Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-495-AC1, 2016

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Moving beyond the cost-loss ratio: Economic assessment of streamflow forecasts for a risk-averse decision maker" *by* Simon Matte et al.

Simon Matte et al.

marie-amelie_boucher@uqac.ca

Received and published: 22 December 2016

We are thankful for your very relevant comments, which will certainly enhance the quality of the paper. We discuss each of your specific comments below and provide our answers. In a similar fashion, we also address your detailed comments. We began preparing a revised version of the manuscript which will address theses comments. We will modify it further once we receive the comments from Reviewer 2.

A) The presentation of the economic framework is too short.

Additional references for the presentation of economic elements, specifically risk aversion and utility theory:

The utility theory of von Neumann and Morgenstern was proposed for the very first time in a book, not in a journal article. This might appear unconventional for hydrologists and we recognize that this might not be very convenient for the reader who wants to read this reference. Still, the generally accepted reference for utility theory is indeed von Neumann and Morgenstern (1944). We acknowledge this issue and now present additional references.

Fishburn (1989) provides a retrospective on von Neumann and Morgenstern theory with extensive excerpts from the book. Furthermore, Fishburn (1989) (which is easy to find) precisely explain the remarkable impact this theory had on the subsequent development of economic theories and also clarifies some of its limits. We will add this reference in the revised version of the manuscript. Fishburn (1989) also cites many other journal papers that the interested reader can rely on to dig deeper into utility theory. As mentioned in our manuscript, utility theory is nowadays considered "standard" in economics and its presentation can be found in numerous textbooks (e.g. Chapters 1 and 2 from Gollier (2004)) and also online. For instance, we recommend Levin (2006) and Werner (2008) which are excellent and review the main concepts (although not peer-reviewed). We will include those references in the revised version of the manuscript.

In addition to the aforementioned references for basic description of utility theory, there exists a immense amount of literature regarding applications in many different fields. For instance, Pope and Just (1991) compare different types of utility functions to represent preferences of farmers for potato acreage and also explain elements of utility theory. Although we could not find references in hydrology where risk-aversion is treated in a similar fashion (hence our contribution), we were able to find examples where the importance of risk-aversion is acknowledged and described, in particular, Krzysztofowicz (1986) and Merz et al. (2009). Shorr (1966) also provides a very intuitive explanation of how the cost-loss ratio implies risk neutrality and attempts a reconciliation with utility theory in the simple context of “protect crops" vs “not protect crops". Cerdá Tena and Quiroga Gómez (2008) also explain that most decision makers are risk-averse when the stakes are high. In their paper, they illustrate how disregarding risk aversion can sometimes lead to misleading conclusions regarding the value of information (such as meteorological or hydrological forecasts). Their framework also involves the CARA utility function. However, the context of their application and the rest of their economic model is different from ours.

Since we wanted to keep the manuscript relatively short, we made the mistake of not providing enough details on the foundations of the economic framework. We agree with you that, since HESS is indeed a journal read mostly by Earth Scientists, some important details, intuitive presentations, as well as references are missing. Accordingly, we will (1) provide more detailed explanations regarding economic elements and (2) provide more references (namely Fishburn (1989), Levin (2006), Werner (2008), Krzysztofowicz (1986), Merz et al. (2009), Shorr (1966) and Cerdá Tena and Quiroga Gómez (2008). We believe that those modifications will clarify the link between risk-aversion and vNM utility functions, without introducing too much technical details and definitions, and while keeping the manuscript reasonably short. However, any additional suggestions are obviously welcome.

About the definition of risk aversion One of the best intuitive illustration of risk aversion and its impact on the computation of “value" is insurance. Insurance companies make sales precisely because people are risk averse. Indeed, the probability of one’s house to burn down is (in general) low. You might never see your house burn down in your life (hopefully). Yet, many people buy insurance for their property. They will pay this amount of money whether or not their house really burns down (or car gets stolen or such). A risk averse person is willing to pay to increase her certainty about the fu-

ture. In the context of insurance, it means that she is willing to pay a certain amount to increase her certainty of not loosing everything if a fire destroys her house, even thought he probability of this outcome is fairly low. The more risk-averse the person is, the more she is willing to pay to “remove" uncertainty or be insured.

About the CARA utility function and reorganization of section 2

We understand that Figure 1 is confusing. As it is in the current manuscript, Figure 1 is a generic representation of typical shapes for utility functions for risk averse, risk seeking and risk neutral individuals. It does not represent the CARA utility function. However, your comments made us realize that it would be better displaying the CARA function instead of generic shapes of utility functions. Consequently, we will reorganize section 2 according to your comments, and we will modify Figure 1 so that it is based (schematically) on the CARA utility function instead of on a generic function (see Figure 1 below).

How μ reflects the decision maker’s preferences regarding uncertainty

In the revised version of the manuscript, we will expose with more details, both in the text and graphically (see Figure 1 below), the link between (1) the fact that risk-averse individuals would be willing to spend money in order to remove risk, and (2) the concavity of the utility function.

Consider the random variable $\tilde{c}$, and its expected value $\bar{c}$. Of course, $\tilde{c}$ has a non-degenerated probability distribution, and thus may take many possible values, while $\bar{c}$ is a known value (not risky). Now suppose that $\tilde{c}$ and $\bar{c}$ can be expressed in terms of monetary units. In the context of a lottery where players have to chose between (A) receiving a random draw $\tilde{c}$ or (B) receiving $\bar{c}$ with certainty, a risk-averse decision maker will prefer receiving $\bar{c}$ with certainty than receiving a random draw from $\tilde{c}$. That is: $U(\bar{c}) > U(\tilde{c})$, or (using the definition of vNM utility functions, as displayed on page 4 of the manuscript) equivalently $\mu(\bar{c}) > \sum_{m=1}^{M} p_m \mu(c_m)$, where p_m is the probability of c_m. This is the mathematical definition of concavity (see also our new Figure 1 below as well as the very intuitive explanations related to Figure 1 of Shorr (1966)).

Note that we can also define $C > 0$, the maximal amount of money that the decision maker would be willing to spend to remove the risk associated with $\tilde{c}$, as follows:

$$\mu(\bar{c} - C) = \sum_{m=1}^{M} p_m \mu(c_m).$$

That is, the individual is indifferent between receiving $\bar{c} - C$ with certainty, or receiving a random draw from $\tilde{c}$. This argument extends directly to any change in risk: any risk-averse decision maker prefers less risky distributions, in the sense of mean-preserving second order stochastic dominance (Rothschild and Stiglitz, 1970). Our new Figure 1 below also presents a graphical version of the above discussion when there are only two states of nature.

[SEE FIGURE 1 ATTACHED]
Also available online at: http://vincentbouchereconomist.com/riskaversion.pdf

Fig. 1. A schematic representation of the utility function for risk-averse individuals. Here, only two states of the world are assumed. The state c_1 is realized with probability α and c_2 is realized with complementary probability. Since μ is concave, we see that the expected utility $U(\tilde{c}) = \alpha\mu(c_1) + (1-\alpha)\mu(c_2)$ is smaller than the utility of the expected value $U(\bar{c}) = \mu(\alpha c_1 + (1-\alpha)c_2)$. In other words, the individual would prefer receiving the certain amount $\bar{c} = \alpha c_1 + (1-\alpha)c_2$ than receiving a lottery $\tilde{c}$ which pays c_1 with probability α and c_2 with probability $1-\alpha$. Equivalently, the individual would be willing to pay *up to* $C = \mu(\alpha c_1 + (1-\alpha)c_2) - [\alpha\mu(c_1) + (1-\alpha)\mu(c_2)] > 0$ to remove the risk associated with this lottery.

We will also add details in Appendix B, which describe the properties of the CARA utility function. In particular, we will show why an increase in A is equivalent to an increase in the level of risk-aversion, and why the level of risk-aversion is independent of the wealth. Cerdá Tena and Quiroga Gómez (2008) also use the CARA utility function in a very simple “protect/not protect" context involving meteorological forecasts.

Finally, your are indeed right in stating that the concavity of μ “reflects the ‘marginal’ interest of the end user in the gain or cost”. This is also true when no risk in involved as described in Shorr (1966). However, when the decision maker faces risky situations, the concavity of μ also reflects preferences toward risk (i.e. risk aversion), for reasons mentioned above.

B) How is the upper tail of the predictive distribution taken into account?

“States of the world" We acknowledge that we did not define the term “state of the world” precisely and that this might have introduce confusion. We will define more precisely what we mean by “states of the world” in the manuscript.

The set of states of the world represent the set of realizations of the random variable for which the decision maker has preferences. For instance, in Cerdá Tena and Quiroga Gómez (2008), there are only two possible states of the world: “adverse weather" and “non adverse weather". In the case of flood forecasting systems, even if streamflow values are continuous, the decision maker may only distinguish between a finite set of implied damages. In practice, this is what we use. The revised version of the manuscript will also include a figure representing the fraction of observations which fall within each damage categories. (See also below for more details.)

Missed events in the database and sufficiency of data to draw full conclusions:

Unfortunately, Montmorency river basin inhabitants experience a fairly high number of flood events. Many properties are located very close to the river. In our paper, we didn’t want to discuss issues such as “freedom space for rivers”, but clearly the question could be raised eventually.

In the paper, we assume that the decision maker has preferences over a finite number of streamflow categories, which correspond to categories of material damage, as defined previously by Leclerc et al. (2001). We assume that the forecasted probability of occurrence of each of these categories (i.e. states of the world) is given by the fraction of the (equiprobable) members that predict streamflow values within each categories. Since the predicted streamflow values for all models cover the set of states of the world, we believe that this approximation is adequate. The revised version of the manuscript will include a new figure. It will display histograms reporting the number of events observed in each category, for each forecasting system.

Regarding missed events, there are indeed missed events in the database. Two types of missed events can be distinguished: the forecasting system can miss the *timing* of the event, the *magnitude*, or both.

Furthermore, we totally agree that the evaluation of the economic value of a forecast needs to be done over a reasonably long period of time. Technically, it relates to how well the empirical distribution of streamflow values represent the true distribution of streamflow values, and is formally represented in our context by the (empirical) expectation $\mathbb{E}_m$, as presented on page 12 of the original manuscript, and as also displayed in equations (6) and (7).

In other words, if one wants to know how "good" a forecast is, one usually looks at the *historical* performance of the forecast. This is indeed an important practical issue, which is not limited to our setting. It also affects the CRPS, as well as any other metric.

C) The discussion consider non scientific issues which some hydrologists and forecasters can disagree with.

We generally agree with this comment, and understand that our discussion may have carried some unwanted meaning. We will carefully rewrite the discussion accordingly. We will follow the suggestion to "emphasize the idea in the paragraph lines 23 to 26 on page 16".

We also want to make the following comments and clarifications.

1. We totally agree that there is a limit to which we should target the forecast to the end-user. In the paper, this is the reason why we didn't want to fix the level of risk-aversion. Our primary message (the upper-tail should be very well predicted) holds for any level of risk-aversion. We implicitly assume that the decision maker is "well trained" and does not suffer from (for instance) cognitive bias, and that he or she has the capacity to interpret the forecast (see also the second bullet point below). We see the training of end-users as an important, but very different issue, which we do not address in the paper.
2. We do not want to convey any indication that the forecasters should "bias" their forecast. On the contrary! Here, we assume that the decision maker completely believes the forecast. In practice, the forecast is not 100% reliable. The main message of the paper is that this is not a problem for forecast members in the lower tail of the predictive distribution. It is, however, a big problem for forecast members in the upper tail of the predictive distribution. If we simply assume that all forecast members are equiprobable (which is what is told the decision maker), when in reality, some of those are likely outliers, then we bias the predictive distribution. This problem is exacerbated for forecast members in the upper tail of the predictive distribution.

Unfortunately, we are not 100% sure to understand the last part of your third general comment (upper portion of page C5).

- **Then, even for risk-averse users, should not the weighting of the missed events and false alarms, accordingly to their true and known (self revealed) risk aversion be an optimal strategy? This can be done in the cost-loss ratio approach (see Verkade and Werner)** Our perception (perhaps wrong?)

is that you consider that the cost-loss ratio can allow for an evaluation of the decision maker's level of risk aversion. If so, this is not the case and this has been demonstrated in the literature (see for instance Cerdá Tena and Quiroga Gómez, 2008), as well as in Appendix A). To that effect, we will add an example following the proof in Appendix A in order to clearly expose how the predictions of the cost-loss ratio differ from the predictions of a model accounting for risk aversion.

We do not pretend that vNM utility theory is perfect. In fact, there exist a vast economic literature generalizing it and proposing alternative or complementary economic theories in contexts when it fails to convincingly predict behaviors. That being said, we believe that for the application at hand, it is the most relevant framework. Verkade and Werner (2011) is a really good paper and still one of the few to address the question of economic value of hydrological forecasts. However, it is not true that the cost-loss ratio can be used to assess the level of risk aversion of a decision maker. As written in Verkade and Werner (2011) for the description of the optimal warning rule: "It is assumed that a decision to issue a warning will only be taken if the *expected value* of the warning response is less than the *expected value* of *not* issuing a warning" (first two italics are ours). By opposition, utility theory would prescribe that "a decision to issue a warning will only be taken if the *expected utility* of the warning response is less than the *expected utility* of *not* issuing a warning". Since the utility function of a risk averse decision maker is concave, the latter statement does not lead to the same decision than the original statement from Verkade and Werner (2011). "Utility" is indeed a concept that might seem unatural at first, but it is very powerful as it encompasses more than monetary value. See for example Fishburn (1989) and Cerdá Tena and Quiroga Gómez (2008).

We believe that the economic assessment of hydrological forecasts could benefit from multidisciplinary collaborations between hydrologists and economists. This would allow for incorporation of more recent economic tools in hydro-economic

studies. However we acknowledge the complex matter of clearly exposing theories from both disciplines. We are extremely grateful for your very relevant comments which will certainly help to improve our explanations of economic theories (see answer to general comment A) to hydrologists.

- **forecasters have to encourage and help end-users training themselves to work with sophisticated forecasts (which is a conclusion shared in the literature).** We completely agree with that. However, we want to restate that risk aversion is a characteristic (of an organization or an individual) that does not disappear with training. It merely describe the preferences toward risk. In our context, a risk-neutral individual would only care about the average predicted streamflow value. This is obviously not the case, given the resources invested by forecasting organizations (includint the DEH) in order to give precise assessments of the uncertainty.

  To that effect (and as a side note), we believe that forecasts should be made assuming well trained users. This is reflected in practice by our focus on vNM utility functions. There exist a well developed literature in so-called "behavioral economics" focusing on individuals' cognitive biases when working with probabilities. That is, in many important economic situations, individuals take decisions that can only be explained by their misunderstanding of probabilities (e.g. pessimism or optimism). If those theories are important to explain how individuals make decisions, we strongly believe that cognitive biases (such as pessimism) *should not* be taken into account in evaluating the value of a forecast. Our focus on vNM utility functions reflect this position since they allows to represent preferences of a well trained individual (i.e. who understand the forecasts) having (well justified) risk averse preferences.

- **A risk-averse end-user who does not know his/her degree of risk-aversion is not an optimal user (for herself or himself)? If she or he knows her/his**

risk aversion, why doesn't she or he quantify it (e.g. in giving comprehensive costs for false alarms and missed events)?

The precise measurement of risk aversion in an individual or an organization is a very complex matter and the subject of ongoing studies. Typical empirical tools such as questionnaires based on the "willingness to pay" are one possibility (i.e. finding $C > 0$ in the description of Figure 1). In our study, we have no mean of measuring the level of risk aversion of the decision maker. Because of this, and since we want our message to hold for all risk averse decision makers, we then compared our results for many values of A.

In fact, it does not really matter that the decision maker herself or himself doesn't know her/his level of risk aversion and it does not make this person a less "optimal" user. vNM utility function are merely a tool to explain or rationalize the decision maker's decisions. In fact, in the paper we adopt the point of view that the decision maker is a (well trained) representative of "the civil security bureau" and the level of risk aversion is that of the organization.

- **even if this framework is interesting as an intellectual tool, it needs yet to demonstrate that it can bring more practical information than the classic cost-loss ratio methodology here**

  We really hope to have convinced you that the proposed framework does, indeed, bring more information than the cost-loss ratio. From our point of view, risk aversion is one such information. We added a concrete (although artificial) example in Appendix A in order to clearly explain how risk aversion affects the prediction of the cost-loss ratio.

  As a side note, we would like to mention that the cost-loss ratio of the civil security for the floods on the Montmorency River seems to be greater than one (they typically spend more that the anticipated losses). This precisely makes the cost-loss ratio framework inapplicable here. Yet, when asked if they *wanted* hydrological

forecasts and if such forecasts were *valuable* (qualitatively) for them, people from the civil security responded "Yes!" without any hesitation. They also mentioned many times that they didn't regret the resources, time and money spent for flood mitigation.

The fact that C/L is greater than one can be explained by risk aversion. Even if the cost is higher than the expected loss, it is possible (i.e. there are states of the world) that the realized loss be much higher than its expected value. Risk aversion implies that the civil security is willing to spend resources to avoid those "bad" states of the world. Although we have not verified, we think possible that C/L greater than one might be more common than we think for real life situations.

Answers to detailed comments

- Page 2:
Line 4: "uncertainty *assessment of* hydrological forecasts conveys important information for decision makers' rather 'uncertainty in hydrological forecasts conveys important information ..."
Yes, of course! Thank you for pointing out this mistake.

Line 7: I agree that the analog forecasts are of common use. However, why quoting this approach first? Is it used by the DEH? How is it relevant for this article? I suggest the authors list the most importance uncertainty sources (in a sorted way) and then present the methodologies which can be used to deal with them. The link between analog forecasts and then ensemble forecasts (line 13) is not clear.
You are perfectly right that the mention of analog forecasts is not relevant here. All occurrences will be removed from the text. We will instead follow your suggestion of

presenting the uncertainty sources.

Line 13 (“ensemble forecasts are superior to deterministic ones"): do the authors focus on ensemble forecasts or is it true as well for probabilistic forecasts? (ensemble forecasts being used as “proto" or substitute of probabilistic forecasts, since probabilistic information is drawn from this kind of forecasts).
Yes it is true of probabilistic forecasts as well. The text has been modified accordingly.

Lines 16-18: I agree that economic value assessment is not straightforward. However, assessing a forecast system by comparing forecasts with corresponding observations is not straightforward either. Indeed, there is not one quality but different qualities (especially for probabilistic (and then ensemble) forecasts). Different end-users would give different weights to these qualities (since they have specific applications).
We fully agree. The text has been modified to reflect that.

Line 26 (“which does not fully exploit the information about forecast uncertainty"): what does “fully" mean here? Verkade and Werner (2011) do take explicitly into account the uncertainty.
This sentence has been rephrased in the new version of the manuscript. Verkade and Werner (2011) do indeed account for uncertainty. What we originally meant by “fully" is that they do not account for risk aversion. Therefore, the hydrological forecasts are probabilistic and account for uncertainty, but the decsion maker is assumed to be risk neutral.

Line 31: check spelling (Neumann / Newman)
This has been corrected in the new version of the manuscript.

Line 32: since the proposed framework is based on the von Neumann and Morgenstern utility function, more references are needed than a single one of 1944. Another reference is given further (page 3, line 30). But it is a book, which may be a “classic" in the economic community, but not the easiest reference to find and read by a hydrologist.
Indeed. Please see our answer to the general comment A.

- Page 3:
Lines 5 and 6: this sentence provides some conclusions of the article. Why here in the introduction?
This sentence was removed from the introduction in the revised version of the manuscript, according to your suggestion.

Line 12 (“Results are presented and discussed in section 6"): in sections 6 and 7. Line 18 (“Most importantly"): do the authors mean “More importantly"?
This has been corrected in the new version of the manuscript.

Line 23: check English (spelling for “weighting")
This has been corrected in the new version of the manuscript.

Line 30: see comment for page 2, line 32.
Done.

- Page 4:
Line 5 (“the curvature of the function mu reflects the decision maker’s preference regarding uncertainty"): why? Some references would be gratefully

welcome.
Indeed. Please see our answer to the general comment A.

Line 9: isn't a reference to Fig. 1 missing here?
Please also see our answer to the general comment A. Figure 1 was initially meant as a generic illustration of utility functions for risk neutral, risk seeking and risk averse individuals. It is now based on the CARA utility function.

Line 20: check English ("teh")
This has been corrected in the new version of the manuscript.

Line 21: check the numerotation of tables (table 3 is referred to before table 1 and 2)
This has been corrected in the new version of the manuscript.

- Page 5:
Lines 32-33 (and lines 1-2 page 6): I did not understand why the HYDROTEL file system is useful for the reader. Are these technical details significant for this study or may they be avoided?
The aforementioned lines have been removed from the modified version of the manuscript, according to your suggestion.

- Page 6:
Lines 31-33: I am not sure that I understood correctly. Is the meteorological forecast ensemble used here computed by the meteorological service of Canada but taken from the TIGGE dataset (for some practical reasons)? If so, it might be clearer if stated this way.

As strange as it might seem, it is indeed much easier to obtain an archive of past Canadian ensemble forecasts through TIGGE than directly from Environment and Climate Change Canada. This is due to a number of practical reasons that we prefer not to detail here. However, according to your suggestion, the text of the manuscript was modified and now reads: "Precipitation and temperature ensemble forecasts from the Meteorological Service of Canada (MSC) covering the 2011âĂŤ2014 period are used. For practical reasons, those forecasts were obtained from the Thorpex Interactive Great Grand Ensemble (TIGGE) database managed by the European Center for Medium Range Weather Forecasts (ECMWF)."

- Page 7:
Lines 10 & 11 ("Thiboult et al. (2016) showed that the [...]"): please be more specific (for this catchment? For this area?...)
The work of Thiboult et al. (2016) was performed on 20 catchment in Quebec. The Montmorency River was not included in this, but it is located in the same general area (Quebec, Canada). The manuscript was modified to include this precision: "In a study involving 20 catchments in Quebec, Thiboult et al. (2016) showed that the uncertainty for initial conditions dominates ..."

Line 16: the additive coefficients for temperature inputs and the multiplicative coefficients for precipitation inputs are huge and I assume that they are much larger than the uncertainty for these inputs. Is the whole range used in practice? Is this manual 'tuning' used for more than reducing the input uncertainty in getting a best guess? Some discussion would be useful here.
Yes they are huge. They are the true operational limits at the DEH. However, it is worth emphasizing that the goal of those perturbations on precipitations and temperature is to (indirectly) affect state variables (soil moisture, snow water equivalent) and correct model uncertainties. They are not intended as to reflect the true uncertainty on

precipitations and temperature. The goal of this manual tuning is indeed to obtain a best guess regarding the initial state of the watershed (under the assumption that the state variables of the model accurately reflect the state of the watershed). However, it might reassure the reviewer (as well as everybody else) to know that those huge limits for perturbations are rarely reached. In our study, the multiplicative coefficient applied to precipitation varied between 0.5 and 2.5. Most additive coefficients for temperature varied between -3 and +2.5, with occasional large coefficient (up to -7 and +7 on 2-3 occasions). Precisions regarding what perturbations were really applied and the limits that were permitted were added in the revised version of the manuscript.

Lines 19-...: the ensemble Kalman filter (as other Kalman filters) is essentially a sequential data assimilation scheme. Here, the 'update' of the M matrix is not described and it is not clear whether it is done (since this data assimilation is made after that a best guess is provided by the human forecaster). If it is not, I am not sure that this scheme may be called an ensemble Kalman filter.
The EnKF that is implemented here follows Thiboult et al. (2016) and Mandel (2006). M is the model error covariance matrix, computed before data assimilation, at each time step of the sequential data assimilation. As such, it is not updated, as it is the model's state variables that are updated according to equation (3). Of course, updating the state variables will affect the model outputs, hence M at the following time step. Thus, in our specific implementation state variables are indeed updated, but not from an open loop simulation. The base line simulation here is the manually assimilated run. This base line simulation is good, but cruelly lacks dispersion. In that context, the purpose of the EnKF is only to consider uncertainty associated to state variables and not to improve the first guess estimate of state variables. The parameters of the EnKF were not fine tuned as in many studies (such as Thiboult and Anctil, 2015), for instance. Random perturbations added to temperature were drawn from uniform distributions U[-8,+8]$^\circ$ and U[0.5, 1.5] (multiplicative) for precip. This choice is coherent with the way the manual data assimilation was performed, but could certainly

be improved. For instance, normal error distributions are most commonly used and the spread of those distribution is calibrated until good agreement with observation is achieved. Again, the goal of the EnKF here is to add spread around best estimate of state variables, in a controlled and systematic manner. We consider that further refinement of the EnKF is outside the scope of our study.

- Page 8:
Line 16: does 's' include the cost of the forecasting system (independently from the money spent for risk mitigation)?
No, it does not. First, we unfortunately don't have this information. Second, and perhaps more importantly, this not the objective of the paper. We focus on the economic value of different forecasts. Of course when the civil security chooses which forecasting system to put in place, it must take into account the value of the system (i.e. how it will affect future spending decisions, and resulting utilities) as well as its cost. The important point is that once the system is in place, its cost should not affect spending decisions. This also motivated our focus on CARA utility functions since they do not depend on "wealth" (which would be affected by the cost of performing the forecast).

Line 21: may the author provide some figures (orders of magnitude?) or some plots?
Unfortunately we probably can't, as these are confidential. We asked civil security of Quebec for the permission to provide figures in terms of percentages or such and we are waiting for their answer.

Line 28 ("these represent relatively small levels of risks of aversion"): may the authors provide some references?
Indeed, we added a reference (Babcock et al., 1993) which provide a review of many

assumed levels of risk-aversion in the literature.

Line 28 (“it is shown that they lead to qualitative changes in the decision makers"): here again, some references would help the non specialist reader.
We meant “in this paper”. We will add a precision in order to avoid confusion. We see from Figures 8 and 9, that a departure from $A = 0$ strongly affects the comparison between the three forecasting systems.

- Page 9:
Line 25: as a non specialist, I was amazed by the range of the psi factor (1.5 to 10). Is this usual?
The range of ψ captures two important aspects. First, in 2014, the civil security spent around 3.5 times more than the realized material damages. This reflects the fact that (perhaps obviously) the decision maker also consider immaterial damages. Since it is extremely hard to evaluate immaterial damages, we let ψ vary to (very) large values. We actually also performed simulations for values much higher than 10. They are not displayed in the current version of the manuscript as the analysis would remain the same, but the graphs would be harder to read.

We believe that $\psi = 10$ is a reasonable value. Recall that immaterial damages include any damage that cannot be easily expressed in monetary value. Those include losses in the “quality of life”, avoiding law suits (including the associated bad press)... and can therefore be quite high.

- Page 10:

Line 16: the accuracy of forecasts is inversely related to lead time. Is it inversely proportional to it?

No, in general they are not.

Line 29: I am not sure that I understood the division of parameter β_m. Why all factors ($2, 1.75, 1.5, ...$) are larger than 1? I would have expected weights whose sum is 1.
Those factors reflect the benefit of early warning. The baseline is the 1-day ahead warning, so any early warning should be more beneficial. In practice, this reflects the fact that the population has time adjust (pack, empty their basements, arrange visits to their relatives...) before being evacuated.

- Page 11:
Line 17: why ’then’? First results provided are the hydrographs (Fig. 3) on which doing the visual inspection.
This sentence has been modified. It now reads: “Firstly, a visual inspection of the forecasted hydrographs is undertaken. This performance assessment also involves the well-known Continuous Ranked Probability Score (...)"

- Page 12:
Lines 24-25: I suggest that the information of Appendix C comes in the main text (it is necessary for the reader).
The content of Appendix C has been placed at the very end of section 4.3 in the revised version of the manuscript. It is introduced by “To summarize, the simulation procedure is as follows:"

- Page 13:
Lines 7 & 8 (“This figure shows that for 1-day forecasts, those based on meteorological ensembles and dressed deterministic forecasts have similar spread"):

this is not obvious for me.
Yes, indeed. Without getting into too much detail, this mistake is an artifact from a previous version of the manuscript. The beggining of section 6.1 has been modified. The sentence now reads "This figure shows that for 1-day forecasts, forecasts based on meteorological ensembles generally have low spread. This is expected, as only the forcing uncertainty is accounted for and this uncertainty requires more than one day to be propagated through the hydrological model. In addition, at short lead times the members of meteorological ensemble forecasts are often very similar. However, before each of the two flood peaks, they display more dispersion than dressed forecasts."

Line 16 ("For very short lead times, the dressed deterministic forecasts outperform those based on meteorological ensembles"): some discussion (interpretation) would be appreciated on this (common) behaviour.
The following sentence was added to the revised version of the manuscript: "As noted above, for short lead times the members of the meteorological ensemble forecasts are often very similar and the forecasts thus have no dispersion. Dressed forecasts, by definition, necessarily have more spread. Since the forecasting system is not perfect, an ensemble with very low spread is at risk of missing the observation."

Line 20: in practice, how does the DEH deal with the very "jumpy" ensemble curves? Are they used by operational forecasters?
No. The DEH doesn't use forecasts based on meteorological ensembles. They use the dressed deterministic forecasts, which are not so "jumpy". This is mentioned on page 3 (line 73-74).

- Page 14:
Lines 16 & 17 ("for higher level of risk aversion [...], the decision maker SHOULD prefer the 'no forecast' situation for low levels of Psi"): doesn't the modal verb

convey a notion of duty? (you are right if you do what you should do). I would rather write that the forecasting system has no (economic value) or usefulness for highly risk-averse users.
Yes, we agree. The text has been changed to include this suggestion.

- Page 16:
Line 12 ("The economic value of a forecasting system is necessarily dependent on the level of risk aversion of the decision maker"): first, it is more the economic value of the forecasts (you have to deduced its cost to get the value of the forecasting system). Then, even if I agree on the fact that it is very common (if not always), is this "necessary"? Can it be shown?
The text has been changed (forecasts instead of forecasting system). This specific comment is in line with general comments A-B-C. We added many new references supporting the importance of considering risk aversion in the evaluation of the economic value of forecasts.

Lines 23-26: this paragraph has to be emphasized. Morevore, communicating the forecasts in a way that the end-users would perfectly understand is a key, but it is totally different from 'overforecast'.
We agree. As mentionned above (see answer to general comment C), this portion of the manuscript will be improved thanks to the reviewer's comments.

- Page 18
Appendix A: it is referred before section 2.1 but it uses the concepts presented in this section.
We thank the reviewer for pointing out this inconsistency. However, we think that it is important to leave the reference to Appendix A in section 2 since it is where we explain the limits of the cost-loss ratio. We also don't want the reader to think that we

make unsupported claims. Consequently, line 20 now reads “Appendix A illustrates a technical presentation that builds on the concepts presented in the next section."

Appendix B. Where is Fig. 1 called?
Figure 1 was modified to avoid confusion and now represents the CARA utility function. Text has been modified accordingly

- Pages 23 & 24:
Tables 1 and 2 might be merged since their comparison is highly teachingful.
We agree. This will be included in the revised version of the manuscript

- Page 25
Table 3 could usefully be replaced by a plot of monthly values (if data is available)
We agree. This will be included in the revised version of the manuscript

- Page 27:
Fig. 1: why is not the utility function plotted for negative values? Because if c < 0, then there is no 'interest' then the utility is 0? If so, why to use it with negative values in appendix A (for example, mu(-d))?
Figure 1 now includes negative values of c. Appendix A has also been improved. Indeed, one of the advantages of working with CARA utility functions is that they are defined for any value of c, and not just for positive values.

References

Babcock, B. A., Choi, E. K., and Feinerman, E.: Risk and probability premiums for CARA utility functions, Journal of Agricultural and Resource Economics, pp. 17–24, 1993.

Cerdá Tena, E. and Quiroga Gómez, S.: Cost-Loss Decision Models with Risk Aversion, vol. 2008, Instituto Complutense de Estudios Internacionales, 2008.

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

Gollier, C.: The economics of risk and time, MIT Press, 2004.

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

Leclerc, M., Morse, M., Francoeur, J., Heniche, M., Boudreau, P., and Secretan, Y.: Analyse de risques d'inondations par embâcles de la rivière Montmorency et identification de solutions techniques innovatrices – Rapport de la Phase I – Préfaisabilité, Tech. Rep. R577, INRS-Eau and Laval University, Quebec, 2001.

Levin, J.: Choice under uncertainty, Lecture Notes, http://web.stanford.edu/%7Ejdlevin/Econ%20202/Uncertainty.pdf, 2006.

Mandel, J.: Efficient implementation of the Ensemble Kalman Filter, Tech. Rep. R1416, University of Colorado at Denver and Health Sciences Center, Denver, 2006.

Merz, B., Elmer, F., and Thieken, A.: Significance of “high probability/low damage" versus “low probability/high damage" flood events, Natural Hazards and Earth System Sciences, 9, 1033–1046, 2009.

Pope, R. and Just, R.: On testing the structure of risk preferences in agricultural supply analysis, Agricultural Journal of Agricultural Economics, 73, 743–748, 1991.

Rothschild, M. and Stiglitz, J. E.: Increasing risk: I. A definition, Journal of Economic theory, 2, 225–243, 1970.

Shorr, B.: The cost/loss utility ratio, Journal of Applied Meteorology, 5, 801–803, 1966.

Thiboult, A. and Anctil, F.: On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments, Journal of Hydrology, 529, 1147–1160, 2015.

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.

von Neumann, J. and Morgenstern, O.: Theory of games and economic behavior, vol. 60, Princeton University Press Princeton, 1944.

Werner, J.: risk aversion, in: The New Palgrave Dictionary of Economics, edited by Durlauf, S. N. and Blume, L. E., Palgrave Macmillan, Basingstoke, 2008.