Answers to Referee #2 comments

First of all the authors would like to thank Referee #2 for his remarks that will allow the authors to improve the original version of the manuscript.

1) **C:** *Section 1.2 "Probabilistic threshold paradigm" is too wordy and an effort should be made to formulate the concept more concisely and clearer, using shorter sentences and/or expressions. A reference to a conference paper does not seem to be appropriate (Todini and Coccia 2010). The statement "dramatically changed the deterministic threshold paradigm" is overstated, as I find it overstated to talk about a "paradigm" in this context.*

A: The authors agree with Referee #2 that Section 1.2 can be improved. It will be shortened and more appropriate references will be chosen, also in accordance to the comments of Referee #1.

2) **C:** *One of my major points of critique is the use of discharge as a forecasting variable in place of the water level, especially as they show "water levels" as critical variable In the explanatory figure 2. Discharge is a derived quantity calculated from the observed water level as primary variable. In my opinion working with observed water levels and transforming the output of hydrological models (Q) into levels reduces the risk of introducing spurious uncertainty with respect to doing it the other way around.*

A: Both Referee #1 and Referee #2 underline that, although the authors point out that the actually occurring water level is the one that causes damages, they use either the observed water level or the estimated discharge as the "predictand".

It is definitely true that observations do not coincide with reality, a measurement error must always be taken into account, nevertheless the authors would like to discuss in detail the choice of the predictand.

1) Water level measurements are affected by relatively small errors (with standard error of the order of 2-3 cm), and it is psychologically fundamental to use them as measures unaffected by measurement errors both because flood decisions have always been essentially based on these measures and because their errors have very small effect on the decisions, given the larger effects of the other sources of uncertainty.

2) Discharge measurements are generally unavailable in real time, although there is a recent tendency to using microwave surface velocity measurements, which could also improve discharge estimates in real time.

3) Classical discharge estimates based on water level measurements and the use of steady state rating curves are affected by errors that may reach 30% (Di Baldassarre and Montanari, 2009) due i) to extrapolation beyond the range of observations used for calibrating the rating curve as well as ii) to the effect of the potential formation of loop, the effect of which is sometimes reduced by correcting formulas, such as for instance the Jones formula or others (Dottori et al., 2009).

4) Hydrological model forecasts are essentially based on predicted discharges, while hydraulic model forecasts can be both in terms of water levels and/ or discharges. In any case they are affected by a wide range of errors.

Therefore, when dealing with real time flood forecasting, the “filtered” water level is in theory the most appropriate quantity to be used as the “predictand”. Please note that we are here talking of a “filtered “quantity because the real occurrence will never be known, but one can reduce the measurement errors by using the classical Kalman Filtering technique (Kalman, 1960; Kalman and Bucy, 1961), which combines measurements and models.

Nonetheless, it is the belief of the authors that, in practice, the “observed water levels” are the “best” operational quantity to be retained as “predictand”: the errors are small and the belief of the “decision makers” is very high, while this is not so for the “filtered” quantities than are “estimated” and not “measured”.
Therefore, whenever possible, one should use the “observed water levels” as the “predictand to be used in any flood predictive uncertainty processor.

In the eventuality that water levels are not available or when one needs to predict inflows to a reservoir or a water detention area, “corrected” and “filtered” discharges should always be used. In other words prior to their use as predictands for the calibration of the hydrological uncertainty processors, improved discharge estimates must be produced both by accounting for the shape of the cross section and by taking into account the loop formation. This will eliminate most of the water level dependent “biases” , while the elimination of the random errors must again be approached by filtering techniques.

In terms of “predictors”, when available from a hydraulic model, the best choice would be the forecasted water levels. Otherwise one can either convert the predicted discharges into predicted water levels using a “corrected” rating curve, as mentioned above, or just use the predicted discharges when this is not available, since the effect of the conversion errors from discharge to levels, may not significantly affect the “order” of the predicted variables, which is what essentially dominates the NQT conversion into the Normal space.

The authors take the point and 1) will introduce this discussion in the paper, 2) will introduce another example, this time an operational one based on water levels, 3) will justify the use of discharges in the case of the Baron Fork river example, due to the fact the no water level observations or rating curve observations were made available to the participants to the DMIP2 Project.

3) **C:** *I do not find it appropriate to present the quantile regression (2.4.1) after introducing the model-conditional processor. I should instead present it before and refer to it afterwards only, ie. Section 2.1.4 should follow Section 2.2. The way it is done (in the middle of the MCP description) is confusing.*

A: The authors completely agree with Referee #2 about the incorrect position of the section 2.4.1, therefore it will be moved, also reflecting the comments of Referee #1. Quantile Regression will be presented with the other existing approaches and the Model Conditional Processor will be described in a different section.

4) **C:** *I absolutely disagree with the authors view that the HUP concept cannot be extended to multiple models as a justification for using MCP.*

A: The authors agree with the reviewer statement that the MCP cannot be justified because the HUP cannot be extended to multiple models. But this was never stated in the manuscript. The choice of MCP must be based on its performances, as it is clearly explained in Todini (2008), where (Figure 12) the results of HUP and MCP, both based on a hydraulic model forecast + an AR1, show that MCP is capable of improved performances, mostly due to the fact that it takes into account the correlation of the hydraulic model and the AR1 forecasts, which is not done in HUP.

5) **C:** *The authors should explain why the approach is to be considered "Bayesian", also if it is called that way in similar contexts such as Bayesian Model averaging. A pure Bayesian approach expresses a posterior conditional probability of a hypothesis A given B in terms of a prior probability of A, the prior probability of B and a conditional probability of B given A. It is not clear to me how Eqn (7) represents an application of Bayes theorem in this sense, as no prior information is being revised, except perhaps the model output being used as prior information.*

A: Referee #2 correctly pointed out that "A pure Bayesian approach expresses a posterior conditional probability of a hypothesis A given B in terms of a prior probability of A, the prior probability of B and a conditional probability of B given A".
Both HUP (which is strictly Bayesian from the definition of Referee #2) and MCP derive:

$$Prob(A|B) = \frac{Prob(B|A)Prob(A)}{Prob(B)}$$

according to Bayes theorem. The point is that in MCP it was acknowledged that there is no need for separately estimating $Prob(B|A)$ as described in Krzystofowicz (1999), but simply to directly estimate

$$Prob(A,B) = Prob(B|A)Prob(A)$$

to give:

$$Prob(A|B) = \frac{Prob(A,B)}{Prob(B)}$$

In the proposed approach (MCP), the variables A and B are referred to as η and $\hat{\eta}$ respectively. Eq. 7 describes how to compute the posterior conditional probability $f(\eta|\hat{\eta}) = \frac{f(\eta,\hat{\eta})}{f(\hat{\eta})}$, where $f(\hat{\eta})$ is the prior probability of $\hat{\eta}$. The joint distribution of η and $\hat{\eta}$ is defined as: $f(\eta,\hat{\eta}) = f(\eta) \cdot f(\hat{\eta}|\eta)$ and it includes the priori probability of η and the conditional probability of $\hat{\eta}$ given η. The difference from this approach to the one described by Referee #2 is that no assumptions are made on the prior probability of η, but the assumption is made directly on the joint distribution on the basis of the historical available data of η and $\hat{\eta}$.

6) **C**: *I suggest to the authors to use a more concise and clearer mathematical notation in 2.3 and 2.4.3 and the Appendix. There is significant room of improvement, while saving space in the presentation and enhancing clarity (see. e.g. Krzystofowicz 1999).*

A: The authors agree with Referee #2 that the mathematical developments can be made clearer. As pointed out also by Referee #1, in Section 2.3 is explained how to compute the conditional distribution from a multivariate normal distribution, which indeed is a well known concept, therefore this explanation will be omitted and the section shortened. Section 2.4.3 will be improved making clearer which of the two truncated normal distributions is being discussed.

7) **C:** *On page 9229 line 22 "Weibull plotting position" should be explained better.*

A: The Weibull plotting position refers to the expected value of the probability of the i^{th} element of a ranked (smallest to largest) sample of size *n* which is $\frac{i}{n+1}$. Its use is recommended when the form of the distribution is unknown and when unbiased exceedance probabilities are desired. A brief explication will be added to the manuscript.

8) **C:** *Page 9237: line 20: "compared each other" should read "compared to each other"*

A: The sentence will be corrected.

Cited Literature

Di Baldassarre, G. and Montanari, A. 2009. Uncertainty in river discharge observations: a quantitative analysis, Hydrol. Earth Syst. Sci., 13, 913–921, doi:10.5194/hess-13-913-2009.
Dottori, F., Martina, M.V.L., Todini, E., 2009. A dynamic rating curve approach to indirect discharge measurement Hydrol. Earth Syst. Sci. Discuss., 6, 859–896.
Kalman R.E. 1960. A new Approach to linear filtering and prediction problems. J. Basic Eng. Trans. ASME, 82 D, 35-45.
Kalman R.E. and Bucy R.S. 1961. New results in linear filtering and prediction theory. J. Basic Eng. Trans. ASME, 83 D, 95-108.
Krzysztofowicz R. 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resour. Res., 35 (9), 2739–2750.
Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. Intl. J. River Basin Management. Vol. 6 (2), 123-137