QUANTIFICATION OF UNCERTAINTY USING BAYESIAN APPROACHES

Prof. Ezio Todini
President Italian Hydrological Society
University of Bologna - Italy

SCOPE

This presentation aims at clarifying the concept of

Predictive Uncertainty (PU)

and at providing a methodological framework for quantifying and using it in flood forecasting and water resources management.

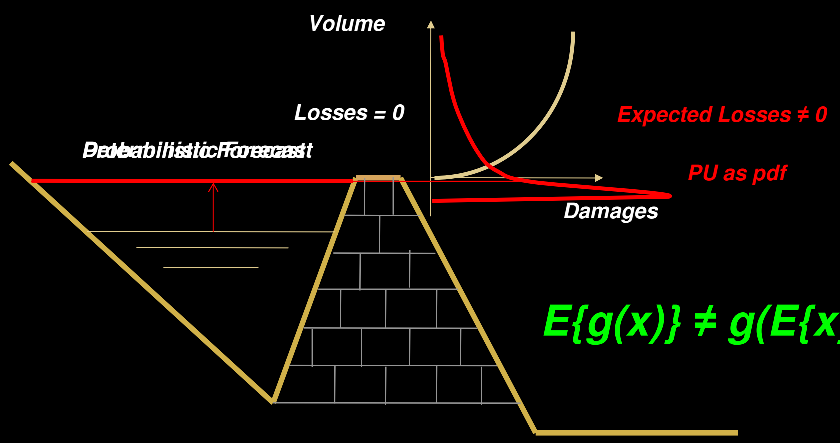
Problems Involved

1. Flood Warning and Evacuation Management
2. Flood Detention and Diversion
3. Real Time Reservoir Management

Etc.

DECISION MAKING UNDER UNCERTAINTY

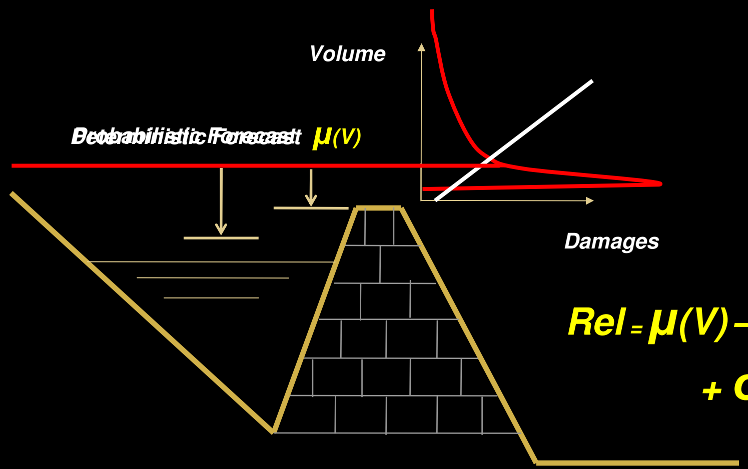
The Reservoir Management Problem



$$E\{g(x)\} \neq g(E\{x\})$$

DECISION MAKING UNDER UNCERTAINTY

The Linear Losses Case



$$Rel = \mu(V) - V_{max} + \sigma(V)\, f(c)$$

THE DEFINITION OF PREDICTIVE UNCERTAINTY

*In order to understand the meaning of **Predictive Uncertainty,** let me pose the following question:*

Flooding damages will occur:

(1) when the forecasted level overtops the dykes?

or

(2) when the actual future water level overtops the dykes?

The obvious answer is

(2) when the actual future water level overtops the dykes

THE DEFINITION OF PREDICTIVE UNCERTAINTY

This answer has a strong implication in the definition of PU

PU is obviously the uncertainty that we have on the occurrence of a real future value, as for instance the water level in 12 hours from now.

This must not be confused with "Validation Uncertainty".

THE DEFINITION OF PREDICTIVE UNCERTAINTY

Following Rougier (2007),

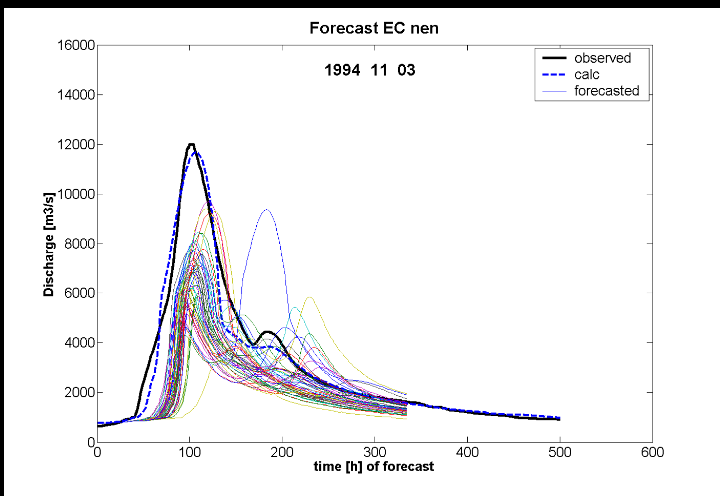
Predictive uncertainty

is the expression of a subjective assessment of the probability of occurrence a future (real) event conditional upon all the knowledge available up to the present (the prior knowledge) and the information that can be acquired through a learning inferential process.

VALIDATION UNCERTAINTY vs PREDICTIVE UNCERTAINTY

Meteorological Ensembles
are a measure of the Validation Uncertainty,
while
Climatological distributions
or
Extreme Value distributions
are measures of Predictive Uncertainty,
although non conditional on real time
information.

An example using Meteorological Ensembles



The definition of Validation Uncertainty

The Validation Uncertainty is the probability that the model predicted value (water level,discharge, water volume, etc) will be smaller or equal to a prescribed value

$$Prob\left(\hat{y}_t \le y^* \middle| y_t, M, \mathcal{D}_{hist}\right)$$

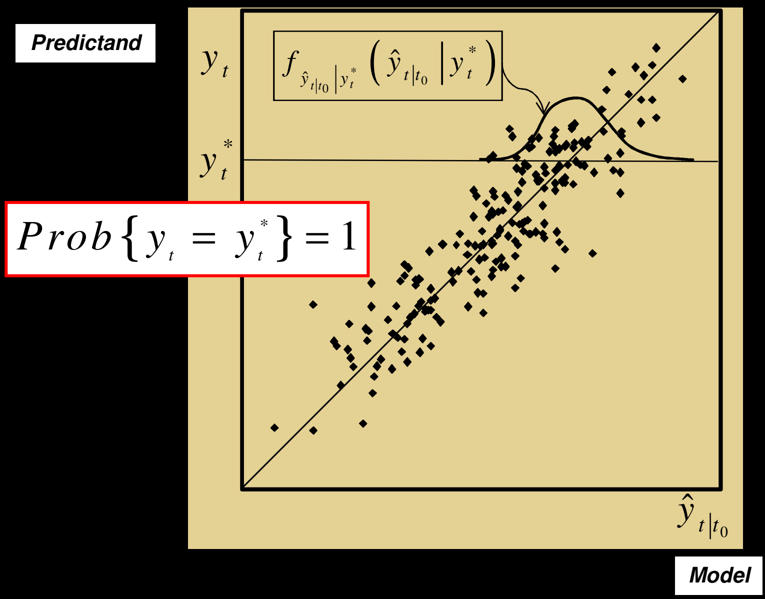
Please note that this expression cannot be used beyond time t because observations are not available

The use of Validation Uncertainty

Assessment of Validation Uncertainty (VU) is essential to evaluate the performances of a model in order to improve it.

Therefore, when dealing with VU, one must also assess and separate the effects of model uncertainty, parameter uncertainty, input and output measurement uncertainty, initial and boundary conditions uncertainty.

Representation of Validation Uncertainty



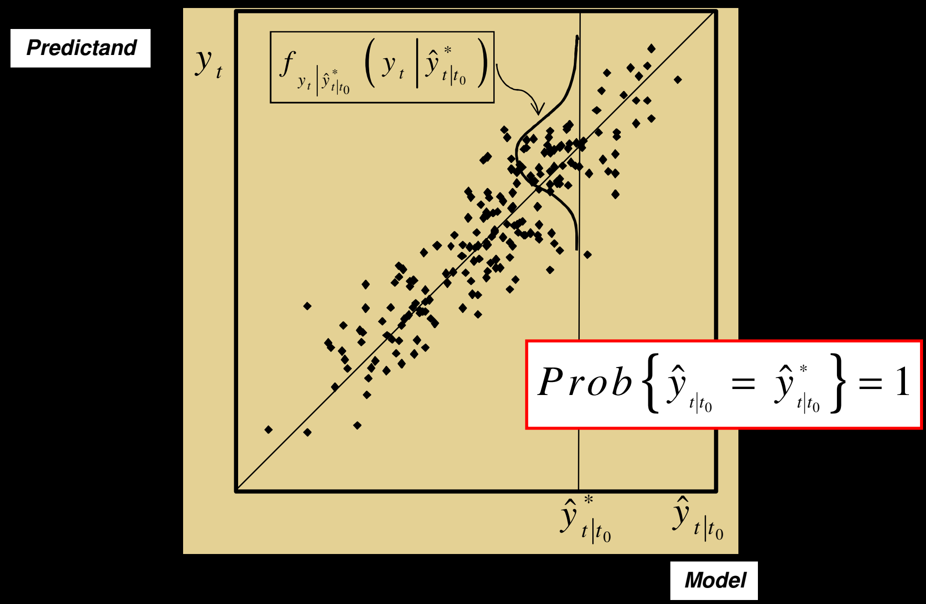
The definition of Predictive Uncertainty

The Predictive Uncertainty is the probability that a future value of the predictand (water level, discharge, water volume, etc) will be smaller or equal to a prescribed value.

$$Prob\left(y_{t+k\Delta t} \le y^* \,\middle|\, \hat{y}_{t+k\Delta t}, M, \mathscr{D}_{hist}\right)$$

given our prior knowledge, all the historical information and the model forecast

Representation of Predictive Uncertainty

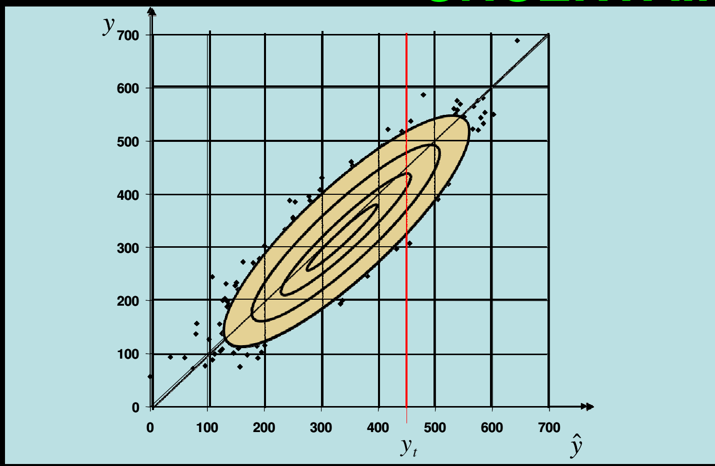


The use of Predictive Uncertainty

Assessment of Predictive Uncertainty is fundamental to take a decision given a model (or several models) forecast.

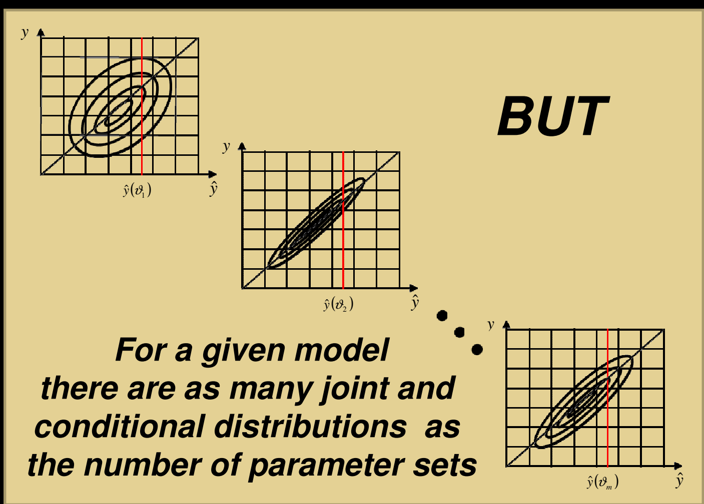
When using PU it is not necessary to assess and separate all the sources of errors if the conditional density used is consistent with the model(s) and all the other sources of uncertainty, which affected its development.

MODEL AND PARAMETER UNCERTAINTY



*For a given **model** and a **set of parameters** one can derive predictand and model **joint/conditional probability densities***

PARAMETER UNCERTAINTY



$$Prob\left(y_{t+k\Delta t} \le y^* \,\middle|\, \hat{y}_{t+k\Delta t}\left(\vartheta_i\right), M, \mathcal{D}_{hist}\right)$$

MODEL AND PARAMETER UNCERTAINTY

Therefore one must derive the "Posterior Density (PD)" of parameters $g_B(\vartheta | M, \mathcal{D}_{hist})$ using the classical Bayesian Inference. This PD is then used to marginalise, namely to integrate out, the effect of parameters.

In a continuous domain:

$$F\left(y_{t+k\Delta t} | M, \mathcal{D}_{hist}\right) = \int_{\Omega_\vartheta} F\left(y_{t+k\Delta t} | \hat{y}_{t+k\Delta t}(\vartheta), M, \mathcal{D}_{hist}\right) g_B\left(\vartheta | M, \mathcal{D}_{hist}\right) d\vartheta$$

or in discrete mode:

$$F\left(y_{t+k\Delta t} | M, \mathcal{D}_{hist}\right) \cong \sum_i F_i\left(y_{t+k\Delta t} | \hat{y}_{t+k\Delta t}(\vartheta_i), M, \mathcal{D}_{hist}\right) g_B\left(\vartheta_i | M, \mathcal{D}_{hist}\right)$$

MODEL AND PARAMETER UNCERTAINTY

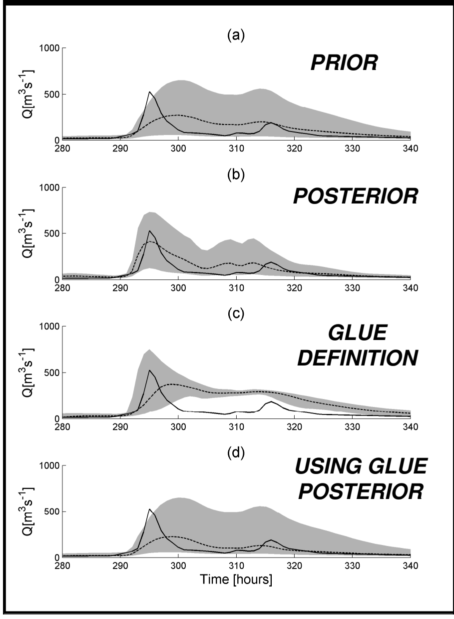
Please note that this is TOTALLY different from what is proposed in Generalized Likelihood Uncertainty Estimation (GLUE), where the definition of PU is given as:

$$P\left(\hat{Z}_t < z\right) = \sum_{i=1}^{B} L\left[M\left(\vartheta_i\right) | \hat{Z}_{t,i} < z\right]$$

where $L = g_G\left(\vartheta_i | M, \mathcal{D}_{hist}\right)$ is nothing else than the posterior parameter density.

Where are the conditional predictive density (???) as well as the marginalisation of parameter uncertainty (???)

MODEL AND PARAMETER UNCERTAINTY



$$F\{y|M,\mathcal{D}_{hist}\} = \sum_{i=1}^{B} F\{y|\hat{y}(\vartheta_i),M\}\, g_0\{\vartheta_i|M,\mathcal{D}_{hist}\}$$

$$F\{y|M,\mathcal{D}_{hist}\} = \sum_{i=1}^{B} F\{y|\hat{y}(\vartheta_i),M\}\, g_B\{\vartheta_i|M,\mathcal{D}_{hist}\}$$

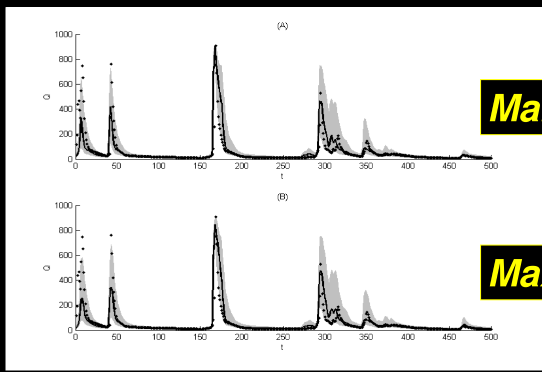
$$F(\hat{y}_t|M,\mathcal{D}_{hist}) = \sum_{i=1}^{B} L\left[M(\vartheta_i)|\hat{y}_{t,i} \le \hat{y}_t\right] = \sum_{i=1}^{B} g_G\{\vartheta_i|M,\mathcal{D}_{hist}\} \quad \forall \vartheta_i|\hat{y}_{t,i} \le \hat{y}_t$$

$$F\{y|M,\mathcal{D}_{hist}\} = \sum_{i=1}^{B} F\{y|\hat{y}(\vartheta_i),M\}\, g_G\{\vartheta_i|M,\mathcal{D}_{hist}\}$$

MODEL AND PARAMETER UNCERTAINTY

Nonetheless, marginalising parameter uncertainty, although statistically correct, does not produce substantial differences from using a best fit parameter set.
This is mostly due to the fact that the nearly best parameters produce predictions that are closely related among them, while the posterior probability of the worst parameter sets is obviously very low.

Predictive uncertainty in hindcast mode (Results on a Chinese catchment)

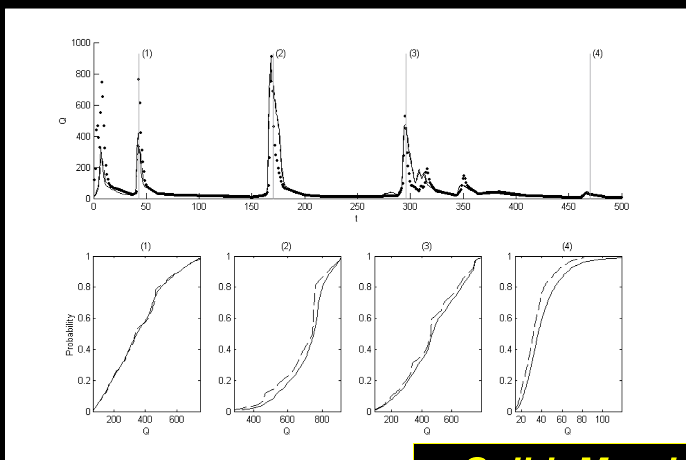


Marginalised

Max Likelihood

The difference between the marginalised density and the one obtained using the "best parameter set" can be relatively small

Predictive uncertainty in hindcast mode (Results on a Chinese catchment)



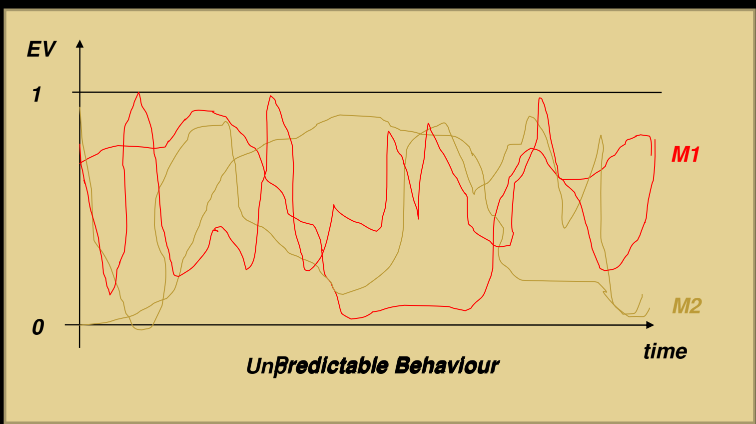
Solid: Marginalised
Dashed: Max Likelihood

MODEL AND PARAMETER UNCERTAINTY

When the behaviour of a set of conditions such as errors deriving from the different sources varies at random in time in an "unpredictable manner" then one can use the "mixture of models" concept.

Please bear in mind that if the conditions ARE predictable then one is better off by using the "model" which best fits the observations under the relevant conditions.

MODEL AND PARAMETER UNCERTAINTY



MODEL AND PARAMETER UNCERTAINTY

This is why it is more interesting to approach the problem in terms of

few alternative models

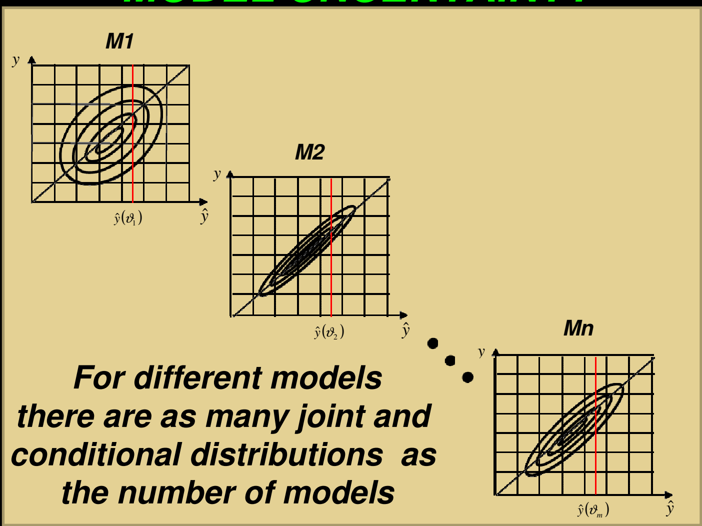
of a widely different nature.

i.e. a physically based model, a conceptual model and a data driven model.

This has given rise to the development of several multi-model

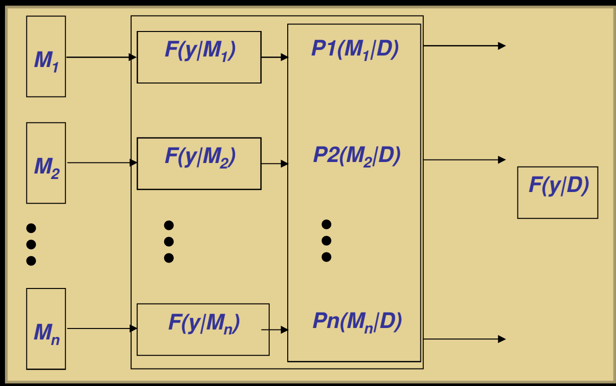
Predictive Uncertainty Processors.

MODEL UNCERTAINTY



$$Prob\left(y_{t+k\Delta t} \le y^{*} \,\middle|\, \hat{y}_{t+k\Delta t}\left(M_i\right), \mathscr{D}_{hist}\right)$$

AVAILABLE PREDICTIVE UNCERTAINTY PROCESSORS



$$F\left(y_t\,|M,\mathcal{D}_{hist}\right)=\sum_{i=1}^{n}F_i\left(y_t\,|M_i\right)Prob\left\{M_i\,|\mathcal{D}_{hist}\right\}$$

THE BINARY RESPONSE PROCESSORS

The Binary Response Processors convert continuous measurements and/or forecasts into discrete 0-1 probability of occurrence of one event.

- *The Logit (based on the Logistic Distribution)*
- *The Probit (based on the Inverse Gaussian Distribution)*
- *The Bayesian Multivariate Binary Processor (BMBP)*
- *The Mixture of Beta Distributions*

Useful tools, but the reliability of the continuous processors seems to be higher.

The Continuous Single or Multi-model Predictive Uncertainty Processors

Hydrological Uncertainty Processor
Krzysztofowicz, 1999; Krzysztofowicz and Kelly, 2000

Bayesian Model Averaging
Raftery et al., 2003;

Model Conditional Processor
Todini, 2008.

....

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour. Res., 35, 2739–2750.
Raftery, A. E., F. Balabdaoui, T. Gneiting, and M. Polakowski, 2003. Using Bayesian model averaging to calibrate forecast ensembles, Tech. Rep. 440, Dep. of Stat., Univ. of Wash., Seattle.
Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. Intl. J. River Basin Management. Vol. 6 (2), 123-137.

AVAILABLE PREDICTIVE UNCERTAINTY PROCESSORS

Krzysztofowicz Bayesian Processor

Krzysztofowicz (1999) approach (HUP) was the first to be developed in hydrological applications.
It combines prior information embedded into an AR1 model with that deriving from a predictive model of unspecified nature (physically based, conceptual, etc.)
Unfortunately

-It has a scalar formulation: only one model can be combined at a time
-The AR1 model is implicitly assumed to be independent from the predictive model

AVAILABLE PREDICTIVE UNCERTAINTY PROCESSORS

Raftery Bayesian Model Averaging

BMA aims at assessing the unconditional mean and variance of any future value of a predictand on the basis of several model forecasts.

$$E\{y|\mathscr{M},\mathscr{D}_{hist}\}=\sum_{i=1}^{n}w_i E\{y|M_i\}$$

$$Var\{y|\mathscr{M},\mathscr{D}_{hist}\}\cong\sum_{i=1}^{n}w_i Var\{y|M_i\}+\sum_{i=1}^{n}w_i\left(\hat{y}_i-\sum_{i=1}^{n}w_i E\{y|M_i\}\right)^2$$

It reformulates the Bayesian mixture equation

$$F(y_t|M_i,\mathscr{D}_{hist})=\sum_{i=1}^{n}F_i(y_t|M_i)Prob\{M_i|\mathscr{D}_{hist}\}$$

by considering the posterior probability as a weight

$$Prob\{M_i|\mathscr{D}_{hist}\}=w_i$$

AVAILABLE PREDICTIVE UNCERTAINTY PROCESSORS

The BMA weights are estimated by constrained maximisation of the Likelihood of observing the predictand

$$\begin{cases}\max_{w_i}\ \log\mathscr{L}=\sum_{t=1}^{T}\log\left(\sum_{i=1}^{n}w_i p_i\left(y_t|\hat{y}_{i,t}\right)\right)\\ s.t.\ \sum_{i=1}^{n}w_i=1; \qquad w_i\geq 0\ \ \forall i=1,\dots,n\end{cases}$$

on the assumption that the probability densities of the observations as well as of the model forecasts are all approximately Gaussian, which is correct if using the Normal Quantile Transform (NQT) to transform the data

AVAILABLE PREDICTIVE UNCERTAINTY PROCESSORS

The Model Conditional Processor

If one makes the hypothesis that all the NQT transformed variables follow a multi-Gaussian joint probability density, a more natural approach can be:

- *To develop a set of models in the real untransformed space (one or more than one)*
- *To build the joint probability density in the Gaussian space (Predictand, a priori model, deterministic model, etc.)*
- *To simply compute the probability of the predictand conditional on ALL the model predictions*

A Useful Property of the Multivariate Normal Distribution

Given a vector $\begin{bmatrix} \mathbf{y} \\ \hat{\mathbf{y}} \end{bmatrix}$ *of Normally distributed random variables, with*

Mean $\boldsymbol{\mu}_x = \begin{bmatrix} \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_{\hat{y}} \end{bmatrix}$ *and Variance* $\boldsymbol{\Sigma}_{xx} = \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{y\hat{y}} \\ \boldsymbol{\Sigma}_{\hat{y}y} & \boldsymbol{\Sigma}_{\hat{y}\hat{y}} \end{bmatrix}$

With marginal distributions $\mathbf{y} \approx N(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_{yy})$ *and* $\hat{\mathbf{y}} \approx N(\boldsymbol{\mu}_{\hat{y}}, \boldsymbol{\Sigma}_{\hat{y}\hat{y}})$

also Normally distributed

Then the conditional distribution is also a Normal distribution

$$N\left(\boldsymbol{\mu}_{y|\hat{y}}, \boldsymbol{\Sigma}_{yy|\hat{y}}\right)$$

with conditional Mean $\boldsymbol{\mu}_{y|\hat{y}} = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{y\hat{y}} \boldsymbol{\Sigma}_{\hat{y}\hat{y}}^{-1} \left(\hat{\mathbf{y}} - \boldsymbol{\mu}_{\hat{y}}\right)$

and conditional Variance $\boldsymbol{\Sigma}_{yy|\hat{y}} = \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{y\hat{y}} \boldsymbol{\Sigma}_{\hat{y}\hat{y}}^{-1} \boldsymbol{\Sigma}_{\hat{y}y}$

AVAILABLE PREDICTIVE UNCERTAINTY PROCESSORS

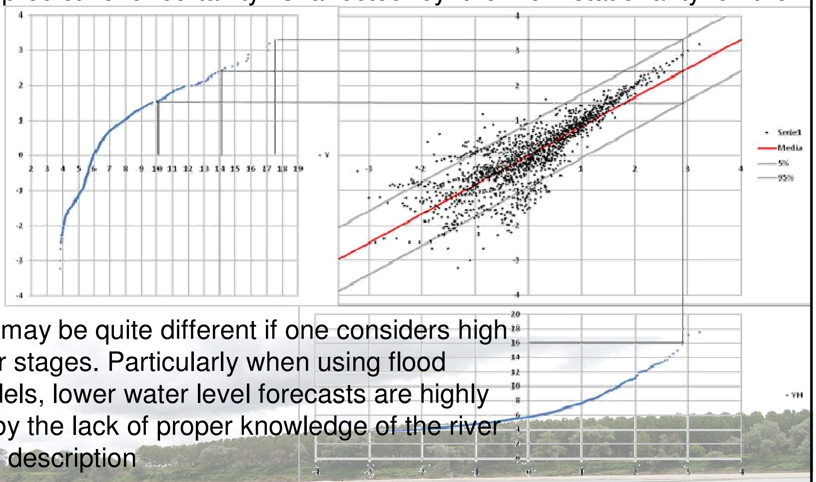
Advantages of MCP

- It allows one to combine together a wide variety of different models without the need of using the constrained optimisation required by BMA
- It accounts for correlation among the predictive models
- It allows one to have multiple outputs, benefitting from spatial correlation (for instance several water levels along the same river)

MODEL CONDITIONAL PROCESSOR

THE NEED FOR USING THE TRUNCATED NORMAL DISTRIBUTION

In many cases the statistical pattern and the correlation between observed and modelled quantities differs from lower to higher values. In this case the estimated predictive uncertainty is affected by the non stationarity of the errors .

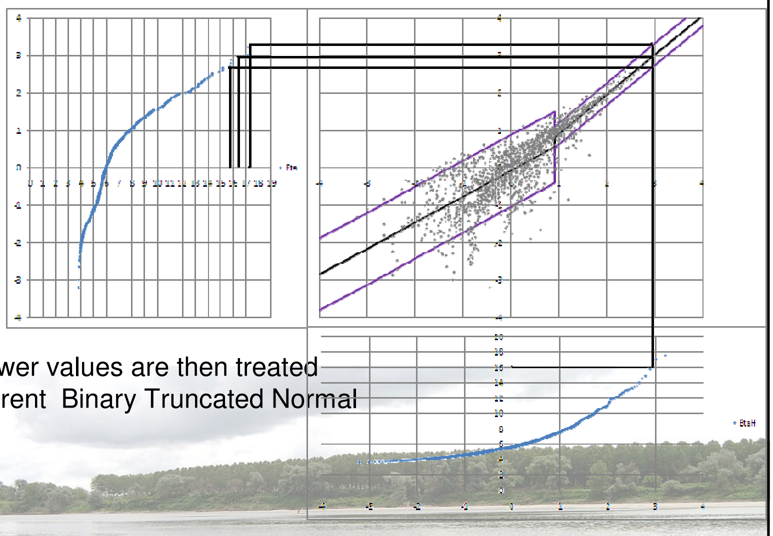


Correlation may be quite different if one considers high or low water stages. Particularly when using flood routing models, lower water level forecasts are highly influenced by the lack of proper knowledge of the river geometrical description

MODEL CONDITIONAL PROCESSOR

THE BIVARIATE JOINT TRUNCATED NORMAL DISTRIBUTION

In order to overcome this problem MCP can also be defined in terms of the Bivariate Truncated Normal Disitribution.

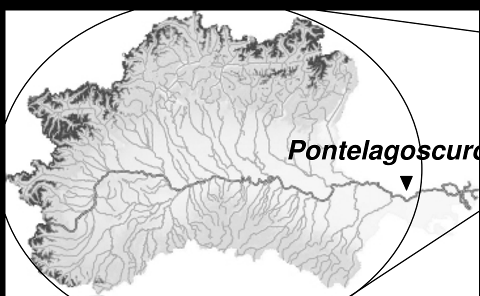


Higher and lower values are then treated using two different Binary Truncated Normal Distributions.

COMPARISON OF DIFFERENT APPROACHES

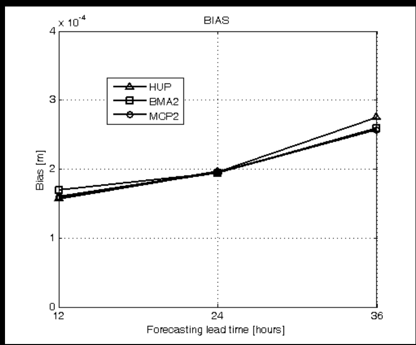
The river Po in Italy

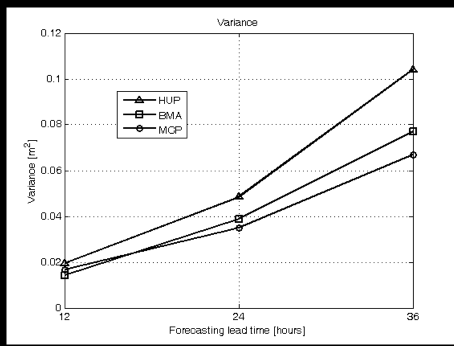
Results obtained in collaboration with ARPA-SIM of Emilia-Romagna



Basin size = 70,000 km2

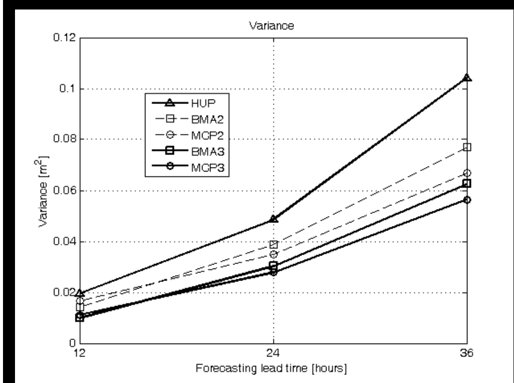
EXAMPLES
Hydraulic model + AR1

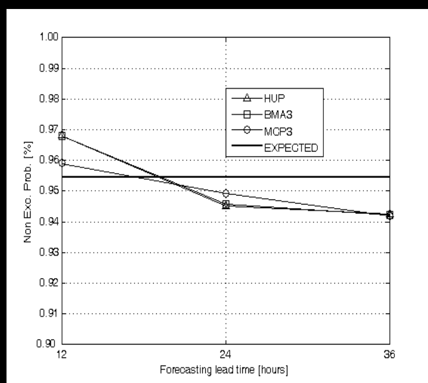




Bias **Variance**

EXAMPLES
Hydraulic model + AR1 + NN



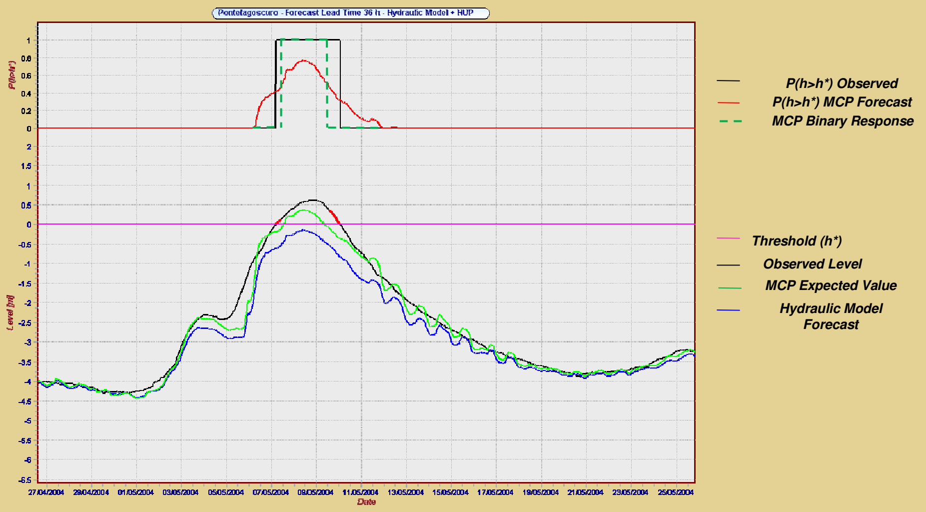


Variance

% observations within limits

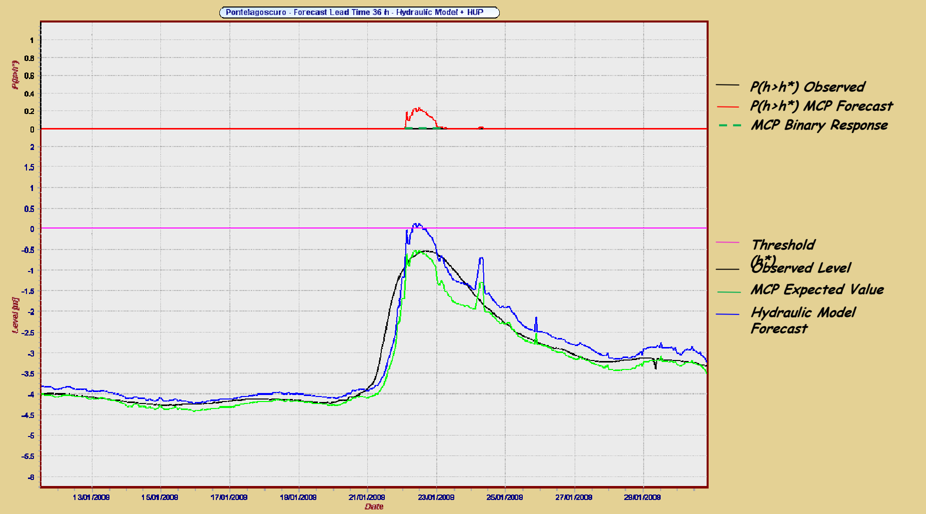
Po at Pontelagoscuro: 36 h forecast Mike11 and MCP processed PU

MISSED ALARMS CORRECTED BY THE MCP



Po at Pontelagoscuro: 36 h forecast Mike11 and MCP processed PU

FALSE ALARM CORRECTED BY THE MCP



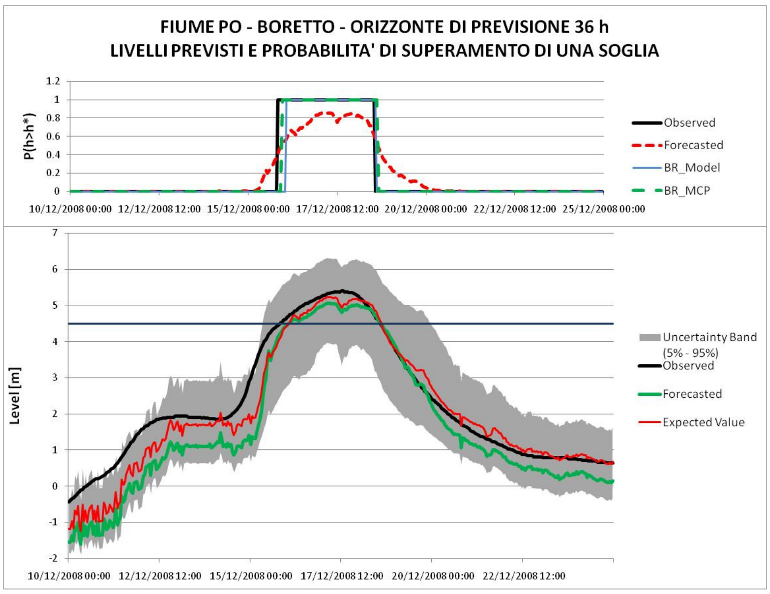
Po at Pontelagoscuro: 36 h forecast
Mike11 + ANN and MCP processed PU

STATISTICAL INDEXES FOR THE VALIDATION PERIOD (01/01/2004 – 30/01/2009)

CONTINUOUS FORECAST STATISTICS		
	ERROR MEAN	ERROR VARIANCE
MODEL	0.1803	7.23E-02
HM + MCP	0.0157	5.09E-02
HM + ANN + MCP	-6.56E-03	3.98E-02

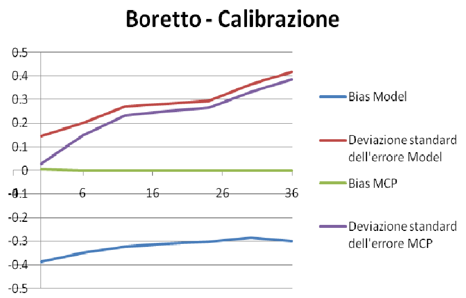
EVENT STATISTICS				
	NEGATIVE HITS	FALSE ALARMS	MISSED ALARMS	POSITIVE HITS
MODEL	42	2	1	3
HM + MCP	43	1	0	4
HM + ANN + MCP	44	0	0	4

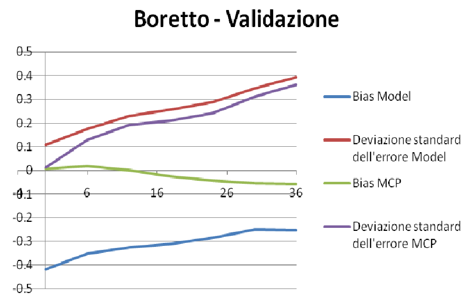
RESULTS
BORETTO (PO) – Forecasting horizon = 36 h



RESULTS

BORETTO (PO)





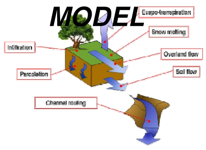
ORE	CALIBRAZIONE					
	FALSI ALLARMI		MANCATI ALLARMI		CORRETTI POSITIVI	
	MODELLO	MCP	MODELLO	MCP	MODELLO	MCP
0	0	0	2	0	16	18
6	1	0	3	0	15	18
12	1	0	3	0	15	18
18	5	3	3	0	15	18
24	0	0	2	0	16	18
30	0	2	2	1	16	17
36	0	2	2	1	16	17

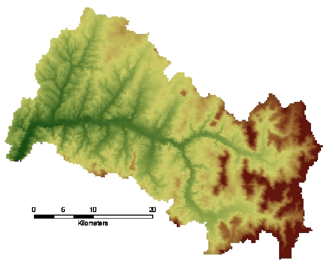
ORE	VALIDAZIONE					
	FALSI ALLARMI		MANCATI ALLARMI		CORRETTI POSITIVI	
	MODELLO	MCP	MODELLO	MCP	MODELLO	MCP
0	0	0	0	0	3	3
6	0	0	0	0	3	3
12	0	0	0	0	3	3
18	1	1	0	0	3	3
24	0	0	0	0	3	3
30	0	0	0	0	3	3
36	0	0	0	0	3	3

MCP: APPLICATION

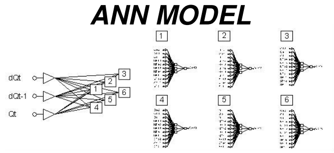
BARON FORK RIVER AT ELDON, OK, USA

TOPKAPI MODEL

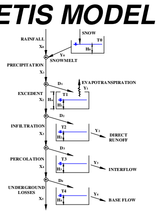




ANN MODEL



TETIS MODEL



01/10/1995 31/05/1997 31/01/1998 30/09/2002

ANN: CALIBRATION — VERIFIC. — VALIDATION

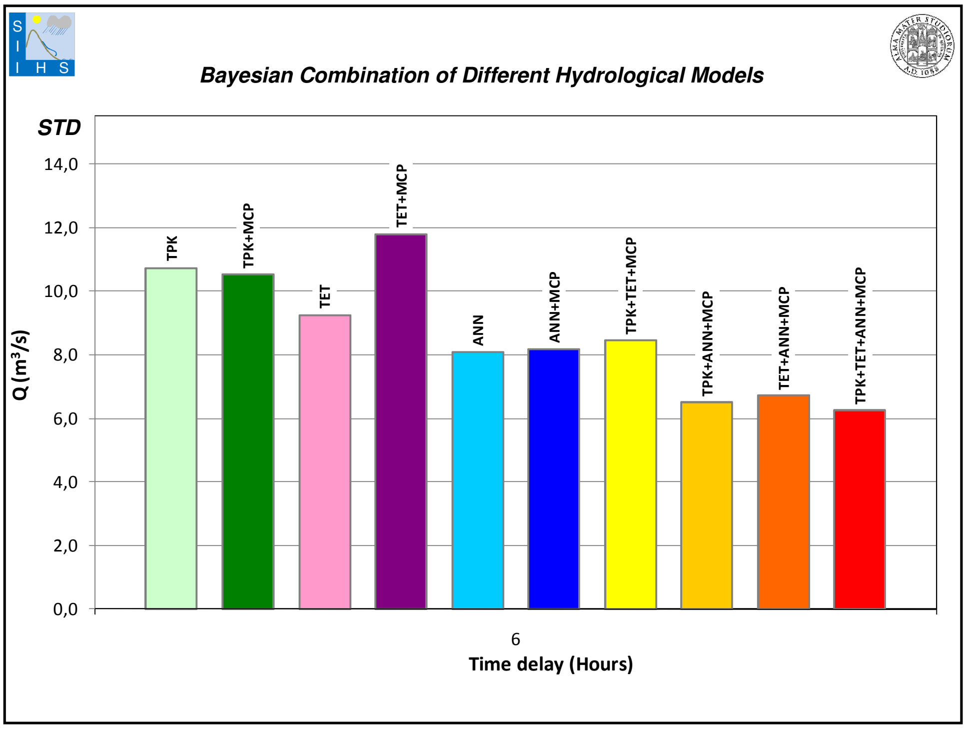
01/10/1995 31/05/1997 01/05/2000 30/09/2002

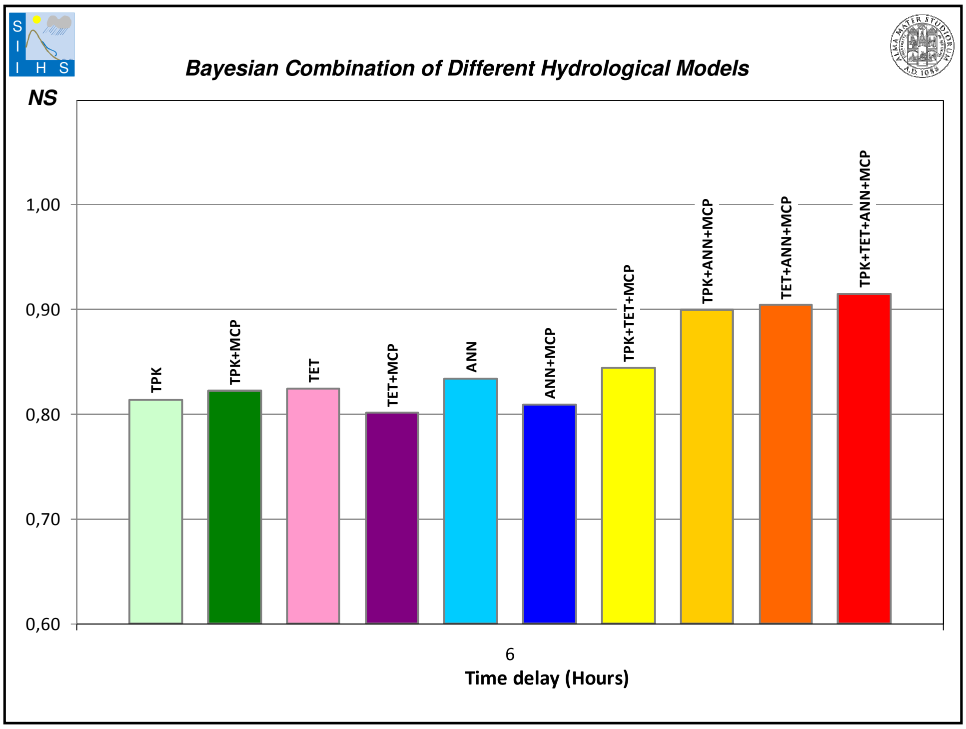
MCP: CALIBRATION — VALIDATION

Available data, provided by the NOAA's National Weather Service, within the DMIP 2 Project:

Gridded hourly precipitation and temperature data

Observed hourly discharge at Eldon


Bayesian Combination of Different Hydrological Models
STD
Q (m³/s)
14,0
12,0
10,0
8,0
6,0
4,0
2,0
0,0
TPK
TPK+MCP
TET
TET+MCP
ANN
ANN+MCP
TPK+TET+MCP
TPK+ANN+MCP
TET+ANN+MCP
TPK+TET+ANN+MCP
6
Time delay (Hours)


Bayesian Combination of Different Hydrological Models
NS
1,00
0,90
0,80
0,70
0,60
TPK
TPK+MCP
TET
TET+MCP
ANN
ANN+MCP
TPK+TET+MCP
TPK+ANN+MCP
TET+ANN+MCP
TPK+TET+ANN+MCP
6
Time delay (Hours)

CONCLUSIONS (1/3)

Time has passed since the concept of Predictive Uncertainty was introduced in Bayesian Statistics.

Unfortunately, limited operational use of PU can be found in the fields of Flood Forecasting, Flood Emergency Management and Water Resources Management mostly due to the widespread confusion on the PU definition and concepts.

CONCLUSIONS (2/3)

Whereas the use of HUPs for operational purposes is in progress, the use of Meteorological QPF for the estimation of PU has not yet reached a resonable level of acceptance.

This is due to two main reasons:

1)The first one is due to the lack of understanding of the operational use and of the real benefits deriving from incorporating PU in the decision process.

2) The second one relates to the "lack of will" shown by the meteorological organisations when requested to re-run their models on past data.

CONCLUSIONS (3/3)

In practice we hydrologists have the following homework: we must convince stakeholders and meteorologists of the real benefits deriving from the estimation and use of

Predictive Uncertainty

in

flood warning, flood emergency and water resources management