Response to the second list of comments by Reviewer 1

Once again, we would like to than Reviewer 1 for his or her very thorough review of our manuscript and for the interesting discussion that arise from his or her suggestions. We are also grateful to the editor for being so pro-active in the reviewing process.

About the link between risk aversion and the optimal cost-loss ratio

One point I still not understand is the possibility of a link between 2 ideas : a) risk aversion: how much an individual is eager to pay to be insured against an event ; b) the optimal ratio of false alarms and missed events for an individual, that is the number of false alarms (which cost a little amount of money c1) in order to avoid more than 1 missed event (which costs a large amount of money c2); this question can be treated by the cost-loss ratio. Since both notions are related to the amount of money an individual is eager to pay to avoid the consequences of any (respectively: more than 1) event, it is not clear for me whether we can make a link between these 2 approaches rather than opposing them.

Indeed, whenever the decision maker's action is binary (alarm or nothing), the optimal decision takes the form of a threshold, where an action is taken if the probability of a flood is greater than some value. In appendix A, we clearly expose such an example and show that risk-aversion effectively reduces the threshold value (action is taken for lower probabilities). However, in the context of the paper, the optimal action is not binary, and represents a level of spending. The same intuition however applies: for any given forecast, as risk-aversion increases, the decision maker is willing to spend more money.

About the number of events in the database

As pointed out by the authors in the discussion and in their revised manuscript, the results of the experiment on the case study are significant only if the data is representative of the catchment behaviour

(see the notion of expectation, page 13). The case study here is driven over 4 years (2011-2014), thus over a (very) limited number of (significant) events. The sufficiency of data (to infer any conclusion) should be explicitly treated. It may be useful to: a) display the observations and the forecasts (at least in the supplementary materials); b) sum up all the events (e.g. discharge over a chosen threshold), and to pay a particular attention to missed events and successful forecasts: I expect that their "ratio" would strongly influence the results of this study.

Figure 4 displays hydrographs of forecasts and observations for one year. In our opinion, this representative year is sufficient to provide an insight regarding the catchment behaviour and the model's behaviour. In addition, the revised version of the manuscript includes Figure 11, which displays the relative frequency of flood events, both forecasted and observed, over the entire period of study. Translated in terms of absolute frequencies, the observed events amount to 36 days of flooding over four years. Those include both minor and major events. This represents at least one flood per year, with various durations. In our opinion this is representative of the catchment's behaviour and sufficient to pursue the economic analysis. The following sentence was also added in the presentation of Figure 11:"Over the four year period, there has been a total number of 36 days of flooding."

Detailed comments

Page 2, lines 5 - 6: if available, some reference would be useful.

Two references were added (Ramos et al. (2013) and Demeritt et al. (2010))

Page 2, line 14 ("the latter gain in importance"): check grammar ("gains"?)

Thank you for pointing out this typo. It was corrected.

Page 2, line 15 ("However, AS there exist many sources of uncertainty in hydrological processes, there also exist many means [...]"): I did not get the logical implication between these 2 ideas.

This sentence was modified and now reads: "However, there exist multiple sources of uncertainty in hydrological processes and there also exist many means of assessing those uncertainties and building an ensemble that convey the associated information."

Page 2 line 15 ("there also exists many means"): check grammar ("exist"?)

Corrected!

Page 2, line 22: since the "value" word is in italic 2 lines further, I suggest to use italic for the "quality" word as well.

The word "quality" is now written in italics also, both on line 22 and line 24.

Page 2, line 25: isn't a point sign missing before the words "In particular"?

Corrected!

Page 2, line 26 ("In particular, the usefulness of a forecast is inherently linked to the decision maker's ability to adapt their behaviour to the information provided"): I totally agree with this sentence but why is this important in the introduction? (why to state this issue here?) In a more general way, it is the ability to use the information which allows a forecast to be useful.

We would prefer to leave this sentence as it is, in the introduction, the objective being to emphasize the difference between forecasts' quality and value.

Page 4, line 29 ("the decision maker may only distinguish between a finite set of implied damages"). I am not sure of what the reader should understand here; is this a practical observation or a more theoretical affirmation?

This is a practical observation. The sentence was modified and now reads: "In the case of flood forecasting systems, even if the streamflow values are continuous, **in practice** the decision maker may only distinguish between a finite set of implied damages".

Page 5, lines 11-12: some references (such as those given in appendix B) could usefully be provided here.

The same references provided in appendix B were added.

Page 6, line 15: since the month of the floods of 2012 and 2014 are given below, to precise in which month the 1964 flood occurred could be useful and consistent.

We included the month, which is November, and we also corrected a typo (the year of the historical flood is 1966 and not 1964).

Page 6, lines 21-22 ("The greatest concern of public authorities occurs when people refuse to evacuate [...]"): this is unfortunately a too common behaviour. However, why is this information useful here to demonstrate the pertinence of the proposed economic assessment?

We consider this to be an important example related to intangible value (or loss). It is rather difficult to assign a monetary value to the fact that people refuse to leave their house and incur risk of injuries. In the paper this is addressed using ψ.

Page 6, lines 25 and following: very little is said about the spatial resolution of the model (how many RHHU? ...). It is not clear how well it is adapted to the data inputs and to the catchment (in particular, to state that it is a physics-based model).

There are 366 RHHU for this catchment. This was added to the manuscript at line 31 on page 6. The number of RHHU is determined by "PHYSITEL", which is a GIS companion software to HYDROTEL. This number depends on many factors: the resolutions of the digital elevation map and on the available information relative to soil type and vegetation.

Page 7, lines 3-5. The model description has been well improved (compared to the first submission). However, I still don't understand why details such as those about the vertical water budget scheme (BV3C) are relevant for this publication.

We consider that the short description of BV3C is important for the reader to understand the general mechanics of HYDROTEL. We would prefer to keep it this way.

Page 8, line 19 ("In a study involving 20 catchments in Quebec"). I acknowledge the fact that the interested reader can easily find Thiboult et al. (2016). However, it could be worth providing a few details about these catchments (can we compare them to the Montmorency River catchment used in this case study?).

The catchments in Thiboult et al. (2016) vary in terms of sizes, river network densities, vegetation, topography, soil composition, etc. They certainly can't be seen as "hydrological twins" of the Montmorency catchment, but they are subject to similar hydro-meteorologic conditions. Even among themselves, they are all different. That is why the definition of "short-term", during which the uncertainty on initial conditions is dominant, varies (1 to 3 days).

The following was added on page 8: "Those catchments vary in size and other physical characteristics, but they are all subject to similar meteorological conditions, which are also shared by The Montmorency catchment."

Page 8, line 20 ("the uncertainty for initial conditions dominate the other sources of uncertainty for short term (1-day to 3-day ahead)"): 3 days seem to be a very long period compared to the response time (12h, according to page 6, line 6). Is this correct?

This is correct for the 20 catchments in Thiboult et al. (2016). The Montmorency catchment is smaller so it is expected that the definition of "short-term" will be different. The following was added on page 8:

"However, the Montmorency catchment has a smaller area than any of the 20 watersheds in Thiboult et al. (2016) and has a shorter response time. Consequently, the uncertainty on initial condition is expected to dominate for less than one day."

Page 9. The description of the "rudimentary" data assimilation scheme is much clearer in this new version. However, I am still not convinced that it can be called an Ensemble Kalman filter (there is no sequential approach here). At least, the absence of a sequential scheme could usefully be pointed out.

Our approach is indeed sequential. The Kalman gain is updated every 24h00 based on the model output and perturbed observations. The following was added in the text: "To do so, a rudimentary version of a sequential updating scheme, namely the Ensemble Kalman Filter (EnKF, Evensen, 2003) was implemented."

In addition, in the presentation of equation 4, it is now mentioned explicitly that the gain is updated sequentially.

The approach is rudimentary in the sense that no experiment was performed to optimize the distribution of the perturbations applied to the observations.

Page 9 ("The inclusion of additive perturbations for precipitation is due to the fact that strong under-captation is suspected for this catchment."): I don't understand this. Do the authors mean "additive positive perturbations"?

Yes, additive perturbations for precipitation were always positive (and small). As written in the manuscript, those perturbations were drawn from a uniform distribution bounded in [0, 0.5].

Page 10, lines 1-3: the 'm' subscript is used for 2 different items: the ensemble members and the damage categories. This could infer some confusion.

Thank you for pointing this out! The text has been modified and now reads "Strictly speaking, streamflow value associated to category m Q_m has a probability of occurrence p_m, and corresponds to a given damage $d(Q_m)$."

Page 13, line 22 ("On the other hand, there can also be various sources of non-stationarity [...]"): this argument is rather specious. This may indeed occur but there are means to detect non-stationarity. The conclusion (lines 23-24) can not (should not?) be inferred from this argument. The representation of the expectation of the utility by its average value could be better discussed (see the main comments).

We believe that it is important to mention those issues. Small samples are very common in hydrology, and yes, there are means to detect non-stationnarity. However, we believe that the available information in the context of hydrometeorological studies is always limited, sometimes more than others. Short samples and non-stationnarities are just examples of such limitations. The question of "truth" has been discussed by some authors, for instance by Weijs and van de Giessen (2011) in the context of forecast quality. We believe that the data available for this study has limitations, but that those limitations, typical of many hydrological studies, do not compromise the validity of the proposed method.

Page 17, line 28 ("We find that risk-averse end-users mainly consider the less favourable scenarios"): is this really a finding of this study? If so, it should be stated "with/in this model, we find [...]".

This has been changed accordingly. It now reads "In this paper, we..."

Page 18, lines 23-24 ("in this paper, we did not address the issue of potential cognitive biases and training issues for end-users"): I agree and thank the authors to point out these issues. In addition, it could be worth clearly stating that risk aversion is not a cognitive bias (see the answers of the authors in the discussion, page C10), e.g. page 10 after lines 21-22.

Indeed. We added the sentence: "However, since risk-aversion is not a cognitive bias, even highly trained decision makers are expected to be risk-averse (c.f. Fishburn (1989), Krzysztofowicz (1986))."

Page 27, Tab. 2: I don't understand the horizontal line under "No limit for a 1-day forecast". Should other lines (above and under "No limit for a 5-day forecast") be added?

For consistency, horizontal lines were added above and below the "No limit for a 5-day forecast".

Page 28, Fig. 1: this figure is a "schematic" representation of the CARA utility function for A ¿ 0. Why not showing a real CARA utility function for A ¿ 0? It could help the reader keeping in mind that the CARA function values are all negatives in such a case?

This is for practical purposes. It is easier to display a schematic representation (which has all the "visible" properties of the CARA) while being able to slightly rescale the picture so that all the relevant quantities (streamflow value, utility, expected utility) can be displayed.

Page 28, Fig. 1 caption: is this equation coherent with the equation 2 (page 5)? Furthermore, I am also uncomfortable with this equation. Since C is the difference between 2 utility values, it is therefore not a money amount (see page 14, line 7: "the actual value of the decision maker's utility has no interpretation). Then, if the equation in the caption is correct, why and how could C be the amount that the individual would be willing to pay up to?

You are correct. Thank you for pointing this out. The correct equation is as in equation 2 on page 5. This has been corrected.

References

Demeritt, D., Nobert, S., Cloke, H., and Pappenberger, F.: Challenges in communicating and using ensembles in operational flood forecasting, Meteorological Applications, 17, 209–222, 2010.

Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynamics, 53, 343–367, 2003.

Fishburn, P.: Retrospective on the Utility Theory of von Neumann and Morgenstern, Journal of Risk and Uncertainty, 2, 127–158, 1989.

Krzysztofowicz, R.: Expected utility, benefit, and loss criteria for seasonal water supply planning, Water Resources Research, 22, 303–312, 1986.

Ramos, M.-H., van Andel, S., and Pappenberger, F.: Do probabilistic forecasts lead to better decisions?, Hydrology and Earth System Sciences, 17, 2219–2232, 2013.

Thiboult, A., Anctil, F., and Boucher, M.-A.: Accounting for three sources of uncertainty in ensemble hydrological forecasting, Hydrology and Earth System Science, 20, doi:10.5194/hess-20-1809-2016, 2016.

Weijs, S. and van de Giessen, N.: Accounting for Observational Uncertainty in Forecast Verification: An Information-Theoretical View on Forecasts, Observations, and Truth, Monthly Weather Reviwe, 139, 2156–2162, 2011.