FROM DATA ASSIMILATION
TO MULTI-TEMPORAL UNCERTAINTY PROCESSORS

TO IMPROVE REAL-TIME FLOOD FORECASTING AND EMERGENCY MANAGEMENT

Ezio TODINI – University of Bologna

Gabriele COCCIA – Idrologia e Ambiente Srl

HYDROPREDICT 2012 – Vienna (Austria) - September 24 -27, 2012

EMERGENCY MANAGEMENT UNDER UNCERTAINTY

Frequently, engineers and decision makers face the problem of taking important decisions, such as preventive releases of water from a reservoir to avoid over-filling or evacuating an area threaten by an incoming flood wave, without perfect knowledge of what will actually happen.

These decisions may have serious consequences in terms of damages and casualties, which requires the decision makers to use all the information they can gather in order to increase their robustness, the reliability and the effectiveness.

PREDICTIVE UNCERTAINTY

Although well known in Statistics and Decision Theory, it is only in the last two decades that the concepts of

PREDICTIVE UNCERTAINTY

have emerged in hydrology and water resources together with the appropriate techniques allowing assessment and incorporation of the uncertainty information into the decision making process.

PLANNING UNDER UNCERTAINTY

Traditionally, planning decisions are taken on the basis of scenarios assumed for the future either as reasonable subjective hypotheses or as the result of some modeling exercise.

For instance climate change effects have been estimated using a series of General Circulation Models, and although their reliability is known to be rather low, adaptation measures are usually planned taking one or the other of the anticipated scenarios.

FROM PLANNING TO MANAGEMENT UNDER UNCERTAINTY

Once operational strategies and plans have being laid, and structural as well as non-structural measures set in place, water resources and flood emergency management will result from the combination of pre-laid plans with shorter terms scenarios or predictions, based on some model forecasts.

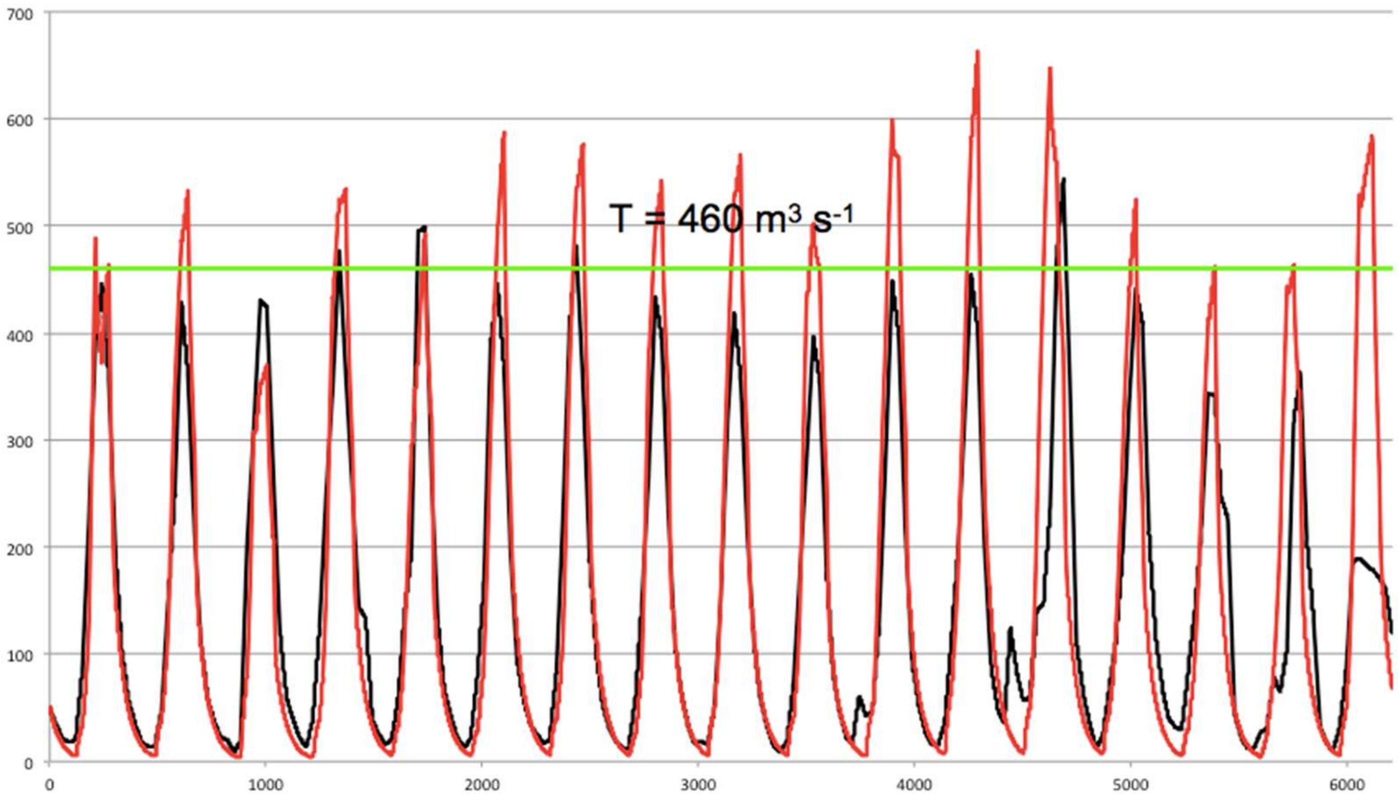
Today, given the vast availability hydrological and hydraulic models, most of planning and management decisions are based on models forecasts.

MODEL FORECASTS AND UNCERTAINTY

Unfortunately, model forecasts, although known to be imperfect, are traditionally used into the decision making processes as “deterministic” quantities without being associated to some cautious measure of uncertainty.

By doing so, decision makers, although aware that model forecasts are not perfectly representing future outcomes, are de facto assuming that the values of discharge, volume, water level, etc. will coincide with what will actually occur.

The threshold example



DECISION

WRONG = 12/17

RIGHT = 5/17

A decision is activated if the river discharge is higher than a pre-set threshold. A model is available. If the decision is taken using the model there is a high likelihood of taking the wrong decision.

MODEL FORECASTS AS VIRTUAL REALITY

The example allows to underline that a model forecast is not "reality", but rather "virtual reality", and as such should never be directly compared to real quantities such as thresholds, set on the basis of measured (real) quantities.

As a corollary of the previous consideration, one must realize that damages as well as benefits resulting from a management decisions only occur when the actual water level (reality) not the forecasted one (virtual reality) overtops the threshold.

RATIONAL DECISION MAKING

Rational decision making must then be based on the expected value of an utility (the Bayesian utility) function expressing the propensity of the decision maker at risk.

$$E\{U(y_t)\} = \int_0^{+\infty} U(y_t) f(y_t) dy_t$$

with

$$U(y_t) = \begin{cases} 0 & \forall y_t \leq y^* \\ g(y_t - y^*) & \forall y_t > y^* \end{cases}$$

or simply

$$Prob\{y_t \geq y^*\} = 1 - \int_{y^*}^{+\infty} f(y_t) dy_t$$

A MEASURE OF PREDICTIVE UNCERTAINTY

Unfortunately, for time $t > t_0$ the density $f(y_t)$ is either unknown or is so flat (climatological distribution) that adds no or very little information.

Alternatively, what can be done is to assess the conditional distribution of the future outcome, based on all the available information, which is generally contained in one or more model forecasts.

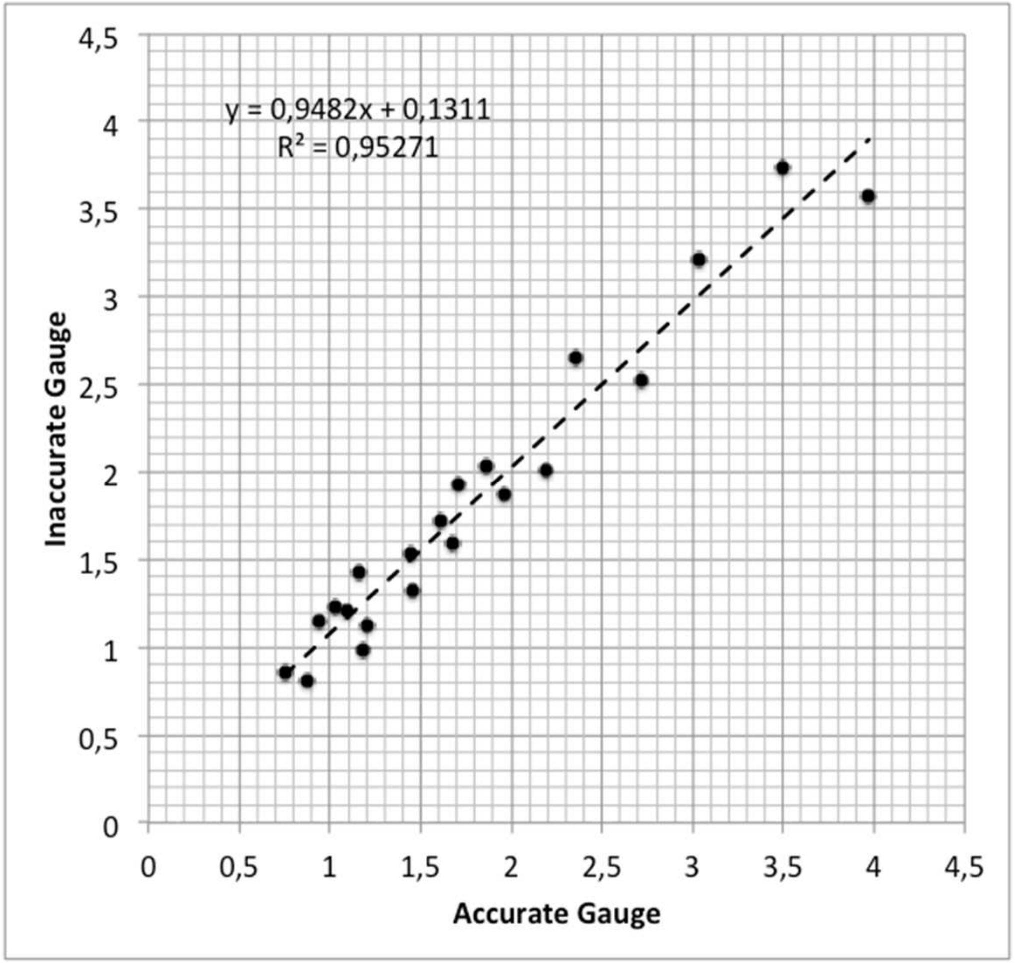
$$f\left(y_t \middle| \hat{y}_{t|t_0}^{(1)}, \hat{y}_{t|t_0}^{(2)}, \dots, \hat{y}_{t|t_0}^{(m)}\right)$$

This is what is called “a measure of predictive uncertainty”

VALIDATION UNCERTAINY

Two gauges are available:

- the first one is very accurate
- the second one is coarse



<u>First Problem</u>

Assess the quality of the coarse instrument using the accurate one.

It can be done via Linear Regression

PREDICTIVE UNCERTAINY

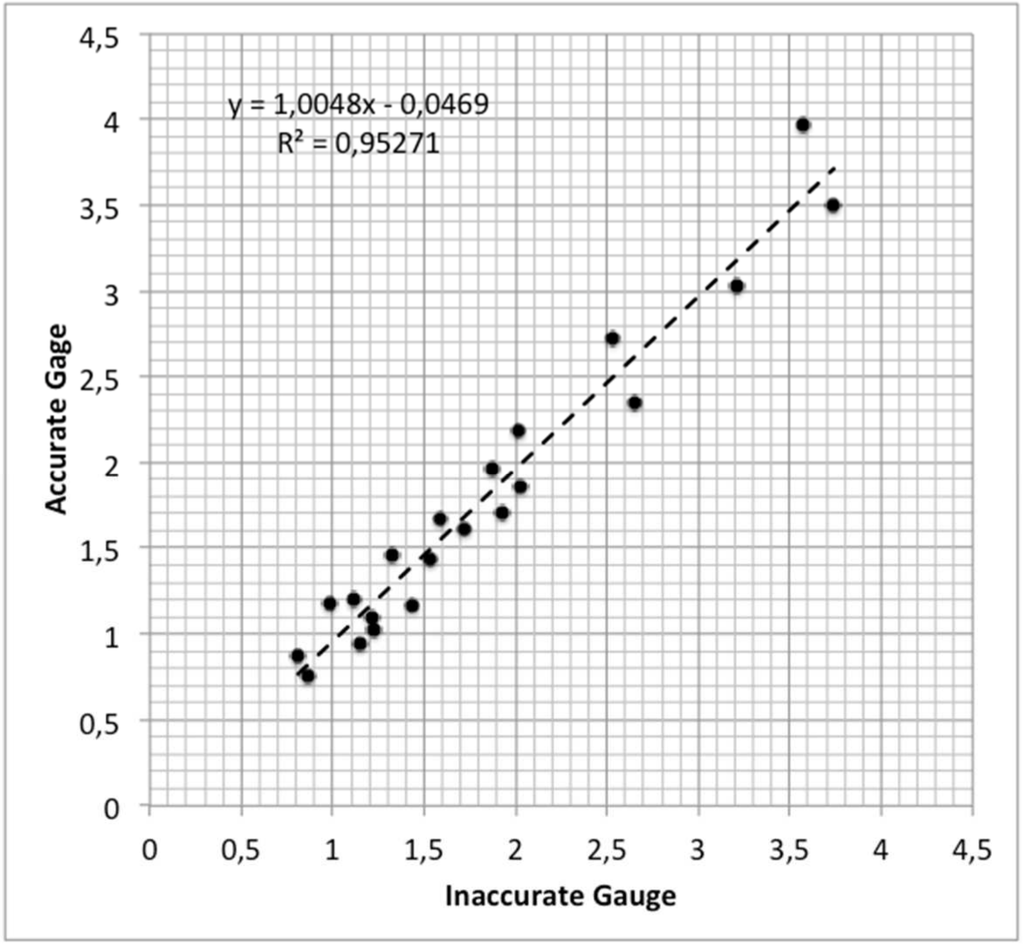
BUT

At a certain point in time the accurate instrument breaks.

Second Problem

Assess the expected true but unknown value using the coarse instrument, and its estimation uncertainty.

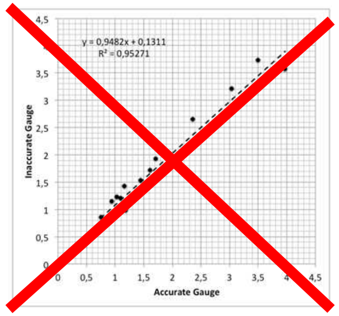
It can still be done via Linear Regression but the other way round



DATA ASSIMILATION

Two gauges are available:

- Both gauges are affected by measurement errors



<u>First and 1/2 Problem</u>

Assess the quality of the coarse instrument using the accurate one.

It cannot be solved with linear regression. One must solve it using data assimilation techniques such as Bayesian combination and or Filters (KF, EnKF, Particle Filter, etc.)

Validation vs Predictive Uncertainty

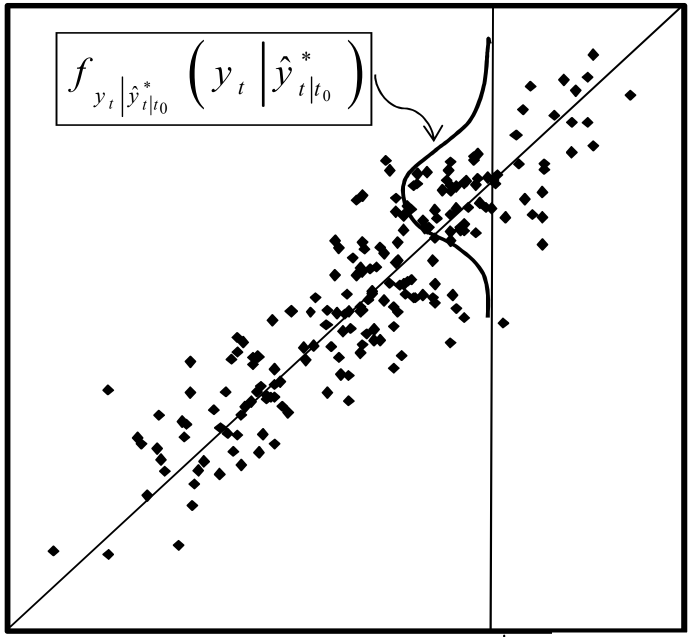
Validation Uncertainty

Uncertainty of model predictions knowing (conditional on) the observations

Predictive Uncertainty

Uncertainty of future occurrences Knowing (conditional on) the model prediction

Observed Values



$\hat{y}^*_{t|t_0}$ Model Prediction

$$\boxed{Prob\left\{\hat{y}_{t|t_0} = \hat{y}^*_{t|t_0}\right\} = 1}$$

Predictive Uncertainty : the definition

A definition of **Predictive Uncertainty** can be the following:

Predictive uncertainty is the expression of **our assessment** of the **probability of occurrence of a future (real)** event conditional upon all the knowledge available up to the present and the information we were able to acquire through a **learning inferential process**.

Validation vs Predictive Uncertainty

Therefore, we must clearly distinguish between two types of uncertainty, namely:

Validation Uncertainty and **Predictive Uncertainty**.

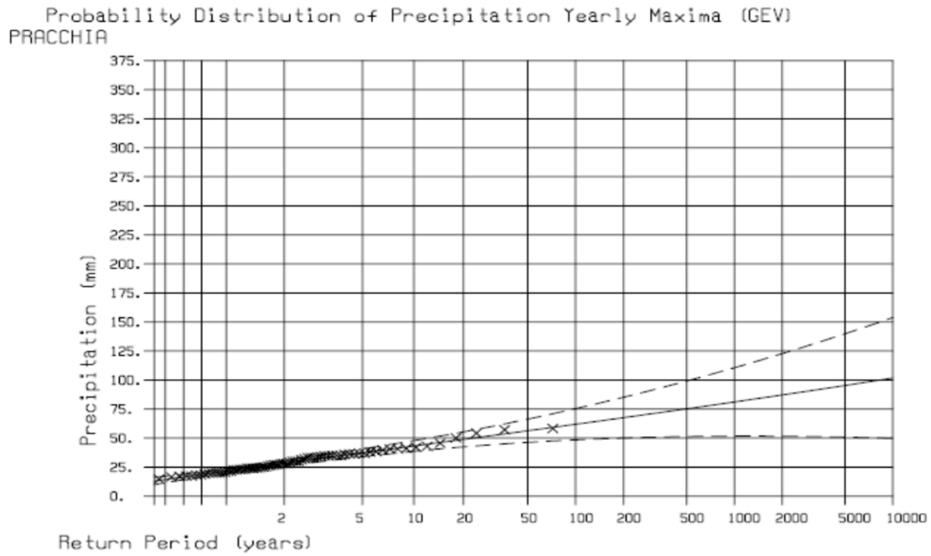
Validation Uncertainty represents how well our model(s) reproduce the observations and is affected by all sorts of errors (measurement, model, parameters, initial and boundary conditions).

Predictive Uncertainty represents the probability of the occurrence of a future event given (conditional to) the observations and the model(s) forecasts, where the model predictions are taken as known and not uncertain quantities.

Validation vs Predictive Uncertainty

Meteorological Ensembles

are a measure of Validation Uncertainty, while



Climatological Distributions

or

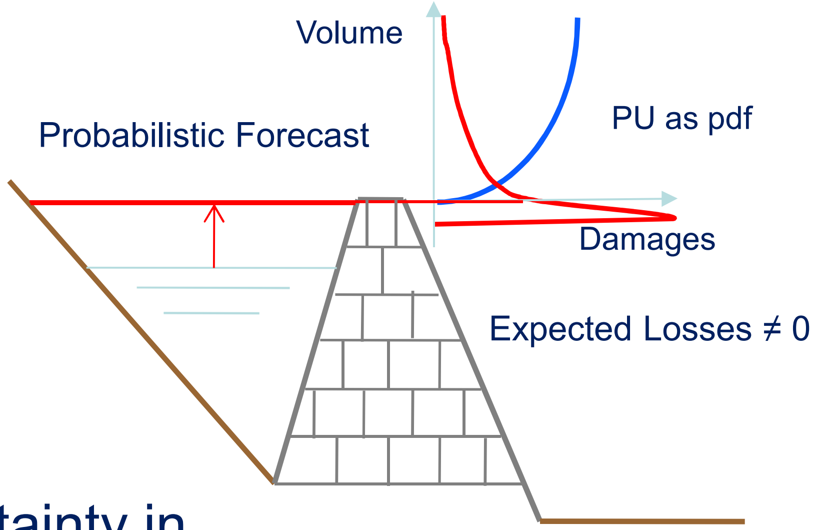
Extreme Value Distributions

are measures of Predictive Uncertainty, although **non conditional on real time information**.

The need for Assessing Predictive Uncertainty
The Reservoir Management Case

In the Reservoir Management Problem it is easy to show that Deterministic Forecasts lead to wrong estimates of losses.

In this simple example losses occur if the reservoir is overtopped. If the Deterministic Forecast reaches the top level of the reservoir the estimated losses are equal to zero.



This is obviously wrong because the uncertainty in the forecast implies that the “expected value” of losses is not null. The “expected value” of losses can be estimated if and when an assessment of Predictive Uncertainty will be available.

Operational use of Predictive Uncertainty

Decision Theory teaches us how to make use of Predictive Uncertainty. When one needs to take decisions under uncertainty he must:

1. Describe the uncertainty, namely by assess PU in terms of a probability distribution function conditional on latest information;
2. Define an utility function, from a description of the decision maker propensity towards risk to more complex loss/ benefit functions involving actual costs;
3. Marginalise the effect of uncertainty by integrating the product of the probability times the utility function: the expected value of the utility function.
4. Use the resulting expected value within the decision making process by trade-off analysis or by optimization.

Most used approaches to PU assessment

1) The Hydrological Uncertainty Processors

Krzysztofowicz, 1999; Krzysztofowicz and Kelly, 2000

2)The Quantile Regression

Koenker and Basset, 1978; Koenker, 2005

3) The Bayesian Model Averaging

Raftery et al., 2003

4) The Model Conditional Processor

Todini, 2008.

1) The Hydrological Uncertainty Processors

Krzysztofowicz, 1999; Krzysztofowicz and Kelly, 2000

After converting the observations and the model forecasts available for the historical period into the Normal space, HUP combines the prior predictive uncertainty (in this case derived as an AR1 model) with a Likelihood function in order to obtain the posterior density of the predictand conditional to the model forecasts via a Bayesian combination approach.

The result is:

$$f\left(y_t \,\middle|\, \hat{y}_{t|t_0}^{(MOD)}, \hat{y}_{t|t_0}^{(AR1)}\right)$$

2)The Quantile Regression

Koenker and Basset, 1978; Koenker, 2005

Instead of estimating the full density, QR estimates one by one the quantiles of the predictive density using linear or non-linear models by appropriately weighting the observations. Parameters are estimated by means of Linear Programming.

Limitations: the large number of parameters to be estimated (at least 2 per quantile).

3) The Bayesian Model Averaging

Raftery et al., 2003

Bayesian Model Averaging (BMA) is not exactly estimating what was defined as Predictive Uncertainty, namely the conditional probability density, the ratio between the joint and the marginal densities,

$$f\left(\eta_t \left| \hat{\eta}_{t|t_0}^{(MOD1)}, \hat{\eta}_{t|t_0}^{(MOD2)}, \dots, \hat{\eta}_{t|t_0}^{(MODm)}\right.\right) = \frac{f\left(\eta_t, \hat{\eta}_{t|t_0}^{(MOD1)}, \hat{\eta}_{t|t_0}^{(MOD2)}, \dots, \hat{\eta}_{t|t_0}^{(MODm)}\right)}{\int_{-\infty}^{+\infty} f\left(\eta_t, \hat{\eta}_{t|t_0}^{(MOD1)}, \hat{\eta}_{t|t_0}^{(MOD2)}, \dots, \hat{\eta}_{t|t_0}^{(MODm)}\right) d\eta_t}$$

But rather related quantity based on Bayesian mixture distributions:

$$g\left(y_t \left| \hat{y}_{t|t_0}^{(MOD1)}, \hat{y}_{t|t_0}^{(MOD2)}, \dots, \hat{y}_{t|t_0}^{(MODm)}\right.\right) = E\left\{f_1\left(y_t \left| \hat{y}_{t|t_0}^{(MOD1)}\right.\right), f_2\left(y_t \left| \hat{y}_{t|t_0}^{(MOD2)}\right.\right), \dots, f_m\left(y_t \left| \hat{y}_{t|t_0}^{(MODm)}\right.\right)\right\}$$

4) The Model Conditional Processor

Todini, 2008

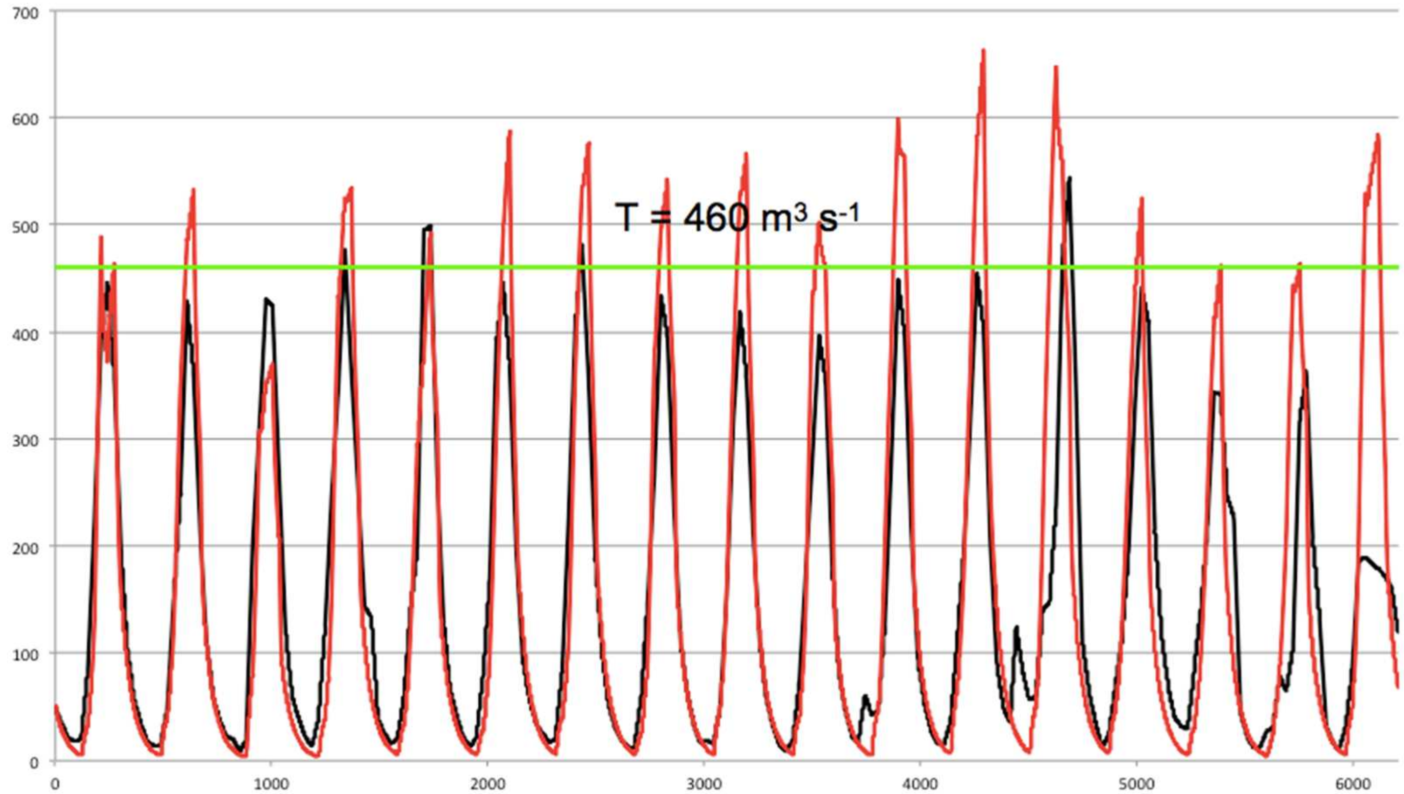
After converting the observations and the models into the Normal space using the NQT, builds the joint density (a multi-Normal density) and analytically derives the conditional density following the definition:

$$f\left(\eta_t \middle| \hat{\eta}_{t|t_0}^{(MOD1)}, \hat{\eta}_{t|t_0}^{(MOD2)}, \dots, \hat{\eta}_{t|t_0}^{(MODm)}\right) = \frac{f\left(\eta_t, \hat{\eta}_{t|t_0}^{(MOD1)}, \hat{\eta}_{t|t_0}^{(MOD2)}, \dots, \hat{\eta}_{t|t_0}^{(MODm)}\right)}{\int_{-\infty}^{+\infty} f\left(\eta_t, \hat{\eta}_{t|t_0}^{(MOD1)}, \hat{\eta}_{t|t_0}^{(MOD2)}, \dots, \hat{\eta}_{t|t_0}^{(MODm)}\right) d\eta_t}$$

The final result is equivalent to a linear regression in the Normal space

The use of Predictive Uncertainty

The threshold example



DECISION

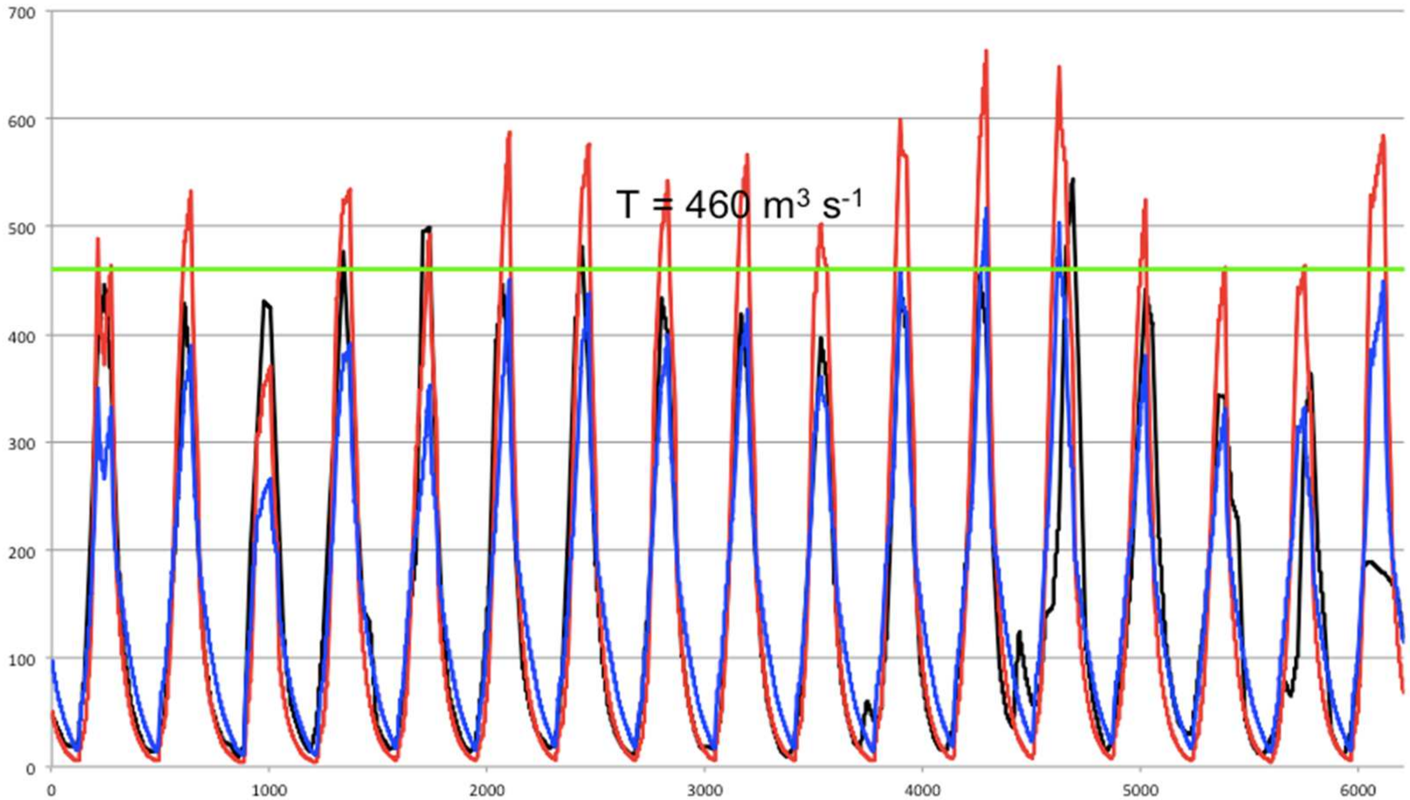
WRONG = 12/17

RIGHT = 5/17

The use of the expected conditional value noticeably improves decision reliability

The use of Predictive Uncertainty

The threshold example



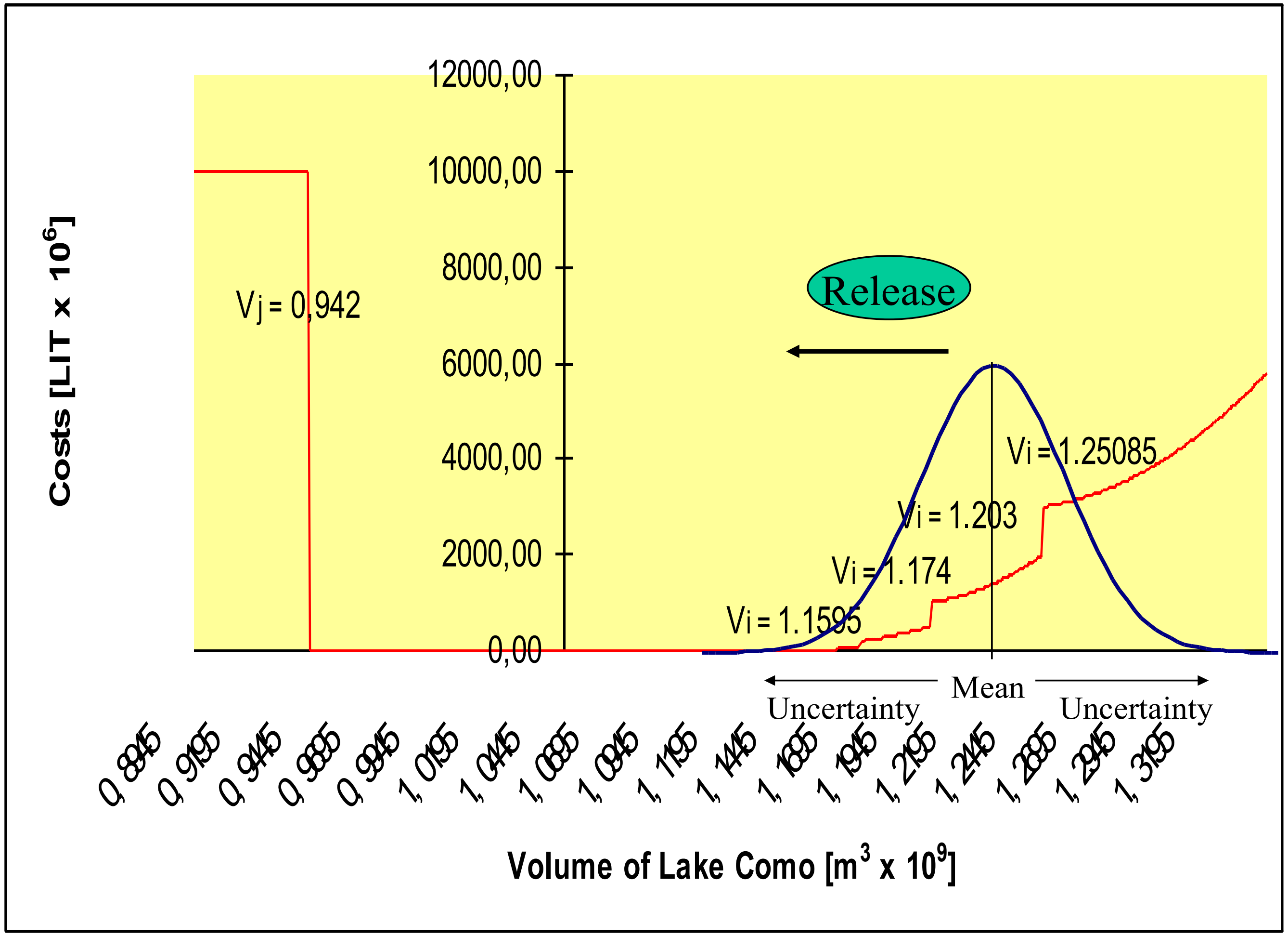
DECISION

WRONG = 4/17

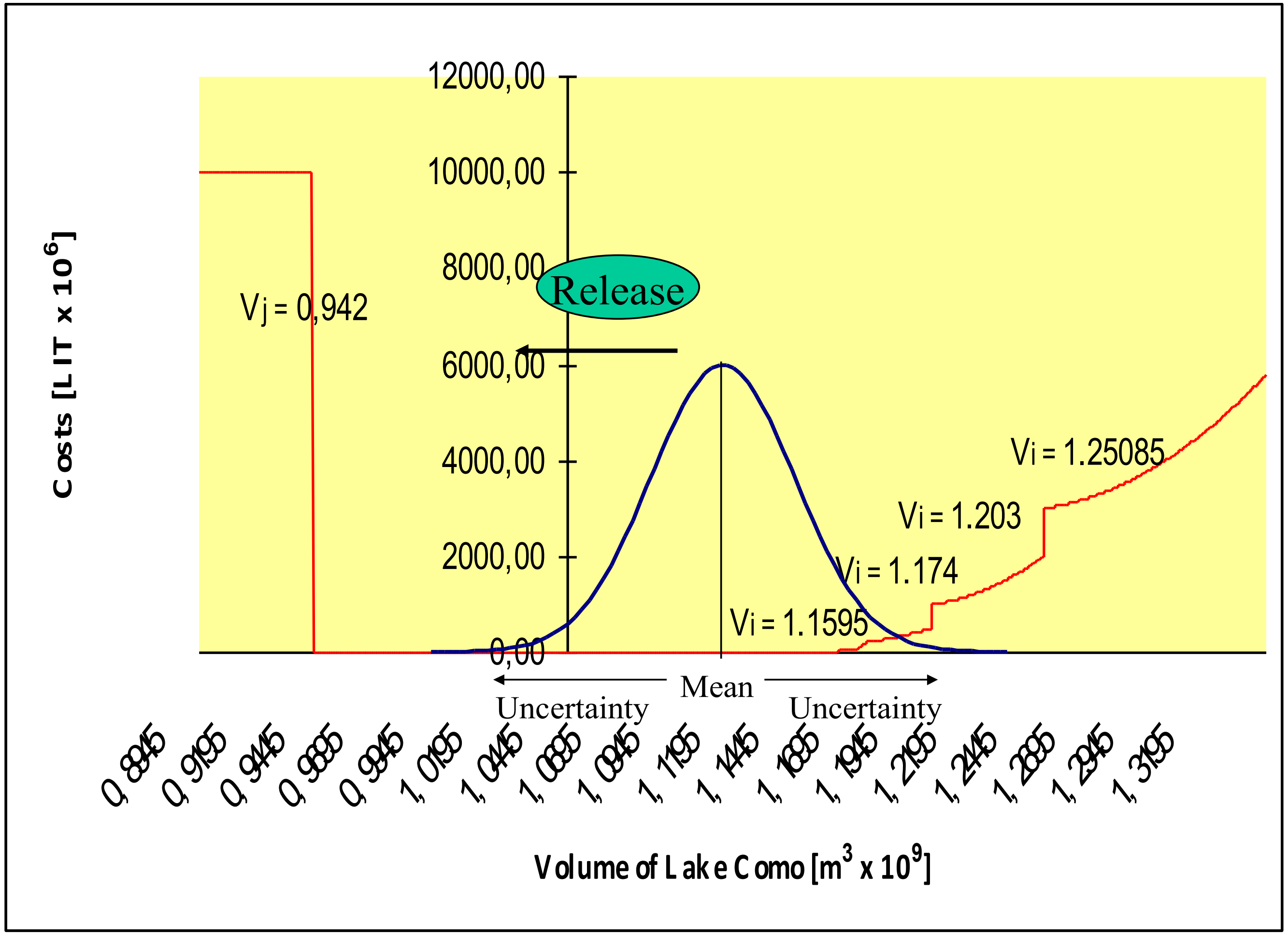
RIGHT = 13/17

The use of the expected conditional value noticeably improves decision reliability

The full use of Predictive Uncertainty

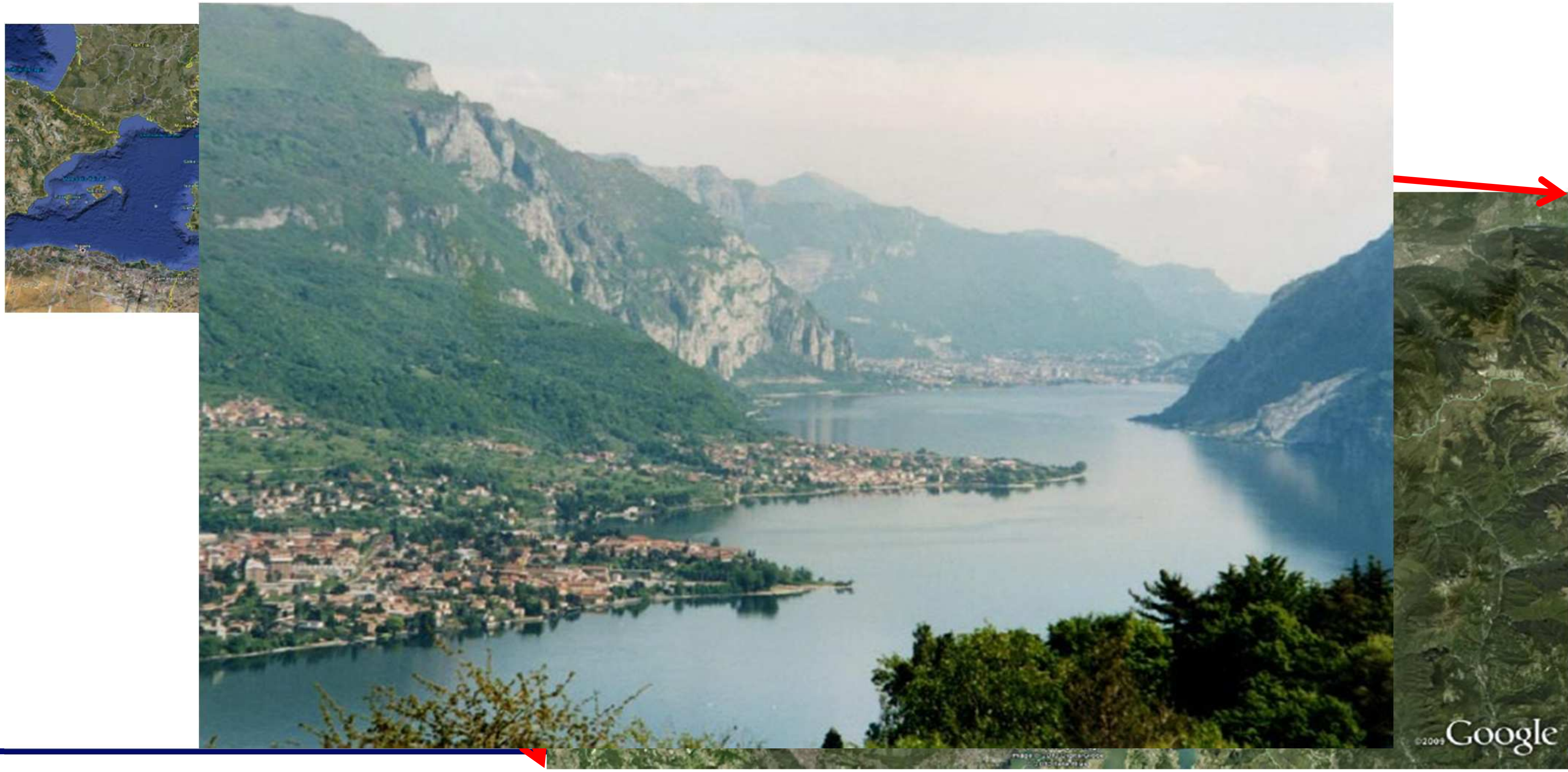


The full use of Predictive Uncertainty



The use of Predictive Uncertainty

To improve management of the Lake Como in Italy



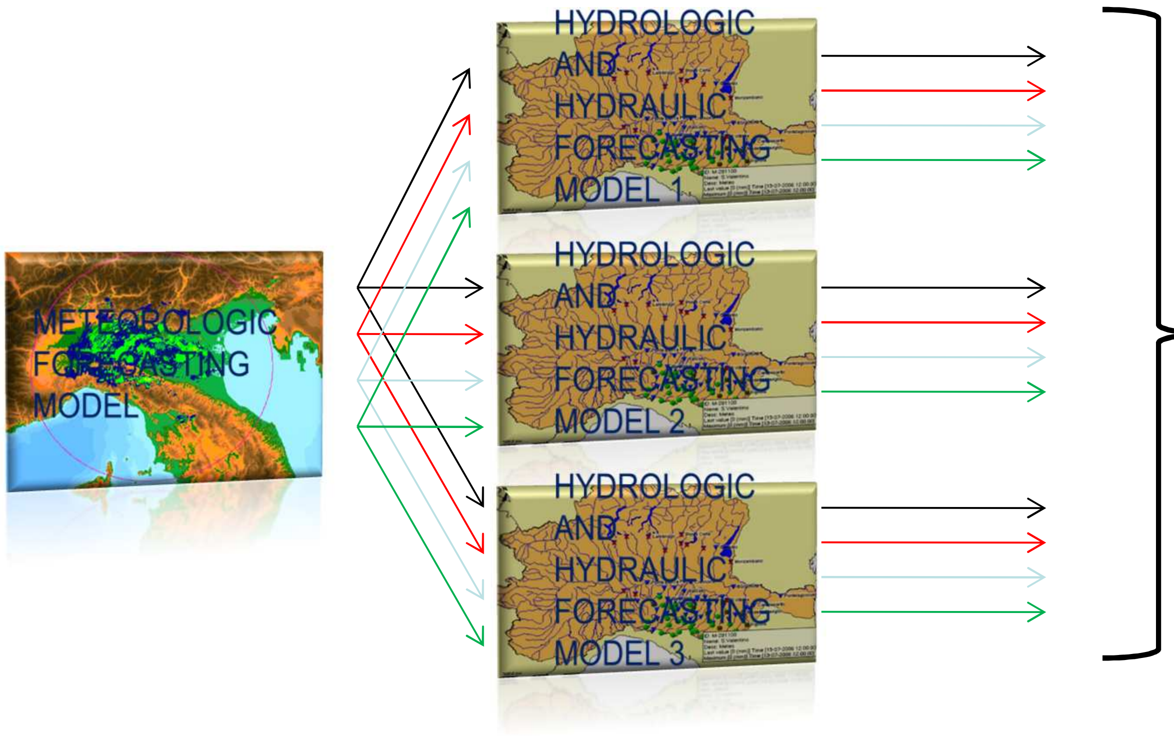
Managing the Lake Como by using Predictive Uncertainty

Results obtained by simulating 15 years of operations from January 1st, 1981 to December 31st, 1995

Water Level	Number of Days	
	Historical	Optimized
<-40 cm	214	0
≥ 120 cm	133	54
≥ 140 cm	71	32
≥ 173 cm	35	11
Water Deficit	890.27 10^6 m^3	694.49 10^6 m^3
Energy Production increased by 3%		

How to reconcile the different hydrological models: the multi-model approach

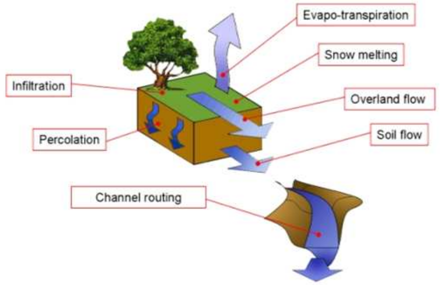
The Uncertainty Processors also allow to sinthesize several models and ensemble forecasts

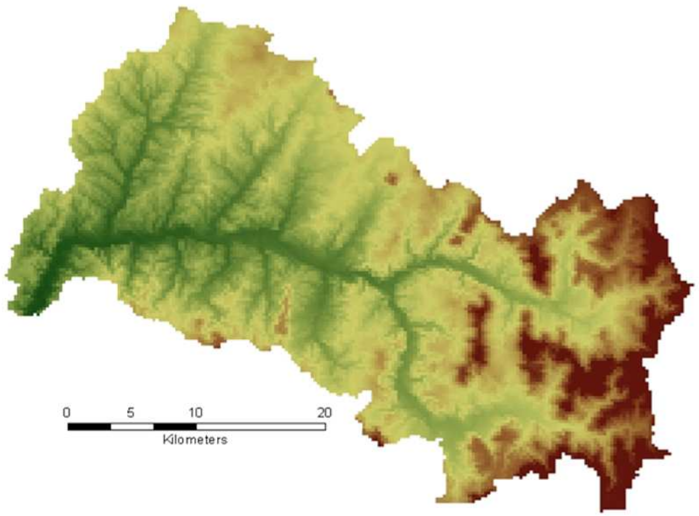


$$f\left(y \middle| \hat{y}_{i,i=1..N}\right)$$

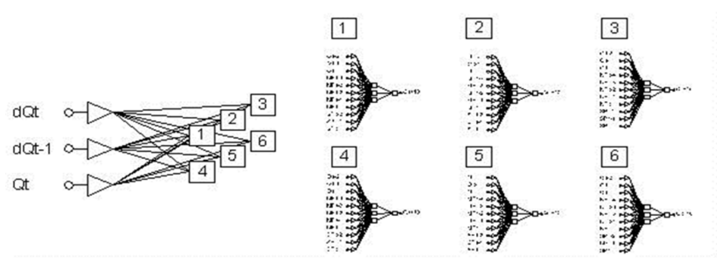
BARON FORK RIVER AT ELDON, OK, USA

TOPKAPI MODEL

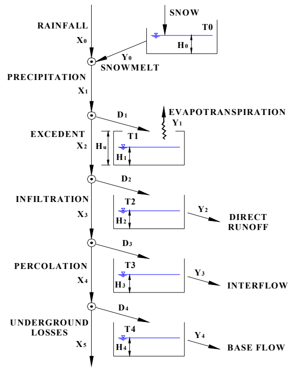




ANN MODEL



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

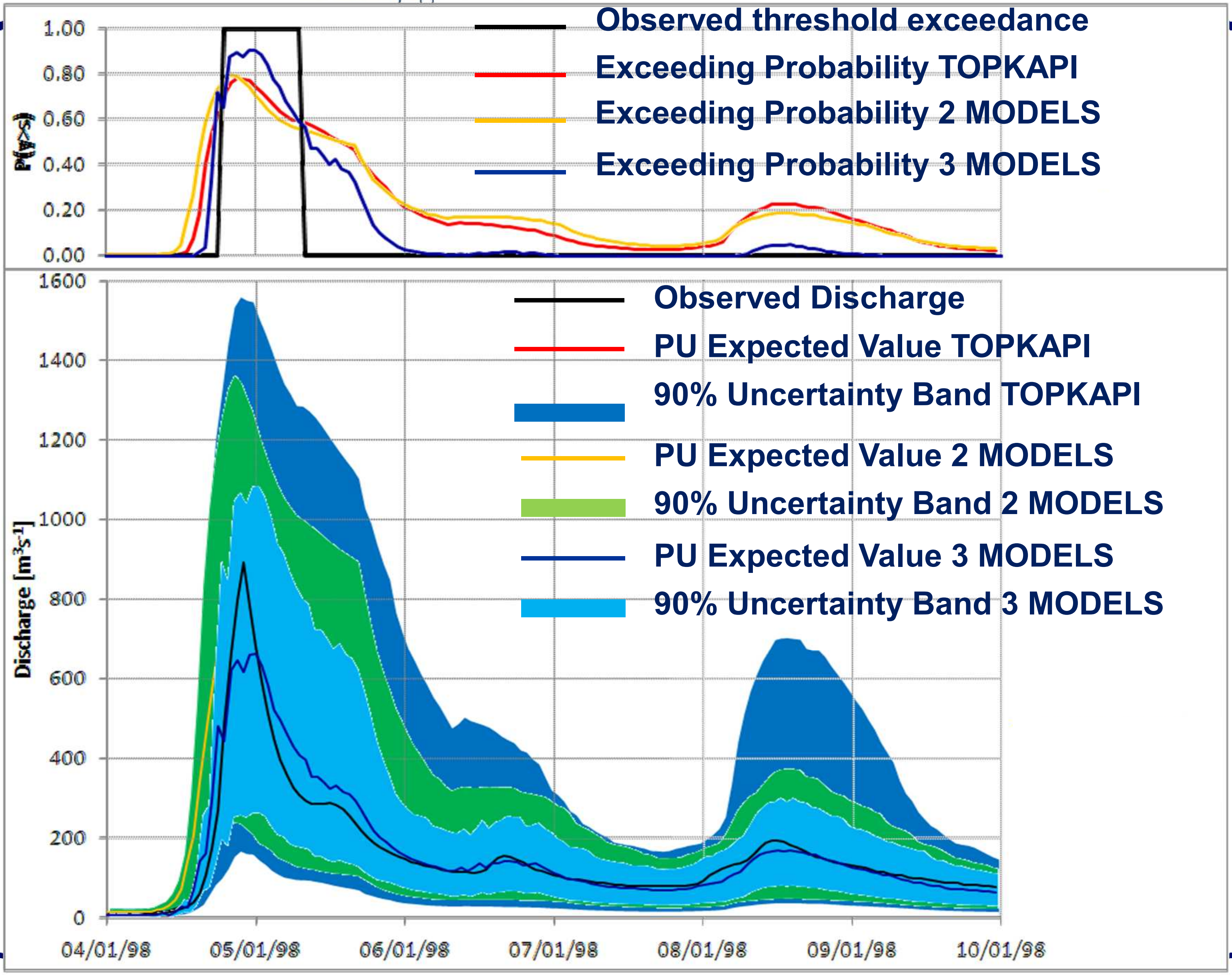
1 MODEL:
TOPKAPI

VS

2 MODELS:
TETIS +
TOPKAPI

VS

3 MODELS:
TETIS +
TOPKAPI + ANN



THE MULTI-TEMPORAL APPROACH

The multi-temporal approach can give an answer to the following important questions:

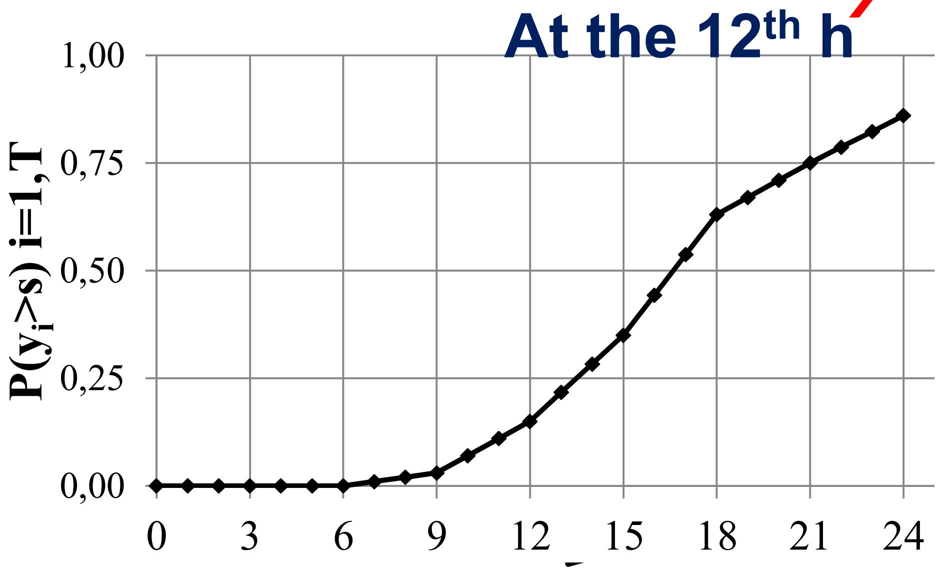
What is the probability of an event in the next 24 hours?

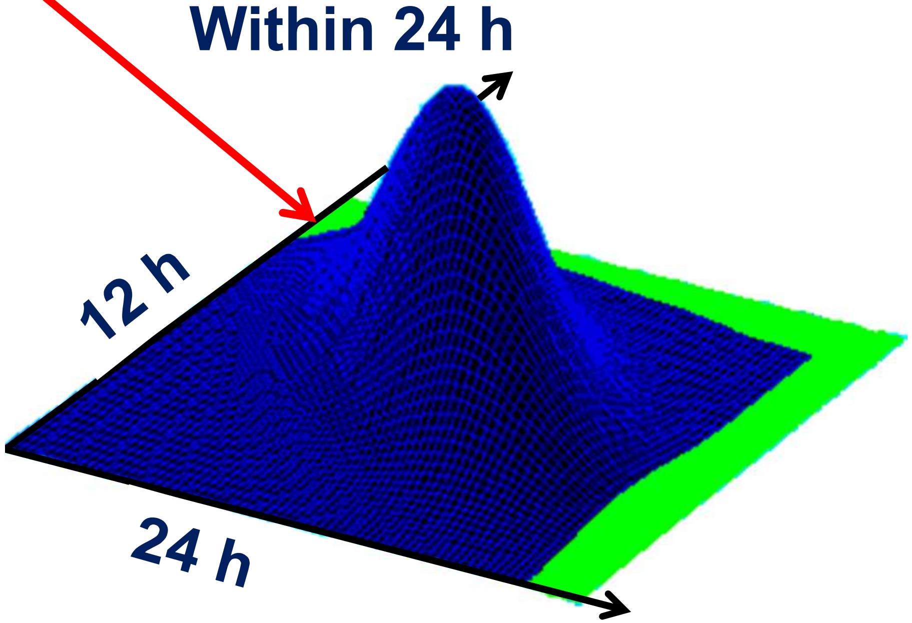
At what time it will most likely occur?

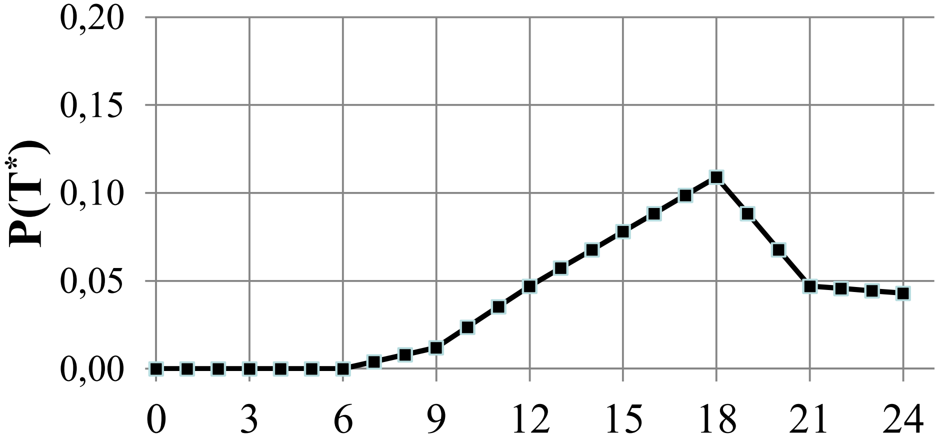
THE MULTI-TEMPORAL APPROACH

$$P\left(y_{t;t=1..T} > a \mid \hat{y}_{t,k;t=1..T;k=1..N}\right) = P\left(y_t > a \mid \hat{y}_{t,k}\right)$$

$$= 1 - \int_{-\infty}^{a} \cdots \int_{-\infty}^{a} f\left(y_{t;t=1..T} \middle| \hat{y}_{t,k;t=1..T;k=1..N}\right) dy_1 \ldots dy_T$$



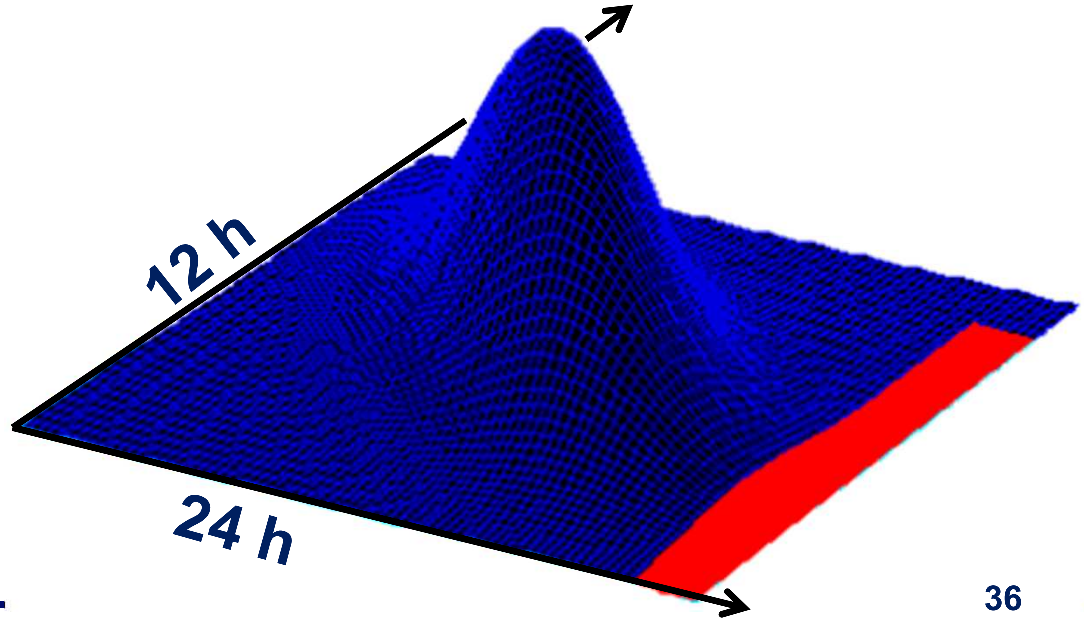




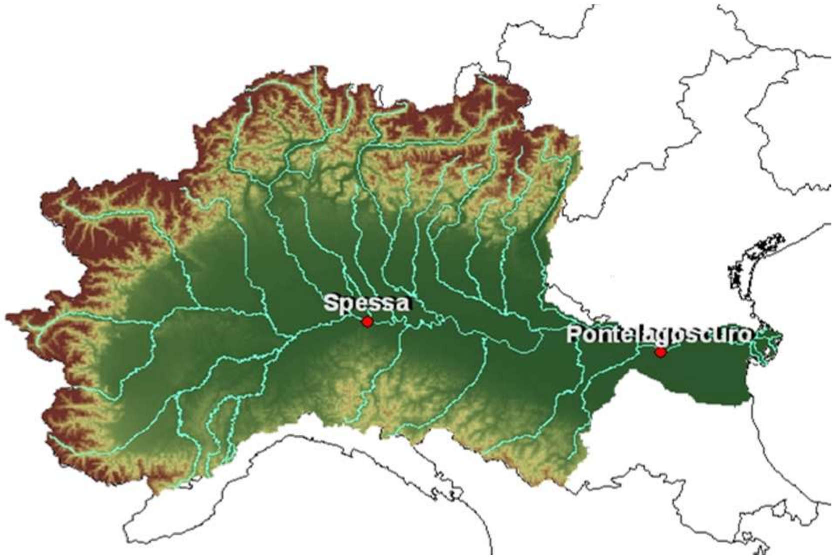
Exact Exceedance time (T*) probability

$$P(T^*) \propto \frac{\Delta P(y_t > a \mid \hat{y}_{t,k})}{\Delta t}$$

Which is the expected time of exceedance within the next 24 hours?



PO at PONTELAGOSCURO and PONTE SPESSA



01/05/2000 30/06/2004 20/01/2009

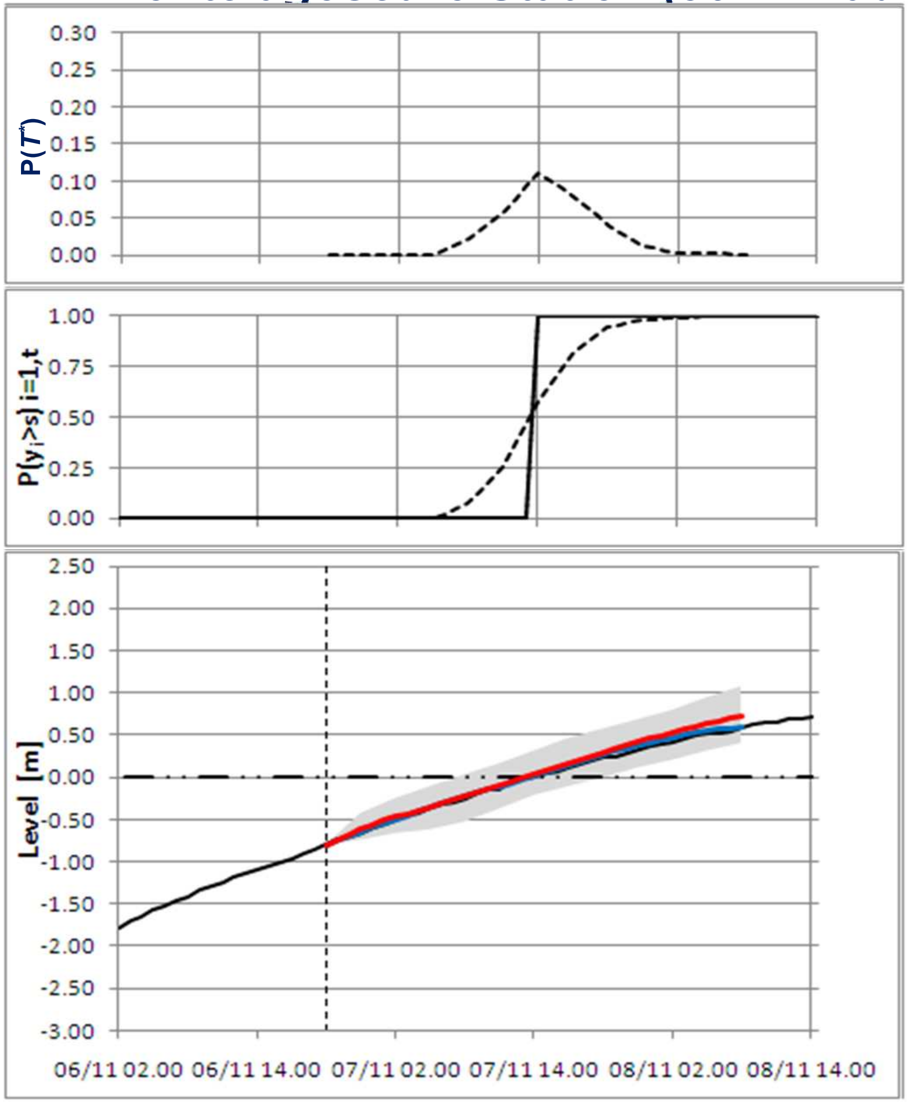
MCP: CALIBRATION
VALIDATION

Available data, provided by the Civil Protection of Emilia Romagna Region, Italy:

Forecasted hourly levels:
forecast lead time 24 h.

Observed hourly levels

Pontelagoscuro Station (36 h in advance)

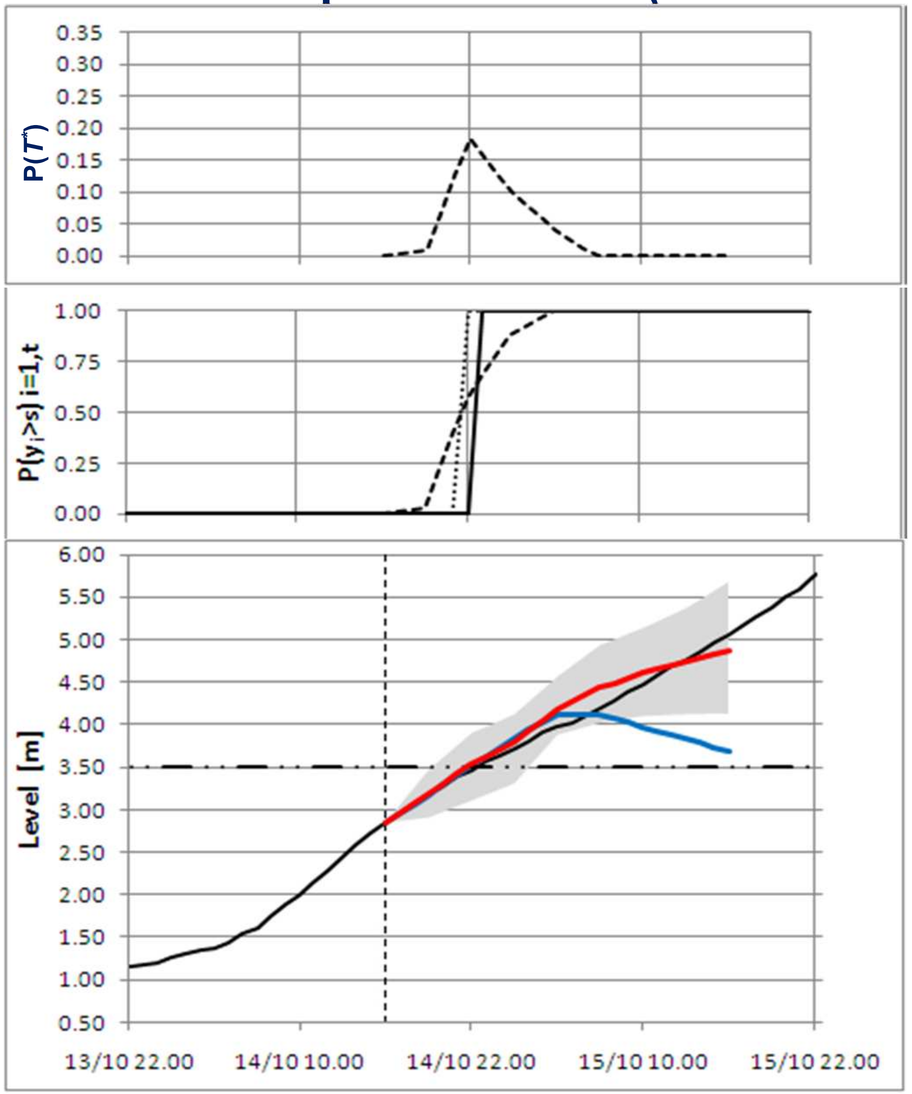


Exact Exceedance Time Probability

Cumulative Exceedance Probability

90% Uncertainty Band with MULTI-TEMPORAL APPROACH

Ponte Spessa Station (24 h in advance)



Exact Exceedance Time Probability

Cumulative Exceedance Probability

90% Uncertainty Band with MULTI-TEMPORAL APPROACH

CONCLUSIONS

- The new probabilistic approaches make use of the forecasting models as incremental pieces of information aimed at allowing decision makers to reliably take correct decisions also using not one but several hydrological models.

- These approaches, and in particular the use of Predictive Uncertainty Processors have produced several successful operational real time flood forecasting and management systems.

- There is still much work to be done in order to guarantee that the produced predictive density is the correct one.

Thank you
for your attention

ezio.todini@unibo.it