APPLICATION OF A BAYESIAN PROCESSOR FOR PREDICTIVE UNCERTAINTY ASSESSMENT IN REAL TIME FLOOD FORECASTING

Gabriele Coccia[1], Felix Francés[2], Juan Camilo Múnera[2] and Ezio Todini[1]
(1) University of Bologna, Bologna, Italy
(2) Polytechnic University of Valencia, Valencia, Spain

European Geosciences Union General Assembly 2010, Vienna, May 2010

DECISION MAKING UNDER UNCERTAINTY

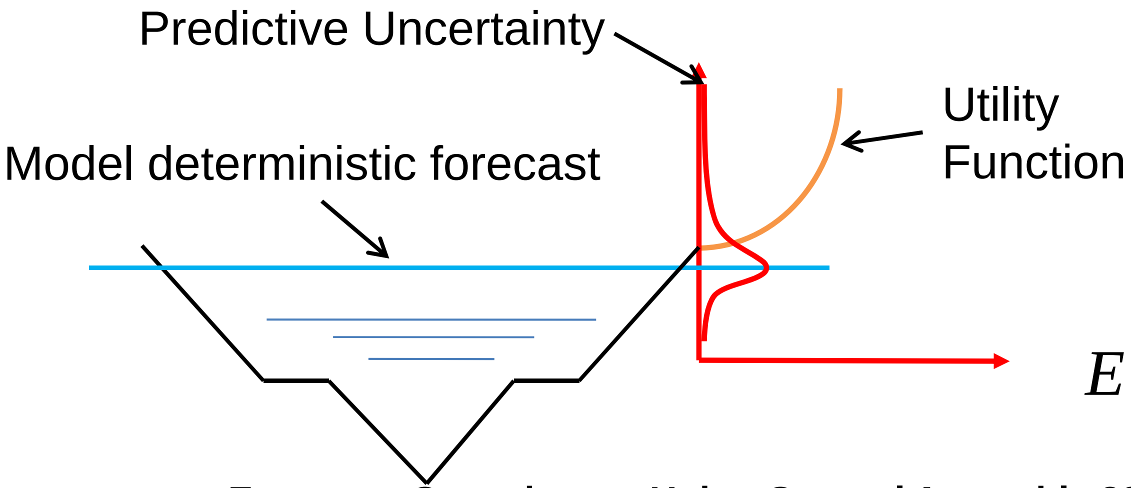
In many operational problems:

- Flood warning;
- Flood emergency management;
- Reservoir management;
- Etc.

decision makers must take important decisions under the uncertainty of future events.

According to the Decision theory, in order to take a rational decision it is necessary to:

1. Define an Utility Function in accordance with the Decision Maker
2. **Quantify the probability density of the future event**
3. Maximize the expected value of the Utility Function



$$U\left(E\{x\}\right)=0$$

$$E\{U\left(x\right)\}\neq 0$$

$$E\{U\left(x\right)\}=\int_0^{+\infty}U(x)\,f(x)\,dx$$

THE DEFINITION OF PREDICTIVE UNCERTAINTY

Predictive Uncertainty can be defined as the probability of occurrence of a future value of a predictand (such as water level, discharge or water volume) conditional on all the information that can be obtained on the future value, which is typically embodied in one or more meteorological, hydrological and hydraulic model forecasts.

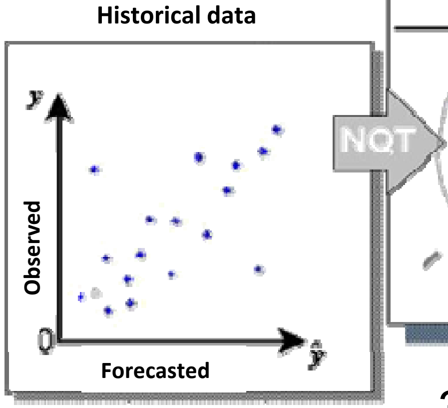
Predictive Uncertainty must be quantified in terms of probability distribution.

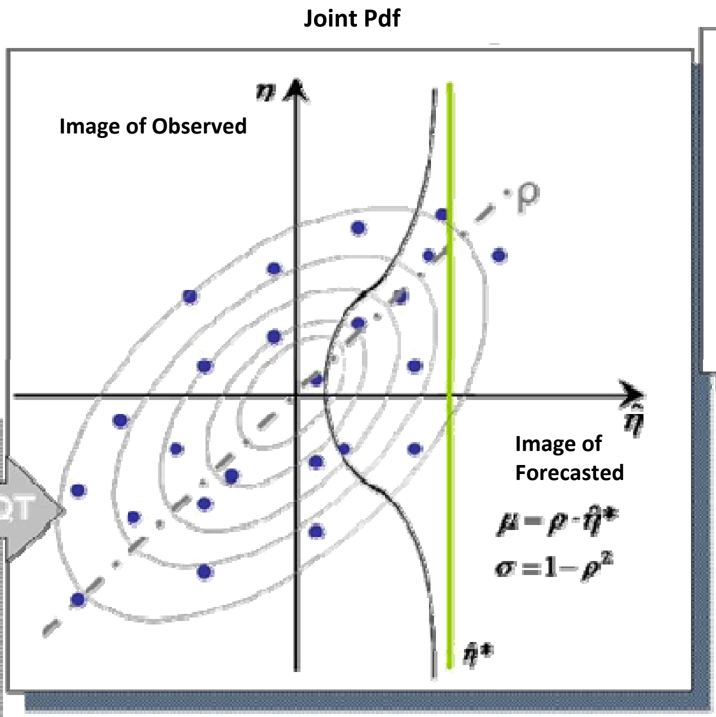
If the available information is a model forecast, Predictive Uncertainty can be written and here will be called as:

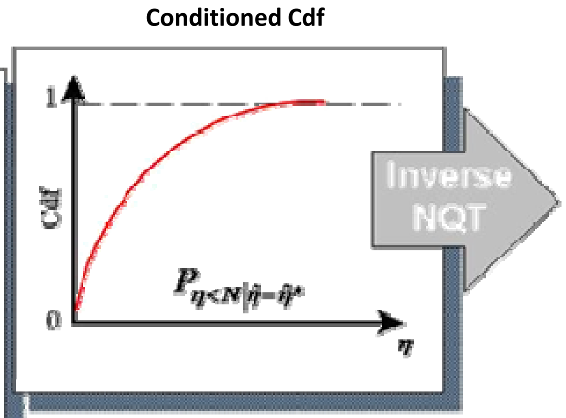
$$f\left(y\middle|\hat{y}\right)$$

MODEL CONDITIONAL PROCESSOR (MCP): METHODOLOGY DESCRIPTION

1) Conversion from the real world to the Normal Space using the NQT







$E\left[y \mid \hat{y} = \hat{y}^*\right]$

2) Joint Pdf is assumed to be a Normal Bivariate Distribution

3) Predictive Uncertainty is obtained by the Bayes Theorem and its mean and variance are:

$$\mu_{\eta|\hat{\eta}} = \rho_{\eta\hat{\eta}} \hat{\eta}^*$$

$$\sigma^2_{\eta|\hat{\eta}} = 1 - \rho^2_{\eta\hat{\eta}}$$

4) The expected value of the Predictive Uncertainty is computed sampling the probability density function in the Normal Space and reconverting the obtained quantiles by the Inverse NQT:

$$E\left\{y \mid \hat{y} = \hat{y}^*\right\} =$$

$$= E\left\{NQT^{-1}\left[f\left(\eta \middle| \hat{\eta} = \hat{\eta}^*\right)\right]\right\}$$

MCP: PROBABILITY TO EXCEED A THRESHOLD

The knowledge of the future event probability distribution allows to easily extrapolate the probability to exceed a threshold value, such as an alert level.

This is an important information when dealing with the decision about giving or not an alarm in emergency managing.

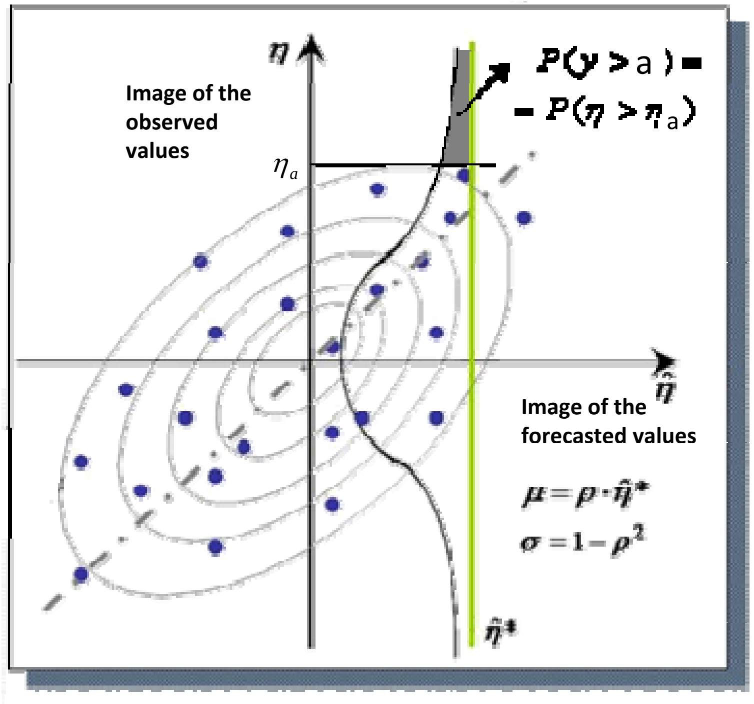
It can be directly computed from the Predictive Uncertainty, as its integral above the threshold *a*.

$$P(y > a \mid \hat{y} = \hat{y}^*) =$$

$$= \int_a^\infty f\left(y \middle| \hat{y} = \hat{y}^*\right) dy =$$

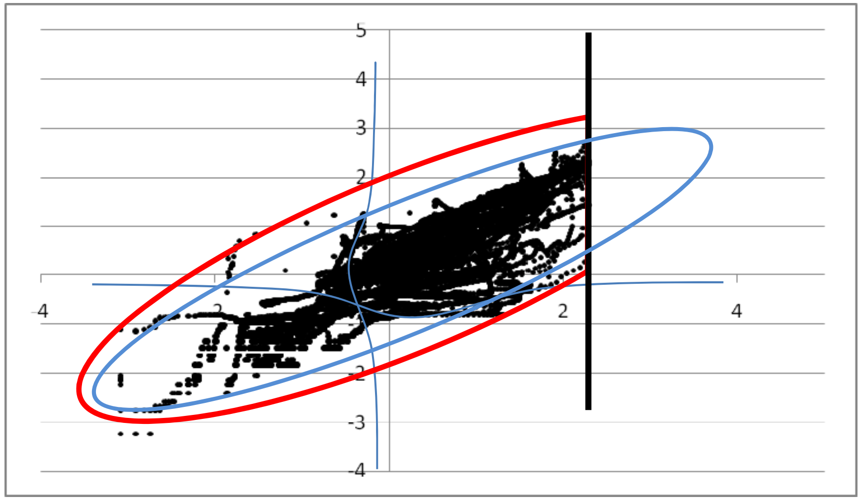
$$= \int_{\eta_a}^\infty f\left(\eta \middle| \hat{\eta} = \hat{\eta}^*\right) d\eta$$

Joint and Conditioned Pdf



MCP: IMPROVEMENT

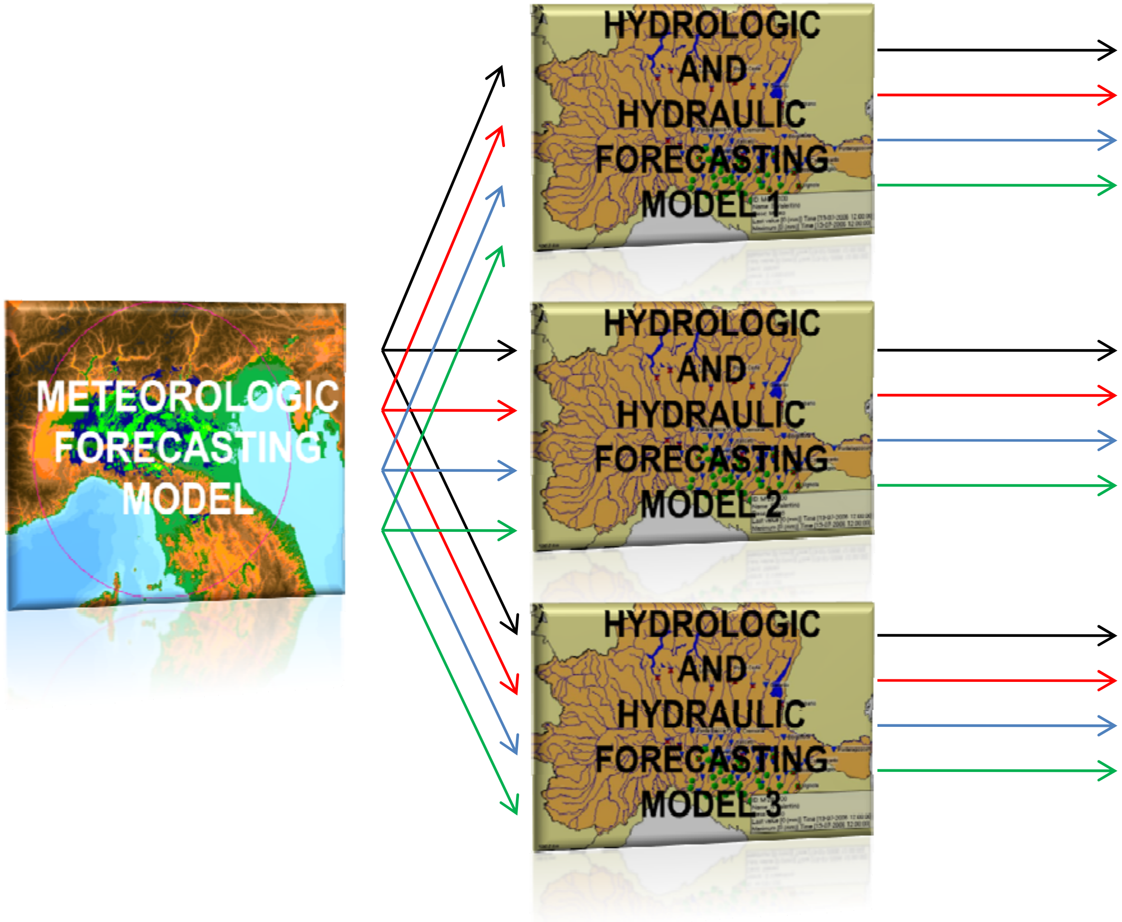
PROBLEM: **The estimate of the correlation coefficient not always well represents the high flow state**, which is due to the different behaviour of the model in reproducing the low and high flows and to the NQT non-linearity combined with the higher frequency of data in low flow state than in high flow state.



PROPOSED SOLUTION: in the Normal Space, data are divided in two samples and each one is supposed to belong to a different Truncated Normal Distribution. Hence, **two Joint Truncated Normal distributions are identified** on the basis of the samples mean, variance and covariance.

MCP: MULTI-VARIATE APPROACH

Usually, a real time flood forecasting system is composed by more than one model chain, different from each others for structure and results.

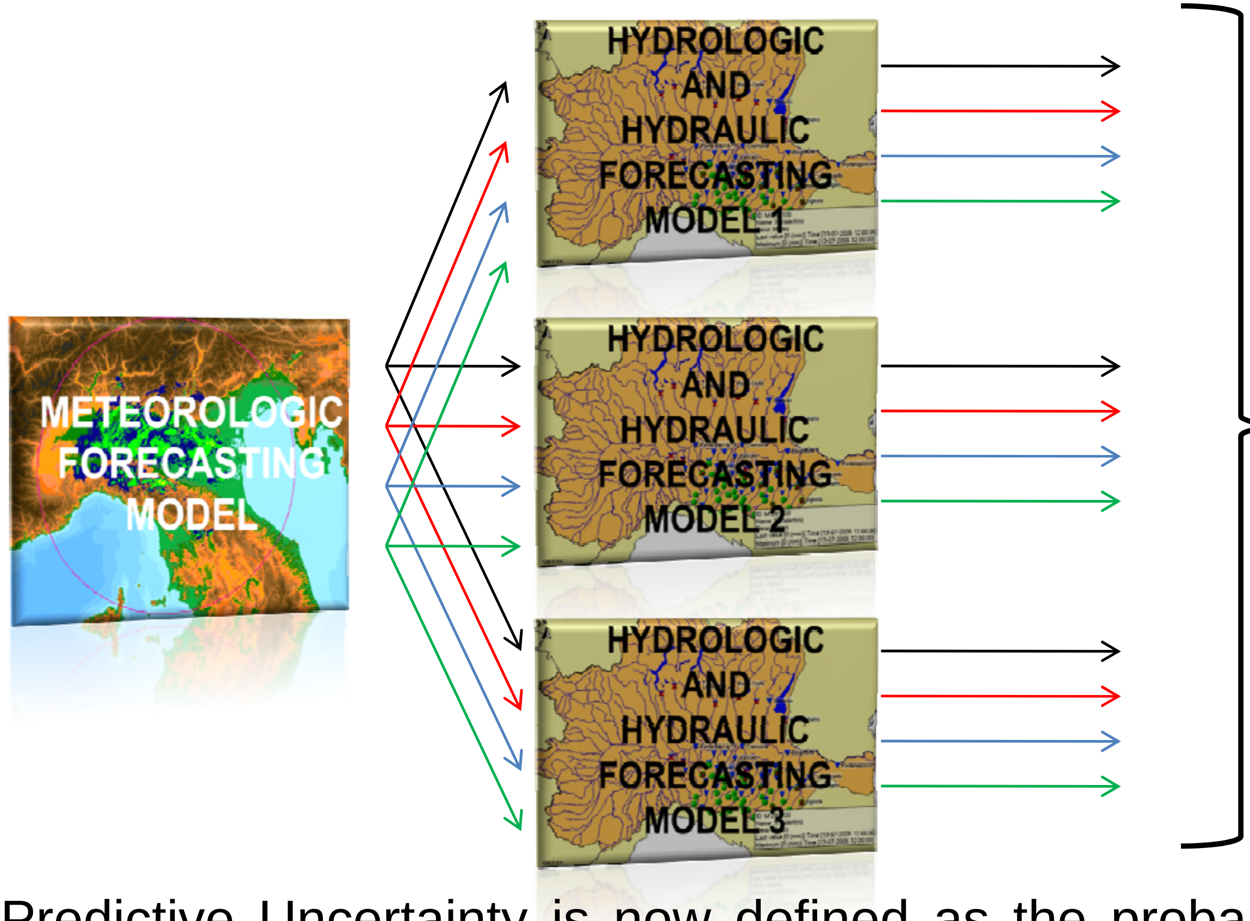


HOW TO DEAL WITH THESE FORECASTS?

WHICH WEIGHT CAN BE ASSIGNED TO EACH ONE?

MCP: MULTI-VARIATE APPROACH

Considering that a model cannot be defined better than another one in absolute terms, the MCP tries to answer these questions combining all the forecasts through a multivariate bayesian analysis.



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

Predictive Uncertainty is now defined as the probability of the real future event conditioned to the forecasts of all the deterministic models

MCP: MULTI-VARIATE APPROACH

Generalizing the previous procedure, if N forecasts are available, the multi-Normal space is composed by N+1 variables; each one is distributed as a Standard Normal and the joint distribution is a Standard Normal (N+1)-Variate,

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

with mean and variance: $\mu = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$$\Sigma = \left[\begin{array}{c|cccc} 1 & \rho_{\hat{\eta}_1 \eta} & \cdots & & \rho_{\hat{\eta}_N \eta} \\ \hline \rho_{\eta \hat{\eta}_1} & 1 & \ddots & & \rho_{\hat{\eta}_N \hat{\eta}_1} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & & \rho_{\hat{\eta}_N \hat{\eta}_{N-1}} \\ \rho_{\eta \hat{\eta}_N} & \rho_{\hat{\eta}_1 \eta_N} & \cdots & & 1 \end{array}\right] = \begin{bmatrix} \Sigma_{\eta\eta} & \Sigma_{\hat{\eta}\eta} \\ \Sigma_{\eta\hat{\eta}} & \Sigma_{\hat{\eta}\hat{\eta}} \end{bmatrix}$$

$\Sigma_{\eta\eta}$ $\Sigma_{\hat{\eta}\eta}$ ${\Sigma_{\hat{\eta}\eta}}^T$ $\Sigma_{\hat{\eta}\hat{\eta}}$

$$f\left(\eta \mid \hat{\eta}_{i,i=1..N}\right) = \frac{f\left(\eta, \hat{\eta}_{i,i=1..N}\right)}{f\left(\hat{\eta}_{i,i=1..N}\right)}$$

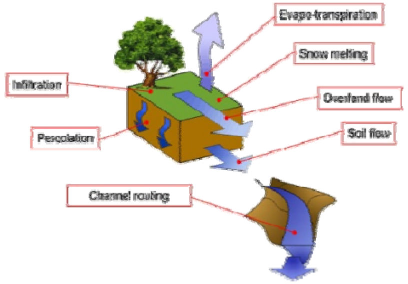
Predictive Uncertainty has mean and variance:

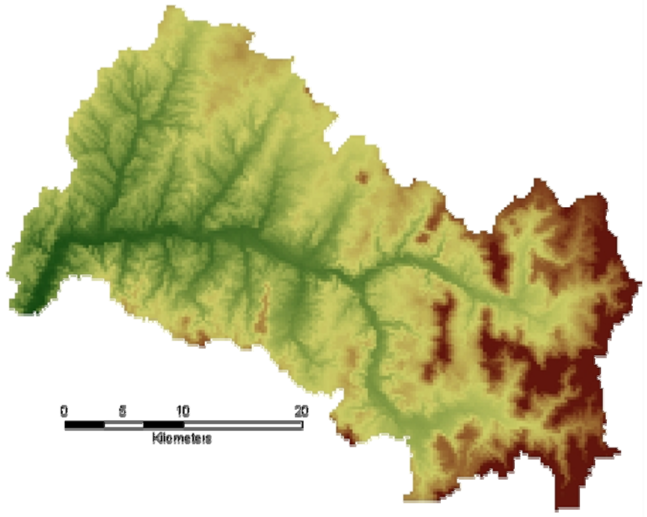
$$\mu_{\eta|\hat{\eta}_1..\hat{\eta}_N} = \Sigma_{\eta\hat{\eta}} {\Sigma_{\eta\hat{\eta}}}^{-1} \left(\hat{\eta}\right)^T \qquad \sigma^2_{\eta|\hat{\eta}_1..\hat{\eta}_N} = 1 - \Sigma_{\eta\hat{\eta}} {\Sigma_{\eta\hat{\eta}}}^{-1} {\Sigma_{\eta\hat{\eta}}}^T$$

MCP: APPLICATION

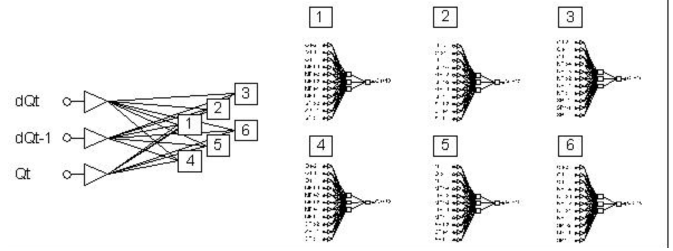
BARON FORK RIVER AT ELDON, OK, USA

TOPKAPI MODEL

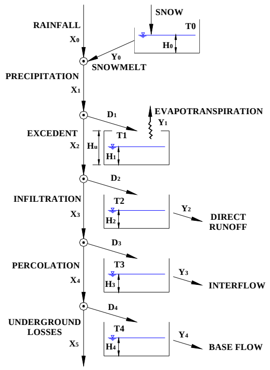




ANN MODEL



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: VALIDATION CALIBRATION

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

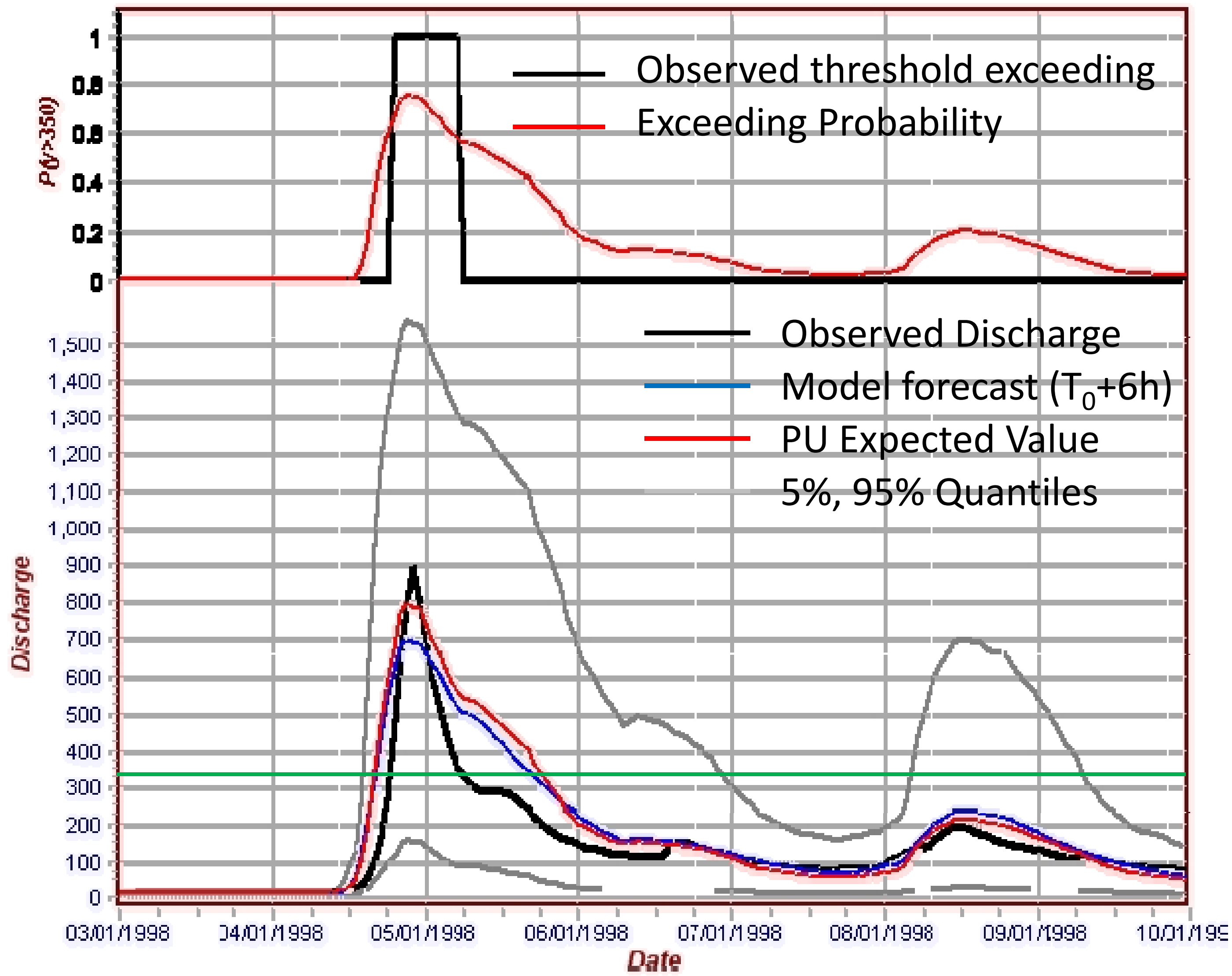
MCP: APPLICATION

The aim of the application was to answer the following questions :

1. Does the MCP assess the Predictive Uncertainty?
2. Does the MCP improve the deterministic forecasts?
3. Does the use of the truncated joint distributions improve the MCP behaviour in reproducing flood event?
4. Does the Multivariate approach reduce the Predictive Uncertainty?

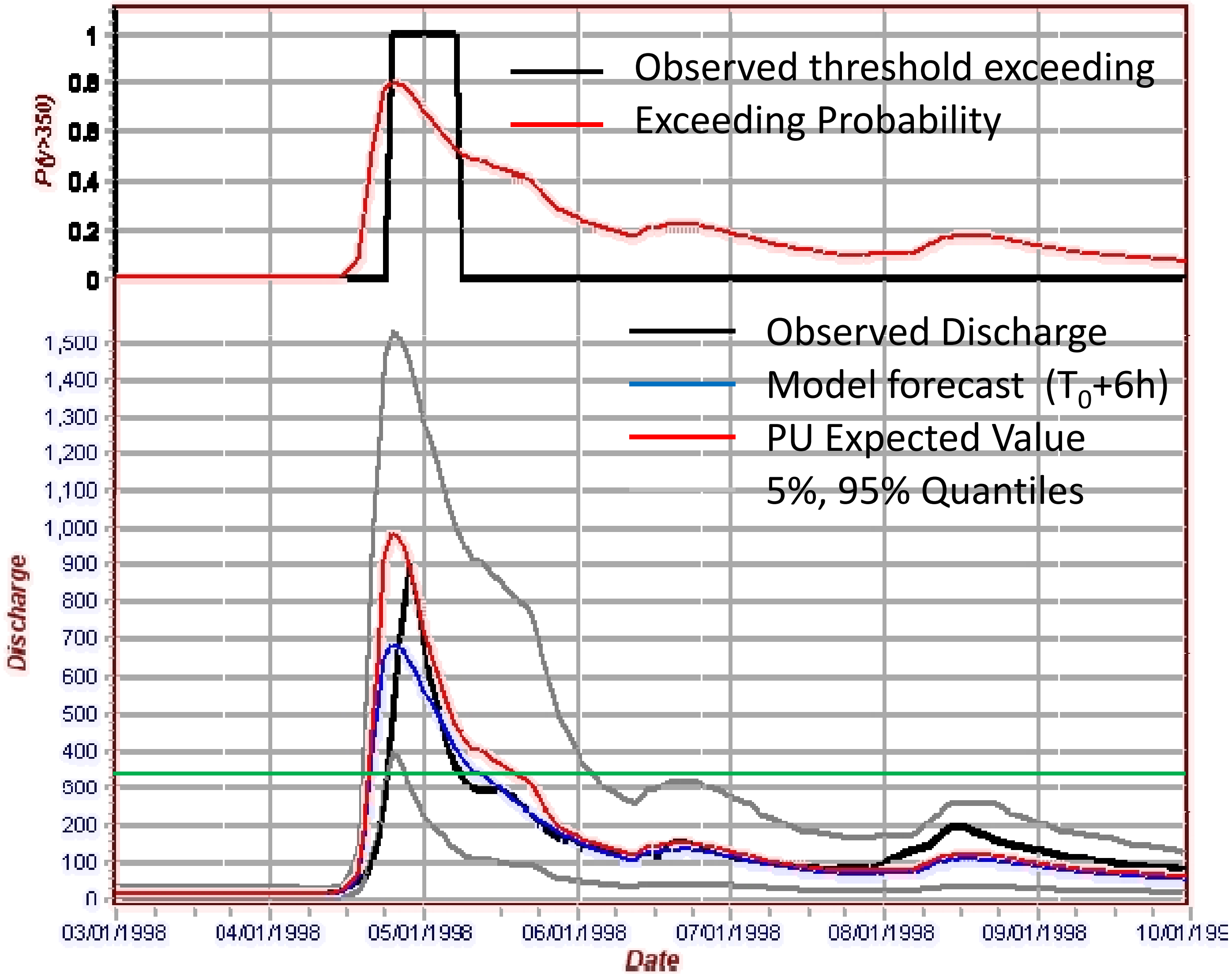
MCP: APPLICATION

1 MODEL: TOPKAPI



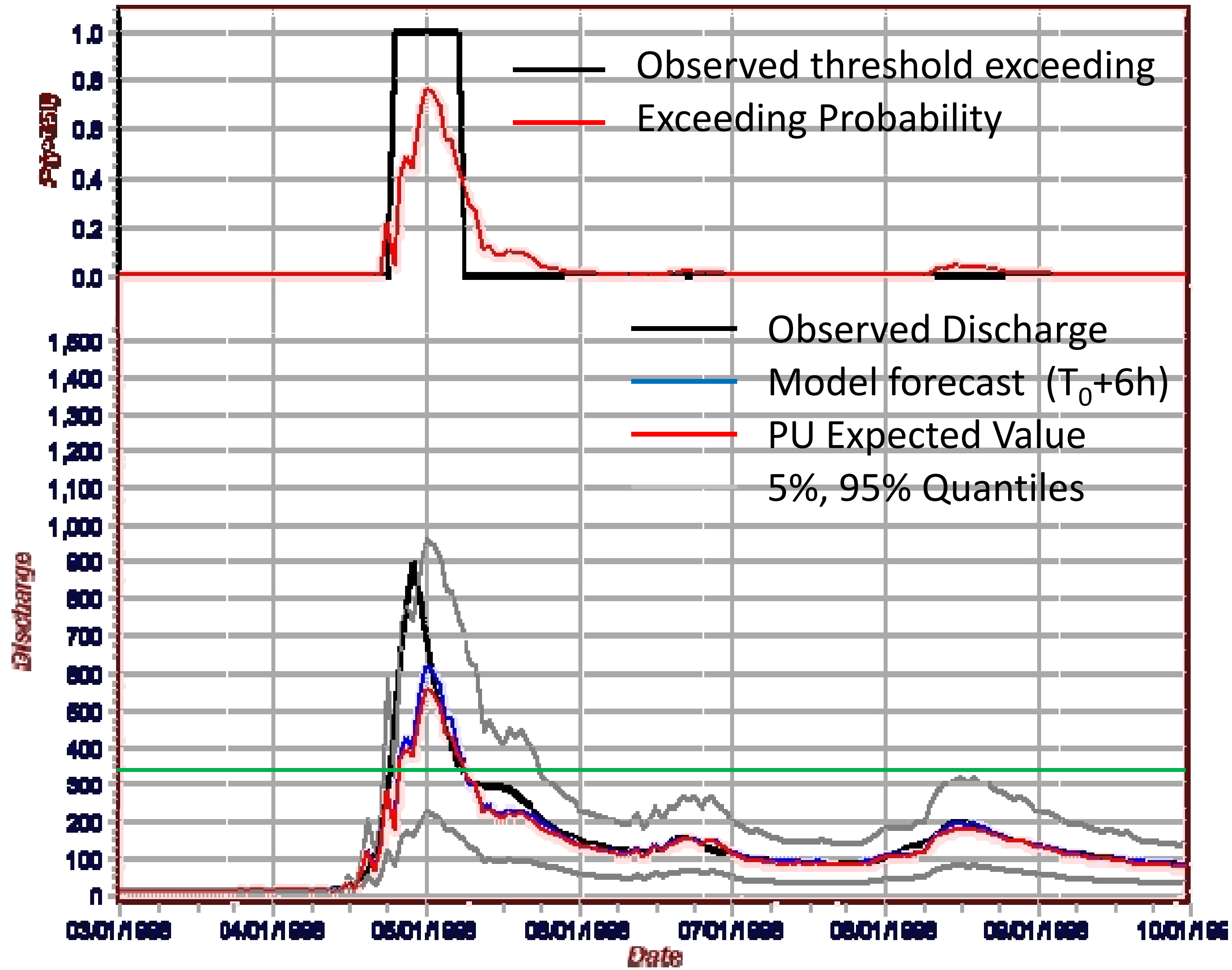
MCP: APPLICATION

1 MODEL: TETIS



MCP: APPLICATION

1 MODEL: ANN



European Geosciences Union General Assembly 2010, Vienna, May 2010

MCP: APPLICATION

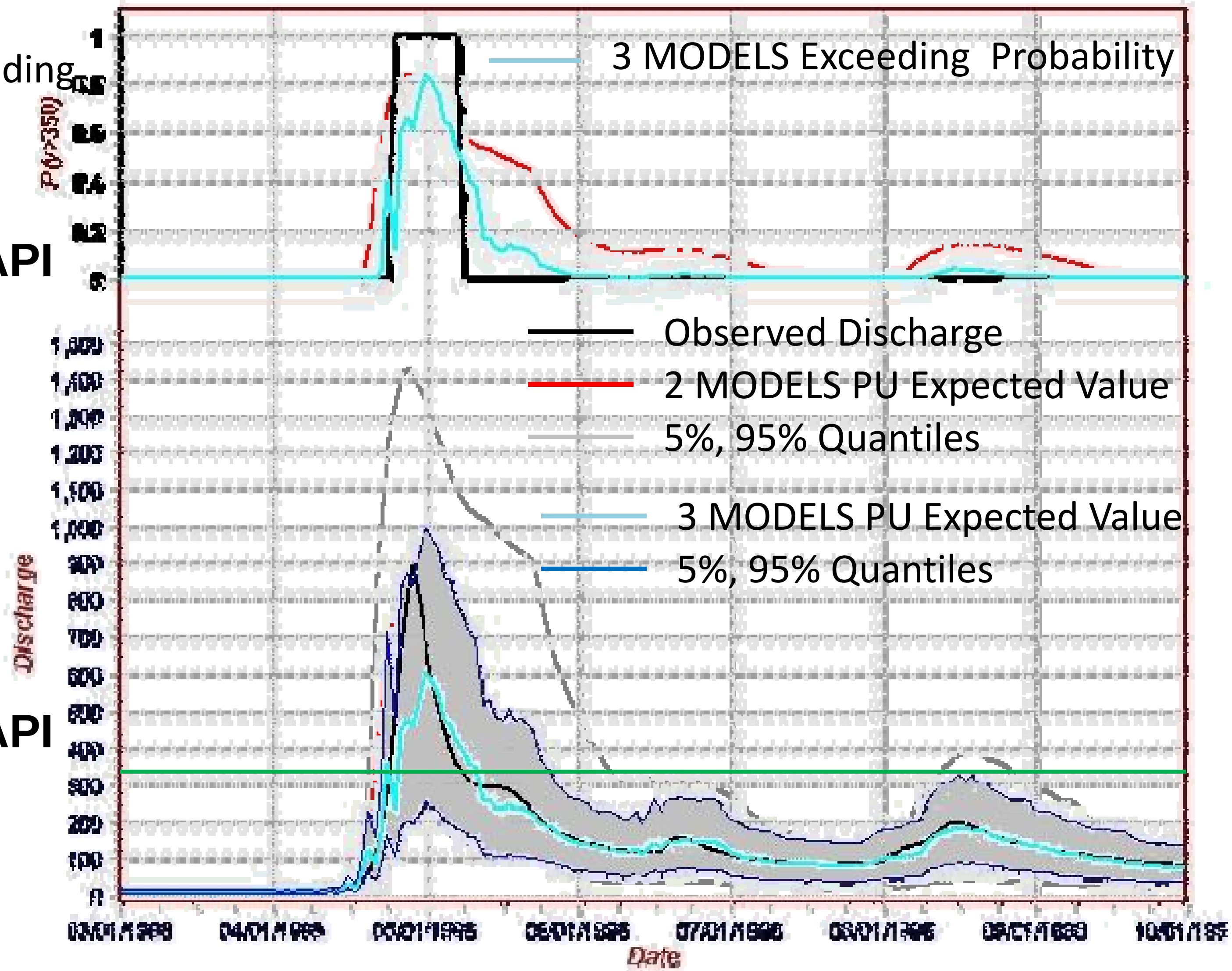
Observed threshold exceeding

2 MODELS Exceeding Probability

2 MODELS: TETIS + TOPKAPI

VS

3 MODELS: TETIS + TOPKAPI + ANN



MCP: APPLICATION

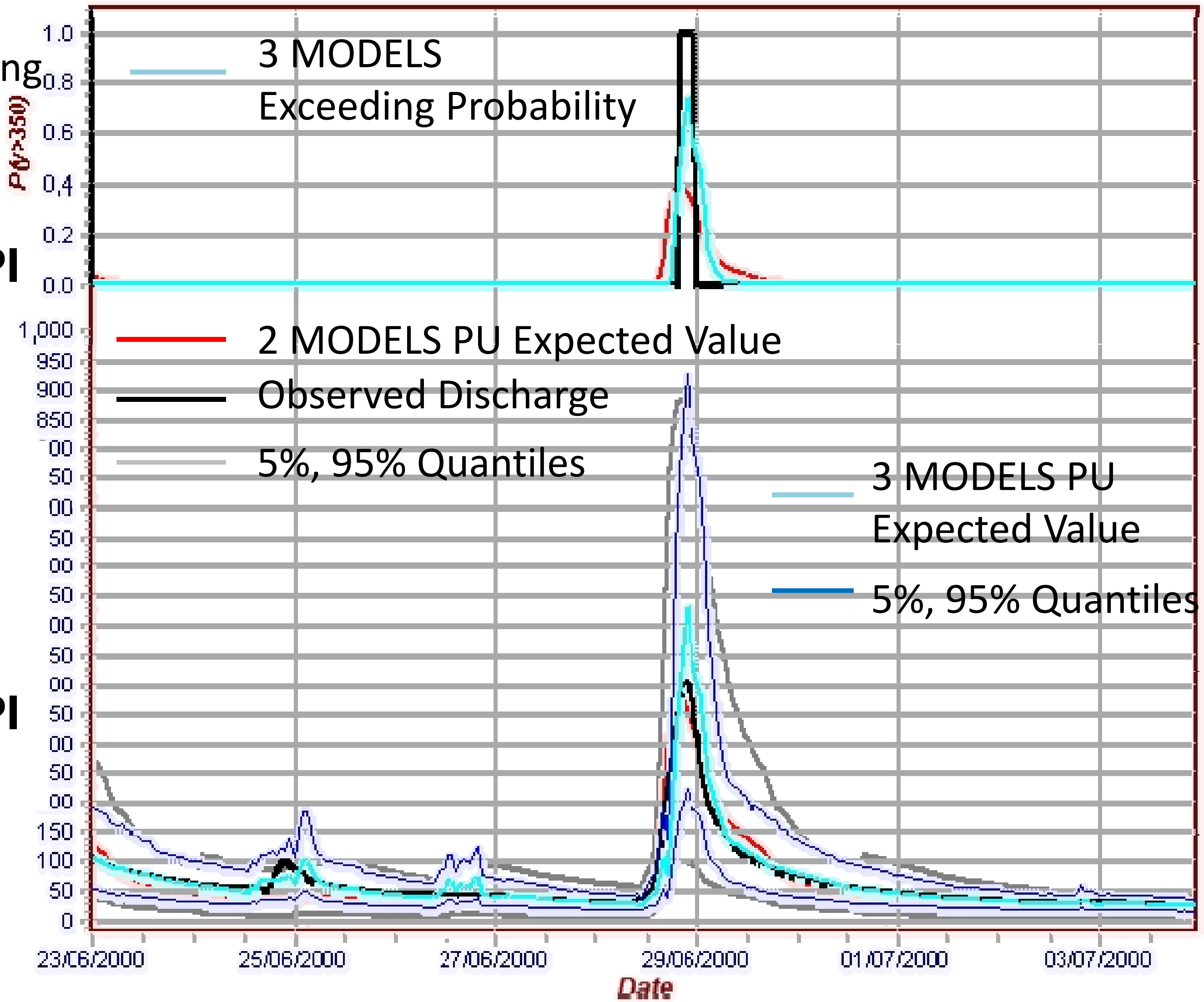
Observed threshold exceeding

2 MODELS Exceeding Probability

2 MODELS: TETIS + TOPKAPI

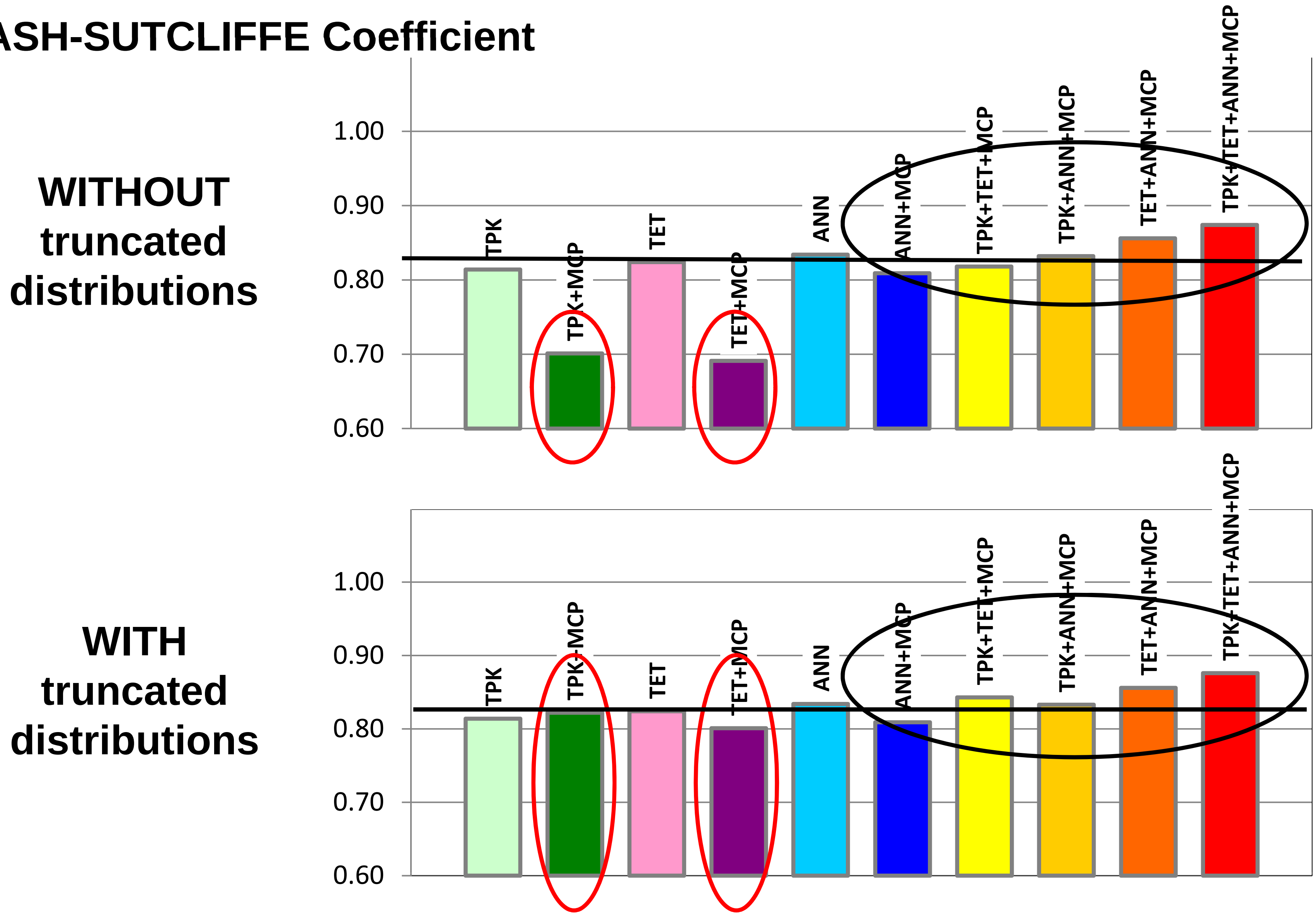
VS

3 MODELS: TETIS + TOPKAPI + ANN



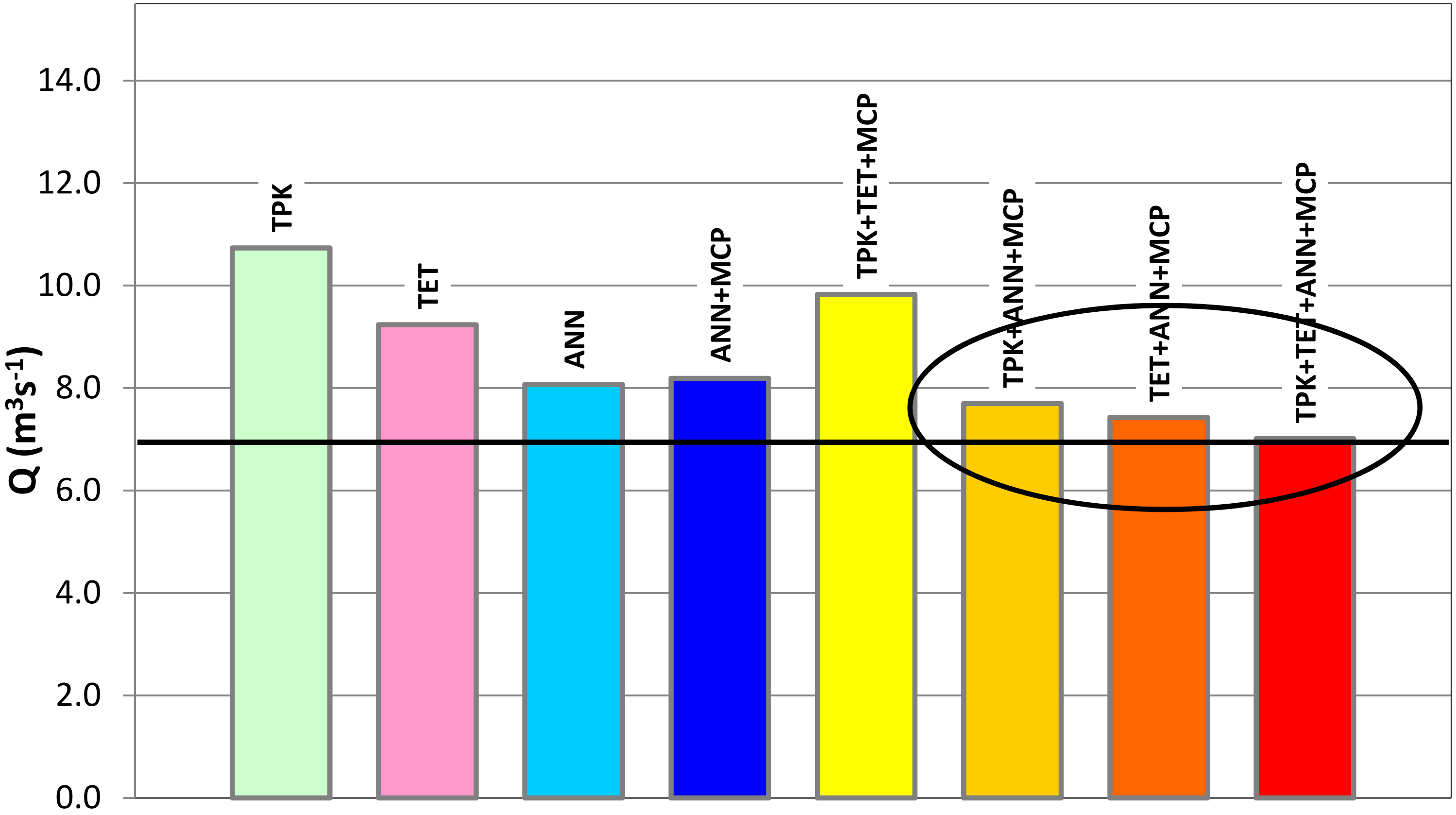
MCP: APPLICATION

NASH-SUTCLIFFE Coefficient



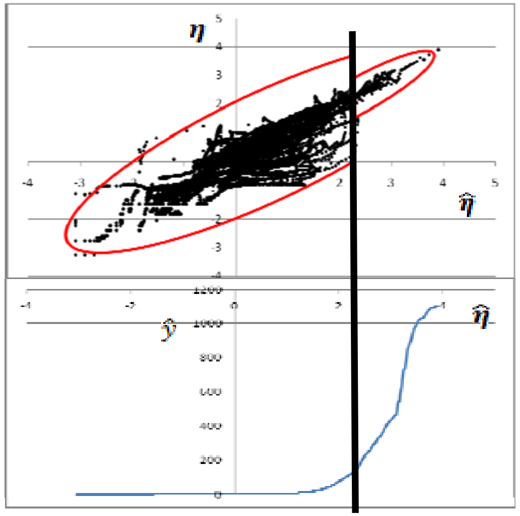
MCP: APPLICATION

ERROR STANDARD DEVIATION WITH the Truncated Joint Distribution



FUTURE DEVELOPMENTS

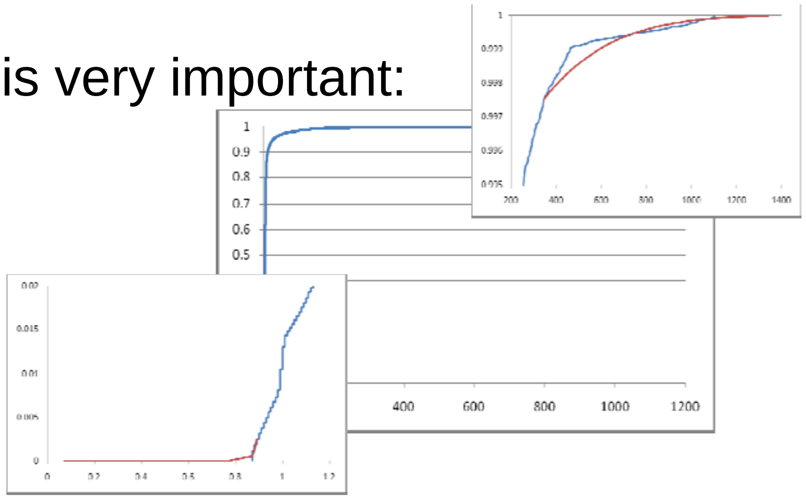
1) The choice of the **threshold using the Joint Truncated Distributions**.



Is it possible to find an objective rule, related to the forecast cdf gradient, to identify this threshold?

Can the use of a Quantile Regression lead to a more realistic uncertainty assessment?

2) A good model fit of the **marginal cdf tails** is very important:



For which probabilities should tails be used?

Which is the best curve?

Would be better the use of tails or to identify a probability model for the whole series?

CONCLUSIONS

Even if improvements are still required, the presented application of the MCP shows that this methodology:

- allows to estimate the Predictive Uncertainty, requiring low computational costs;

- allows to combine different models forecast, reconciling physically based and data driven models gaining from the benefits of both approaches.

Furthermore,

- the use of the truncated distributions allows to better reproduce the flood events;

- the assessment of the probability to exceed an alert level allows to deal in probabilistic terms with the emergency management and it can lead to identify probability thresholds instead of the deterministic ones commonly used.