Uncertainty assessment of streamflow simulation on ungauged catchments using a distributed model and the Kalman filter based MISP algorithm

Authors:

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

(1) **Technical University of Valencia (Spain)**
Research Institute of Water Engineering and Environment
Research Group of Hydrological and Environmental Modeling
http://lluvia.dihma.upv.es

(2) **University of Bologna (Italy)**
Earth Sciences and Environment Department
http://www.geomin.unibo.it/

EGU Leonardo Conference Series on the Hydrological Cycle
FLOODS IN 3D: PROCESSES, PATTERNS, PREDICTIONS
Bratislava, Slovakia, 23-25 November 2011

Introduction

- Important role of Predictive Uncertainty (PU) on flood forecasting and decision making:
  - Flow simulations and forecasting made from models are not error free
  - Growing interest in assessing uncertainty of predictions/simulations
  - Severe economic and social consequences derived from flood emergencies

- Developed methods and tools to assess flood forecasting uncertainty at gauging stations:
  - Hydrologic Uncertainty Processor (Krzysztofowicz, 1999)
  - Bayesian Model Averaging (Raftery, 1993; Raftery et al, 2003; 2005)
  - Model Conditional Processor (Todini, 2008)
  - And others statistical approaches…

- How to estimate PU at ungauged sites?

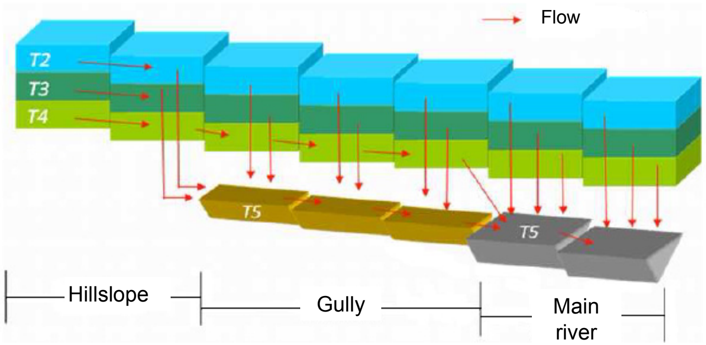
Introduction

- Our proposal: Combining simulations of the hydrological distributed model **TETIS** (Vélez et al, 2001; Francés et al, 2007) with the **MISP** technique (*Mutually Interactive State-Parameter Estimation)* (Todini, 1978)
  - Distributed models: advantage of simulating flows at any point throughout the spatial domain
  - **MISP** is a Kalman filter based algorithm, which:
    - Performs the state-parameter estimation of a discrete time dynamic system
    - Makes alternative use of two interacting filters in parallel, both with minimum variance

- Filtering Scheme:
  - Model predictions at an ungauged site are assumed as imperfect observations (as random variables)
  - Incorporating observations and simulations at a gauging station as on-line instrumental variables correlated with flow at the ungauged site
  - Log transformation of all input data to improve Gaussian assumptions of the K.F.
  - Inclusion of a cross covariance term in the covariance matrix of the measurement error (assumption of spatially correlated errors)

Hydrological model TETIS

- Developed at Technical University of Valencia since 1994
- Spatially distributed in regular cells
  - Reproduces spatial variability of hydrological processes
  - Reduces spatial scale effects with regard to lumped models
  - Allows to exploit all available physical and environmental spatial information

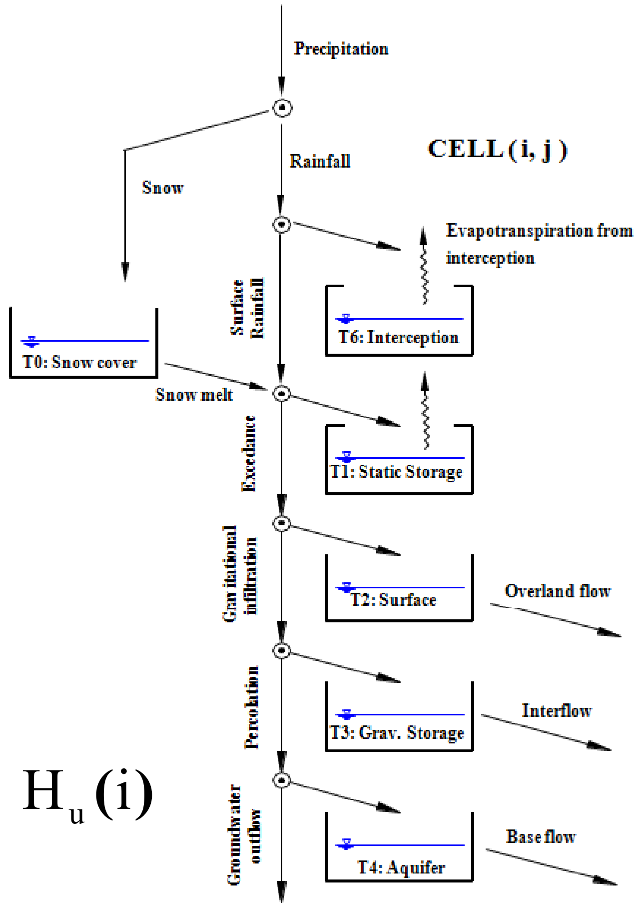


- Separate runoff modeling on slopes and channels
  - Each tank drains into the topographical downstream corresponding tank.
  - Drainage area thresholds defined for each type of tank
- Nonlinear channel routing scheme (geomorphologic kinematic wave).

Hydrological model TETIS

- Robust and parsimonious model:
  - Adequate simulation of initial state (includes balance at all times)
  - 6 storage tanks (state variables)
  - 5 external outflows (3 horizontal responses)

- Potential problem in distributed models:
  - Calibration of high number of parameters in each cell from an outflow hydrograph.
  - Proposed solution: Split Effective Parameters Structure (Frances et al, 2007):
    - Phase I: Parameter estimation from all physical and environmental information at each cell
    - Phase II: Global correction factors for each parameter map (effective parameters)



$H_u(i)$

$H_u(i) \cdot R_1$ → **Calibration**

Production parameters at each cell (v.7)

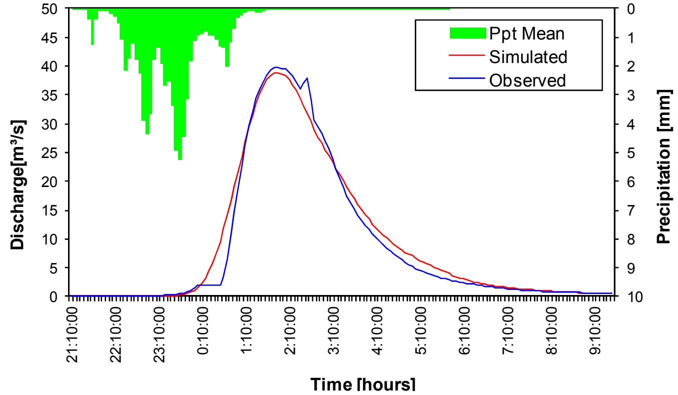
- Vegetation cover index: $\lambda(t)$ — *Crop factor*
- Maximum static capacity: H_u — *Initial abstractions + upper soil capillary capacity*

- Overland flow velocity: v — *Hill slope stationary velocity*
- Interflow velocity: k_{ss} — *Horizontal macropore upper soil permeability*
- Base flow velocity: k_b — *Upper aquifer permeability*

- Infiltration capacity: k_s — *Vertical upper soil permeability*
- Percolation capacity: k_p — *Vertical deep soil permeability*
- Underground losses capacity: k_{pp} — *Lower aquifer permeability*

Calibration Process

- Parameter estimation normally through comparison between simulated and observed values of some state variables
  - Traditionally: Discharge at the basin outlet



- Model performance assessment:
  - Graphic comparison
  - Statistical Objective Functions:
    - Balance Error $$BE(\%) = \frac{\sum (S_i - Q_i)}{\sum Q_i} \times 100$$
    - Root mean square Error $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_i - S_i)^2}$$
    - Efficiency Index (Nash-Sutcliffe) $$NSE = 1 - \frac{\sum (Q_i - S_i)^2}{\sum (Q_i - Q_m)^2}$$
- Automatic calibration of correction factors: SCE-UA (Duan et al., 1992; Sorooshian et al, 1993)

Mutually interactive parameter estimation (MISP)

- The first filter performs the minimum variance state estimation, given the parameters set :

$$\mathbf{x_t} = \mathbf{\Phi_{t/t-1}} \mathbf{x_{t-1}} + \mathbf{\Gamma_{t-1}} \mathbf{w_{t-1}}$$

$$\mathbf{z_t} = \mathbf{H_t} \mathbf{x_t} + \mathbf{v_t}$$

<u>where</u>:

$\mathbf{x}_t$: State vector (nx1)

$\mathbf{z}_t$: Measurement vector (mx1)

Φ: State transition matrix (nxn)

Γ, $\mathbf{H}$: Compatibility matrices

$\mathbf{v}_t$, $\mathbf{w}_t$: Normal independent processes

- The second one performs the minimum variance parameter estimation, given the state estimate, both at the present and previous time steps.

$$\theta_{\mathbf{t}} = \theta_{\mathbf{t-1}} + \mathbf{\Gamma^*_{t-1}} \mathbf{w^*_{t-1}}$$

$$\mathbf{z^*_t} = \mathbf{H^*_t} \theta_{\mathbf{t}} + \mathbf{v^*_t}$$

<u>where</u>: θ_t: Parameter vector (px1)

Mutually interactive parameter estimation (MISP)

- Equations required to accomplish a calculation cycle:

State update:

$$\mathbf{v}_t = \mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1} - \overline{\mathbf{v}}_{t-1}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}_t^T\left(\mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}_t^T + \mathbf{R}_{t|t}\right)^{-1}$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t\mathbf{v}_t$$

$$\mathbf{P}_{t|t} = \left(\mathbf{I} - \mathbf{K}_t\mathbf{H}_t\right)\mathbf{P}_{t|t-1}$$

Time update:

$$\hat{\mathbf{x}}_{t+1|t} = \Phi_{t+1|t}\hat{\mathbf{x}}_{t|t} + \Gamma_{t+1}\overline{\mathbf{w}}_t$$

$$\mathbf{P}_{t+1|t} = \Phi_{t+1|t}\mathbf{P}_{t|t}\Phi_{t+1|t}^T + \Gamma_{t+1}\mathbf{Q}_{t|t}\Gamma_{t+1}^T$$

Parameters update:

$$\nu_t^* = \mathbf{H}_t\mathbf{K}_t\nu_t$$

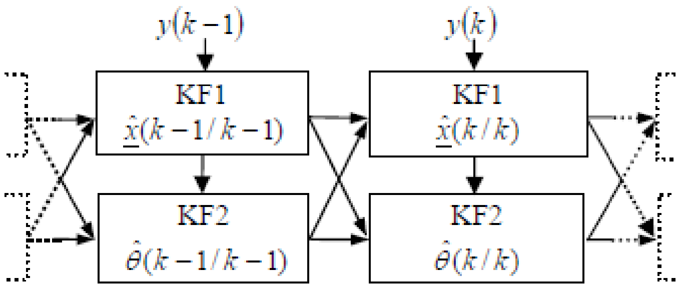
$$\mathbf{R}_t^* = \mathbf{H}_t\mathbf{K}_t\mathbf{C}_{t|t}^{\circ}\mathbf{K}_t^T\mathbf{H}_t^T$$

$$\mathbf{C}_t^{\circ *} = \mathbf{H}_t^*\mathbf{P}_{t|t-1}^*\mathbf{H}_t^{*T} + \mathbf{R}_{t|t}^*$$

$$\mathbf{K}_t^* = \mathbf{P}_{t|t-1}^*\mathbf{H}_t^{*T}\left(\mathbf{C}_t^{\circ *}\right)^{-1}$$

$$\hat{\theta}_{t+1|t} = \hat{\theta}_{t|t-1} + \mathbf{K}_t^*\nu_t^* + \Gamma_{t+1}^*\overline{\mathbf{w}}^*$$

$$\mathbf{P}_{t+1|t}^* = \mathbf{P}_{t|t-1}^* - \mathbf{K}_t^*\mathbf{C}_t^{\circ *}\mathbf{K}_t^{*T} + \Gamma_{t+1}^*\mathbf{Q}^*\Gamma_{t+1}^*$$



Mutually interactive parameter estimation (MISP)

- Equations required to accomplish a calculation cycle:

State update:

$$\mathbf{v}_t = \mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1} - \overline{\mathbf{v}}_{t-1}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t^T \left(\mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \mathbf{R}_{t|t} \right)^{-1}$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \mathbf{v}_t$$

$$\mathbf{P}_{t|t} = \left(\mathbf{I} - \mathbf{K}_t \mathbf{H}_t \right) \mathbf{P}_{t|t-1}$$

Time update:

$$\hat{\mathbf{x}}_{t+1|t} = \Phi_{t+1|t} \hat{\mathbf{x}}_{t|t} + \Gamma_{t+1} \overline{\mathbf{w}}_t$$

$$\mathbf{P}_{t+1|t} = \Phi_{t+1|t} \mathbf{P}_{t|t} \Phi_{t+1|t}^T + \Gamma_{t+1} \mathbf{Q}_{t|t} \Gamma_{t+1}^T$$

Parameters update:

$$\nu_t^* = \mathbf{H}_t \mathbf{K}_t \nu_t$$

$$\mathbf{R}_t^* = \mathbf{H}_t \mathbf{K}_t \mathbf{C}_{t|t}^{\circ} \mathbf{K}_t^T \mathbf{H}_t^T$$

$$\mathbf{C}_t^{\circ *} = \mathbf{H}_t^* \mathbf{P}_{t|t-1}^* \mathbf{H}_t^{*T} + \mathbf{R}_{t|t}^*$$

$$\mathbf{K}_t^* = \mathbf{P}_{t|t-1}^* \mathbf{H}_t^{*T} \left(\mathbf{C}_t^{\circ *} \right)^{-1}$$

$$\hat{\theta}_{t+1|t} = \hat{\theta}_{t|t-1} + \mathbf{K}_t^* \nu_t^* + \Gamma_{t+1}^* \overline{\mathbf{w}}^*$$

$$\mathbf{P}_{t+1|t}^* = \mathbf{P}_{t|t-1}^* - \mathbf{K}_t^* \mathbf{C}_t^{\circ *} \mathbf{K}_t^{*T} + \Gamma_{t+1}^* \mathbf{Q}^* \Gamma_{t+1}^*$$

- The estimation of the state vector provided by the first filter ($\hat{\mathbf{x}}_{t|t}$) is used as observations, in order to estimate the parameter vector θ.
- Optimality is reached after a series of runs through the historical data, as model residuals become less and less correlated.
- At the same time, it is possible to estimate the unknown noise statistics: $\overline{\mathbf{w}}, \overline{\mathbf{v}}, \mathbf{Q}, \mathbf{R}$

MISP implementation

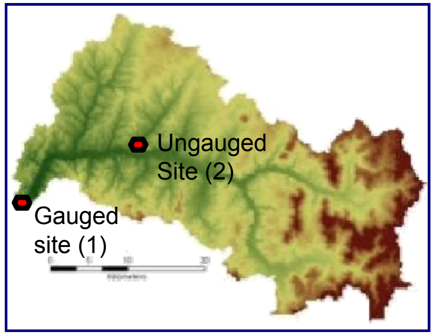
- System equations (matrix form):

$$\mathbf{x}_t = \Phi_{t/t-1}\mathbf{x}_{t-1} + \Gamma_{t-1}\mathbf{w}_{t-1} \Rightarrow \begin{bmatrix} Q_{2(t)} \\ Q_{2(t-1)} \\ Q_{1(t)} \\ Q_{1(t-1)} \\ Q_{s1(t)} \\ Q_{s1(t-1)} \end{bmatrix} = \begin{bmatrix} \alpha_1^1 & \alpha_2^1 & \beta_1^1 & \beta_2^1 & \chi_1^1 & \chi_2^1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \beta_1^2 & \beta_2^2 & \chi_1^2 & \chi_2^2 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \chi_1^3 & \chi_2^3 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Q_{2(t-1)} \\ Q_{2(t-2)} \\ Q_{1(t-1)} \\ Q_{1(t-2)} \\ Q_{s1(t-1)} \\ Q_{s1(t-2)} \end{bmatrix} + \Gamma_{t-1} \begin{bmatrix} w_{2(t-1)} \\ w_{1(t-1)} \\ w_{s1(t-1)} \end{bmatrix}$$

$$\mathbf{z}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t \Rightarrow \begin{bmatrix} z_{1(t)} \\ z_{2(t)} \\ z_{3(t)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} Q_{2(t)} \\ Q_{2(t-1)} \\ Q_{1(t)} \\ Q_{1(t-1)} \\ Q_{s1(t)} \\ Q_{s1(t-1)} \end{bmatrix} + \begin{bmatrix} v_{1(t)} \\ v_{2(t)} \\ v_{3(t)} \end{bmatrix}$$

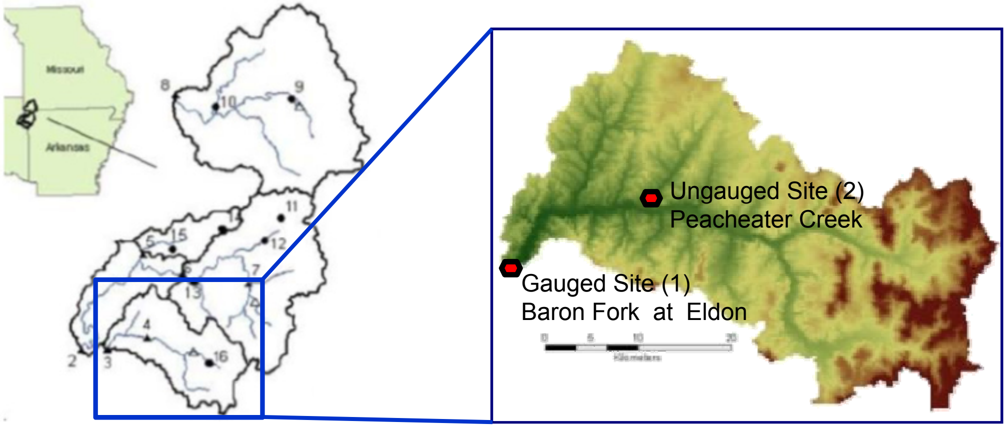
- Parameter equation:

$$\theta_t = \theta_{t-1} + \Gamma_{t-1}^*\mathbf{w}_{t-1}^*$$
$$\mathbf{z}_t^* = \mathbf{H}_t^*\theta_t + \mathbf{v}_t^* \Rightarrow \theta_t = \begin{bmatrix} \alpha_1^1 \\ \alpha_2^1 \\ \vdots \\ \chi_1^3 \\ \chi_2^3 \end{bmatrix}_t$$



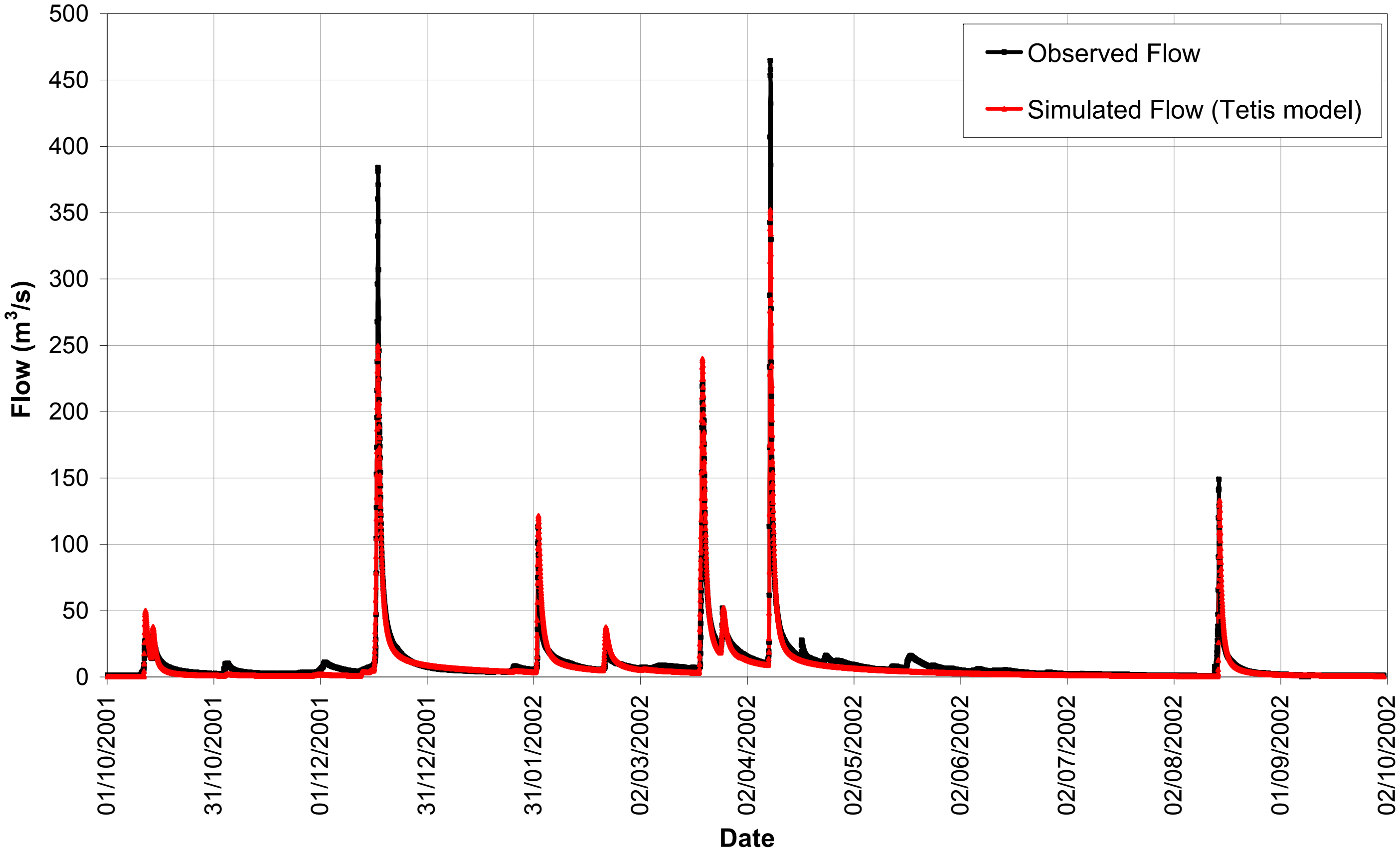
Case study: simulation at Baron Fork

- Distributed Model Intercomparison Project (DMIP2), NOAA/NWS.
  - Series of experiments to guide NOAA/NWS research into advanced hydrologic models for river and water resources forecasting.
- Study basins: Baron Fork river and Peacheater Creek (tributary)
  - Complete description (Smith et al, 2004)
  - Availability of cartographic information of physical and environmental parameters
  - Concurrent time series (1995-2002) of discharges, radar precipitation (NEXRAD), temperature and ETP from Reanalysis (NCEP-NCAR)
- Basin areas
  - Baron Fork: 795 km^2
  - Peacheater: 65 km^2



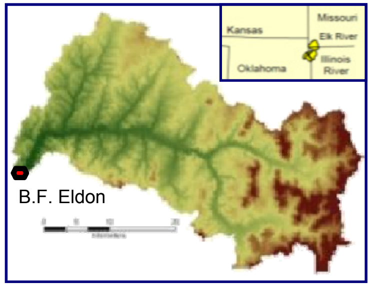
TETIS model calibration

- Calibration results, Baron Fork at Eldon (oct/2001-sep/2002)



Statistical Index

BE (%):	-14.6
RMSE (m^3/s):	6.61
NSE:	0.91



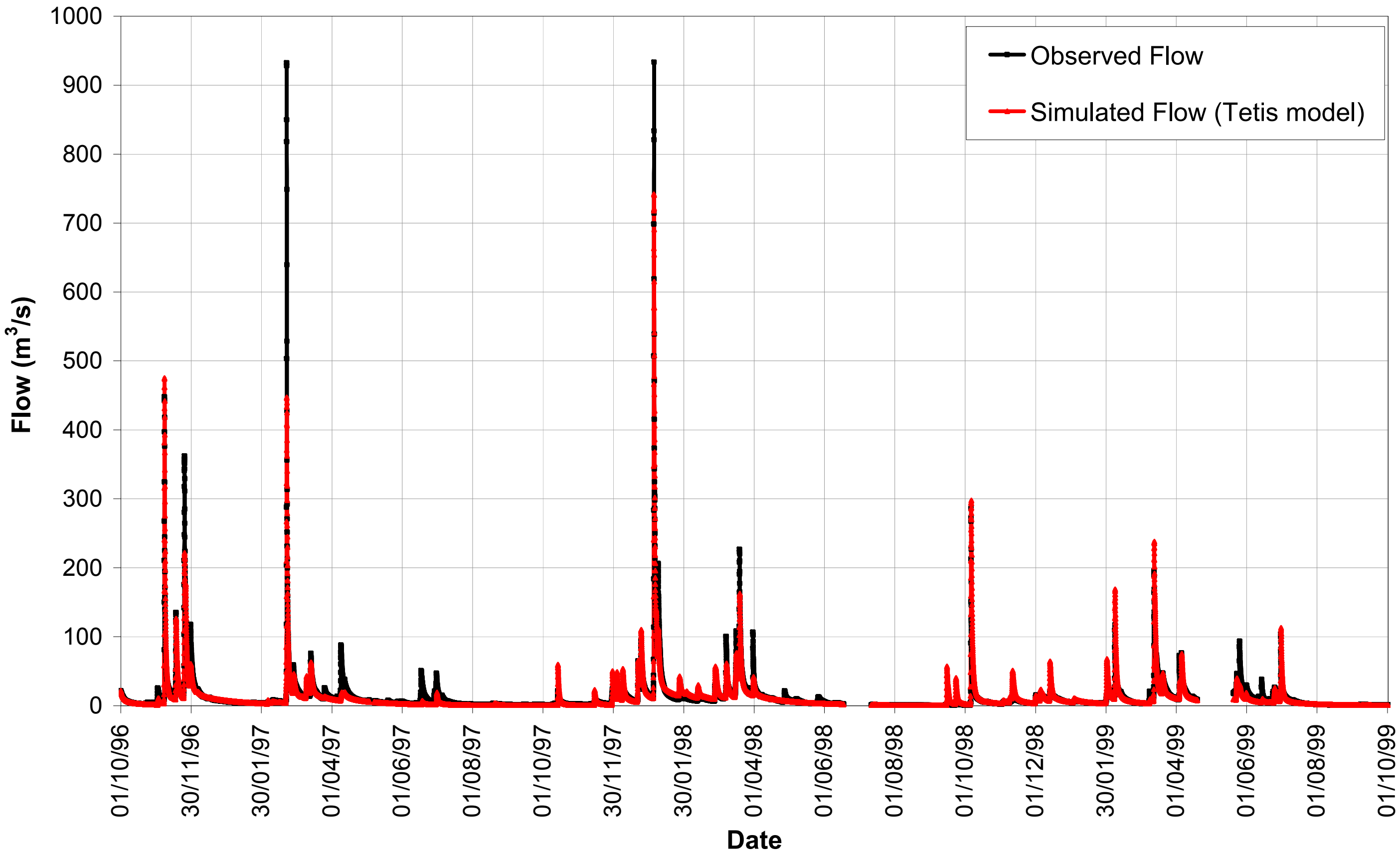
TETIS model calibration

- Calibrated correction factors (cell parameters) for the DMIP2 case study

Cell parameter	TETIS Param. decomposition	Calibrated values of R
Static storage capacity	$h_u^* = R_1 \cdot h_u$	0.3831
Index of monthly vegetation density coverage (i=1, 2, … , 12) for ET.	$\lambda_i^* = R_2 \cdot \lambda_i$	0.7232
Infiltration capacity at constant rate	$k_s^* = R_3 \cdot k_s$	1.2594
Direct Runoff hillslope velocity (linear reservoir)	$u_{OF}^* = R_4 \cdot u_{OF}$	2.0
Percolation capacity at constant rate	$k_p^* = R_5 \cdot k_p$	0.3329
Interflow velocity (linear reservoir)	$k_{if}^* = R_6 \cdot k_s$	30.00
Groundwater loss capacity at constant rate	$k_{pp}^* = R_7 \cdot k_p$	0.0
Baseflow velocity (linear reservoir)	$k_{bf}^* = R_8 \cdot k_p$	114.43
Streamflow velocity in the river network	$u_{CF}^* = R_9 \cdot u_{CF}$	0.2009

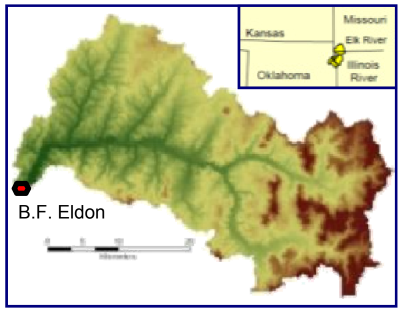
TETIS model validation

- Temporal validation results, Baron Fork at Eldon (oct/1996-sep/2001)



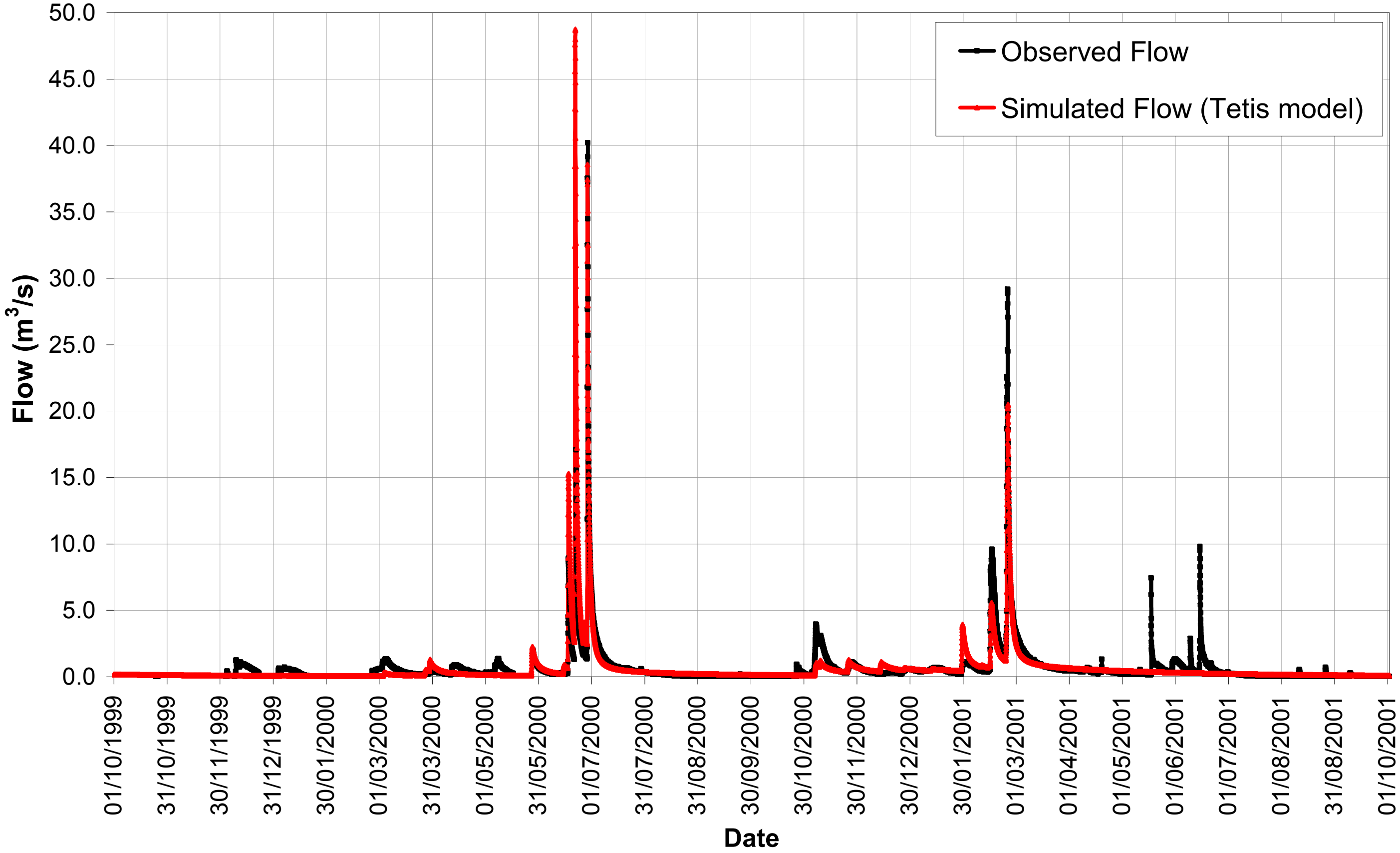
Statistical Index

BE (%):	-11.2
RMSE (m^3/s):	14.89
NSE:	0.83



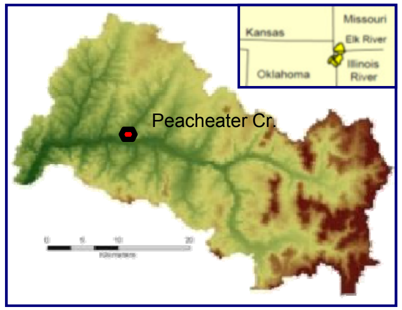
TETIS model validation

- Spatial validation results, (Peacheater Cr. At Christie (oct/1996-sep/2002)



Statistical Index

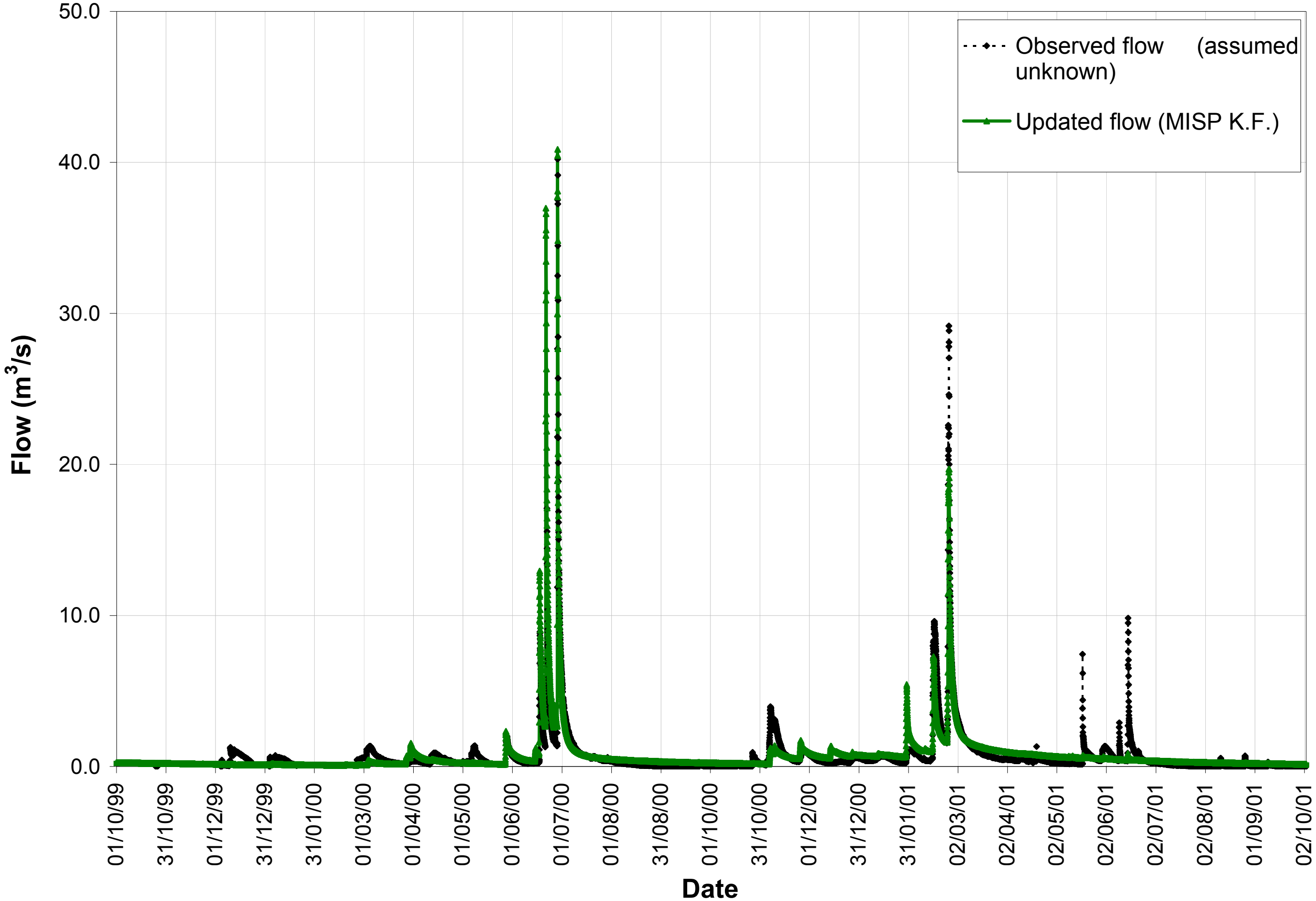
BE (%):	9.3
RMSE (m^3/s):	0.88
NSE:	0.67



EGU Leonardo Conference Series on the Hydrological Cycle
FLOODS IN 3D: PROCESSES, PATTERNS, PREDICTIONS
Bratislava, Slovakia, 23-25 November 2011

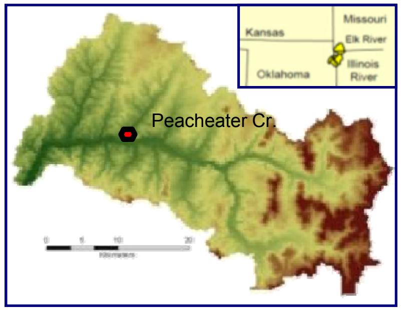
Results, application of MISP

- MISP results, Peacheater Cr. (oct/1996-sep/2002)



Statistical Index

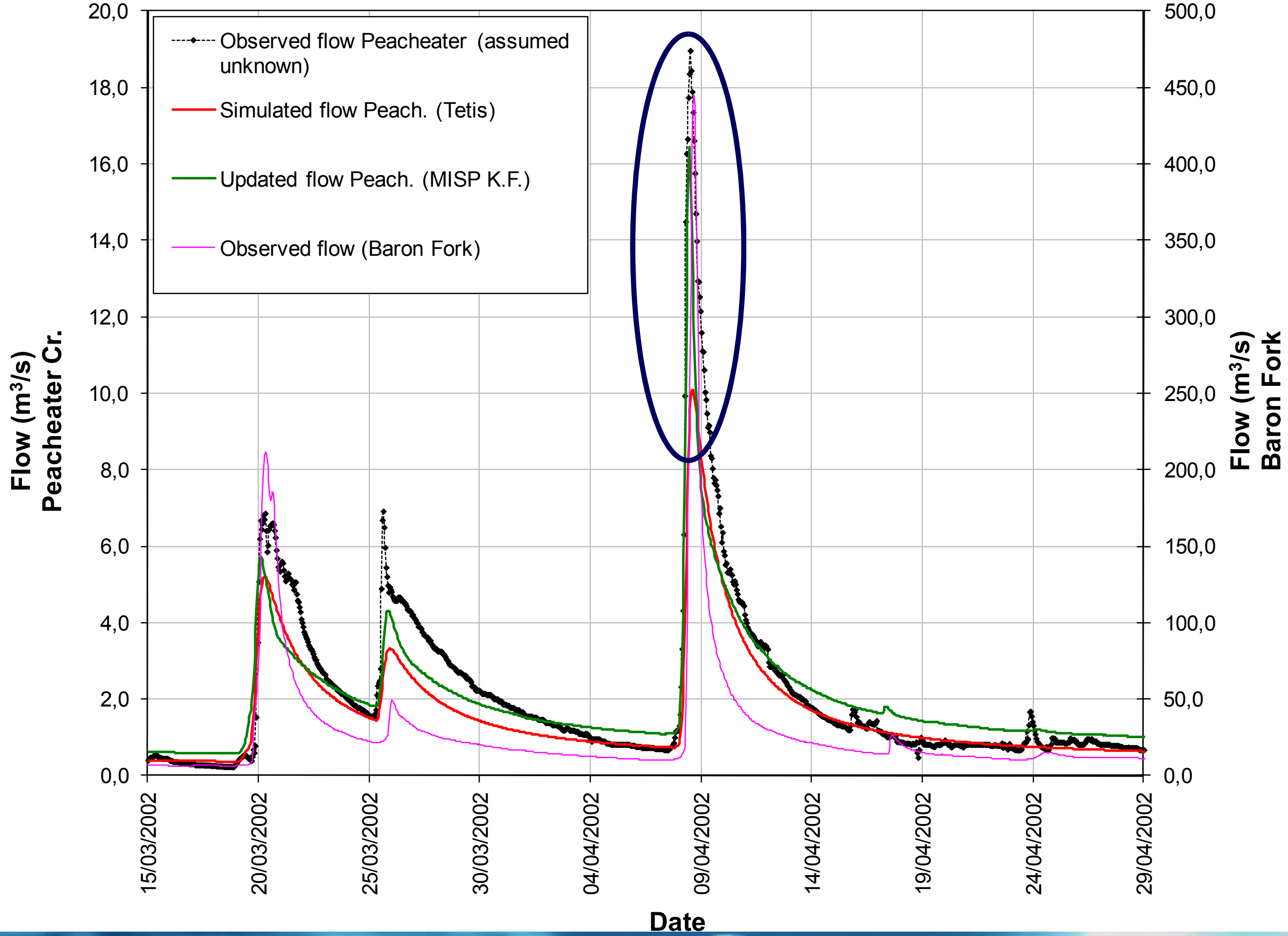
BE (%)	9.3
RMSE (m^3/s)	0.83
NSE	0.71

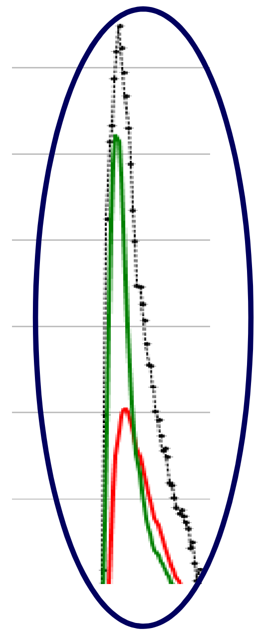


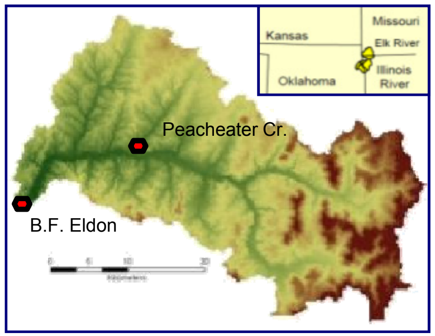
EGU Leonardo Conference Series on the Hydrological Cycle
FLOODS IN 3D: PROCESSES, PATTERNS, PREDICTIONS
Bratislavia, Slovakia, 23-25 November 2011

Results, application of MISP

- Peacheater Creek (15/03/2002 – 29/04/2002)



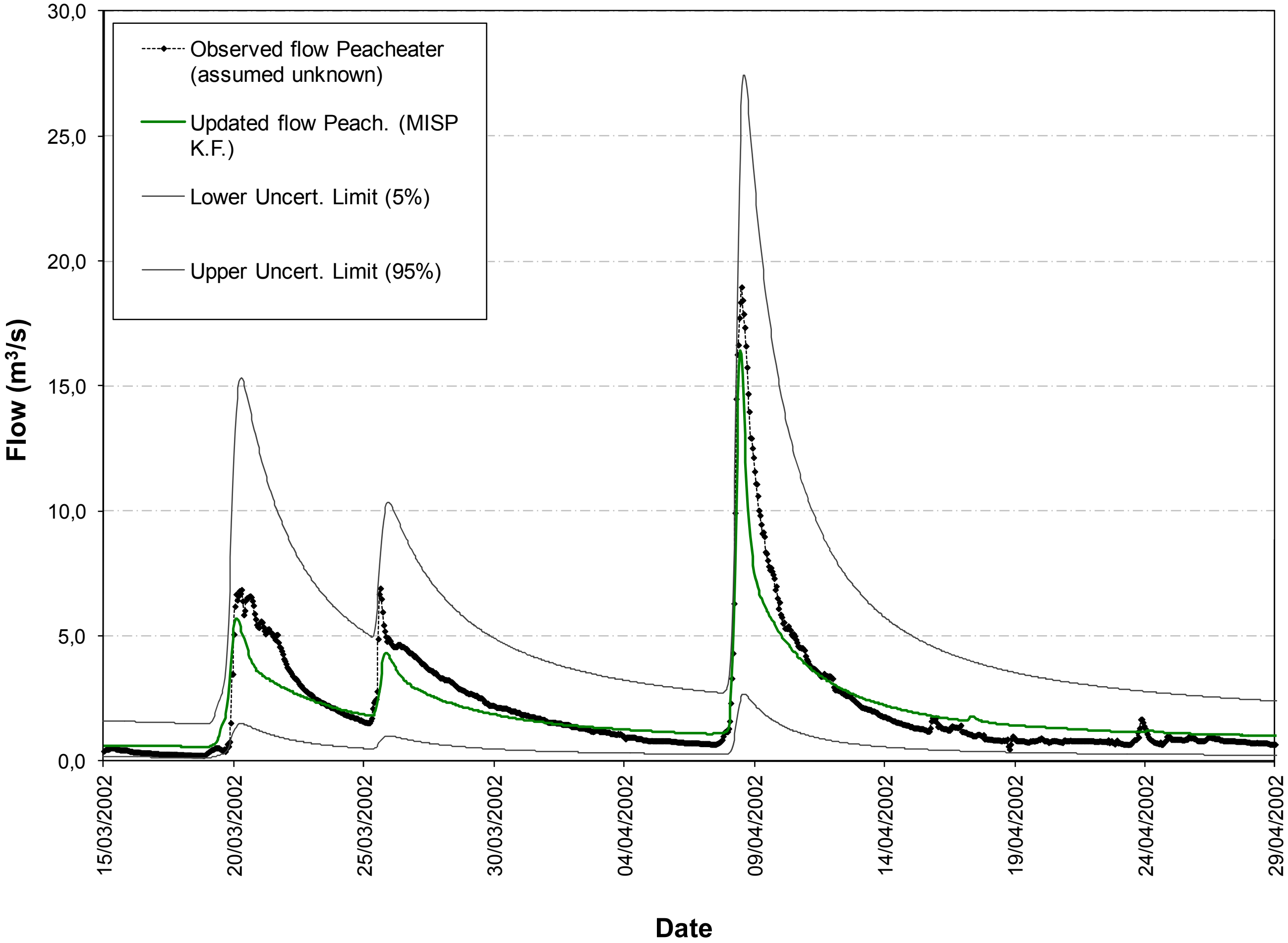


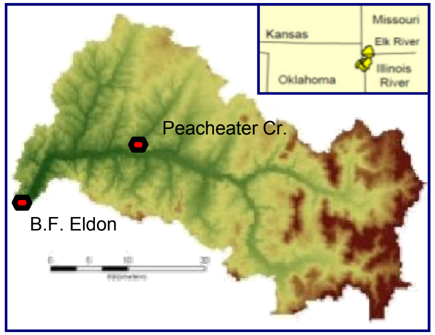


EGU Leonardo Conference Series on the Hydrological Cycle
FLOODS IN 3D: PROCESSES, PATTERNS, PREDICTIONS
Bratislava, Slovakia, 23-25 November 2011

Results, application of MISP

- Peacheater Creek, uncertainty bounds (15/03/2002 – 29/04/2002)

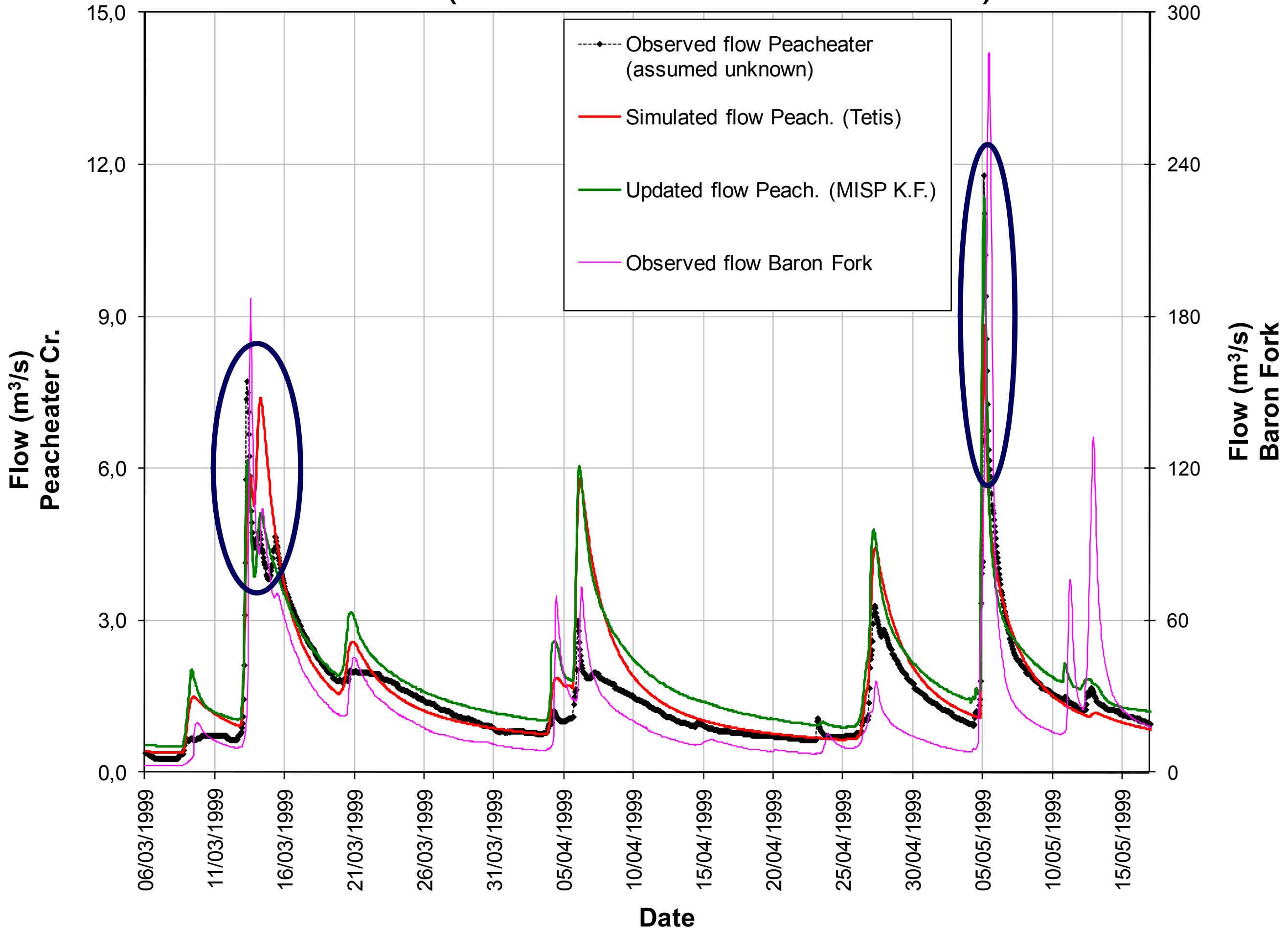


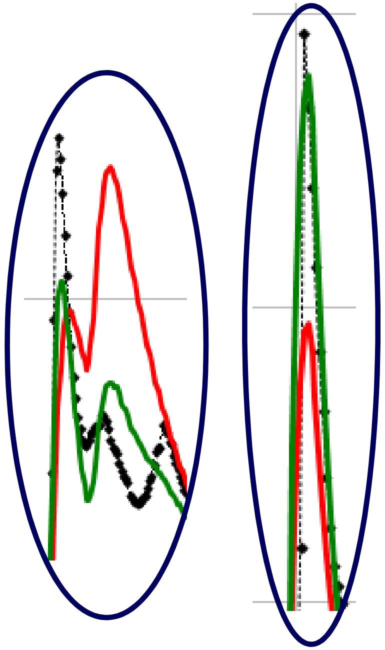


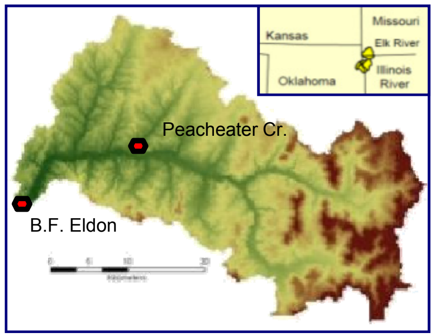
EGU Leonardo Conference Series on the Hydrological Cycle
FLOODS IN 3D: PROCESSES, PATTERNS, PREDICTIONS
Bratislava, Slovakia, 23-25 November 2011

Results, application of MISP

- Peacheater Creek (06/03/1999 – 17/05/1999)



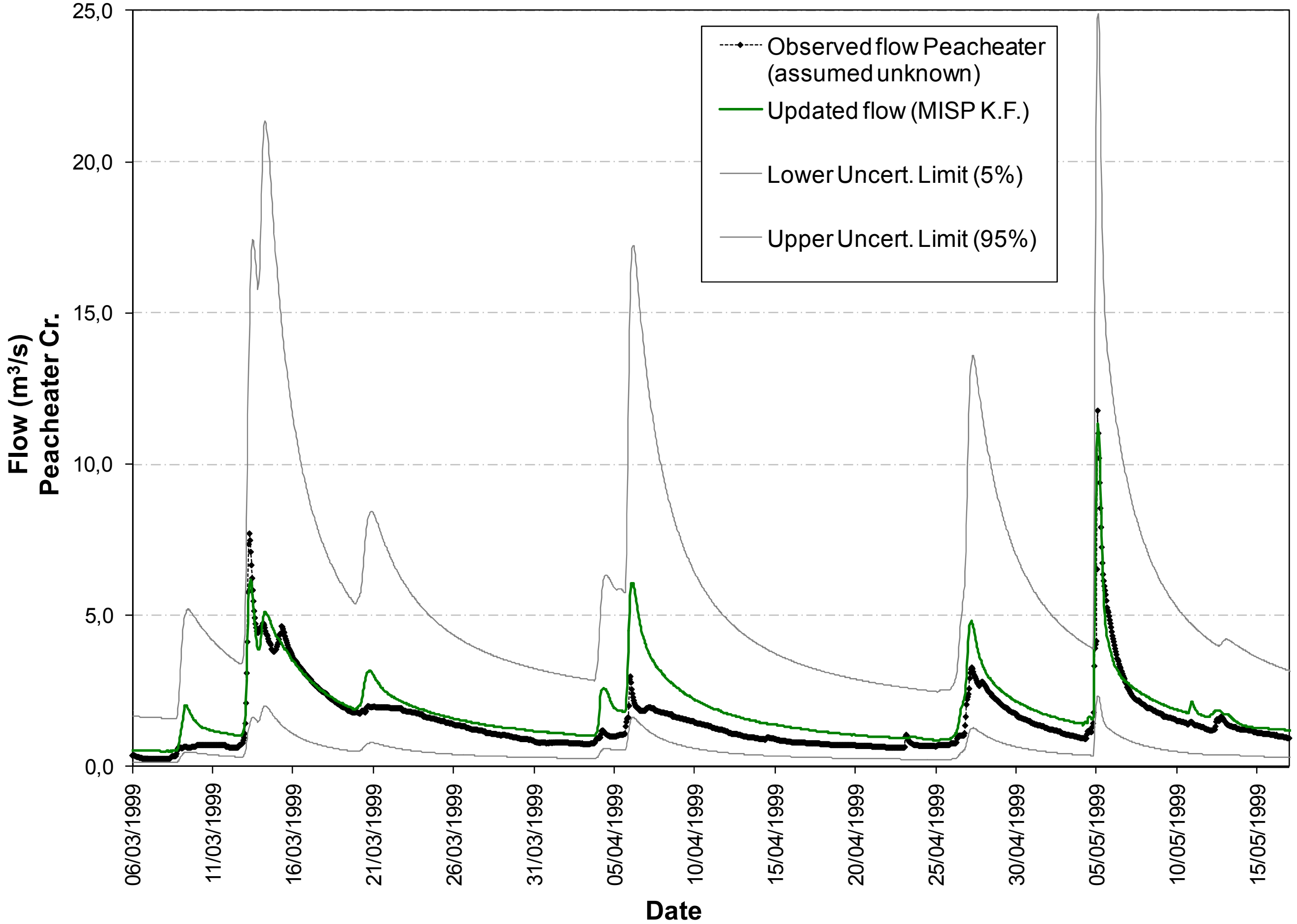


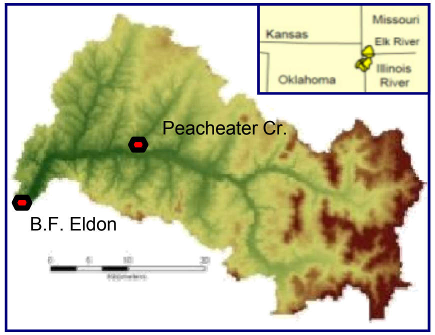


EGU Leonardo Conference Series on the Hydrological Cycle
FLOODS IN 3D: PROCESSES, PATTERNS, PREDICTIONS
Bratislava, Slovakia, 23-25 November 2011

Results, application of MISP

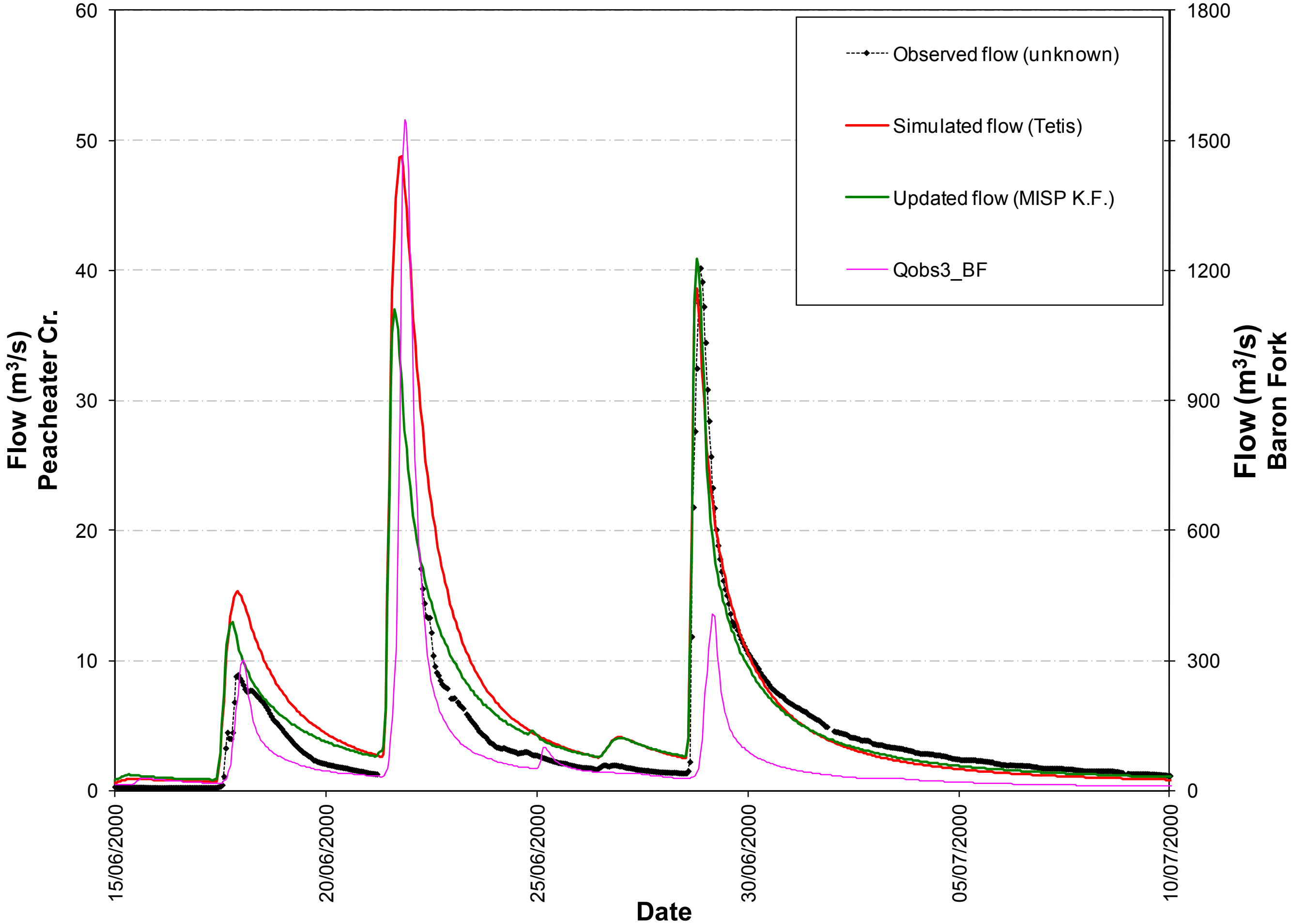
- Peacheater Creek, uncertainty bounds (06/03/1999 – 17/05/1999)

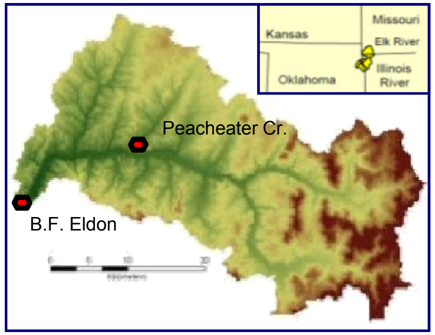




Results, application of MISP

- Peacheater Creek (15/06/2000 – 10/07/2000)

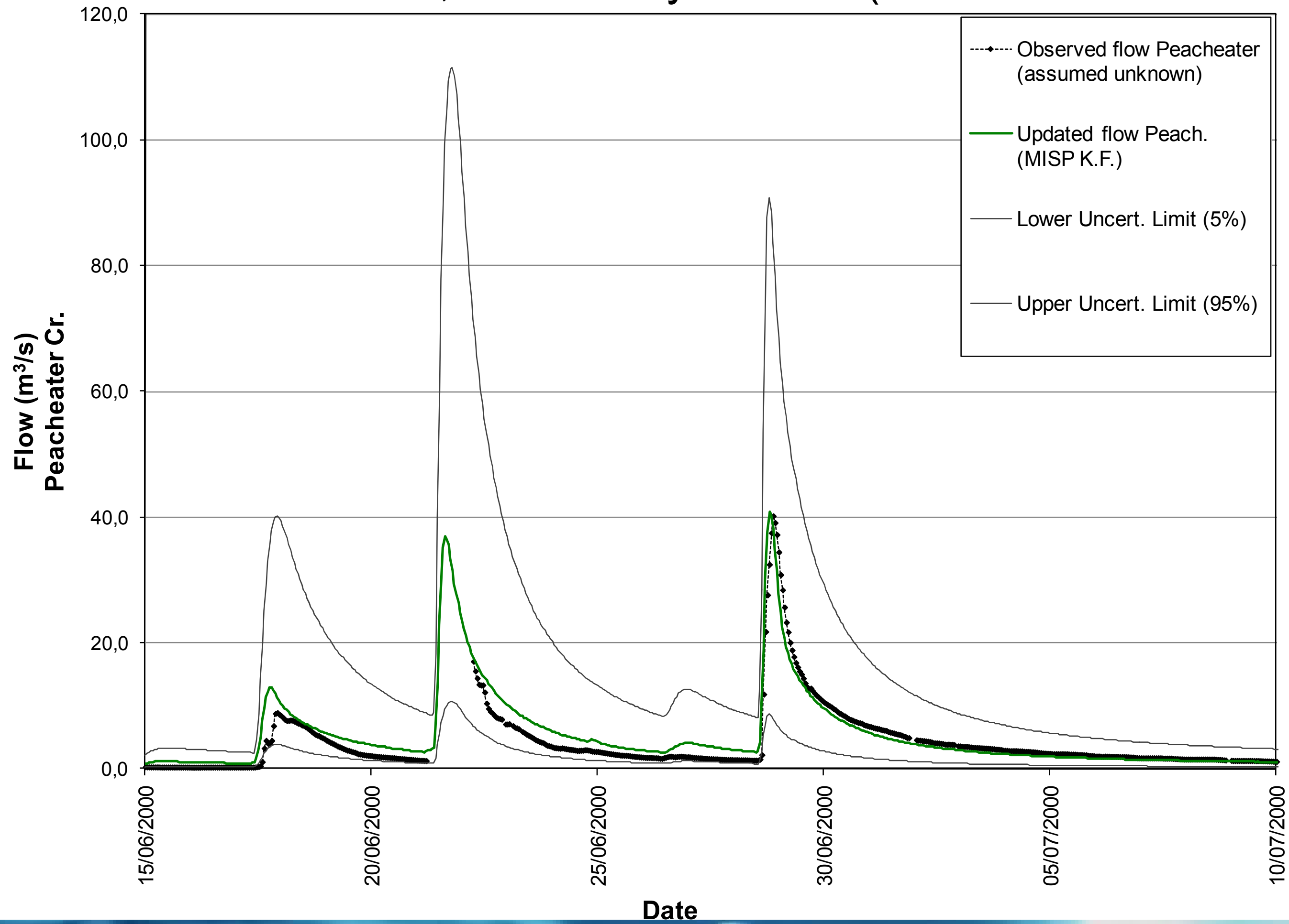


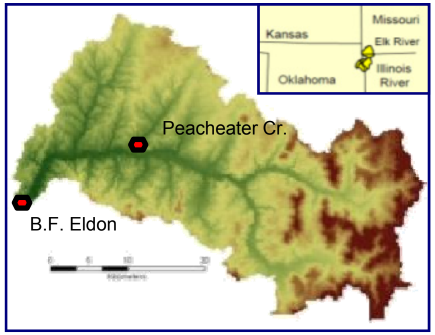


EGU Leonardo Conference Series on the Hydrological Cycle
FLOODS IN 3D: PROCESSES, PATTERNS, PREDICTIONS
Bratislava, Slovakia, 23-25 November 2011

Results, application of MISP

- Peacheater Creek, uncertainty bounds (15/06/2000 – 10/07/2000)





Conclusions

- A methodology based on the Kalman filter MISP algorithm is presented, aimed to improve predictions performed with a distributed hydrological model at ungauged sites and to estimate the related predictive uncertainty.
- The observed and simulated discharges at the basin outlet are used as instrumental variables in the Kalman filter implementation.
- Importance of including a cross covariance error term in matrix **R**.
- The results of the proposed approach are considered very satisfactory. The NSE was improved from 0.67 to 0.71 for the Peacheater Creek (assumed as ungauged site).
- The Kalman filter based **MISP** approach allows assess the predictive uncertainty associated to the model prediction (simulation or forecasting mode).
- The uncertainty band is strongly affected by model performance, and is expected that any improvement of the model would reduce predictive uncertainties.

Thank you for your attention!

Contact:

Juan Camilo Múnera

E-mail: juamues1@doctor.upv.es
juancmunera@yahoo.es

Technical University of Valencia (Spain)
Research Institute of Water Engineering and Environment
Research Group of Hydrological and Environmental Modeling
http://lluvia.dihma.upv.es