UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Instituto de Ingeniería del Agua y Medio Ambiente

On the influence of Error Model in the good performance of a hydrological model (for the right reasons)

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*

HIC 2014 – 11th International Conference on Hydroinformatics
New York, USA August 17 – 21, 2014

Introduction

- Problem: hydrological models provide predictions, which are not lacking of uncertainty

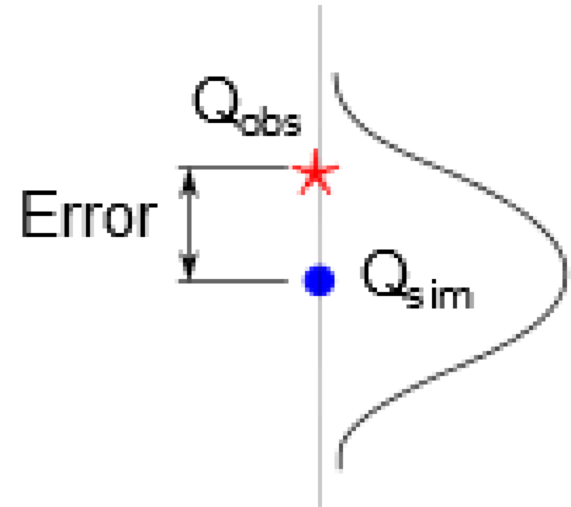
- **Uncertainty sources:**
  - Observed data: forcings, state variables, characteristics and boundary conditions
  - Model structure
  - Parameters estimation

- If calibration, calibrated parameters are a **sinkhole of uncertainty**

- Additive uncertainty model: for the state variable of interest, the difference (Error) between observed and simulated values can be modeled

$$\tilde{Q}_t = E_t\left(\boldsymbol{\theta}_h, \boldsymbol{\theta}_e, \tilde{\mathbf{X}}_{1:t}, \tilde{\mathbf{S}}_0\right) + \varepsilon_t$$

  - **Classical approach**: i.i.d. Gaussian Errors
    - It is equivalent to **Standard Least Squares** (**SLS**) calibration

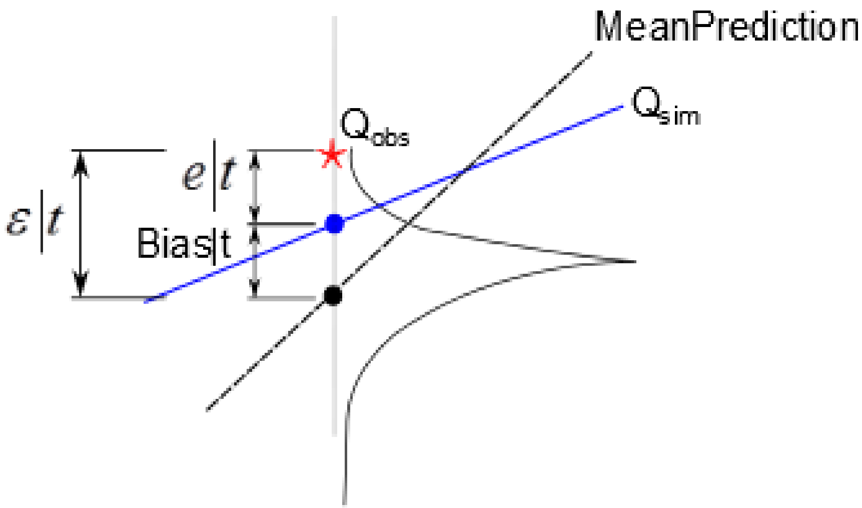


Introduction

- Additive uncertainty model: for the state variable of interest, the difference (Error) between observed and simulated values can be modeled

$$\tilde{Q}_t = E_t\left(\boldsymbol{\theta}_h, \boldsymbol{\theta}_e, \tilde{\mathbf{X}}_{1:t}, \tilde{\mathbf{S}}_0\right) + \varepsilon_t$$

  - **Errors in Hydrology** don't satisfy the **SLS** hypothesis
    - Non-Gaussian
    - Biased (even non-stationary)
    - Heteroscedastic
    - Autocorrelated



- Aims of this work:

  - Infer **Specific Error Model (EM2)** that best fit Hydrological Model Additive Errors in a case study (model + catchment)

  - Compare the performance of **SLS vs EM2**

EM2 description (1/4)

- **Formal Bayesian Joint Inference approach:**
  - Target: The Posterior of Hydrological and Error models parameters

$$p\left(\boldsymbol{\theta}_{h,e} \middle| \tilde{\mathbf{Q}}\right) \propto p\left(\tilde{\mathbf{Q}} \middle| \boldsymbol{\theta}_{h,e}\right) p\left(\boldsymbol{\theta}_{h,e}\right)$$

  - Development of an appropriate Likelihood function

$$l\left(\boldsymbol{\theta}_{h,e} \mid \boldsymbol{\varepsilon}\right) = p\left(\tilde{\mathbf{Q}} \mid \boldsymbol{\theta}_{h,e}\right) = p\left(\boldsymbol{\varepsilon} \mid \boldsymbol{\theta}_{h,e}\right)$$

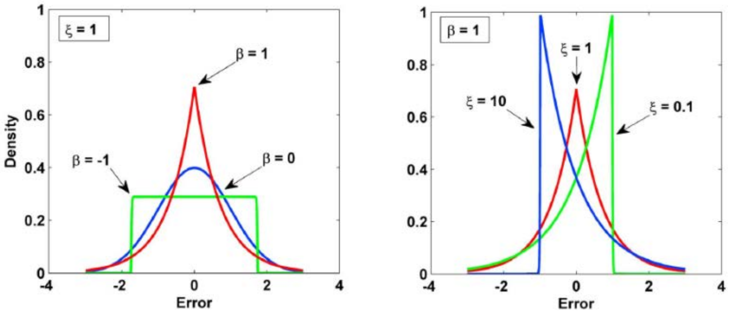
- Non-Informative Priors for the inferred parameters

- **MCMC** Posterior sampling algorithm: ***DREAM_ZS***
  - C. Ter Braak, J. Vrugt, *Differential Evolution Markov Chain with snooker updater and fewer chains*, Stat. Comput. 18 (2008) 435–446. doi:10.1007/s11222-008-9104-9

EM2 description (2/4)

- Obtaining the **(stationary) Likelihood**

  - Modeling the shape of Errors Conditional Distributions
    - Skew Exponential Power (SEP)



Picture taken from: Schoups, G., and J. Vrugt (2010*), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors*, Water Resour. Res., 46(10), 1–17, doi:10.1029/2009WR008933

  - Modeling the dependence: AR models

=> A renewed Generalized Log-Likelihood Function

$$L\left(\boldsymbol{\theta}_{h,e} \mid \boldsymbol{\varepsilon}\right) \cong N \log \frac{2\sigma_{\xi} w_{\beta}}{\sigma_{\mathbf{z}}\left(\xi+\xi^{-1}\right)} - \sum_{t=1}^{N} \log \sigma_{\varepsilon_t} - c_{\beta} \sum_{t=1}^{N} \left|a_{\xi,t}\right|^{\frac{2}{1+\beta}}$$

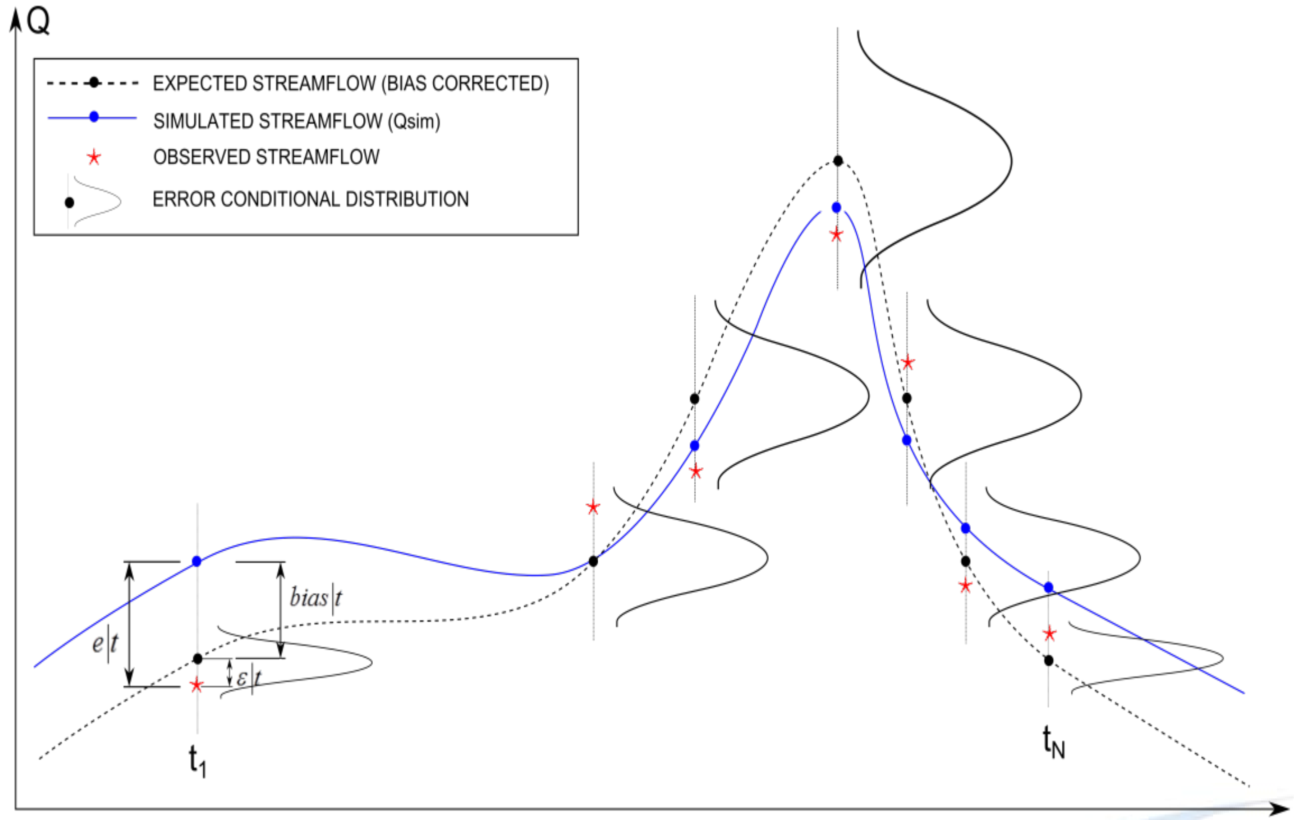
EM2 description (3/4)

- **Introducing the Non-Stationarity in the Likelihood**
  - For the Variance: $$\mathrm{f}_v = \theta_1^e + \theta_2^e Q_{tsim}$$
  - For the Bias:
$$\mathrm{f}_b = \theta_3^e + \theta_4^e Q_{tsim} \qquad Q_{tsim} \le \theta_5^e$$
$$\mathrm{f}_b = \theta_6^e + \theta_7^e Q_{tsim} \qquad Q_{tsim} > \theta_5^e$$
  - Variance and Bias parameters are not free, because **Total Laws must be enforced**
    - Total Variance Law (**TVL**) $$V(\tilde{\mathbf{Q}}) = E_t\left[V(\tilde{Q}|t)\right] + V_t\left[E(\tilde{Q}|t)\right] = E_t\left[V(\tilde{Q}|t)\right] + V_t\left[(Q_{sim} + bias)|t\right]$$
    - Total Expectation Law (**TEL**) $$E(\tilde{\mathbf{Q}}) = E_t\left[E(\tilde{Q}|t)\right] = E_t\left[(Q_{sim} + bias)|t\right]$$

EM2 description (4/4)

- Prediction model: Hydrological model + EM2

$$Q_{tpred} = Q_{tsim} + b_t + \sigma_{\varepsilon_t}\left[\phi_p^{-1}(B)\left[\sigma_z a_t\right]\right]$$



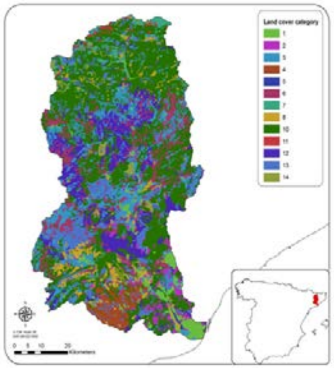
Hydrological Model

- **Distributed Hydrological Model**

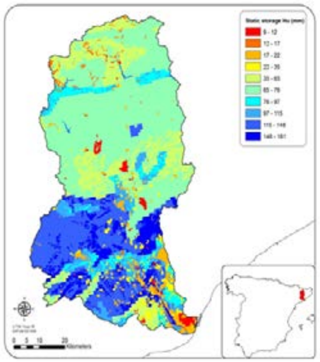
TETIS

http://lluvia.dihma.upv.es

- Runoff production and propagation:
  - Traditional schemes used in modeling, but adapted to cell scale
- Effective Parameter Structure divided in two components:
  - Estimated Value at each cell: parameter maps
  - **Global Calibrated Correction Factor** applied to each map



x F1



x F2

....

Model divergence phenomenon

		SLS		
		CALIB	**VALID**	**% CHANGE**
HYDRO MODEL	**NSE**	0.93	0.86	**7%**
	RMSE	2.62	3.48	**33%**
	ErrVol (%)	2.40	-4.5	**88%**

Prediction performance

		SLS			EM2		
		CALIB	VALID	% CHANGE	CALIB	VALID	% CHANGE
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%
MEAN PREDICTION	NSE				0.91	0.85	7%
	RMSE				2.92	3.60	23%
	ErrVol (%)				0.01	-3.70	

Parameters physical meaning

PARAMETER		MAP MEAN VALUE	SLS CORRECTION	EM2 CORRECTION	SLS VALUE	EM2 VALUE
Hu (mm)	Soil Capillary Storage	214.49	X 3	X 0.81	**643.47**	**173.74**
Kss (mm/day)	Interflow hydraulic cond.	1.8	X 2508	X 4913	4514.40	8843.40
Ksa (mm/day)	Aquifer hydraulic cond.	0.74	X 68	X 1038	**50.32**	**768.12**

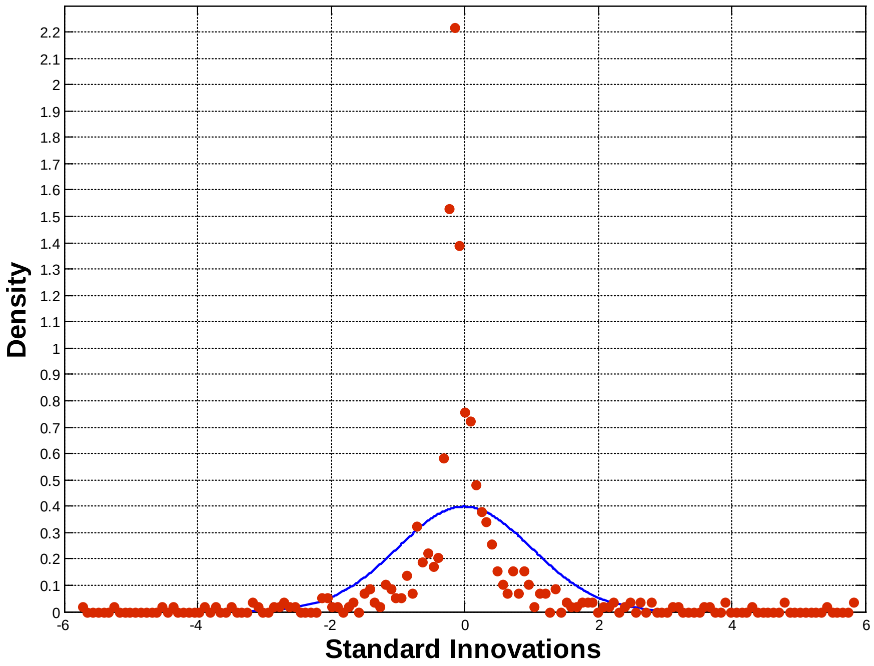
- A 643 mm for Soil Capillary Storage isn't realistic
- SLS underestimates the aquifer response

=> SLS converts a conceptual hydrological model into a data driven-model

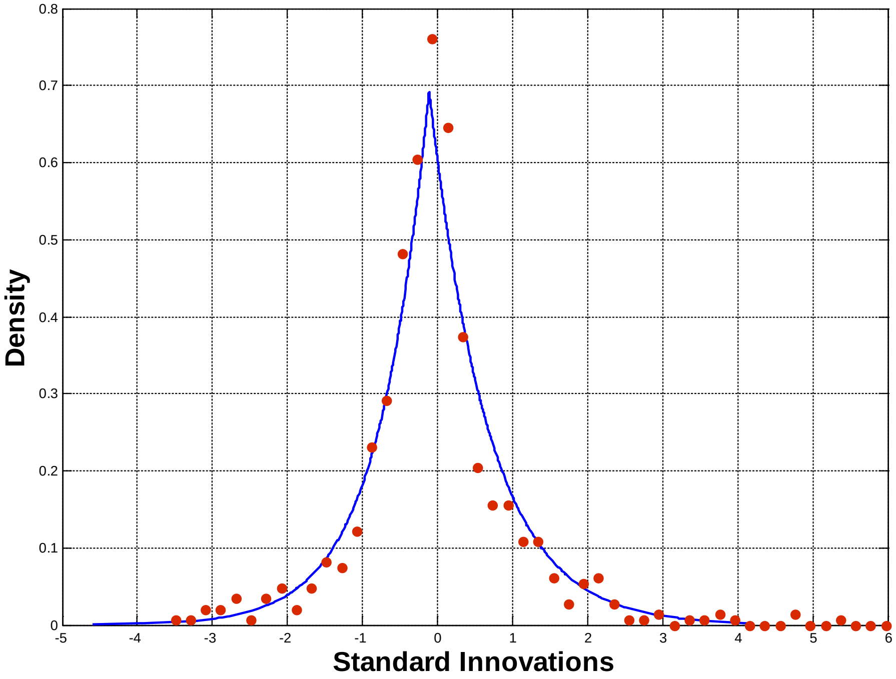
Fulfillment of Error Models hypothesis

Standar innovations *Normality*

SLS

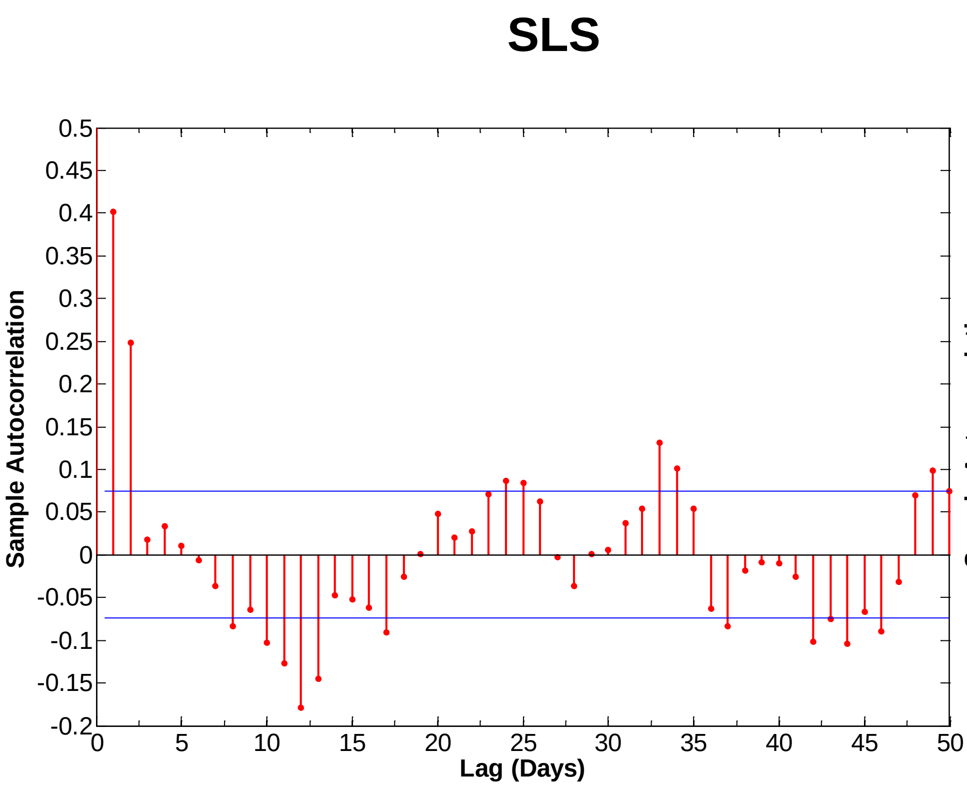


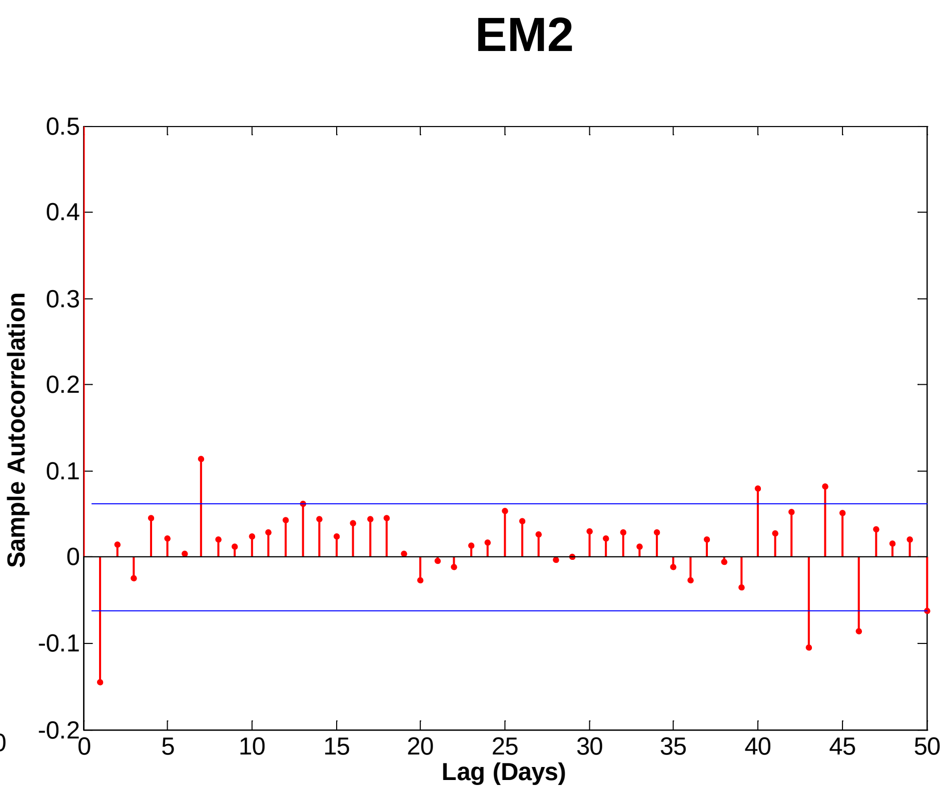
EM2



Fulfillment of Error Models hypothesis

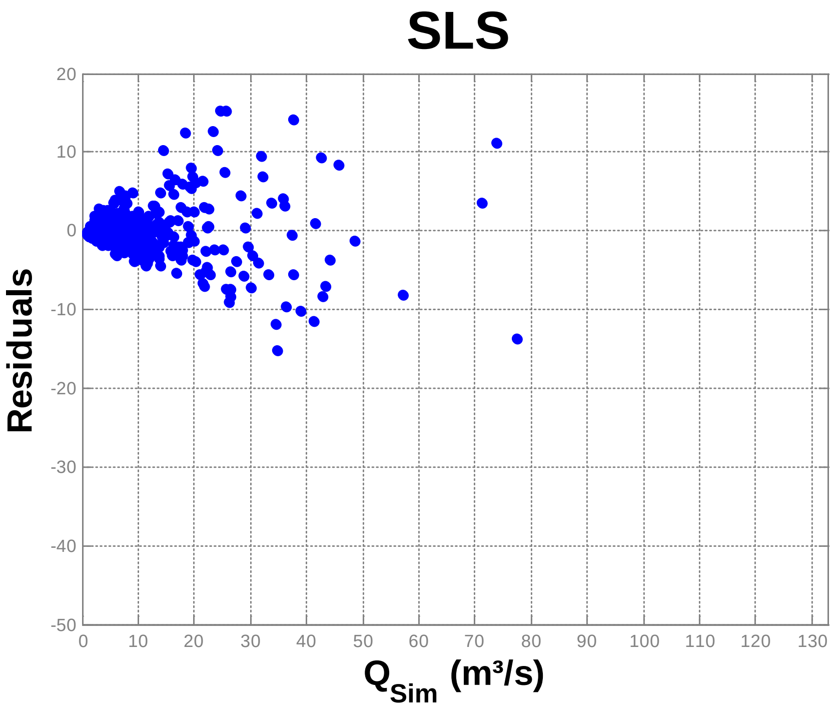
Independence

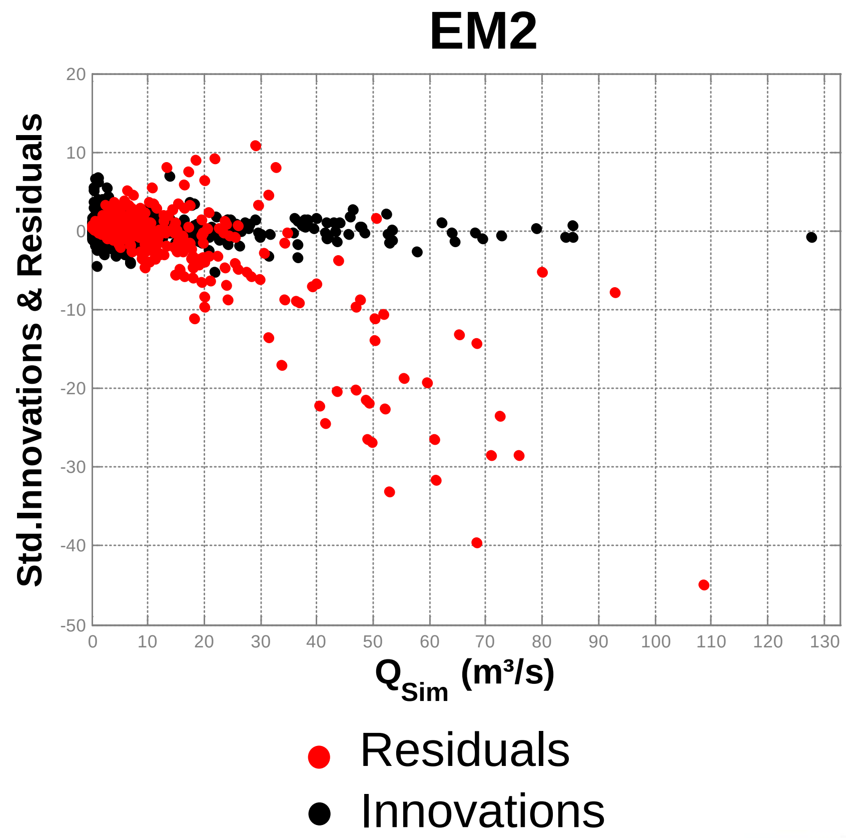




Fulfillment of Error Models hypothesis

Homoscedasticity

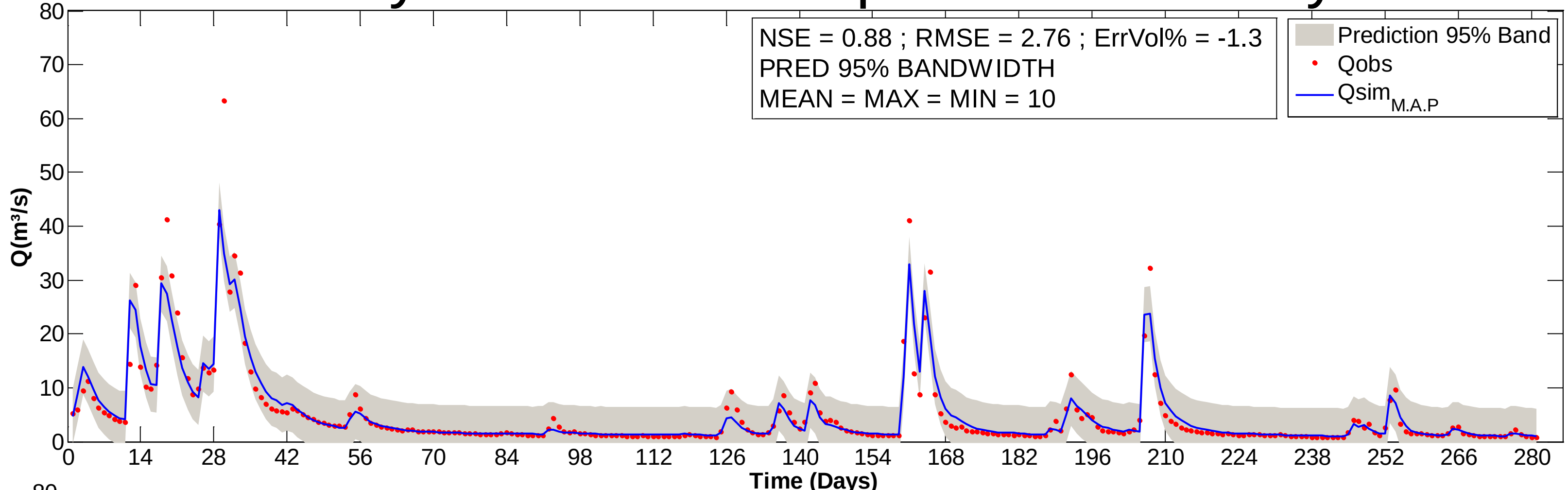




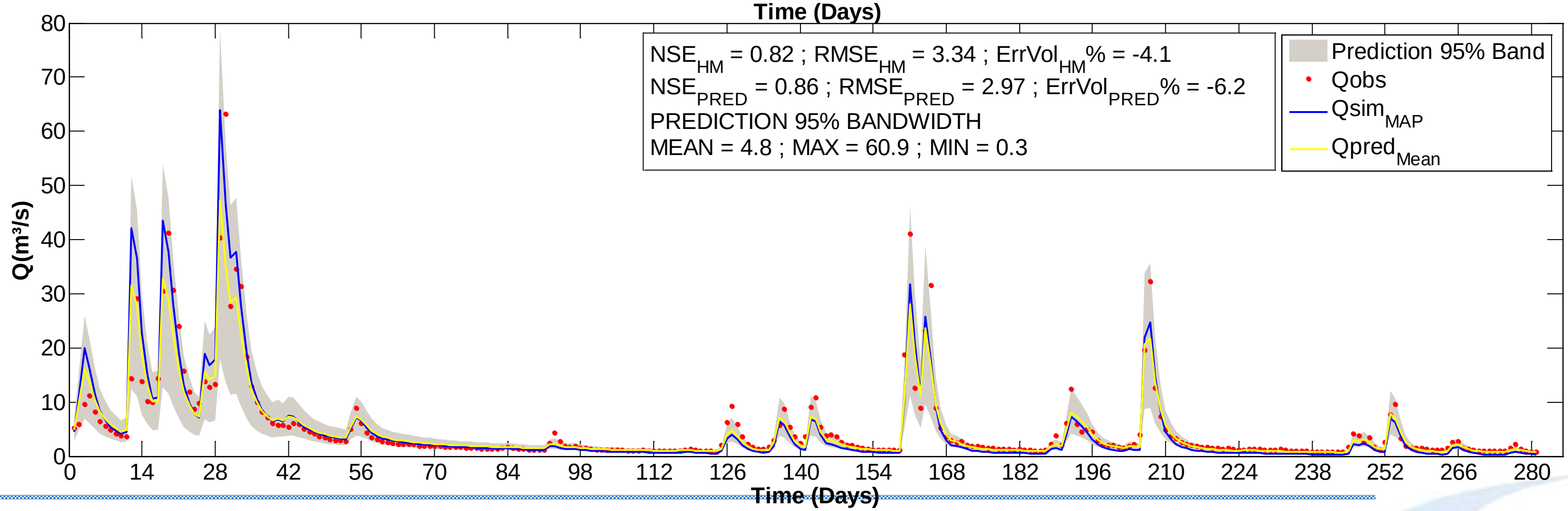
Prediction uncertainty reliability

- 95% uncertainty band for a sub-period of 280 days

SLS

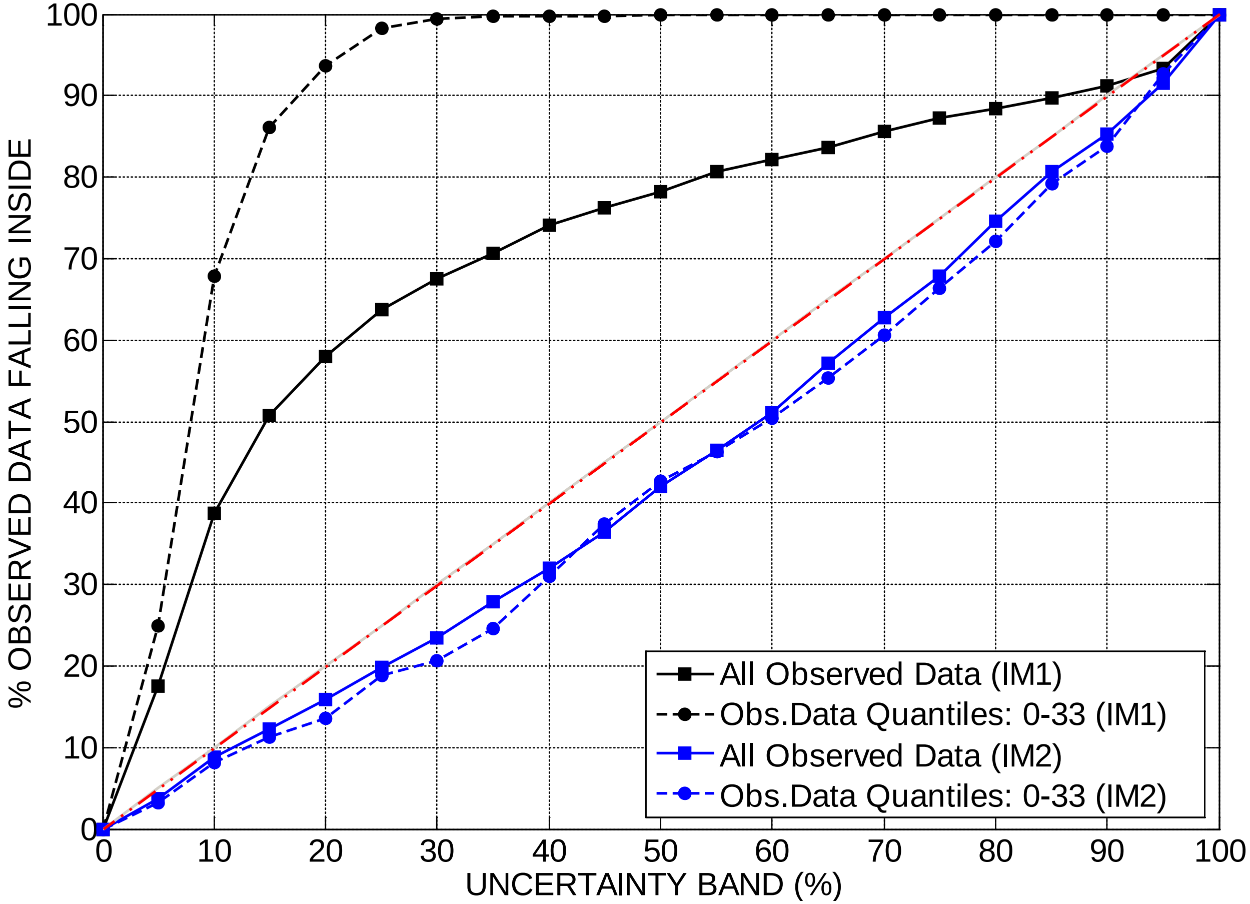


EM2



Prediction uncertainty reliability

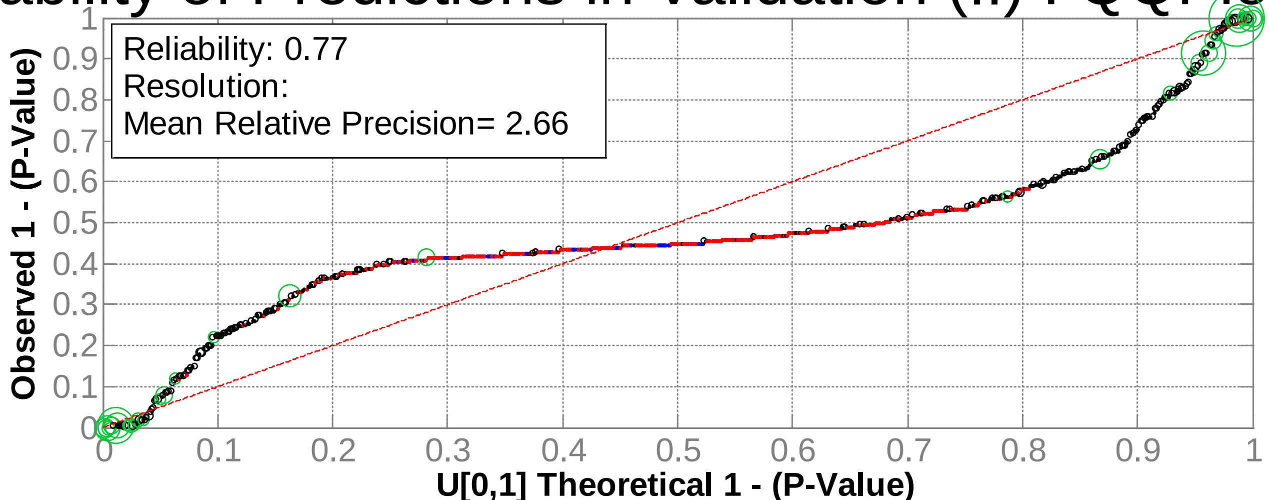
- Reliability of Predictions in Validation (I)



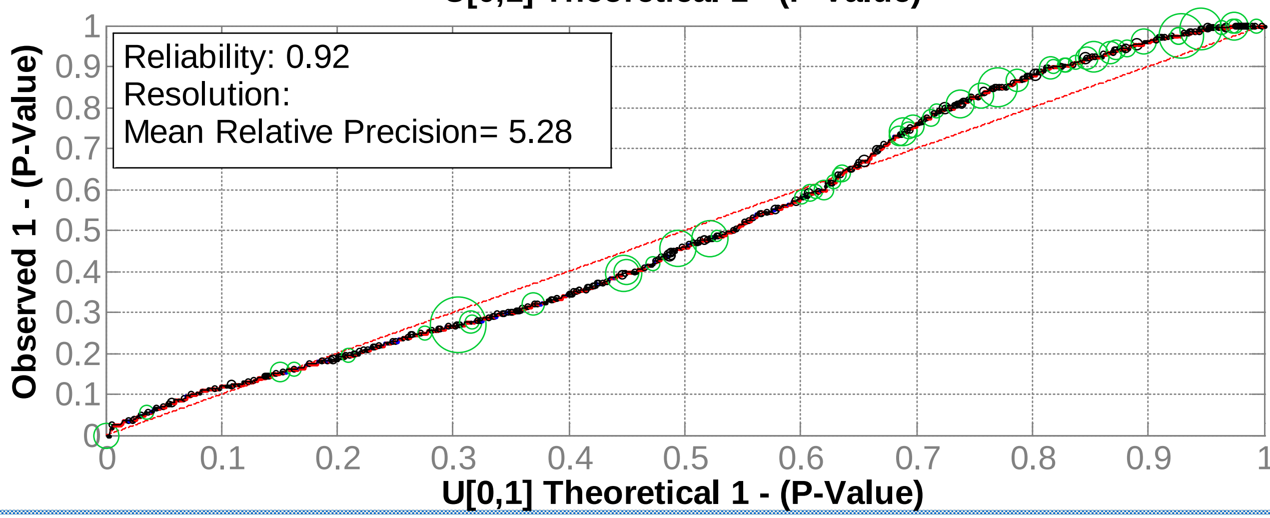
Prediction uncertainty reliability

- Reliability of Predictions in Validation (II) : QQPlots

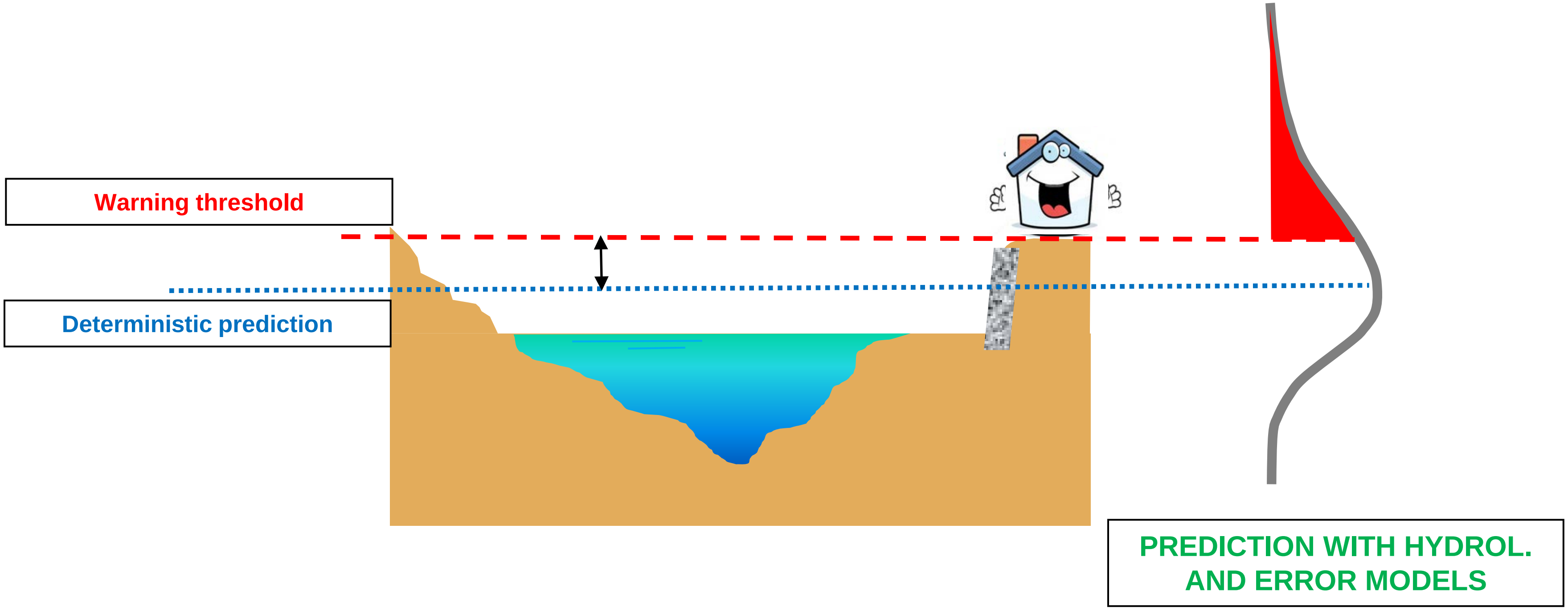
SLS



EM2



One example: decision making



Conclusions

- Errors in Hydrology very often **don’t satisfy** the **SLS** hypothesis. Consequences:
  - Biased parameter values and loss of physical meaning
  - Poor Parameter Uncertainty estimates
  - Incorrect estimation of Predictive Uncertainty

=> With SLS calibrated parameters, model can work for the wrong reasons

iiama

Conclusions

- Errors in Hydrology very often **don’t satisfy** the **SLS** hypothesis

- It is possible to develop Error Models for a distributed Hydrological Model in a Bayesian Joint Inference framework, easily with a **Split Effective Parameter Structure**

- Non-Stationary Error Models must consider the **Total Laws** for the Expectation and Variance

- Be positive: uncertainty is reflecting our **knowledge**

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Instituto de Ingeniería del Agua y Medio Ambiente

Thanks for your attention

Prof. Félix Francés (ffrances@hma.upv.es)

Research Group of Hydrological and Environmental Modelling

http://lluvia.dihma.upv.es

Research Institute of Water and Environmental Engineering

http://iiama.upv.es

Universitat Politècnica de València, Spain

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)