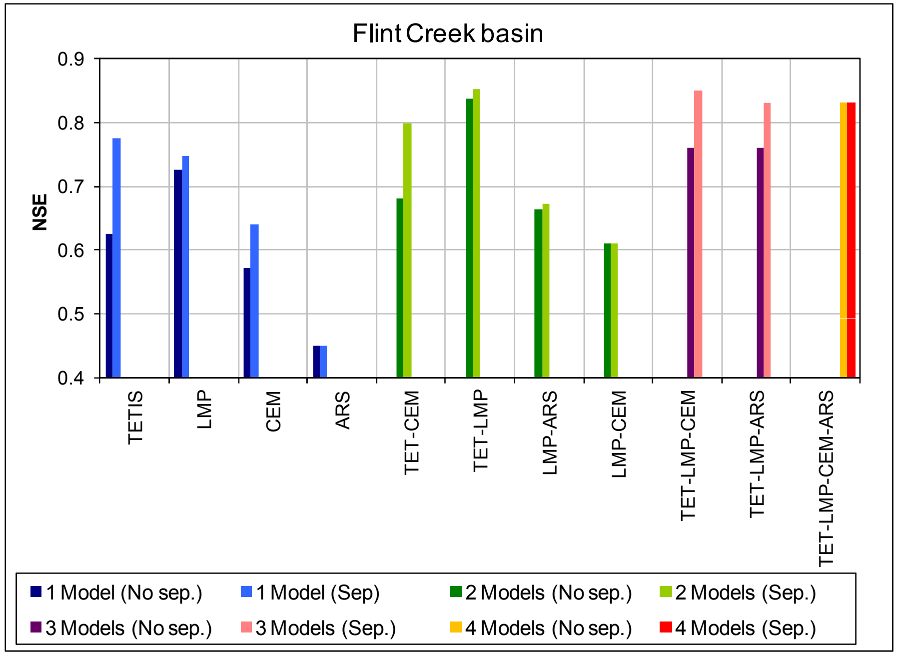


Results at gauged sites

- Nash Sutcliffe efficiency index

One single model (blue) and combination of two (green), three (purple-pink) and four (orange-red) models.



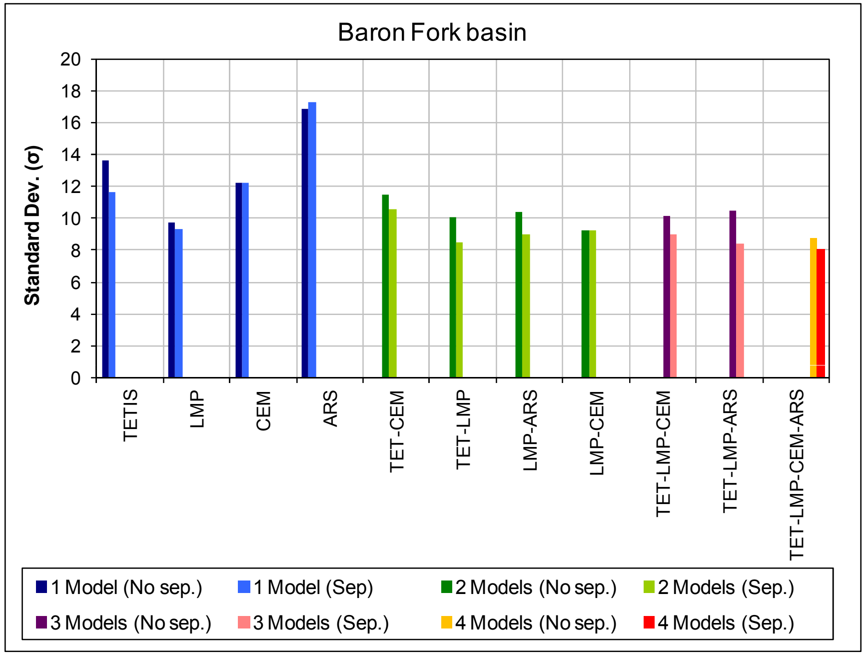


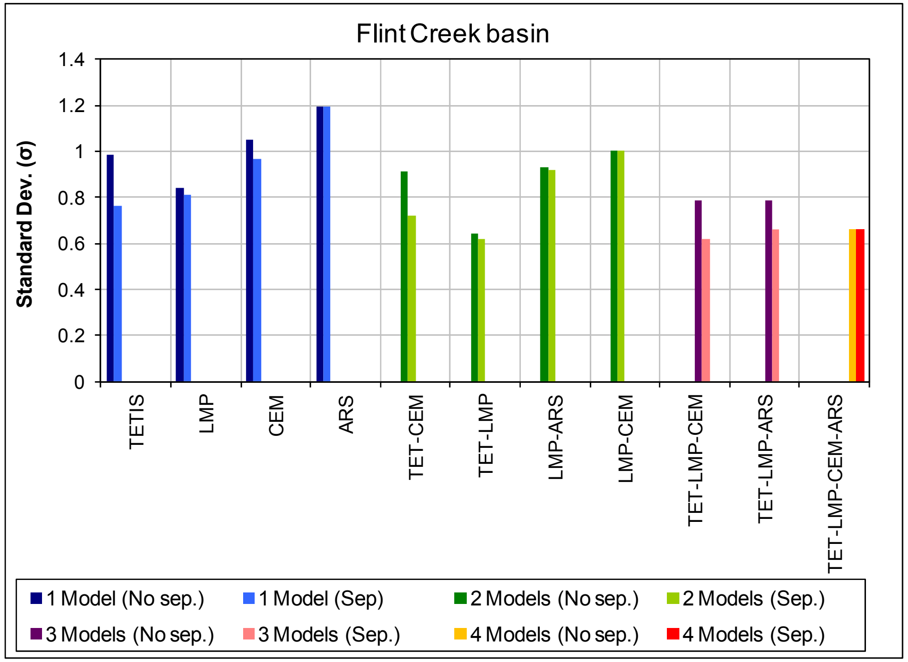
<u>Left bar</u>: all data in a single sample

<u>Right bar</u>: separation of data in two Truncated Normal Joint distributions

Results at gauged sites

- Standard deviation (σ) of model errors





<u>Left bar</u>: all data in a single sample

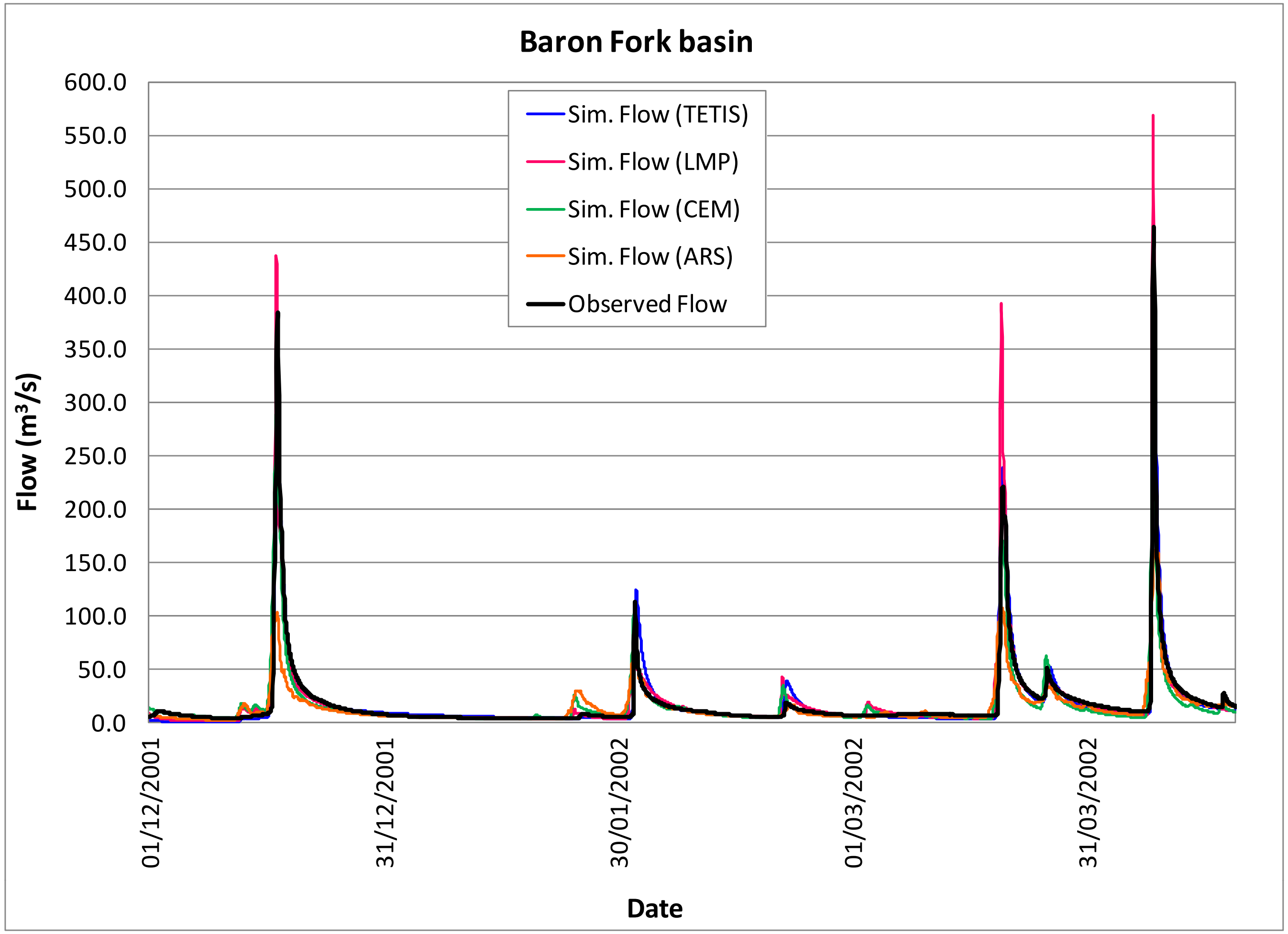
<u>Right bar</u>: separation of data in two Truncated Normal Joint distributions

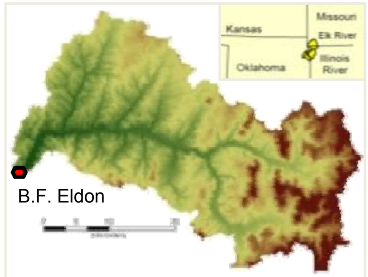
Instituto de Ingeniería del Agua y Medio Ambiente

Results at gauged sites

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

- Single models at the outlet point: calibration period



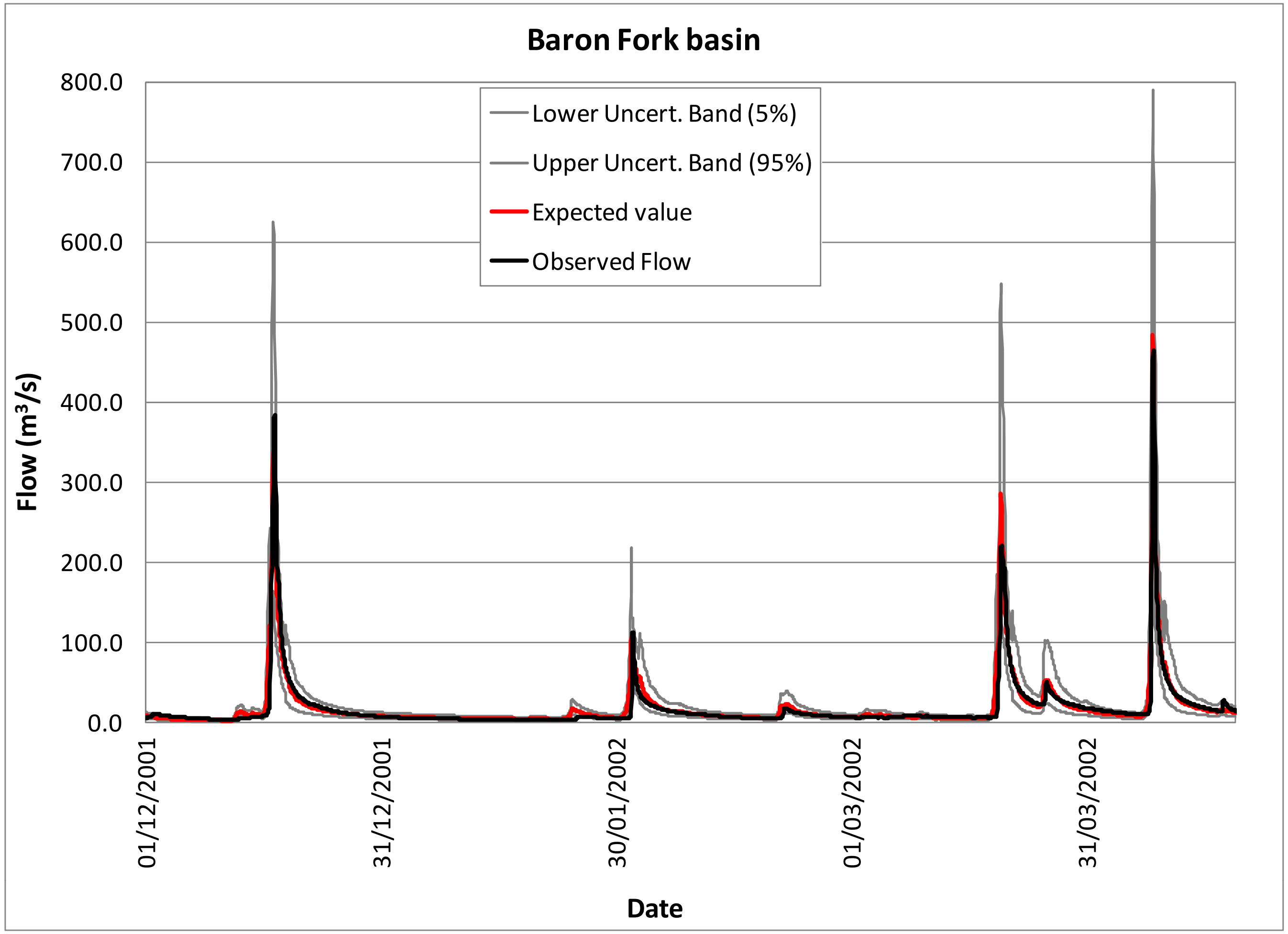


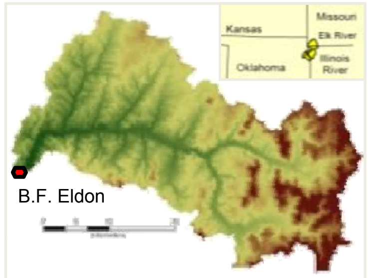
Instituto de Ingeniería del Agua y Medio Ambiente

Results at gauged sites

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

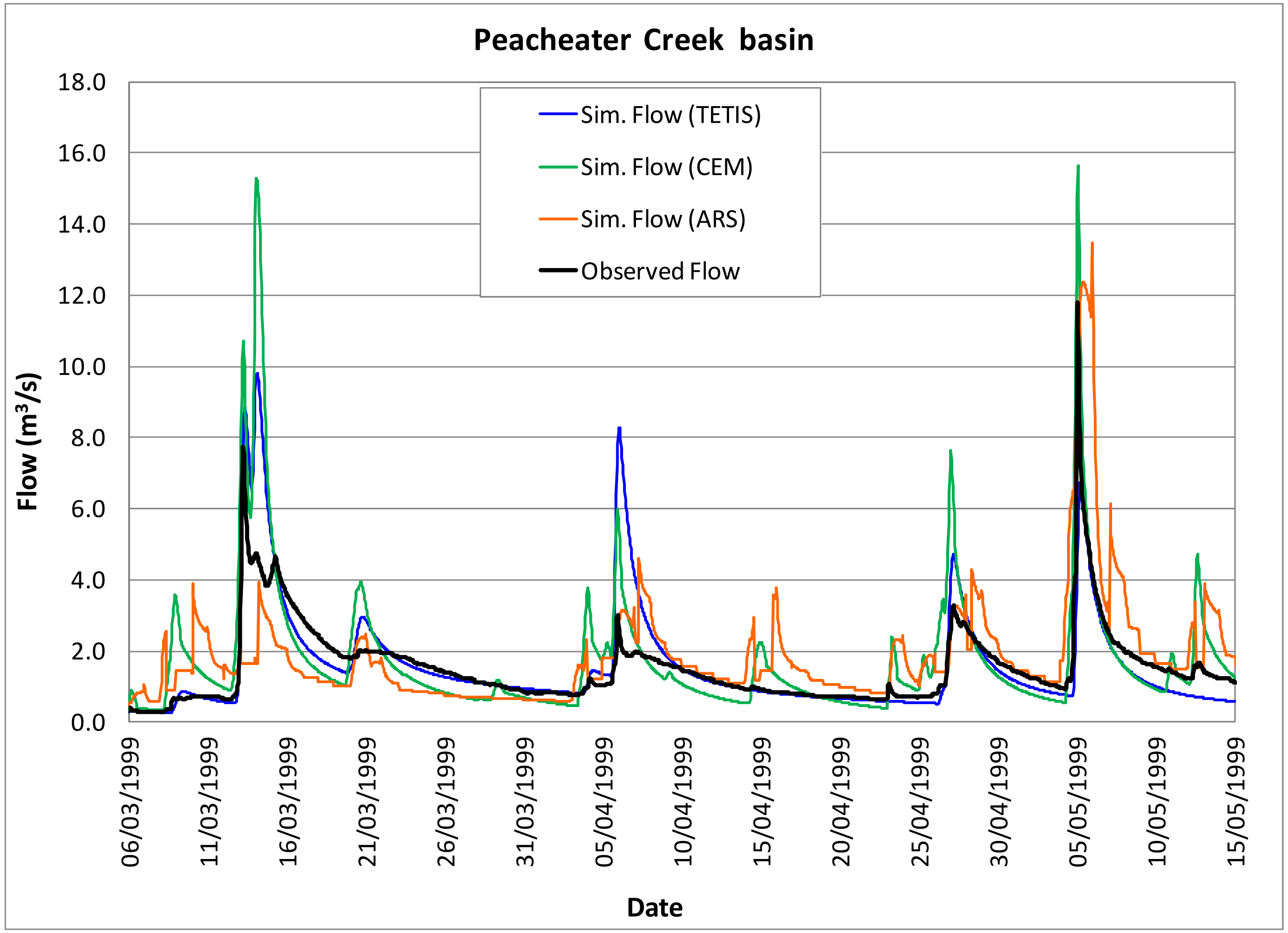
- Predictive Uncertainty Band (90 %) for the combination of 4 models

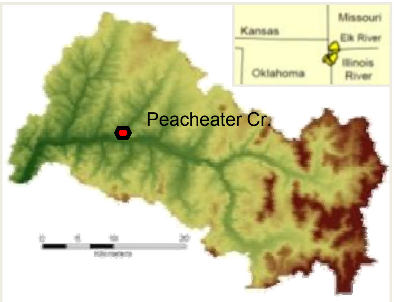




Results at gauged sites

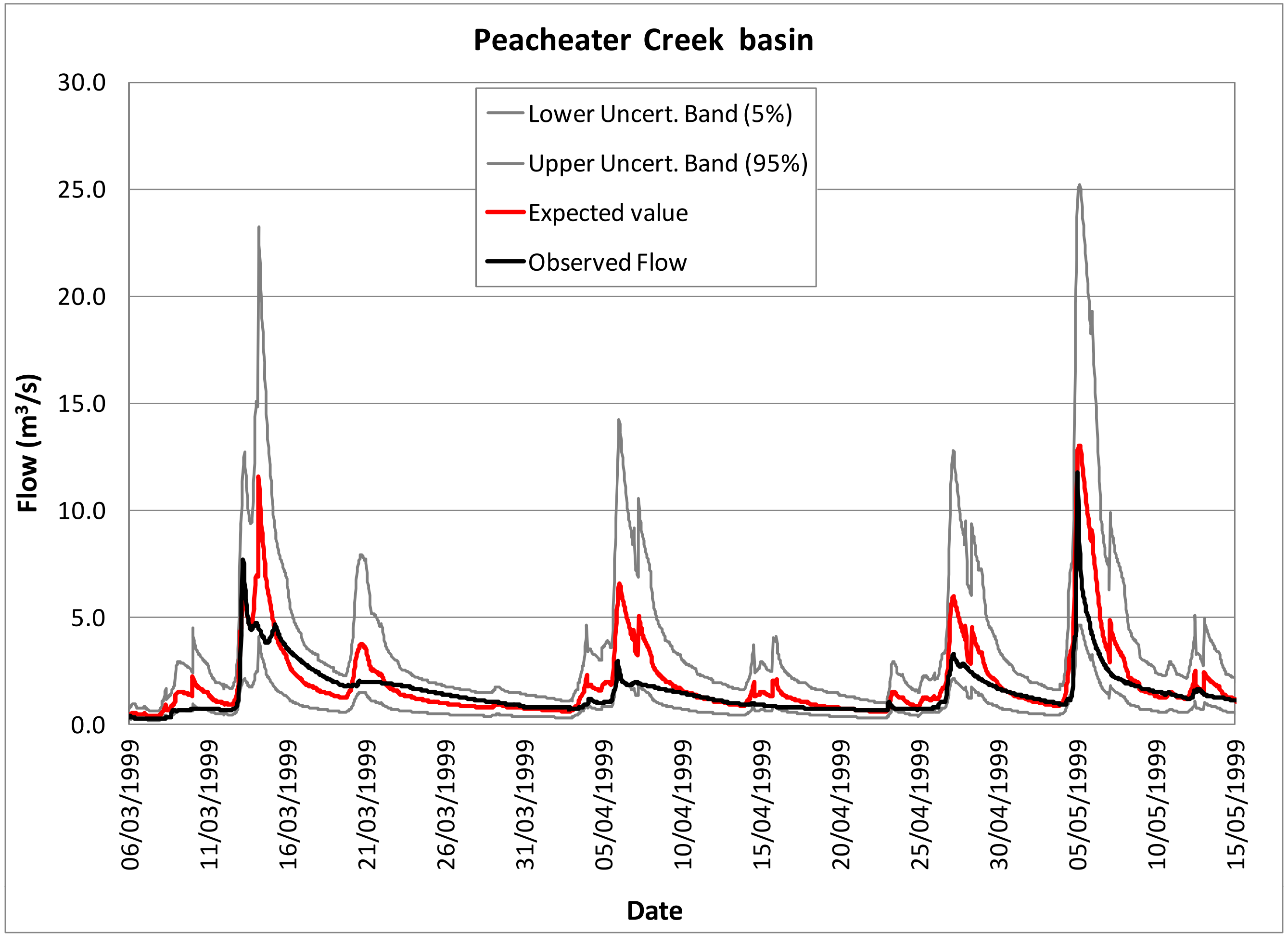
- Model predictions (interior point): spatial and temporal validation

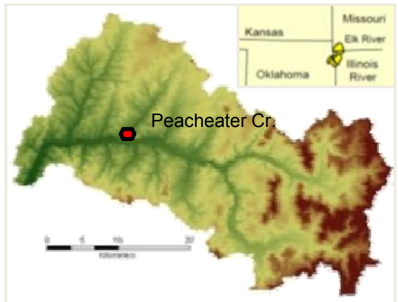




Results at gauged sites

- Predictive Uncertainty Band (90 %) with the combination of 3 models



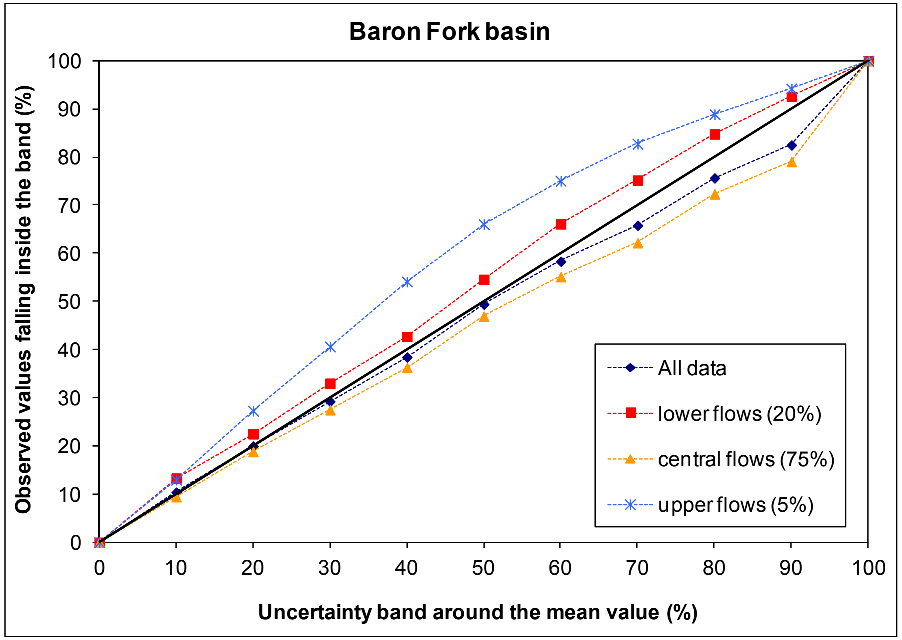


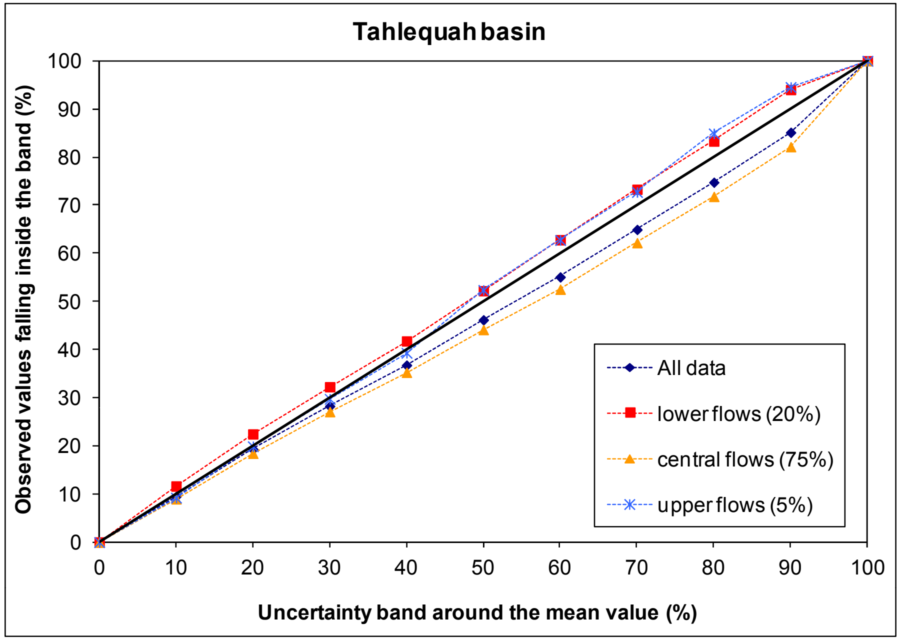
Instituto de Ingeniería del Agua y Medio Ambiente

PU band reliability assessment

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

- Percentage of observed data that fall inside the uncertainty band at different probability levels. Combination of 4 models.





Extension for PUB

- Reduce uncertainty by models combination using MCP approach

- Regionalization to extend the MCP approach to **ungauged sites:**
  - Use of an analytical CDF instead of the empirical Weibull Plotting Position
    - Parameter estimation by L-moments (Hosking, 1990; Hosking and Wallis, 1995)
    - L-moments estimation at ungauged sites by regression with simulated L-moments and basin descriptors
  - Regression of "correlation coefficients" (ρ) in the Gaussian space vs "drainage area"

Instituto de Ingeniería del Agua y Medio Ambiente

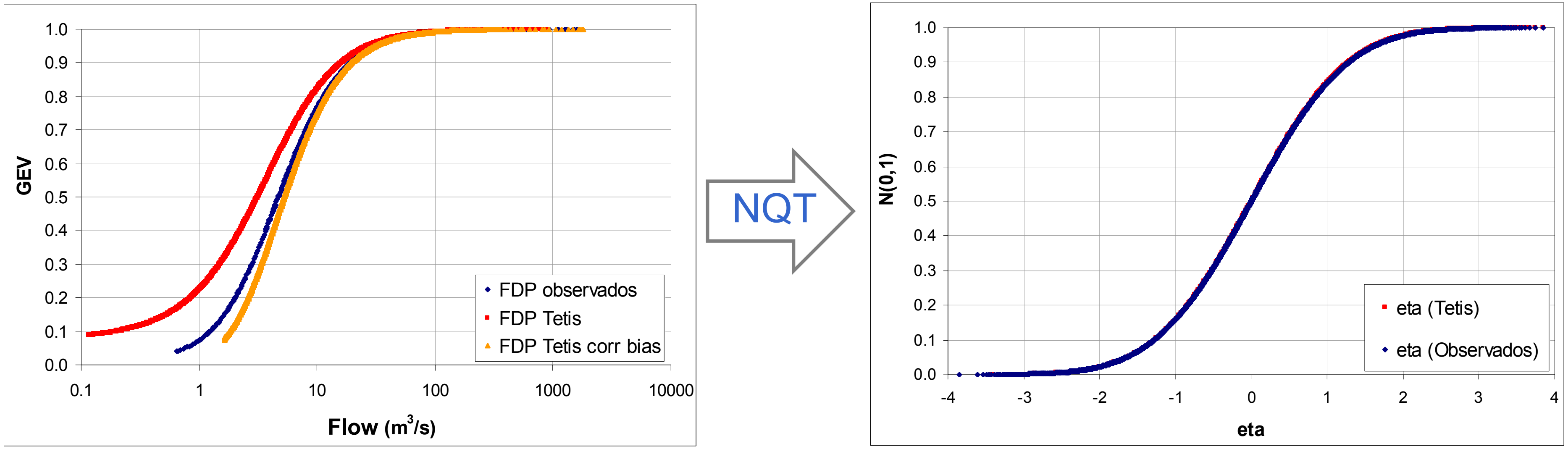
Analytical NQT transformation

DIPARTIMENTO di SCIENZE della TE[RRA] e GEOLOGICO-AMBIENTALI

- Chosen CDF: General Extreme Value (GEV):

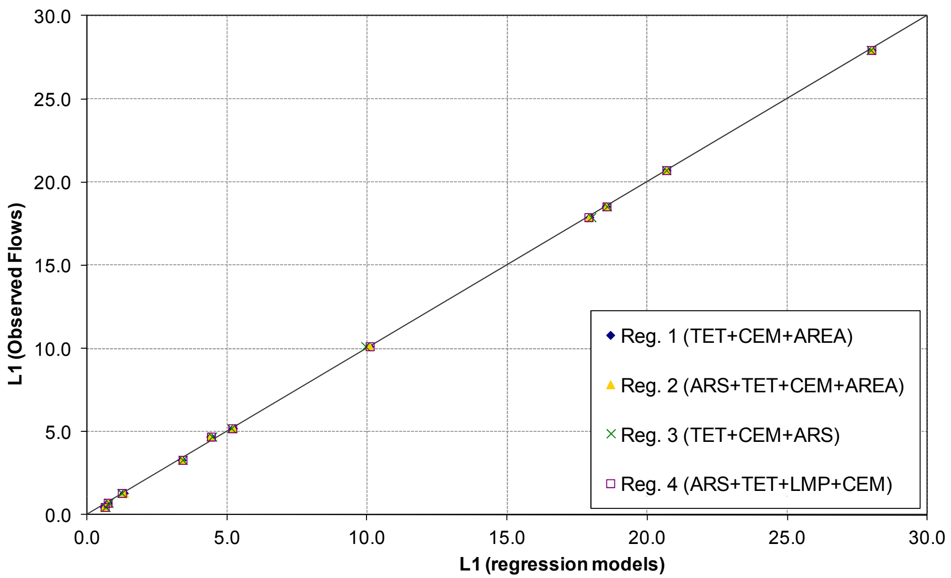
$$F = \exp\left[-\left\{1-k(x-\xi)/\alpha\right\}^{1/k}\right]$$

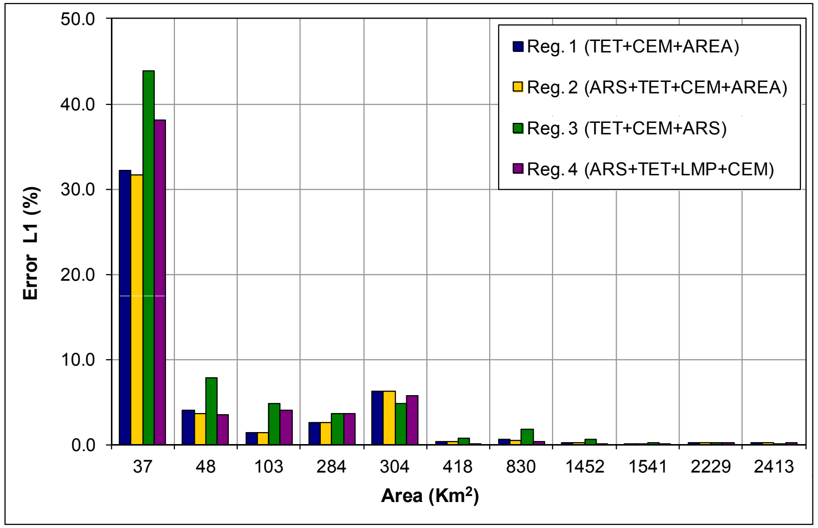
$$x = \xi + \frac{\alpha\left\{1-(-\log F)^{k}\right\}}{k}$$



L-moments regression

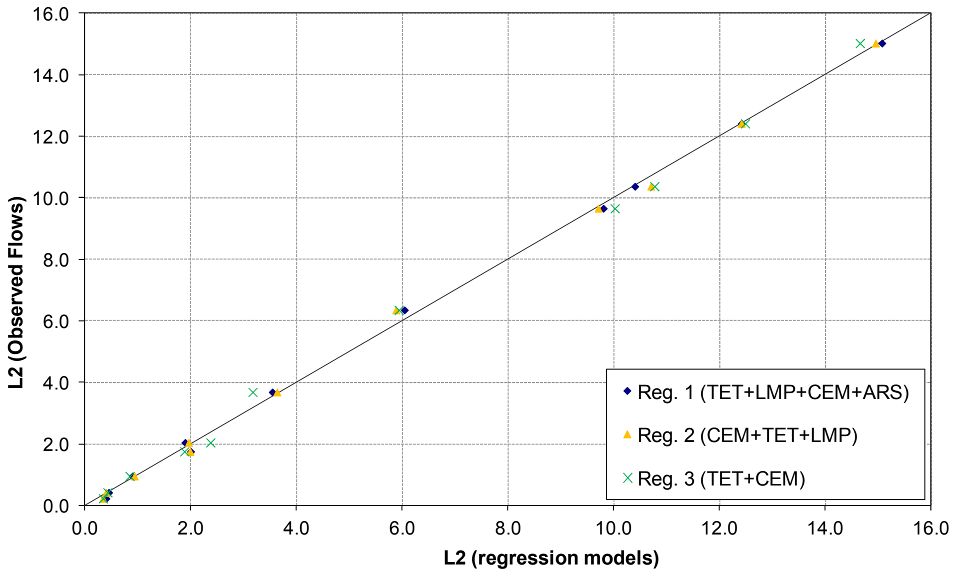
First L-moment (L1)

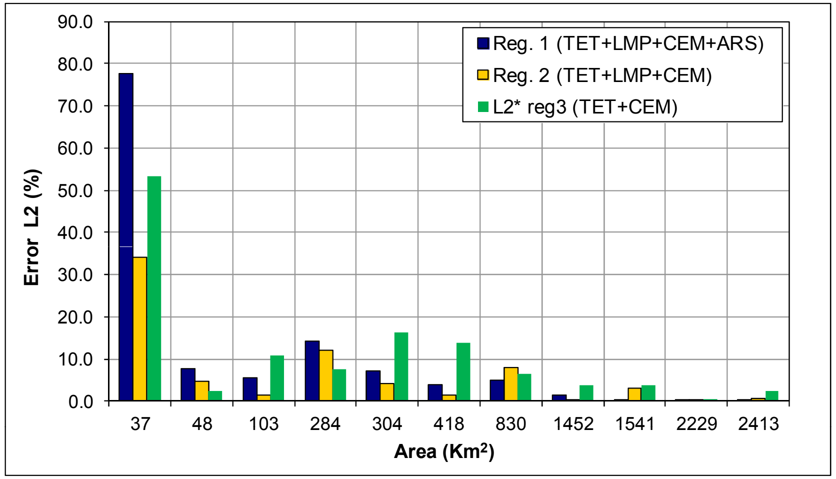




L-moments regression

- Second L-moment (L2)



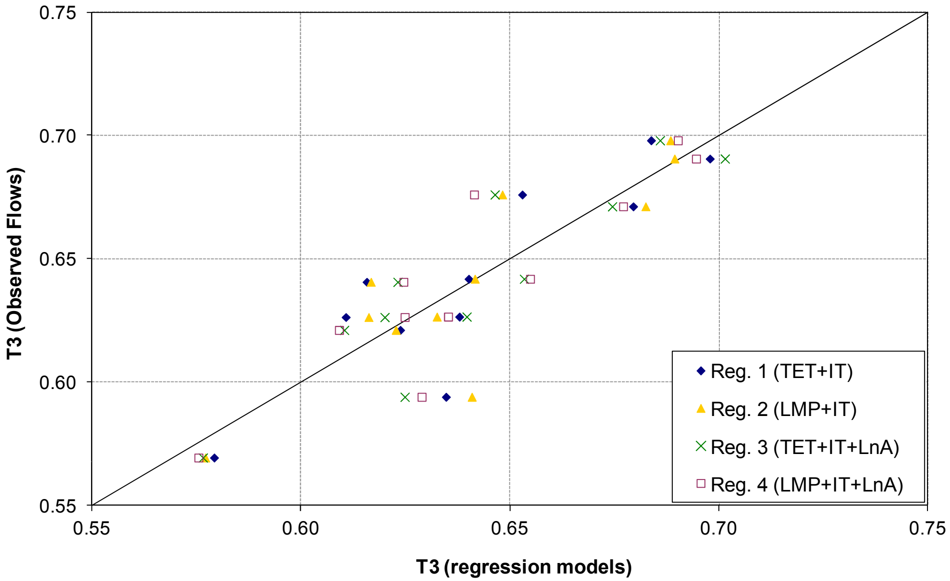


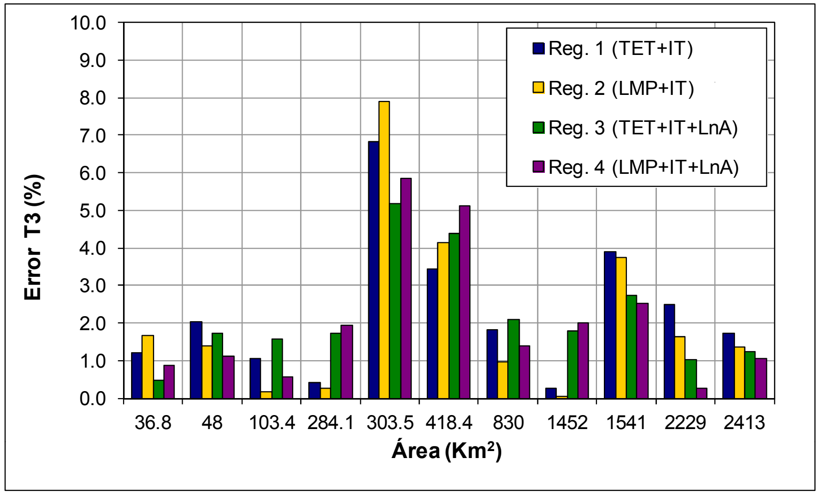
Instituto de Ingeniería del Agua y Medio Ambiente

L-moments regression

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

- L-Skewness (T3)





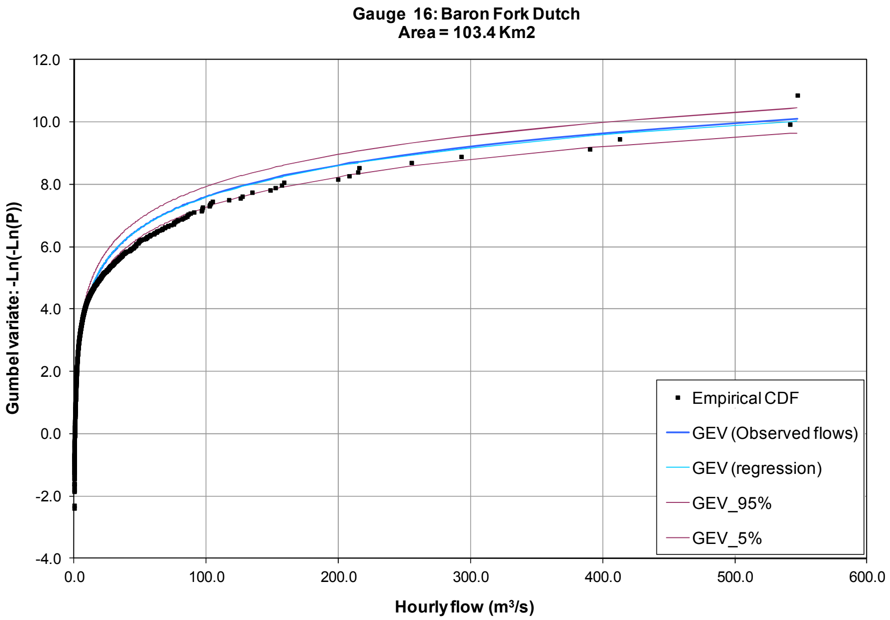
CDF estimation by regressions

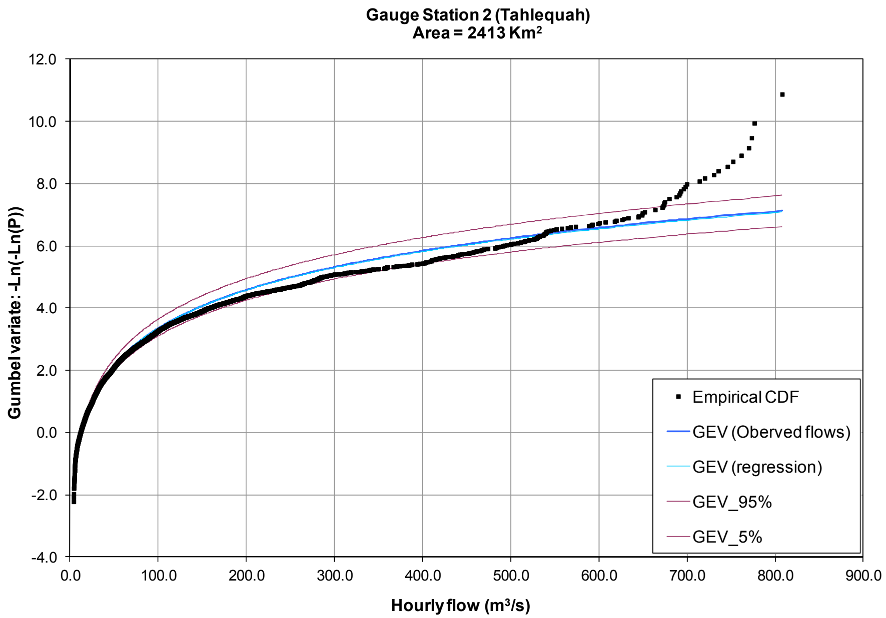
Instituto de Ingeniería del Agua y Medio Ambiente

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

- Empirical, adjusted and estimated (L-moments statistical regression) CDF of hourly flows.

General Extreme Value Function (GEV)



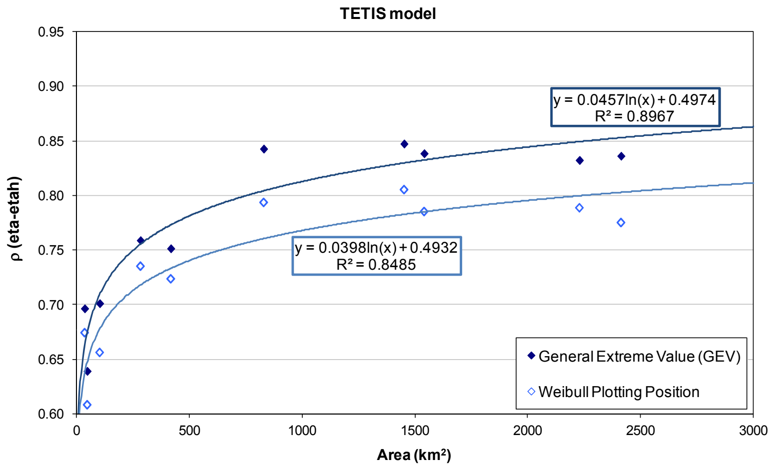


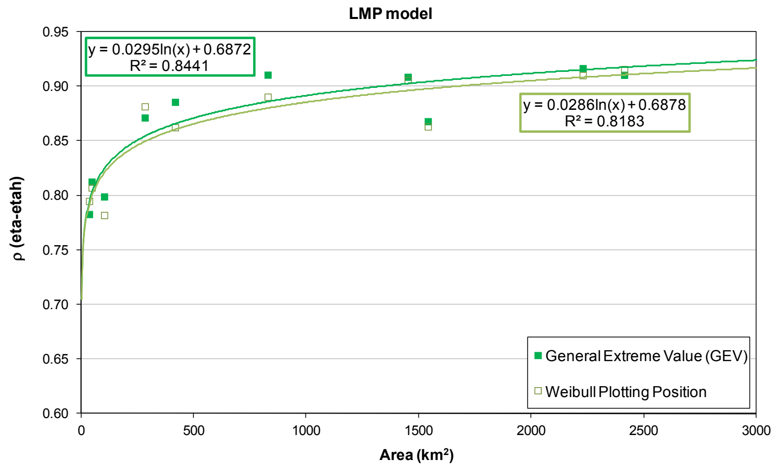
Instituto de Ingeniería del Agua y Medio Ambiente

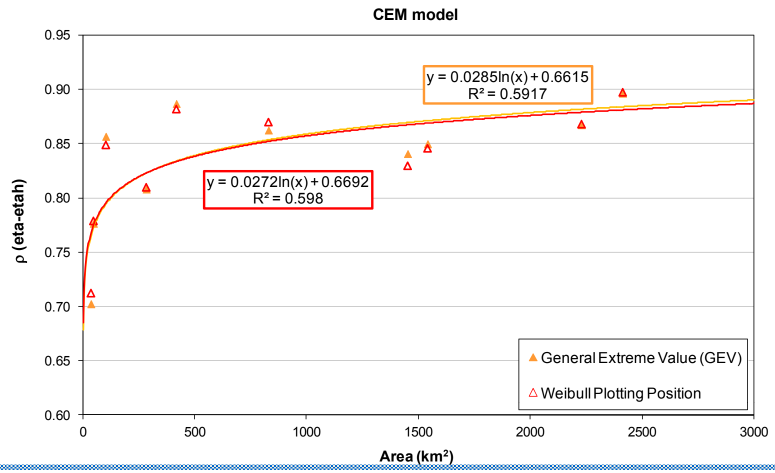
Correlation coefficients regressions

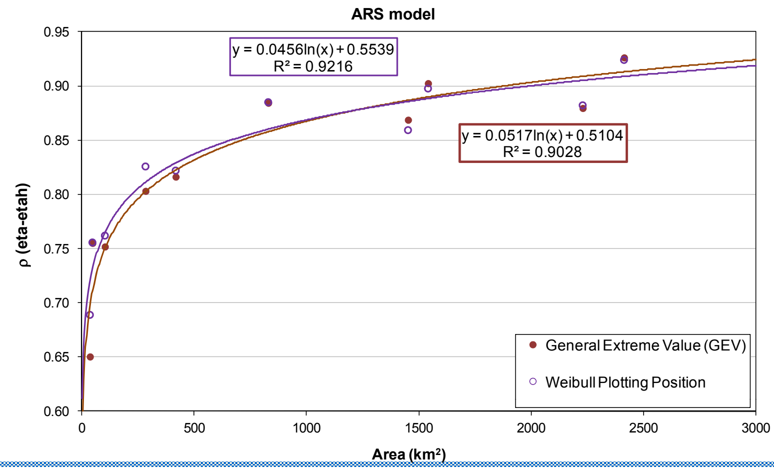
DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

- Regression of “correlation coefficient” (ρ) between observed and simulated flows in the Gaussian field vs “drainage area”



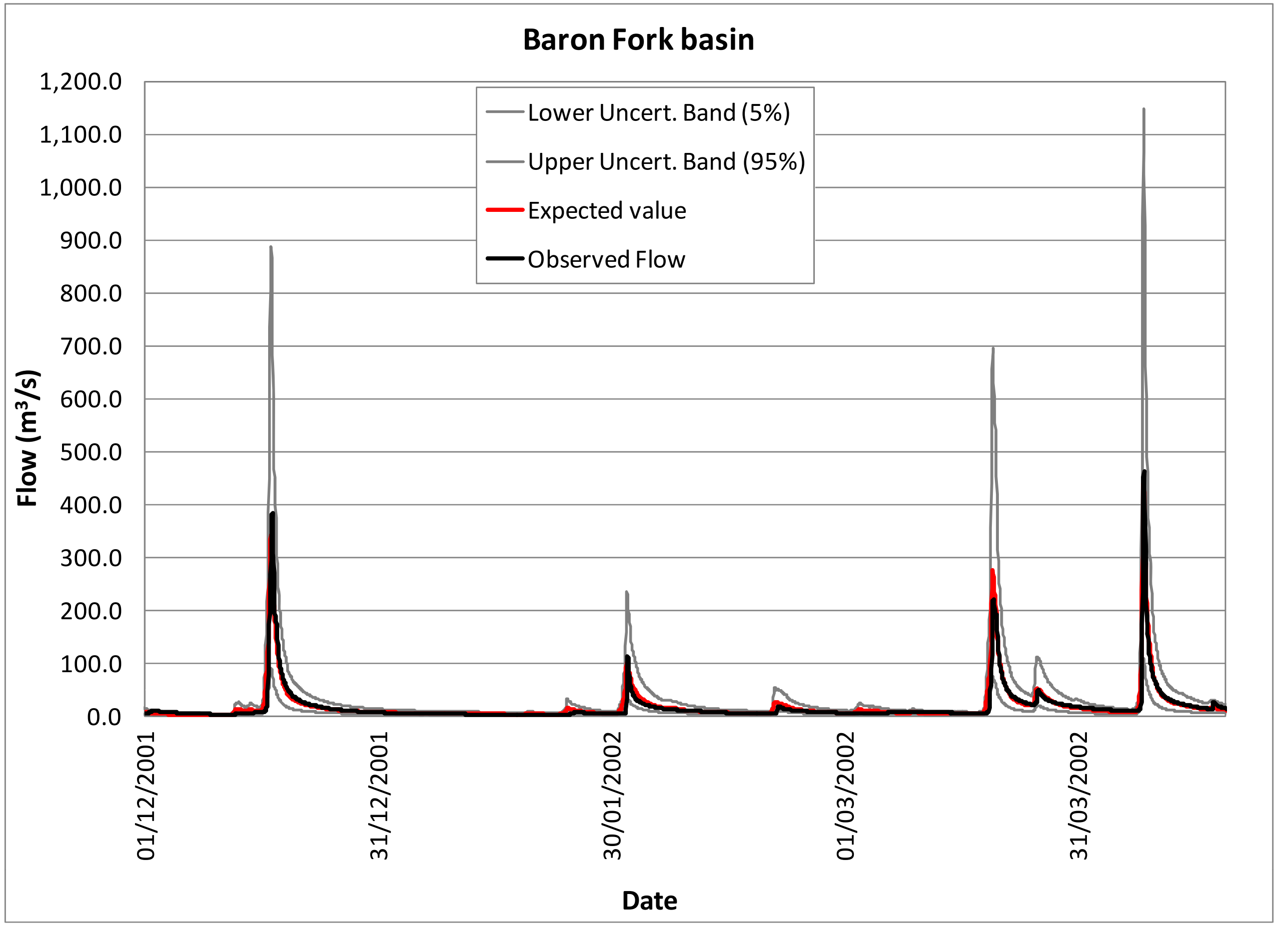


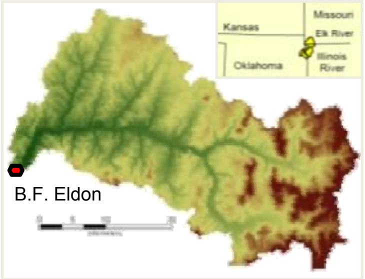




Multi-model combination

- PU band (90 %). Regression approach, combination of 4 models at Baron Fork => using regressions at a gauge site



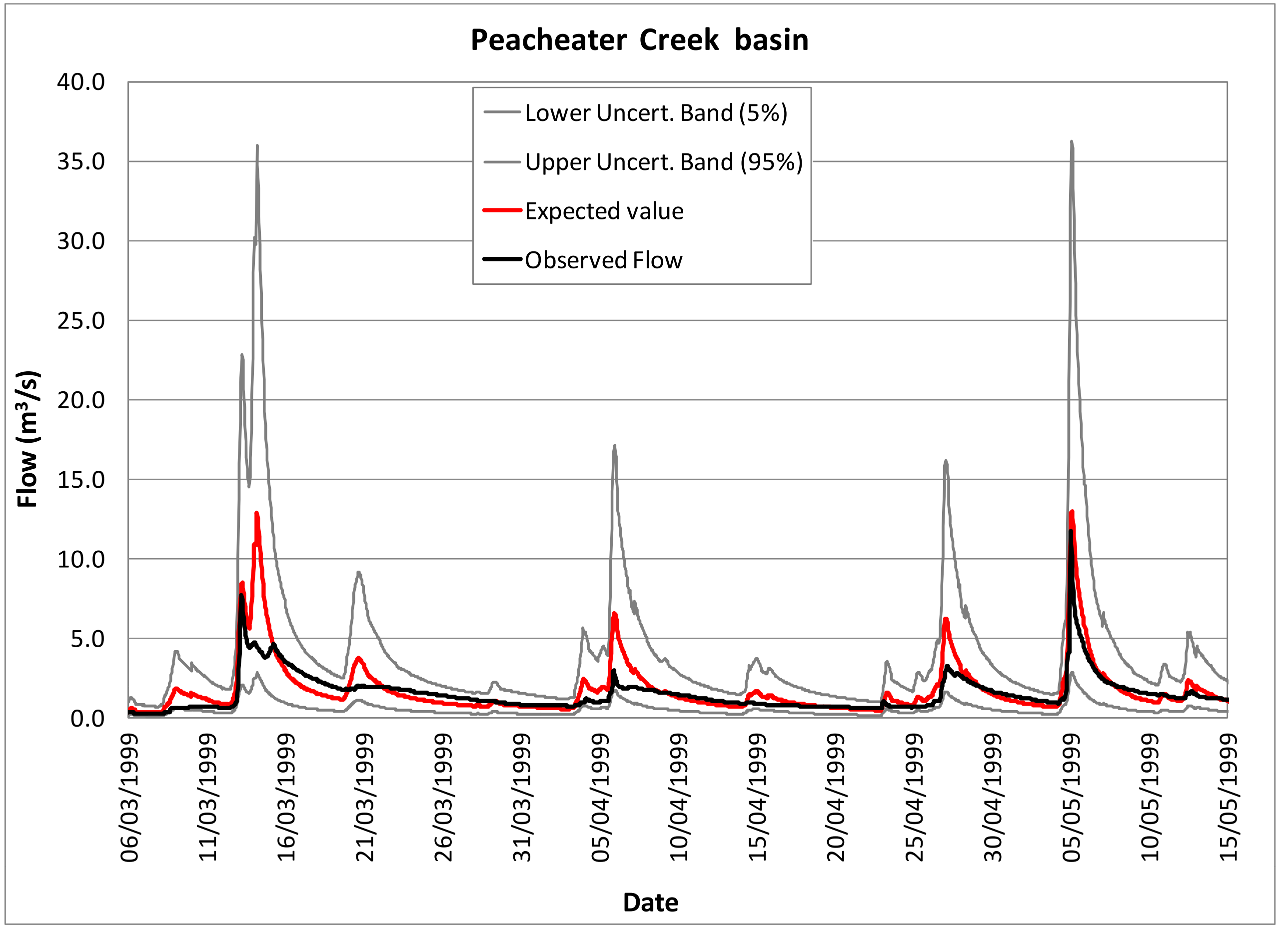


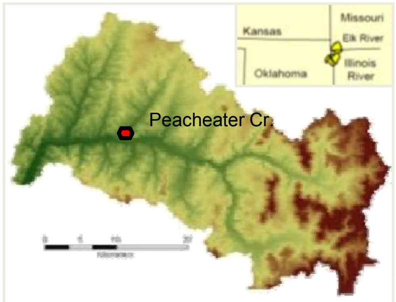
Instituto de Ingeniería del Agua y Medio Ambiente

Multi-model combination

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

- PU band (90 %). Regression approach, combination of 3 models at Peacheater => using regressions at an ungauged site



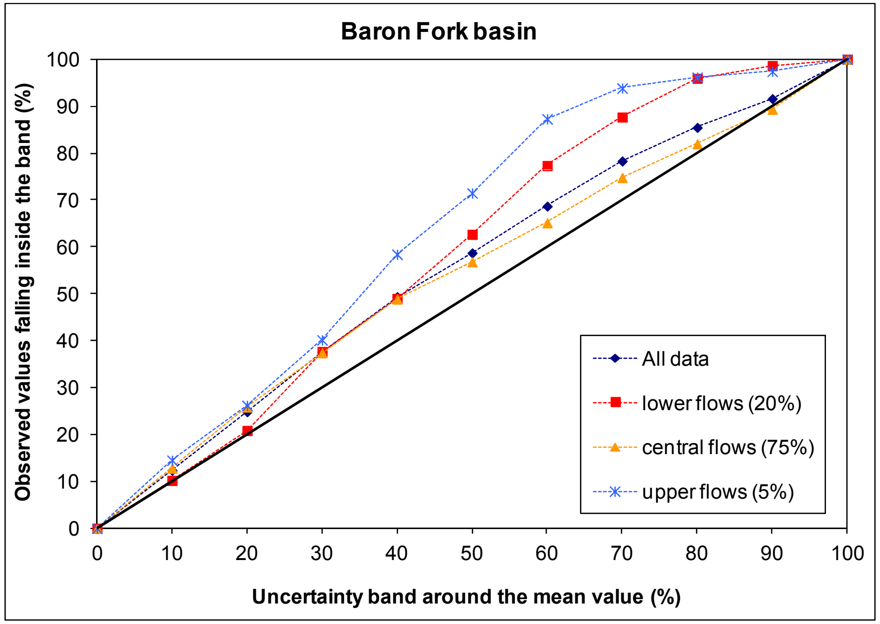


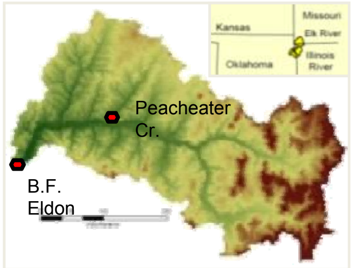
PU band reliability assessment

Instituto de Ingeniería del Agua y Medio Ambiente

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

- Percentage of observed data that fall inside the uncertainty band at different probability levels.



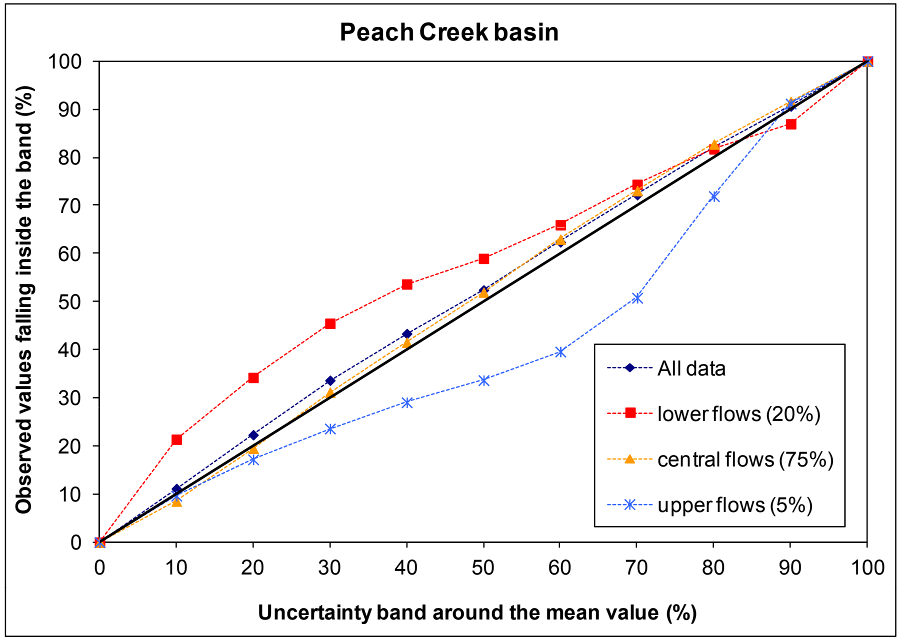


Baron Fork: combination of 4 models.

Overestimation of PU for all flows.

Peacheater Creek: combination of 3 models.

Underestimation of PU for high flows and overestimation for low flows.



Conclusions

- **Models combination** reduces Predictive Uncertainty at gauged and ungauged sites.

- It has been proposed to estimate the PU at ungauged sites: models combination using the **bayesian MCP post-processor** with L-moments and correlation **regionalization by regressions**.

- The reliability of the estimated PU at gauged and ungauged sites was verified through a “PU band assessment” for probabilities 10 (10) 90%, with **satisfactory results**.

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Instituto de Ingeniería del Agua y Medio Ambiente

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO di SCIENZE della TE e GEOLOGICO-AMBIENTALI

Thanks for your attention!

Félix Francés (ffrances@hma.upv.es)

Research Group of Hydrological and Environmental Modelling

http://lluvia.dihma.upv.es

IAHS AISH