Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-495-RC1, 2016

Hydrology and Earth System Sciences
Open Access
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.

Anonymous Referee #1

Received and published: 2 November 2016

This is the review of the article entitled "Moving beyond the cost-loss ratio: Economic assessment of streamflow forecasts for a risk-averse decision maker" and submitted to Hydrology and Earth System Sciences by Matte et al.. It puts forward a framework for economic assessment of forecasts, in order to go beyond the cost-loss ratio analysis. It argues that this "classic" scheme does not take into account the possible risk aversion of the end-user (the decision maker). It uses the economic notion of "utility functions" which is presented as very classical in the economic field.

*** GENERAL COMMENTS ***

To the best of my knowledge, this article is quite innovative in the hydrological field. It presents a methodology intended to help answering a question which is common in

Printer-friendly version

Discussion paper

many operational forecasting centres. One important and very positive aspect of the study is that it goes beyond a simple binary decision scheme but it takes into account the fact that crisis management is a series of decisions made at different lead times. This is a clear progress in economic assessment of forecasting system.

It is rather well written and most often clear (except for a couple of sentences, see the detailed comments below). The hydrological part of the article appears to be rather sound, even if not novel (but it is not the objective of the authors).

Yet, there are 3 issues that I would like to discuss (see the specific comments section). That is why I recommend to work on sections 2 & 7.

*** SPECIFIC COMMENTS ***

A) The presentation of the economic framework is too short

Thus, this article offers an economic study to hydrologists and forecasters. It is submitted to a Journal which is mostly read by Earth scientists. All of them are not specialists of economics and therefore, an extra care should be brought to the presentation of the economic elements of the framework. As a hydrologist (not particularly aware of economic theories), I was embarrassed by the lack of explanations about the key notions and the lack of references on which I could lean on (except for the studies on the application of the cost-loss approach for flood forecasting systems which are well referenced). In my opinion, the presentation of the notion of risk aversion is too short. Risk aversion is defined as "the fact that, given the opportunity, a decision maker would be willing to spend money (or resources) to reduce the amount of uncertainty they face." (page 2, line 28). I am not convinced that the risk aversion which is discussed further (section 2...) matches this general theoretical definition. Indeed, the fact that the mu functions reflect "the decision maker's preference regarding uncertainty" (page 4) is not obvious (at least for me). As a non specialist, my naive interpretation is that it reflects the 'marginal' interest of the end user in the gain or cost. But since the uncertainty of this gain or cost is not obviously related to the gained or lost amount (it could be as uncertain to avoid 1 M$ damage as to avoid 1k$), I don't make a clear logical link here. Furthermore, section 2 would benefit from a reorganization. Thus, appendix A is referred before the "utility" function is introduced (subsection 2.1) but it deals with this notion.

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

There is another issue in the methodology that I don't understand completely. In section 2, the utility function is the weighted sum of the outcomes of the different "states of the world" (approximated by a (I assume sufficiently large) number of realizations of the random outcome c); these outcomes being possibly modified by the mu function. In section 4 (eq. 4), the utility is computed as the sum of the weighted outcomes of the members of an ensemble. But an ensemble, especially with a low number of members, is an approximation of the whole distribution. It does not give all the different "states of the world". How is taken into account the fact that in this study, an ensemble forecast can completely miss an event (e.g. all members can be (far) under the future observation, even if the ensemble correctly approximates the [p=1/(N+1); p=N/(N+1)] part of the distribution, N being the number of members)?

In the case study, the evaluation of the benefits of the forecasts is computed over 4 years (2011-2014). I assume (I hope) that the Montmorency river basin inhabitants experienced a small number of flood events which called for mitigation actions (the authors indicate 2 floods in section 3: March 2012 and April 2014). I wonder if there is at least one flood (among all of the floods experienced during this period) for which the forecasts totally missed the event (that is, all ensemble members were (far) below the observation). If not, can we think that the results of the study would have been quite different if there were at least one missed event? Then is the data sufficient to draw full conclusions?

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

I have some reluctances to share completely the discussion which made me feel uncomfortable. According to me, the interpretation of the results is sometimes not crystal clear but it suggests some dangerous conclusions or behaviour in practice. The interpretation of the case study results goes far and there is a need for a stronger frame (what is the job of a forecaster? What is the job of the end user?). One main idea is that the economic value of the forecasts is the one given by the users. I agree with this idea, but aren't there two issues in practice here: - their economic value for the actual end users; - their potential economic value (for well trained users)? Both issues interest the operational forecasting community. But the answers to these questions are different and lead to different actions. I am not sure that practical conclusions should go further than:

1) risk-averse end-users mainly consider the upper tail of the predictive distribution (assessed through the largest member of the ensemble). Therefore, any 'outlier' would lead to costly actions and then the forecasts would be of low or null economic value if these outliers too often drive to false alarms. 1a) A consequence is that forecasters may be extra careful about the forecast values for high probability of non exceedance: this is a warning on the fact that equiprobable members of an ensemble (even if presented as so) can be seen with different weights by many actual end-users. In my opinion, this first conclusion should not go further than this simple warning. Otherwise, this could lead to suggestions to not present the whole predictive distribution (not all the ensemble members). Therefore, I suggest the authors to emphasize the idea in the paragraph lines 23 to 26 on page 16. It is very important not to suggest any 'biased' communication of the forecasts by the forecaster. The responsibility issue should also be discussed. Furthermore, it may be stressed that the interpretation of the ensemble can also be biased: an outlier (which is too often taken as a clue for a false alarm) is easily spotted, while the absence of any outlier (which can lead to a missed event) can not be detected! 1b) Furthermore this situation only conveys the fact that false alarms are seen as much less costly than missed events. 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).

2) forecasters have to encourage and help end-users training themselves to work with sophisticated forecasts (which is a conclusion shared in the literature).

Finally, there might a confusion between uncertainty aversion and risk (of damages) aversion. 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)? (it is not the forecaster's job to infer them, even if she or he may help to do so) Therefore, 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.

*** 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 ...'

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.

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).

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).

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.

Line 31: check spelling (Neumann / Newman)

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.

- Page 3

Lines 5 and 6: this sentence provides some conclusions of the article. Why here in the introduction?

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"?

Line 23: check English (spelling for "weighting")

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

- 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.

Line 9: isn't a reference to Fig. 1 missing here?

Line 20: check English ("teh")

Line 21: check the numerotation of tables (table 3 is referred to before table 1 and 2)

- 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?

- 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.

- Page 7

Lines 10 & 11 ("Thiboult et al. (2016) showed that the [...]"): please be more specific (for this catchment? For this area?...)

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.

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.

- Page 8

Line 16: does 's' include the cost of the forecasting system (independently from the money spent for risk mitigation)?

Line 21: may the author provide some figures (orders of magnitude?) or some plots?

Line 28 ("these represent relatively small levels of risks of aversion"): may the authors provide some references?

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.

- 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?

- Page 10

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

Line 29: I am not sure that I understood the division of parameter beta_m. Why all factors (2, 1.75, 1.5,...) are larger than 1? I would have expected weights whose sum is 1.

- Page 11

Line 17: why 'then'? First results provided are the hydrographs (Fig. 3) on which doing the visual inspection.

- Page 12

Lines 24-25: I suggest that the information of appendix C comes in the main text (it is necessary for the reader).

- 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.

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.

Line 20: in practice, how does the DEH deal with the very "jumpy" ensemble curves? Are they used by operational forecasters?

- 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.

- 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?

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’.

- Page 18

Appendix A: it is referred before section 2.1 but it uses the concepts presented in this section.

Appendix B. Where is Fig. 1 called?

- Pages 23 & 24

Tables 1 and 2 might be merged since their comparison is highly teachingful.

- Page 25

Table 3 could usefully be replaced by a plot of monthly values (if data is available)

- 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))?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-495, 2016.