On how to avoid input and structural uncertainties corrupt the inference of hydrological parameters using a Bayesian framework

M.R. Hernández and **F. Francés**

Research Group on Hydrologic and Environmental Modeling
Research Institute of Water and Environmental Engineering
***Universitat Politècnica de València**, Spain*

European Geosciences Union
General Assembly 2015
Vienna | Austria | 12 - 17 April 2015

Introduction (1/2)

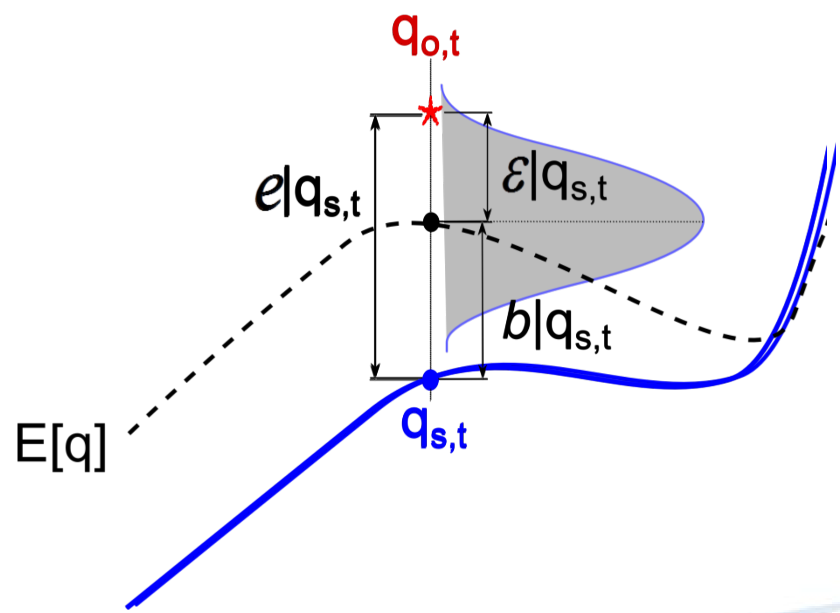
- **Problem:** Hydrological models provide extrapolations or predictions, which are not lacking of uncertainty
  - Model State Variables, as streamflow, do not match observations

$$q_{o,t} = q_{s,t} + e \mid q_{s,t}$$

- We need to **model the 2 components of the error term**:

$$q_{o,t} = q_{s,t} + b \mid q_{s,t} + \varepsilon \mid q_{s,t}$$

$$q_{o,t} = E_{q,t}\left(\boldsymbol{\theta}_h, \boldsymbol{\theta}_e, \tilde{\mathbf{X}}_{1:t}, \tilde{\mathbf{S}}_0\right) + \varepsilon \mid q_{s,t}$$



iiama

Introduction (2/2)

- **Classical Approach for modeling the Error Term**
  - **equivalent to Std. Least Squares calibration**
    - This method does not account for Bias: $E_{q,t}\left(\boldsymbol{\theta}_h, \boldsymbol{\theta}_e, \tilde{\mathbf{X}}_{1:t}, \tilde{\mathbf{S}}_0\right) = q_{s,t}$
    - Considers Errors are serially uncorrelated (iid) - White Noise
    - With Gaussian Distribution
    - Constant Conditional Variances (Homoscedastic errors)
- **Errors in Hydrology** do not satisfy the **SLS** hypothesis
  - **Causes** are mainly the Input errors and an unsuitable H. Model structure
  - **Consequences**
    - Biased or corrupted **parameter values** produce the ***Divergence Phenomenon*** and a **loss of physical meaning**
    - An incorrect estimation of the **predictive uncertainty**

Aims of this research

- 1- Inferring a **Specific Error Model** that best fits Hydrological Model Errors
  - **Inference must be a JOINT INFERENCE to avoid biased parameters in both models**
- 2- Compare Performance of **SLS vs Specific Error Model**
  - Comparison Criteria
    - Fulfillment of **Error Model Hypothesis**
    - **Performance of Prediction** in Validation based on NSE, RMSE, and VE% indexes
    - **Reliability of the Predictive Distribution** in Validation
    - **Assessment of Model Divergence Phenomenon**, is to say, the deterioration of the H. Model performance between Calibration and Validation
    - **Physical meaning** of hydrological model parameter values

iiama

Error Model description (1/3)

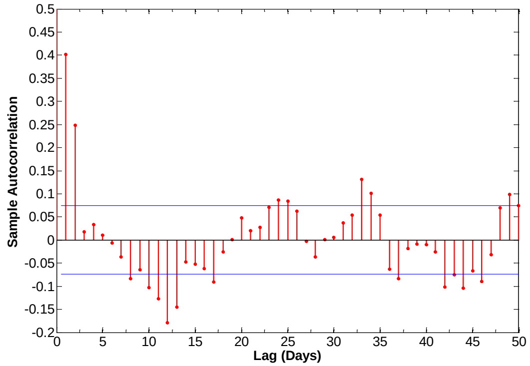
- **Research follows a Formal Bayesian Procedure**
  - Main target is Posterior of Hydrologic and Error parameters

$$\underbrace{p(\theta_h, \theta_e | \mathbf{q_o})} \propto \underbrace{p(\mathbf{q_o} | \{\theta_h, \theta_e\})} \; \underbrace{p(\theta_h, \theta_e)}$$

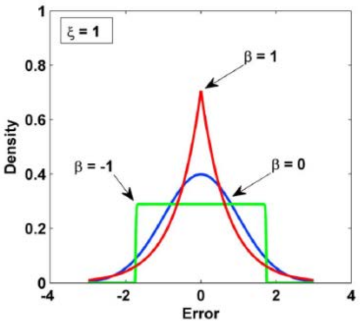
Posterior sampled with the **MCMC** algorithm ***DREAM_ZS*** [*Ter Braak and Vrugt (2008)*]

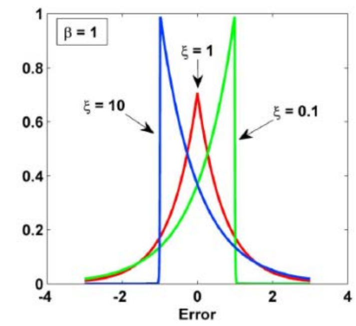
  - Non-Informative (Uniform) Priors for the inferred parameters
  - Developed Formal Likelihood function is based on:

Modeling the Errors Dependence through an **AR(p)** model



Modeling innovations through the flexible Skew Exponential Distribution (**SEP**)





[*Schoups and Vrugt (2010)*]

iiama

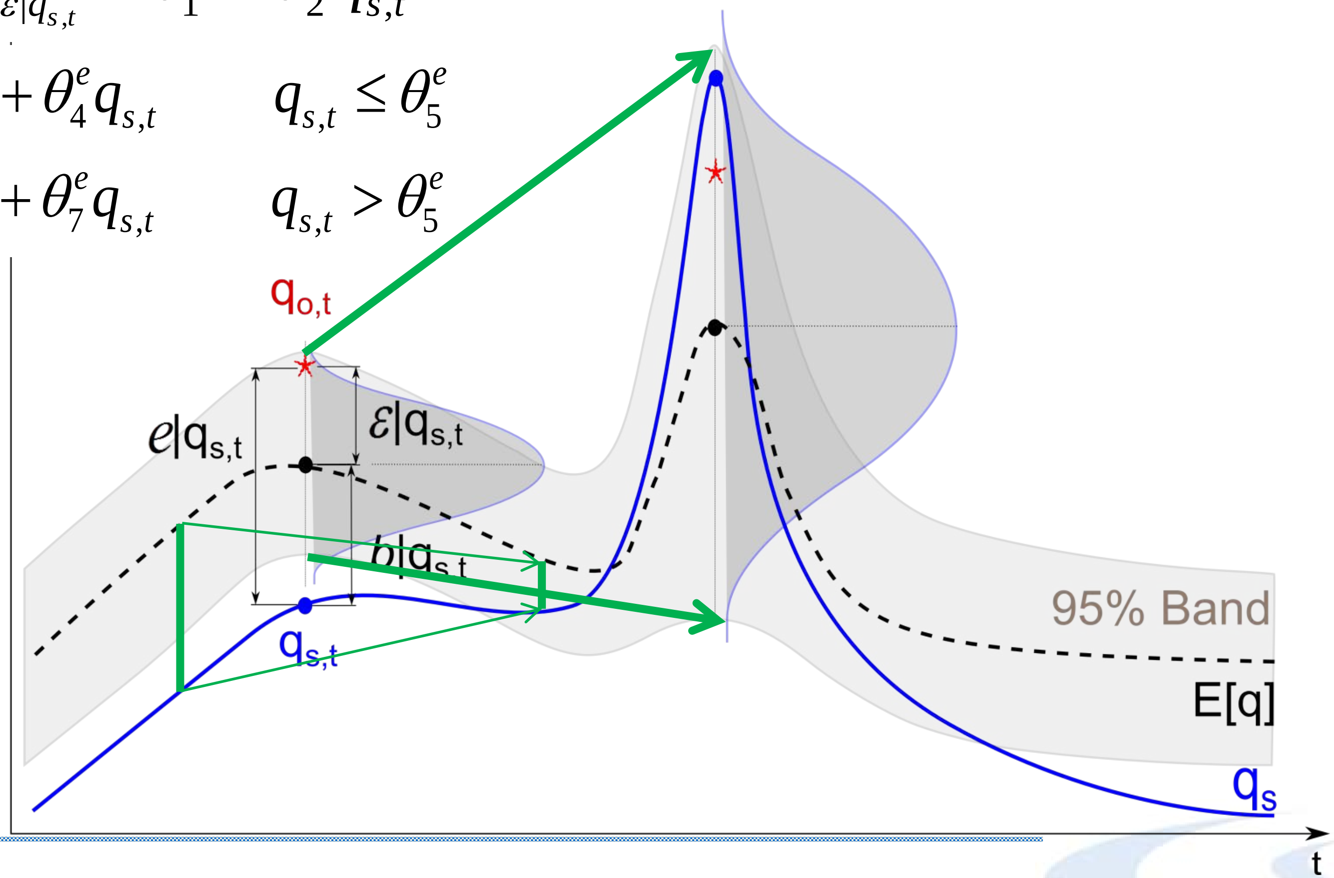
Error Model description (2/3)

❑ Introducing the Time-Variation in the Error Model

- Modeling **Variance & Bias** of the Error Conditional Distributions

Variance $\sigma_{e|q_{s,t}} = \sigma_{\varepsilon|q_{s,t}} = \theta_1^e + \theta_2^e q_{s,t}$

Bias $\begin{cases} b_{e|q_{s,t}} = \theta_3^e + \theta_4^e q_{s,t} & q_{s,t} \le \theta_5^e \\ b_{e|q_{s,t}} = \theta_6^e + \theta_7^e q_{s,t} & q_{s,t} > \theta_5^e \end{cases}$



iiama

Error Model description (3/3)

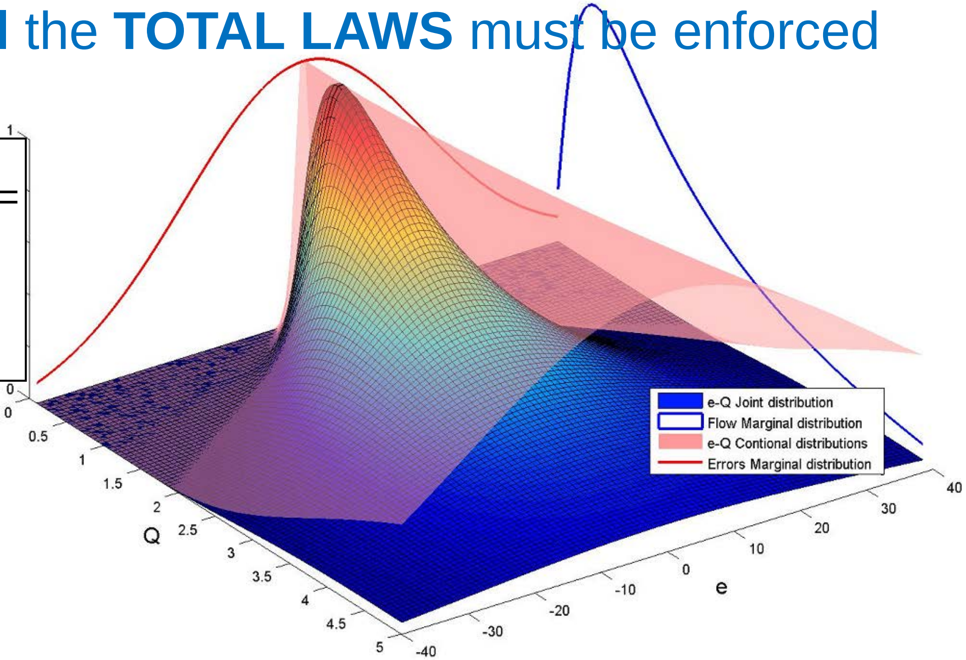
- **Parameters of Variance & Bias functions are not free:**
  - **Why?**
    - **Marginal and Conditional Error Distributions** are linked by Total Variance Law (**TVL**) and Total Expectation Law (**TEL**)
    - For the correct implementation of the **JOINT INFERENCE with a Time-Varying Error Model** the **TOTAL LAWS** must be enforced

TVL

$$V(\mathbf{e}) = E_Q\left[V\left(e|q_{s,t}\right)\right] + V_Q\left[E\left(e|q_{s,t}\right)\right] = \\ = E_Q\left[V\left(\varepsilon|q_{s,t}\right)\right] + V_Q\left[\,b|q_{s,t}\right]$$

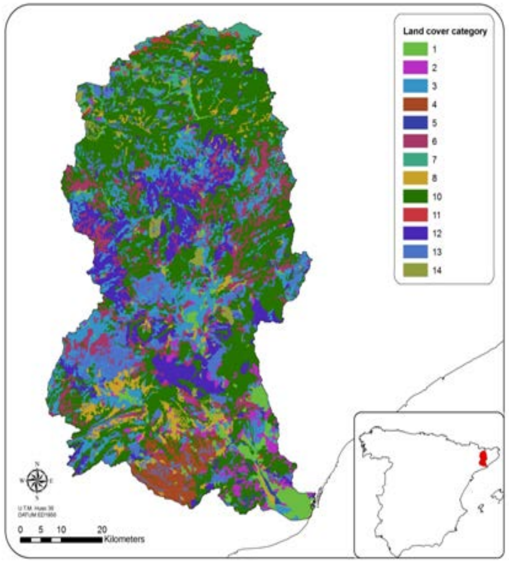
TEL

$$E(\mathbf{e}) = E_Q\left[E\left(e|q_{s,t}\right)\right] = E_Q\left[b\,|\,q_{s,t}\right]$$

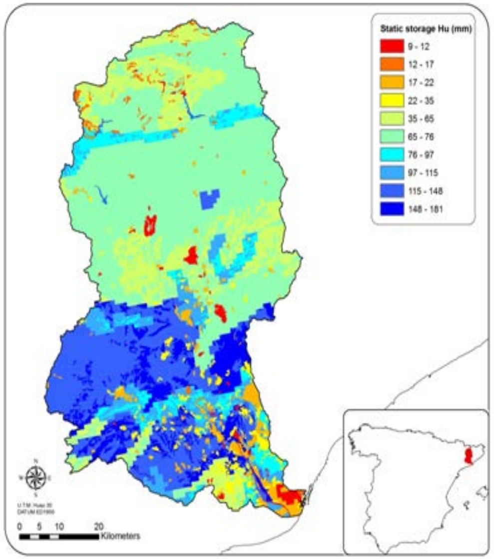


Case Study

- **Distributed Hydrological Model (on a Spanish humid catch.)**
  - **TETIS** **http://lluvia.dihma.upv.es/EN/software/software.html**
  - Effective Parameter Structure divided in two parts:
    - An estimated Value in each cell setting-up the **Parameter Maps**
    - **Regularization Function**: **Global calibrated correction factor** applied to each parameter map



x F1

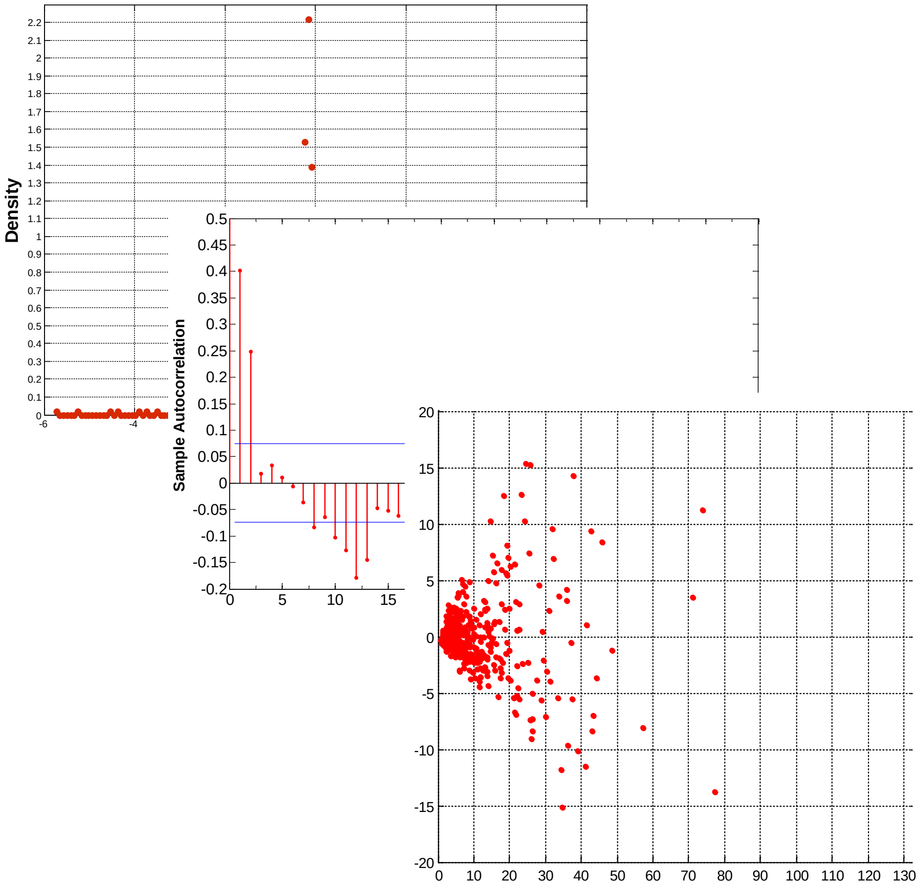


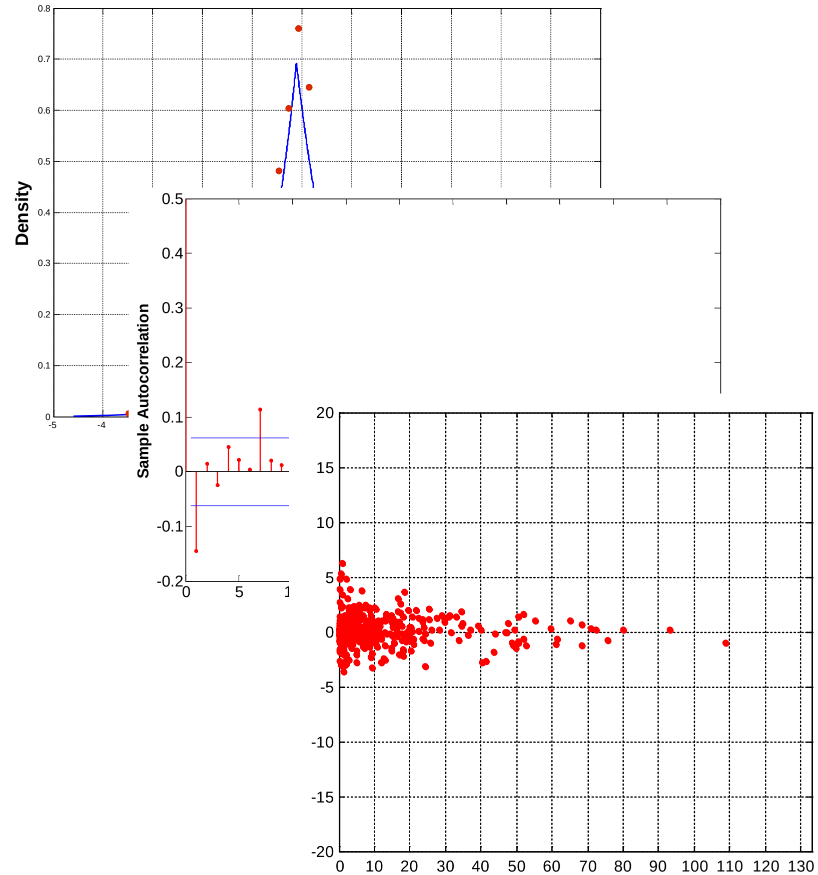
x F2 **.......**

Fulfillment of Error Models hypothesis

Normality, Independence, Homoscedasticity

SLS **EM2**





iiama

Prediction Performance in Validation

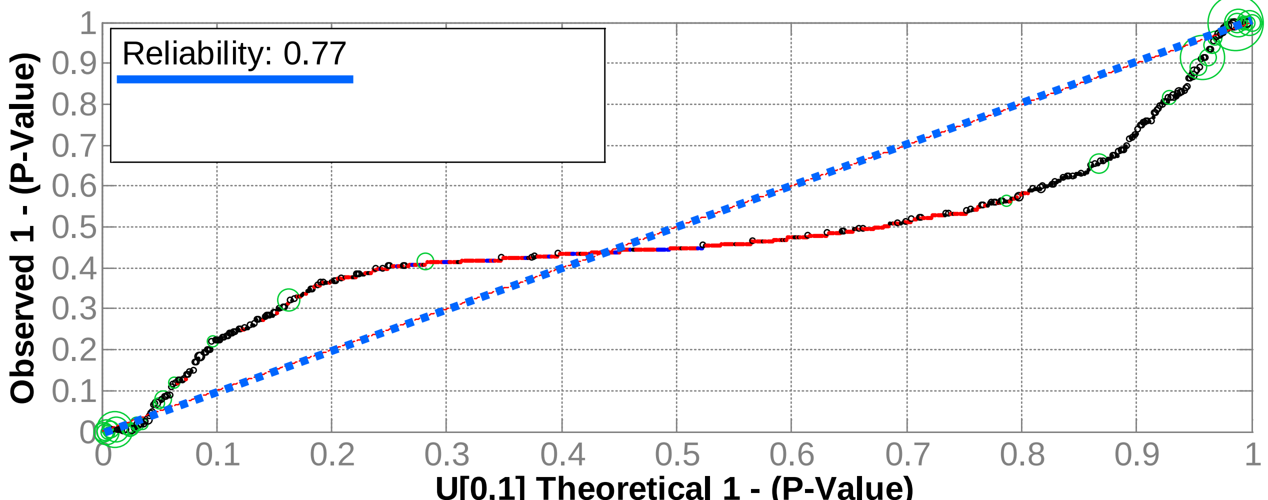
- Comparing the **performance indexes for Prediction**, not for the Simulation, albeit with SLS inference both are the same, …

		SLS		EM2	
		CALIB	VALID	CALIB	VALID
HYDRO MODEL	NSE	0.93	0.86	0.74	0.72
	RMSE	2.62	3.48	5.00	4.99
	ErrVol (%)	2.40	-4.5	9.90	2.70
PREDICTION	NSE			0.91	0.85
	RMSE			2.92	3.60
	ErrVol (%)			0.01	-3.70

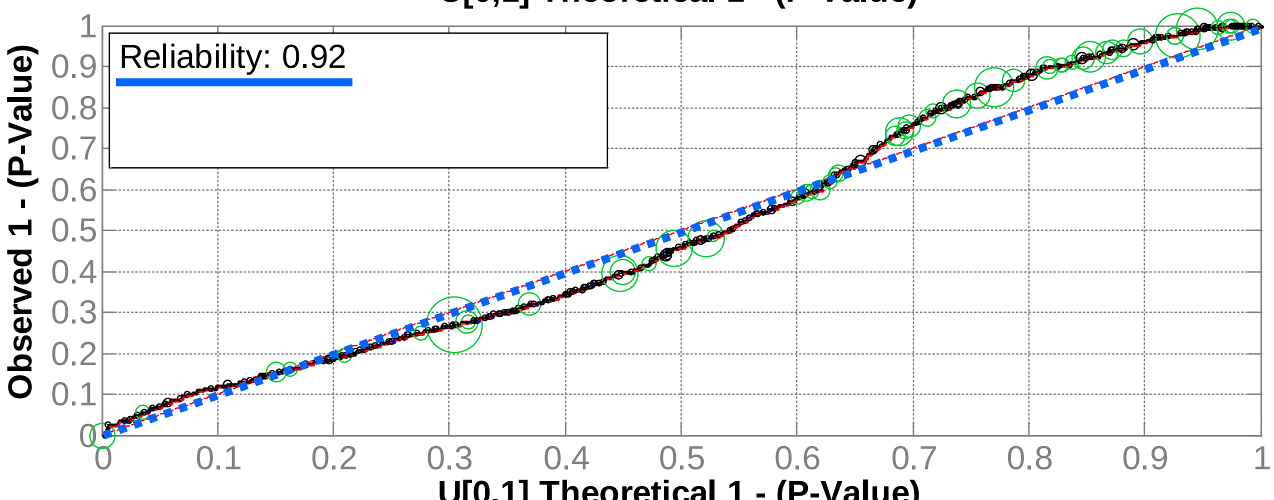
Predictive Distribution in Validation

Reliability of Predictions in Validation: PP-Plots

SLS



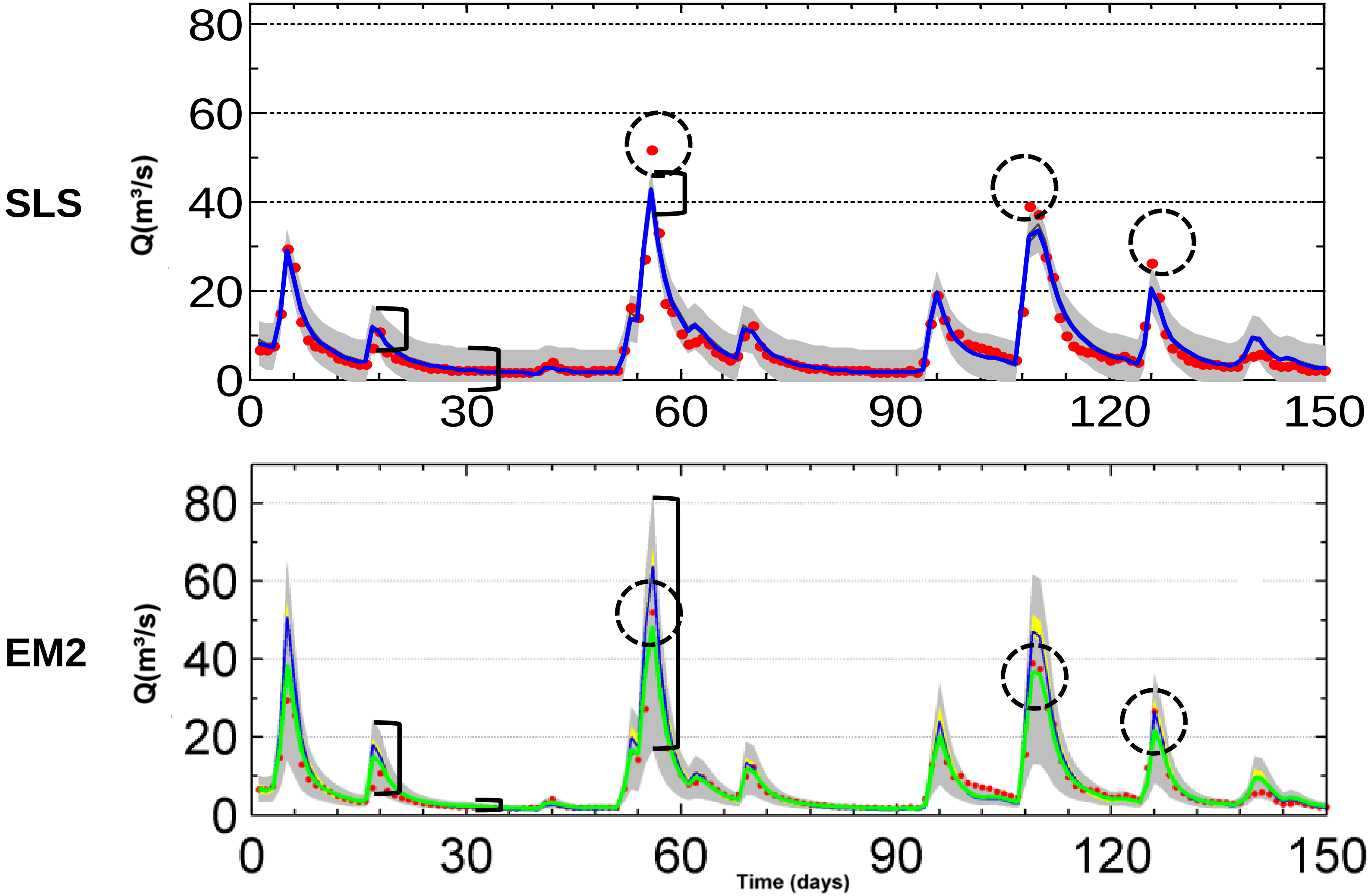
EM2



iiama

Performance of Predictive Distribution

95% Band in Valid.



iiama

Model divergence phenomenon

- Comparing to what extent Hydrological Model results in Validation period are worse than results in Calibration …

		SLS			EM2		
		CALIB	VALID	% CHANGE CALIB-VALID	CALIB	VALID	% CHANGE CALIB-VALID
HYDRO MODEL	NSE	0.93	0.86	7%	0.74	0.72	3%
	RMSE	2.62	3.48	33%	5.00	4.99	0%
	ErrVol (%)	2.40	-4.5	88%	9.90	2.70	73%

- In short, in our case study, **model divergence phenomenon with SLS is stronger** than with EM2

Parameters physical meaning

PARAMETER		MAP MEAN VALUE	SLS CORRECTION	EM2 CORRECTION	SLS VALUE	EM2 VALUE
Hu (mm)	Maximum Static Storage	214.49	X 3	X 0.81	**643.47**	**173.74**
Ksa (mm/day)	aquifer hydraulic cond.	0.74	X 68	X 1038	**50.32**	**768.12**

- In the case study, **SLS overestimates** the maximum capillary water content in the upper soil
  - **643 mm it is not a fair value with physical meaning**
- **SLS underestimates** the behavior of the aquifer flows for the analyzed humid watershed
  - **50.32 mm/day it is not a fair conductivity value from our knowledge of the catchment**

iiama

Conclusions

Error Model	Shape Distribution	Dependence	Variance	Bias
SLS	Gaussian	----	constant	----
EM2: Especific	Generalized Likelihood Skew Exponential Power	AR(2)	Linear (2 par.)	Double Linear (4 par.)

- In the analyzed case study **the prediction performance is similar** with both SLS and EM2 Error Models, but…
- **The Use of appropriate error models allows**
  - Get unbiased (or less corrupted) hydrological parameters
  - Correctly assess the uncertainty of the prediction
- An Error Model will work properly only if it is calibrated together with the hydrological model: **JOINT INFERENCE**
- Time-Varying Error Models must consider **THE TOTAL LAWS** (**TVL** and **TEL**)

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

iiama Instituto de Ingeniería del Agua y Medio Ambiente

Thank you

for your attention

Mario R. Hernández (maherlo@hma.upv.es)

Research Group on Hydrologic and Environmental Modeling

lluvia.dihma.upv.es

The study was funded by the Spanish Ministry of Economy and Competitiveness through the research projects SCARCE (CSD2009-00065) and ECOTETIS (CGL2011-28776-C02-01), and by the Universitat Politècnica de València through the Research and Development Grants Program (PAID).