DIPARTIMENTO di SCIENZE della TERRA e GEOLOGICO-AMBIENTALI

Probabilistic Forecasts Within a Time Horizon and Exact Flooding Time Probability

Gabriele Coccia[1][2] and Ezio Todini[1]

[1] University of Bologna, Bologna, Italy

[2] Idrologia e Ambiente s.r.l, Napoli, Italy

PREDICTIVE UNCERTAINTY (PU)

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 (Krzysztofowicz,1999)

PREDICTIVE UNCERTAINTY MUST BE QUANTIFIED IN TERMS OF PROBABILITY DISTRIBUTION.

If the available information is a model forecast, the Predictive Uncertainty can be denoted as:

$$f\left(y_t \middle| \left(\hat{y}_t \mid x_{t_0}, \vartheta\right)\right) = f\left(y \middle| \hat{y}\right)$$

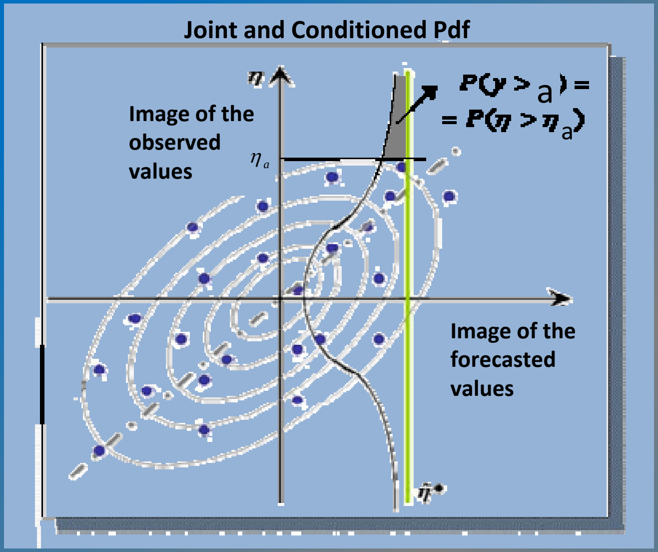
Krzysztofowicz, R.: Bayesian theory of probabilistic forecasting via deterministic hydrologic model, Water Resour. Res., 35, 2739-2750, 1999.

MODEL CONDITIONAL PROCESSOR: BASIC CONCEPTS

1) Conversion from the Real Space to the Normal Space using the NQT

2) Joint Pdf is assumed to be a Normal Bivariate Distribution or composed by 2 Truncated Normal Distributions

3) Predictive Uncertainty is obtained by the Bayes Theorem



BI-VARIATE

UNI-VARIATE

$$f(\eta|\hat{\eta}) = \frac{f(\eta,\hat{\eta})}{f(\hat{\eta})}$$

UNI-VARIATE

4) Reconversion of the obtained distribution from the Normal Space to the Real Space using the Inverse NQT

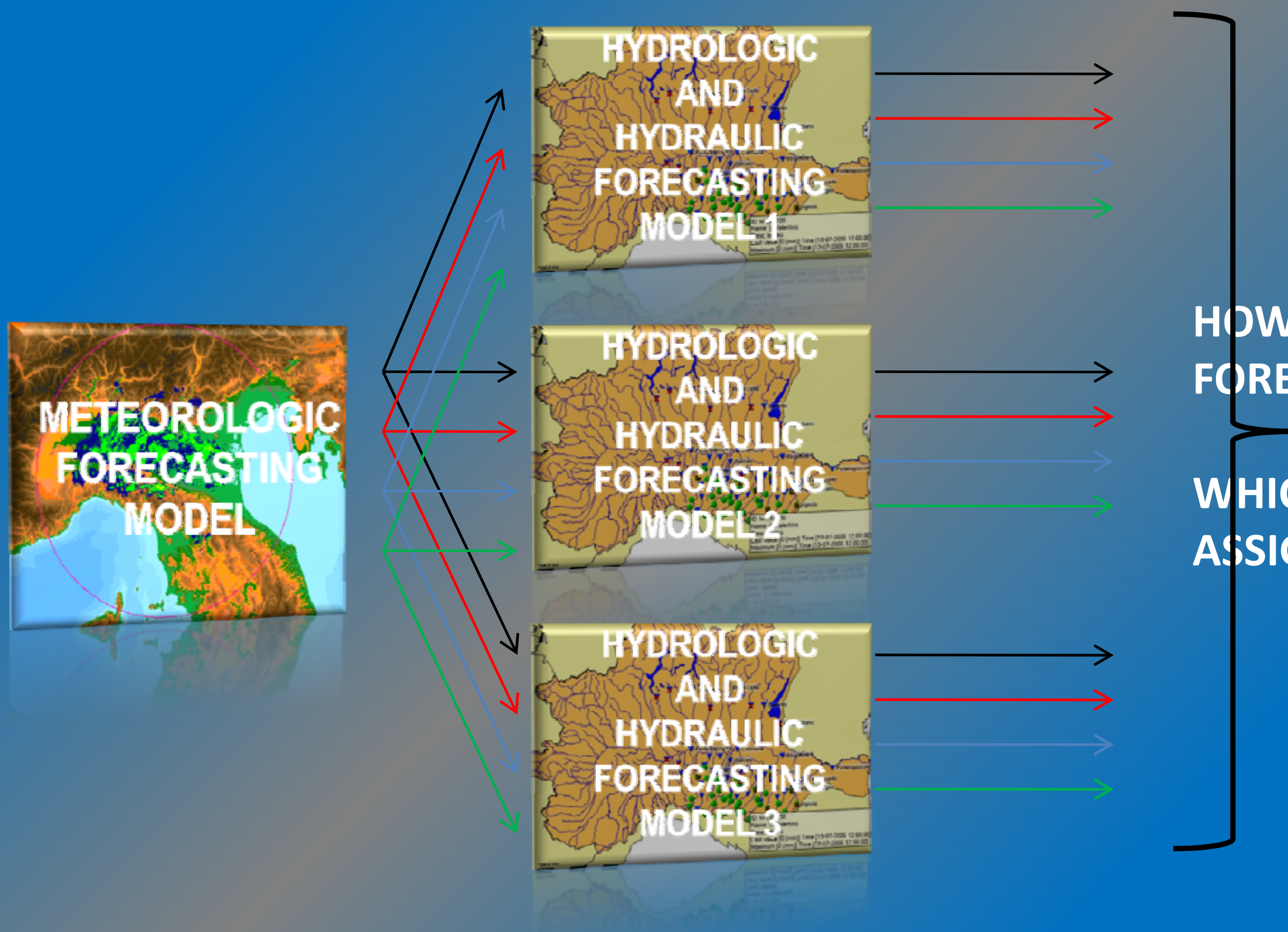
$$f(\eta|\hat{\eta}) \longrightarrow f(y|\hat{y})$$

Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, Intl. J. River Basin Management, 6 (2), 123-137, 2008.

G. Coccia and E. Todini: Recent Developments in Predictive Uncertainty Assessment Based on the Model Conditional Processor Approach, HESSD, 7, 9219-9270, 2010

MULTI-MODEL APPROACH

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



BAYESIAN COMBINATION

HOW TO DEAL WITH THESE FORECASTS?

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

WHICH WEIGHT CAN BE ASSIGNED TO EACH ONE?

MULTI-MODEL APPROACH

THE BASIC PROCEDURE CAN BE EASILY EXTENDED TO MULTI-MODEL CASES

N+1-VARIATE

UNI-VARIATE

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

N-VARIATE

N = NUMBER OF MODELS

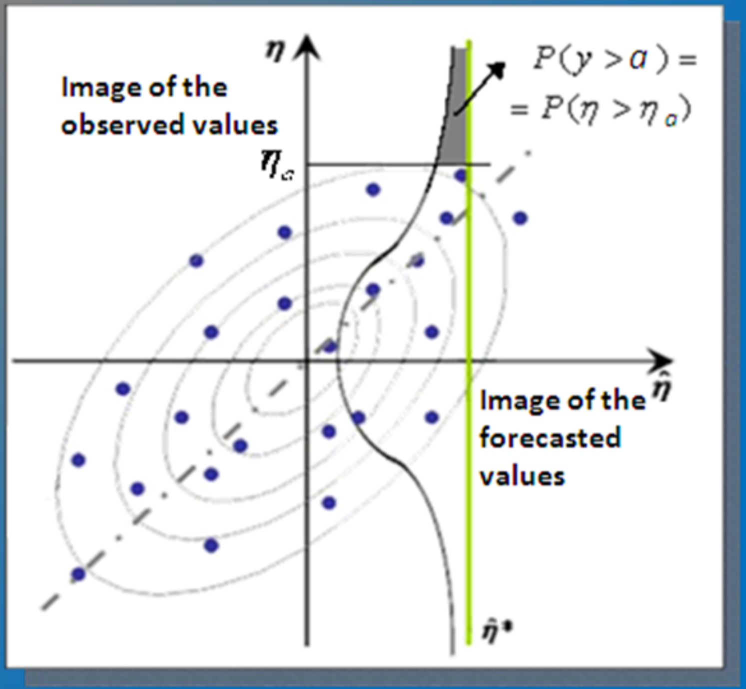
FLOODING PROBABILITY AT A SPECIFIC TIME HORIZON

The knowledge of the Predictive Uncertainty allows to easily extrapolate the **probability to exceed a threshold value**, such as the dyke level.

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

$$P(y_{\Delta t} > a|\hat{y}_{\Delta t} = \hat{y}^*_{\Delta t}) =$$
$$= \int_a^\infty f(y_{\Delta t}|\hat{y}_{\Delta t} = \hat{y}^*_{\Delta t})dy =$$
$$= \int_{\eta_a}^\infty f(\eta_{\Delta t}|\hat{\eta}_{\Delta t} = \hat{\eta}^*_{\Delta t})d\eta$$

Joint and Conditioned Pdf



In the decision making process, this methodology allows to answer to the following question:

Which is the probability that the water level will be higher than the dykes at the hour 24th?

Probably a more interesting question may be:

Which is the probability that the river dykes will be exceeded within the next 24 hours?

To be able to answer to this question it is necessary to known the correlation between the predicted variable at the different time steps of the prediction

MULTI-TEMPORAL APPROACH

Following Krzysztofowicz (2008), the procedure can be generalized including in the bayesian formulation all the available forecasts within the entire horizon time.

A MULTI-VARIATE PREDICTIVE DISTRIBUTION is obtained, which accounts for the joint PU of the observed variable at each time step.

Krzysztofowicz, R.: Probabilistic flood forecast: exact and approximate predictive distributions, Research Paper RK0802, University of Virginia, September 2008.

MULTI-TEMPORAL APPROACH

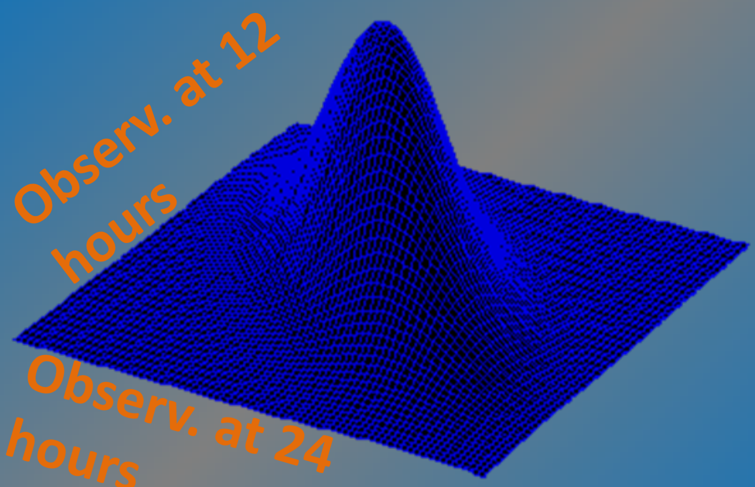
With respect to the multi-model approach, the dimension of all the distributions is multiplied by the number of the time steps.

((N+1) · T) - VARIATE

T - VARIATE

$$f\left(\eta_{j(j=1..T)} \middle| \hat{\eta}_{ij(i=1..N,j=1..T)}\right) = \frac{f\left(\eta_{j(j=1..T)}, \hat{\eta}_{ij(i=1..N,j=1..T)}\right)}{f\left(\hat{\eta}_{ij(i=1..N,j=1..T)}\right)}$$

(N · T) - VARIATE



N = NUMBER OF MODELS

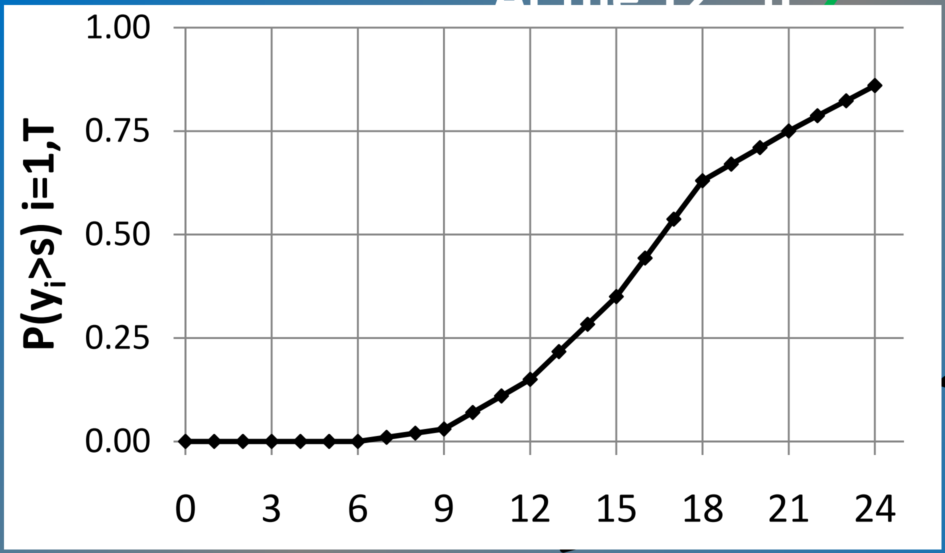
T = NUMER OF TIME STEPS

FLOODING PROBABILITY WITHIN THE TIME HORIZON OF *T* TIME STEPS

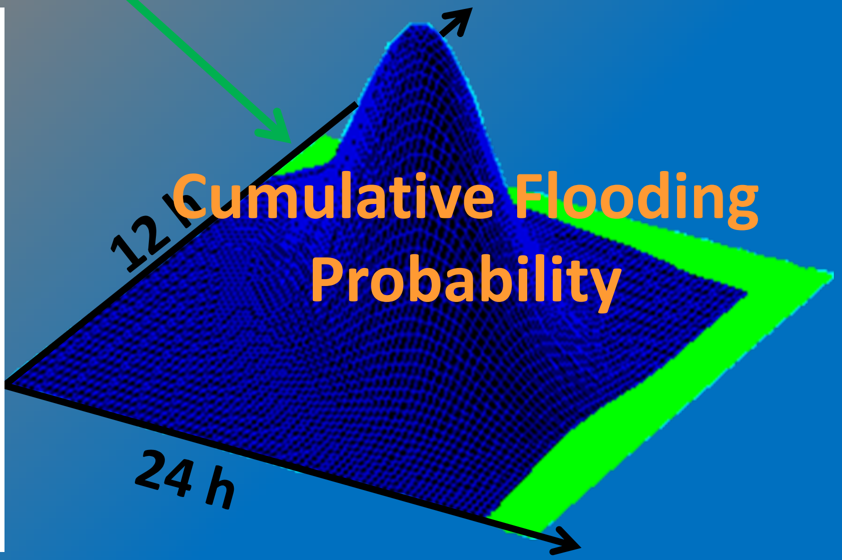
$$P(y_{t;t=1..T} > a \mid \hat{y}_{t,k;t=1..T;k=1..N}) = P(y_t > a \mid \hat{y}_{t,k})$$

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

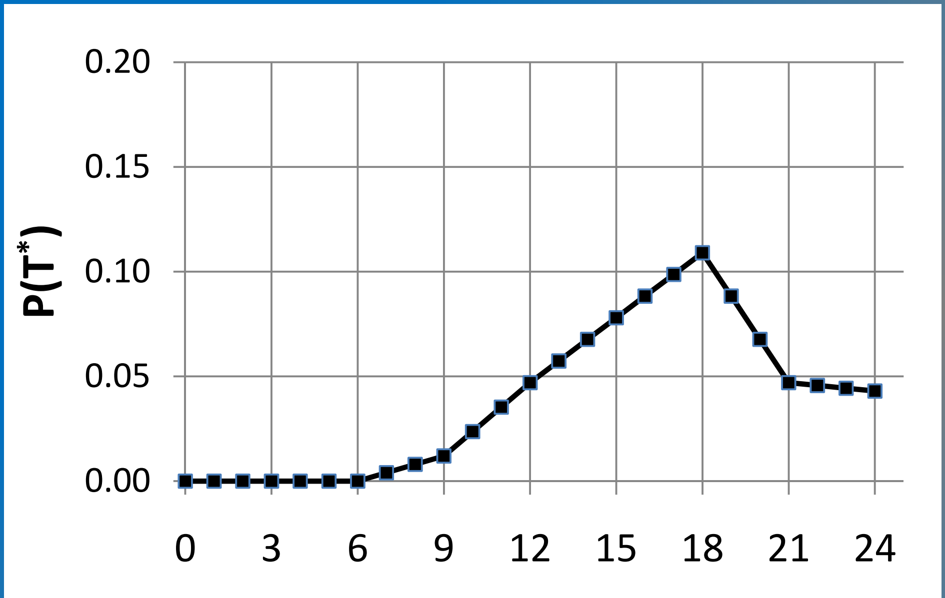
At the 12th h



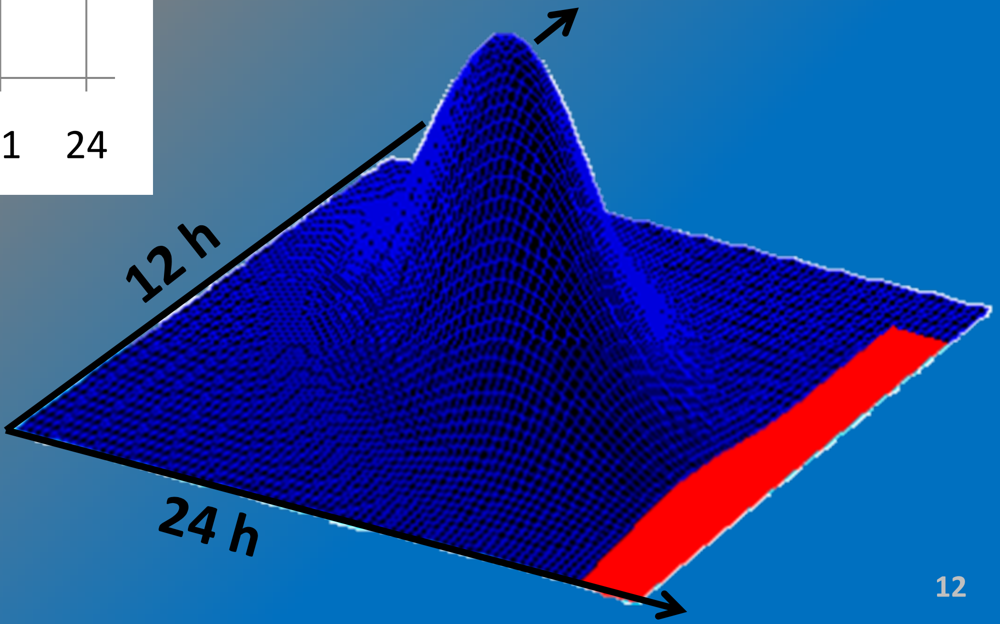
Within 24 h



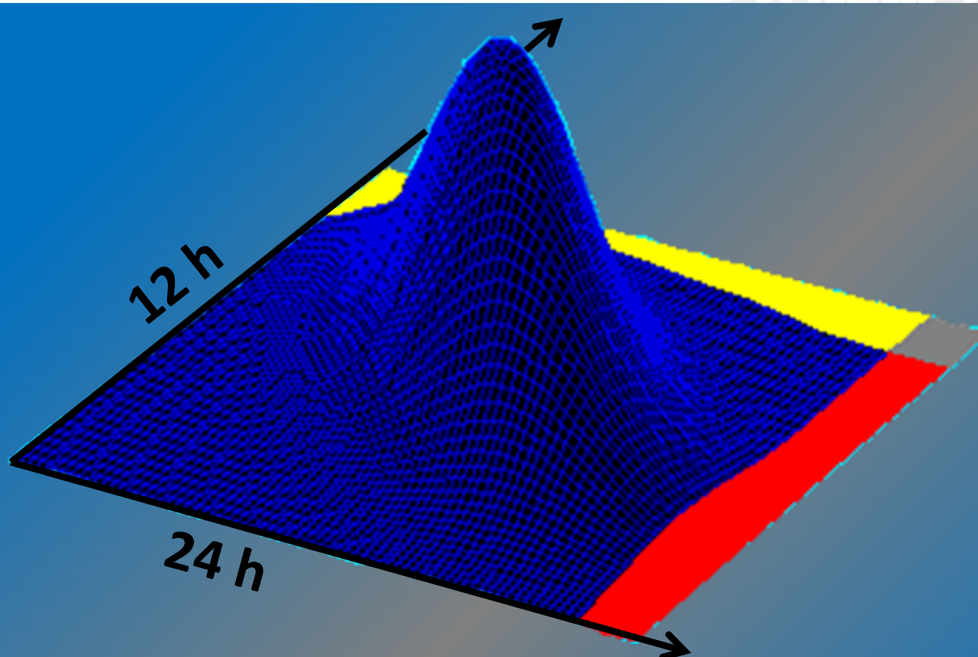
Which is the probability that the river dykes will be exceeded exactly at the hour 24th?



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



EXACT FLOODING TIME (*T) PROBABILITY**



Which is the probability that the water level will be higher than the dykes one at the hour 24th? → **RED + GREY**

Which is the probability that the river dykes will be exceeded within the next 24 hours? → **RED + GREY + YELLOW**

Which is the probability that the river dykes will be exceeded exactly at the hour 24^{th}? → **RED**

Can be obtained also with the basic and multi-model approaches since it does not depend on the state of the variable at 12 hours

PO RIVER AT PONTELAGOSCURO and PONTE SPESSA

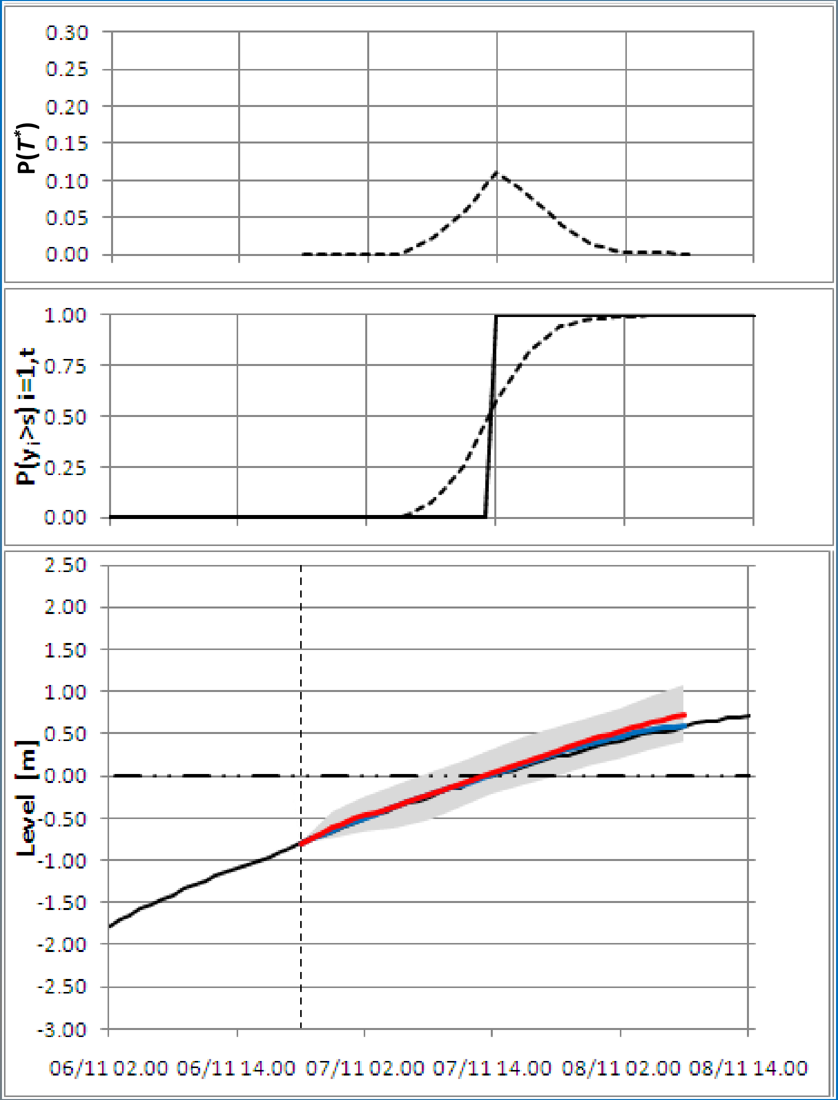
01/05/2000	30/06/2004	20/01/2009
MCP:	CALIBRATION	VALIDATION

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

Forecasted hourly levels with a time horizon of 24 and 36 h

Observed hourly levels

Pontelagoscuro Station (36 h)

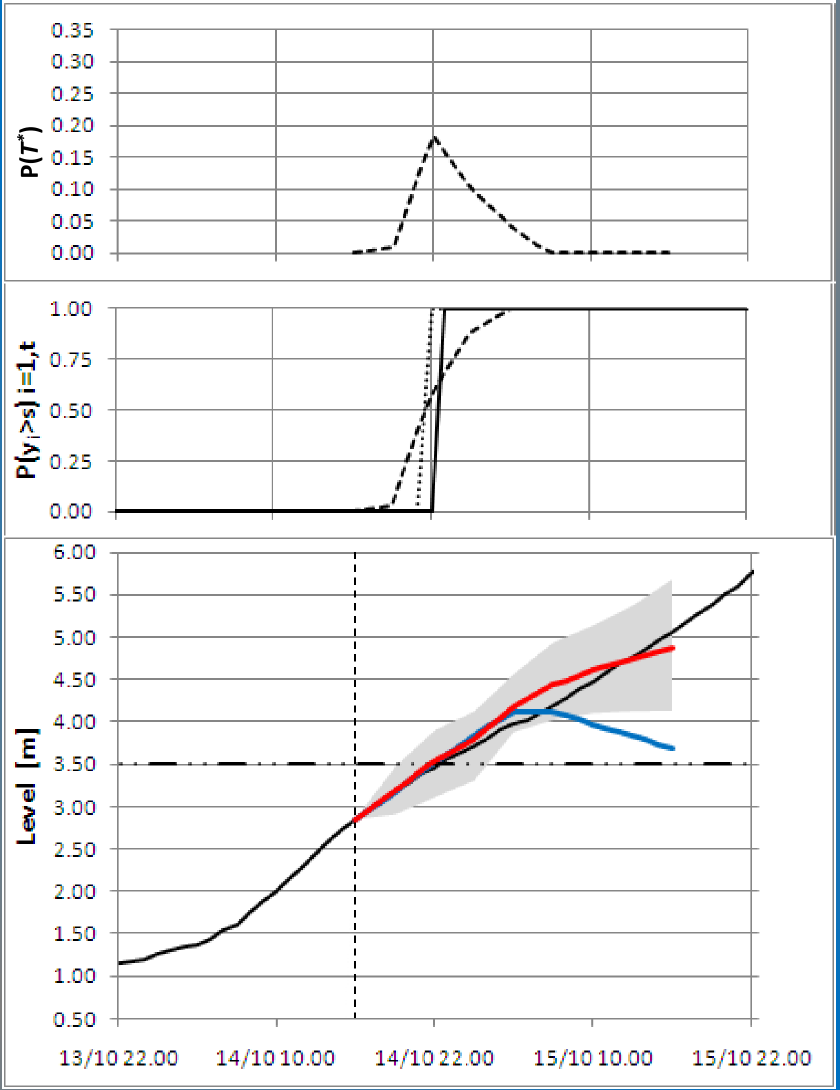


Exact Flooding Time Probability

Cumulative Flooding Probability

90% Uncertainty Band with MULTI-TEMPORAL APPROACH

Ponte Spessa Station (24 h)



Exact Flooding Time Probability

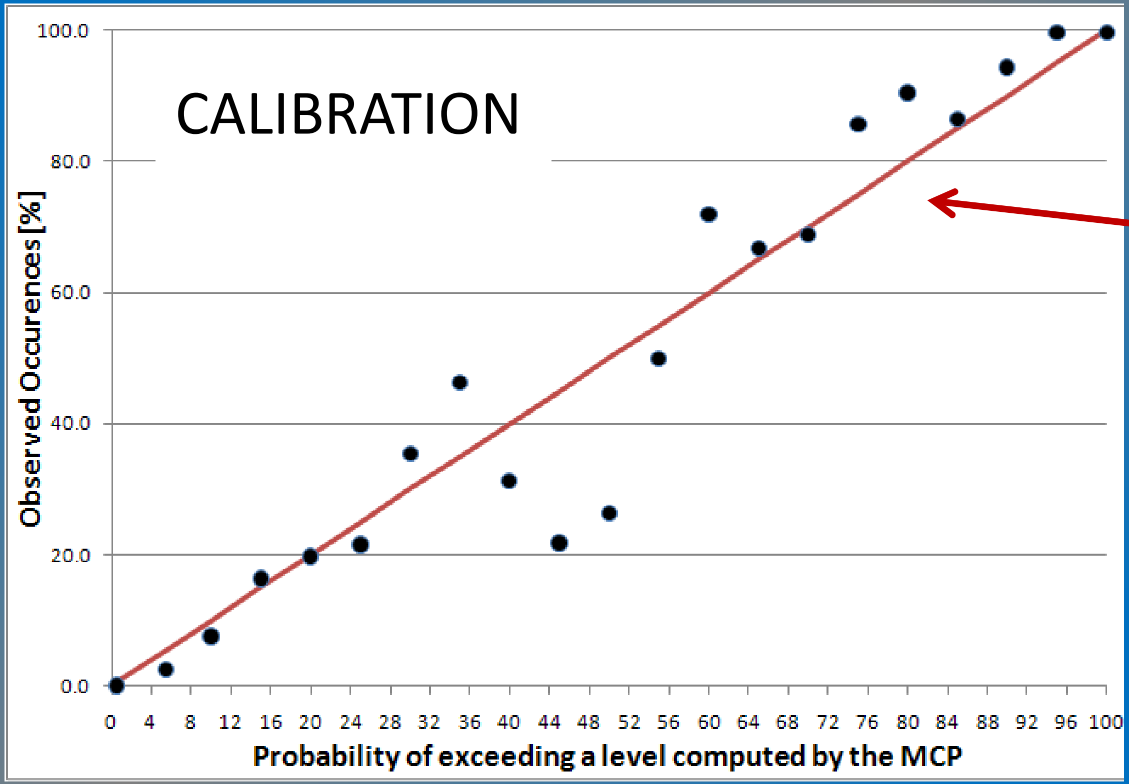
Cumulative Flooding Probability

90% Uncertainty Band with MULTI-TEMPORAL APPROACH

FLOODING PROBABILITY ASSESSMENT VERIFICATION

If the value provided by the processor is correct, considering all the cases when the computed exceeding probability takes value P, the percentage of observed exceeding occurrences must be equal to P.

Ponte Spessa Station (24 h)



Red Line = Perfect beahviour

Computed with a 5% discretization

CONCLUSIONS

- Most of the existing Uncertainty Processors do not account for the evolution in time of the forecasted events

- The correlation between the predicted variable at different time steps of the prediction should be taken into account

- The presented multi-temporal approach allows to identify the joint predictive distribution of all the forecasted time steps

CONCLUSIONS

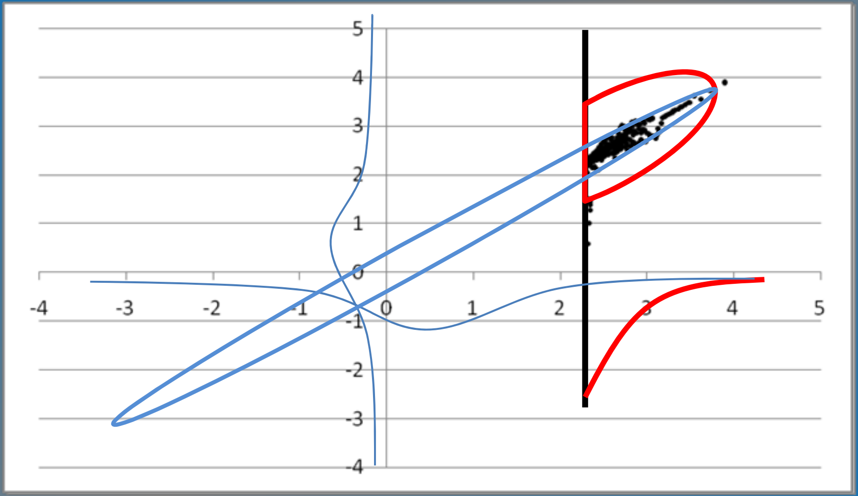
- This procedure allows to recognize and reduce the systematic time errors and it gives important information, such as the probability to have a flooding event within a specific time horizon and the exact flooding time probability

- The comparison of predicted and observed flooding occurrences verified that, a part small errors due to the unavoidable approximations, the methodology computes the flooding probability with good accuracy

THANK YOU FOR YOUR ATTENTION AND YOUR PATIENCE

HETEROSCEDASTICITY OF THE ERROR: THE TRUNCATED NORMAL DISTRIBUTIONS

The assumption of the **homoscedasticity of errors** leads to a lack of accuracy, especially for high flows.

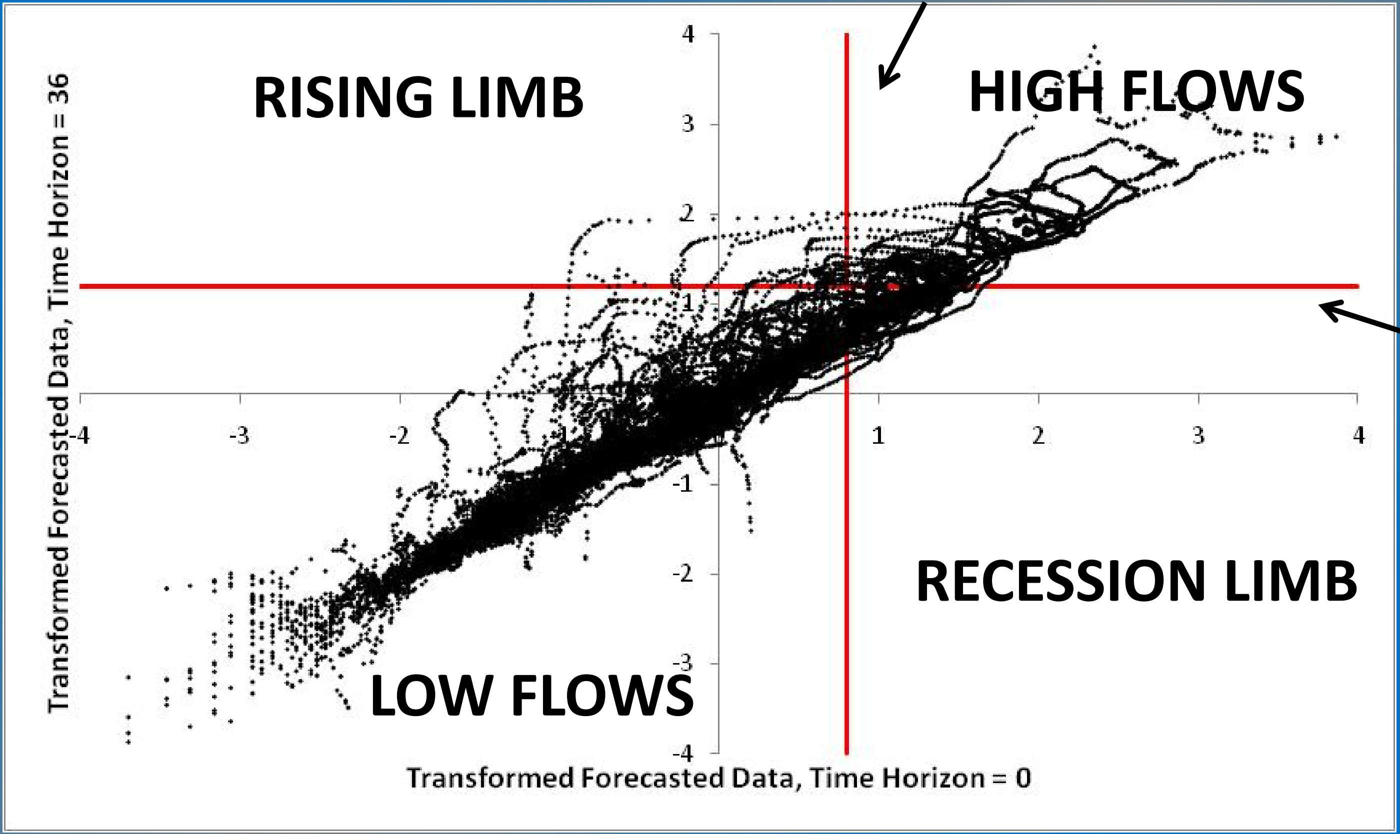


In the Normal Space **the data are divided in two (or more) samples and each one is supposed to belong to a different Truncated Normal Distribution**. Hence, two Joint Truncated Normal distributions (TNDs) are identified on the basis of the samples mean, variance and covariance.

Coccia, G. and Todini, E.: Recent developments in predictive uncertainty assessment based on the model conditional processor approach, Hydrol. Earth Syst. Sci. Discuss., 2010.

HOW TO USE THE TNDs IN THE MULTI-TEMPORAL APPROACH?

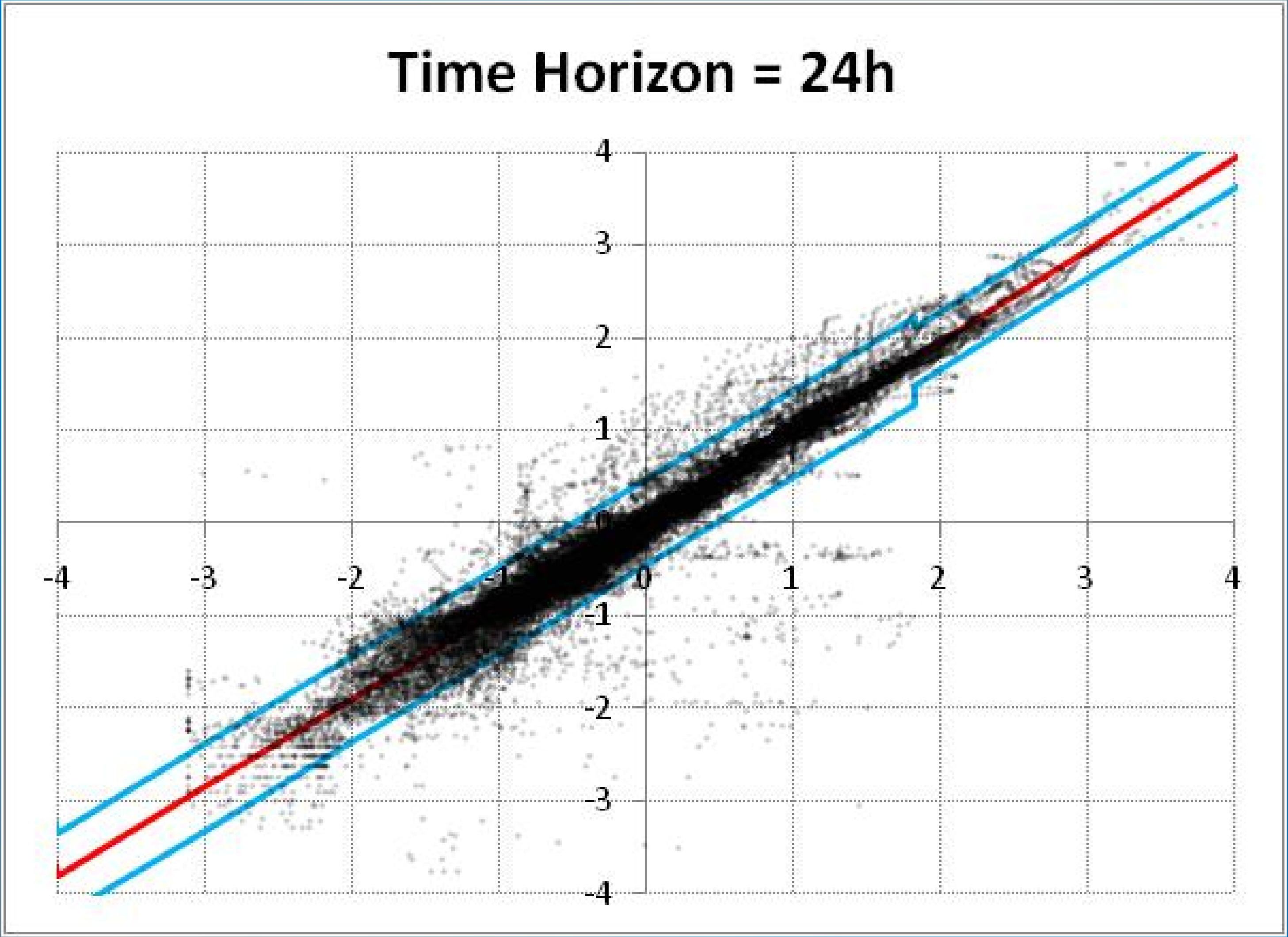
THRESHOLD FOR $T=t_0$



THRESHOLD FOR $T=t_0+36$

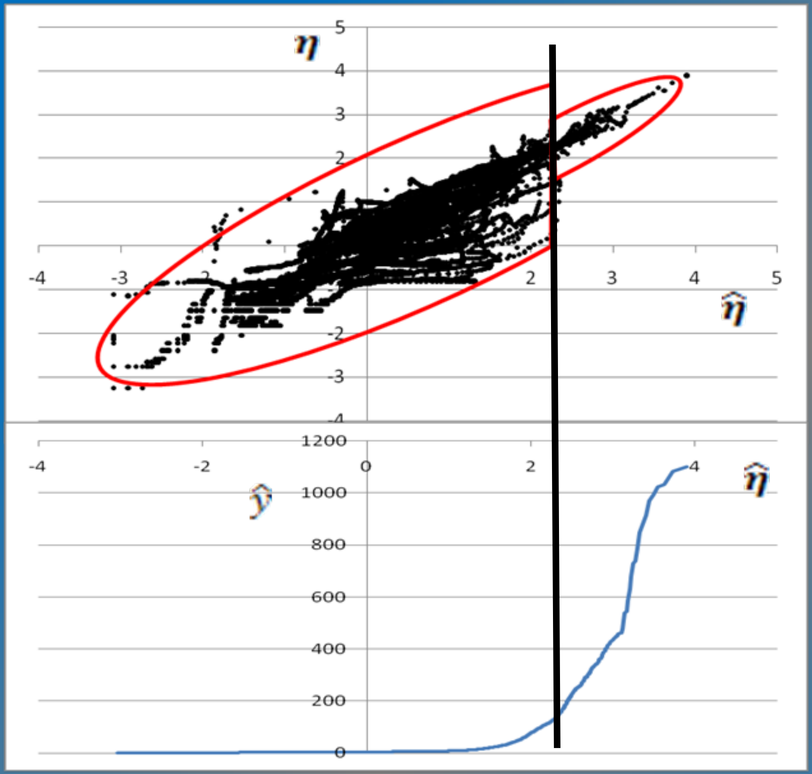
DATA OF THE MODEL THAT BETTER PERFORMS IN HIGH FLOWS

PONTE SPESSA



OPEN QUESTIONS

The choice of the threshold using the Joint Truncated Distributions.



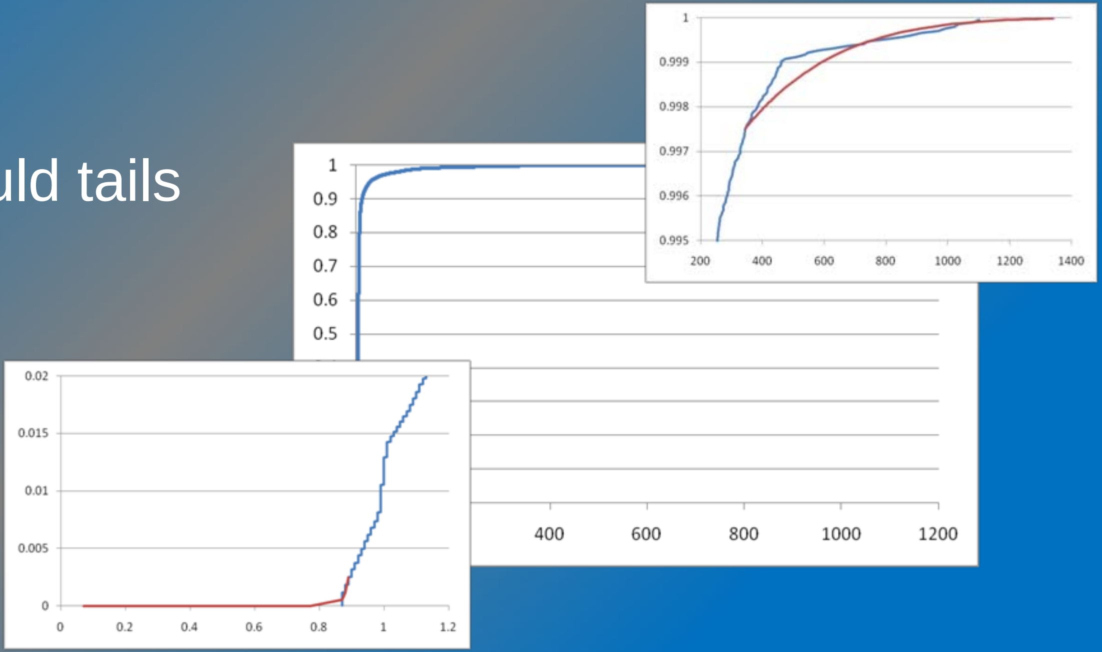
Is it possible to find an objective rule, related to the forecast cdf gradient, to identify this threshold?

How many TNDs must be used? Are 2 enough?

OPEN QUESTIONS

A good model fit of the marginal distribution 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?