Hydrol. Earth Syst. Sci. Discuss., 3, S799–S804, 2006
www.hydrol-earth-syst-sci-discuss.net/3/S799/2006/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Skill and relative economic value of medium-range hydrological ensemble predictions" *by* E. Roulin

Anonymous Referee #2

Received and published: 29 August 2006

Review Comments

I believe this paper makes a significant contribution by introducing the concept of using a prototype decision problem, common in meteorology, for evaluating the potential economic value of ensemble hydrologic forecasts. The paper is well organized, but needs significant editing for English usage. Overall, I enjoyed this paper, and would hope the author would consider the following comments in any revision.

The use of a climatology forecast as a reference for measuring the skill of medium-range flow predictions has some limitations. Compared to climatology, a persistence forecast would be very skillful in this situation. In particular, the relatively high skill of the deterministic forecasts (see Fig. 9) suggests that persistence alone should have

much higher accuracy than a climatology forecast. I would recommend including the skill of a persistence forecast (in addition to the other deterministic forecasts), and adding a statement that the climatology reference can overstate the skill realized with the hydrologic ensemble system, since other simple reference forecasts can easily be generated that are more accurate than climatology forecasts.

A related comment is that the same climatology forecast was assumed for all times of the year. However, due to the annual cycle of flows (see Fig. 2), the climatological probability of a defined event in seasons with lower (higher) flows could be much less (greater) than the unconditional climatological probability used as a reference. In this situation, a climatology forecast will be biased and not very accurate. I recognize that with only five years of record, it is probably necessary to pool the forecasts for the entire year to reduce sampling uncertainty. However, I would at least mention the problem with the changing climatological probability of events during a year, and the limitation of using a fixed climatology forecast throughout. A recommendation that skill and value be evaluated for subintervals within a year, if sufficiently long verification data records are available, also seems warranted.

I found the use of a static cost-loss scenario to be a very useful way to interpret the quality of hydrologic ensemble forecasts in terms of forecast value. Although treating decision-making for a short-to-medium range flood situation as a static problem might be viewed as overly simplistic, I think its ease in application makes it a quick and insightful way to screen forecast information that could be used in decisions. On the other hand, my opinion is that the extension to the two-stage decision problem is less useful as an assessment tool. Qualitatively, the resulting relative values would lead to the same inferences as with the static problem. Therefore, if screening with the simpler static problem indicates to a user that the forecasts could be valuable, I think the logical next step for a user is to move ahead and build the much more complex dynamic decision models that would be employed to ingest probabilistic forecast information in actual applications. While the two-stage problem used tries to build in more of the

HESSD

3, S799–S804, 2006

complexities, its unrealistic treatment of the dynamic elements of decisions, and the need for more specific information on the nature of the costs associated with decisions at different times, simply reveal the limitations of prototype scenarios (which would never be implemented in practice for a situation like this). These observations lead me to the following conclusion — the static cost-loss scenario is a sufficient prototype problem for initial screening to indicate the value of probabilistic forecasts, and that users encouraged by the results should take the next step of directly evaluating forecast value in the context of the way they intend to use the forecasts for decision-making (using the hindcast information).

Technical Comments

The relative measures are not defined for the CR and LBR factorizations. I gather that these were obtained by substituting the CR and LBR factorizations into equation (1) for the BSS (see Page 1376). If so, this needs to be mentioned for clarity.

The description of the LBR decomposition results does not correspond to what is shown in Fig. 7 and 8. As an example, RDIS does not increase significantly with increasing threshold value (Page 1382, Line 13). Note that for higher skill, sharper forecasts also need to have higher discrimination, but instead, the RDIS is nearly constant (Fig. 8a).

The critical need for developing hindcasts in order to use forecasts in decision-making is not stated strongly enough (especially in the conclusions). Without hindcasts, there is no way to select optimal thresholds for the decision problem (see comment on Page 1384, Line 22). With the evidence that suboptimal thresholds result in much lower relative value, one conclusion is that hindcasts need to be routinely generated to facilitate the use of forecasts in decision-making.

Note that one could also “optimize” the deterministic forecasts, by selecting a flow threshold (not necessarily equal to the actual flood threshold) that produces the best relative economic value. With hindcasts, this could be done for deterministic forecasts,

HESSD
3, S799–S804, 2006

EGU

or even persistence forecasts, which would have no value without threshold selection. This approach would also provide a fairer comparison with probabilistic forecasts, which have the advantage of being optimized (see comment on Page 1385, Line 29).

Please comment on how the analysis accounted for forecasts made when the river is currently exceeding a flood threshold, or has just recently experienced an exceedance event, i.e., cases where the costs and losses were already incurred. It would seem that some screening for events would be necessary to avoid inflating the economic value of the forecasts with these situations.

The statements in the first paragraph of the conclusions (Page 1388, Lines 6-10) that the system could be improved by using a timestep of 6 h or streamflow updating may be reasonable speculation, but they are not conclusions supported by the results presented. If presented, they should be mentioned (later in the section?) as hypotheses to examine as part of future development efforts.

The impact of the paper could be enhanced by strengthening the conclusions section. It is mostly a summary and statement of selected results. Some discussion of the implications of the results would help.

Editorial Comments

Page 1370, Line 17: Suggested wording: “better physics parameterizations”

Page 1370, Line 19: Suggested wording: “... of these computing resources at several centres is the development of ensemble methods. ”

Page 1370, Line 21: Suggested wording: “... all assumed equally probable ...”. The weather ensembles are only assumed to be equally probable.

Page 1370, Line 23: Suggested wording: “... of ECMWF, whose archives will be used in this study, has been operational ...”

Page 1371, Line 2: Spelling: “skillful”.

HESSD

3, S799–S804, 2006

Page 1371, Line 9: Suggested wording: “that hydrological ensemble prediction ...”

Page 1371, Line 20: Suggested wording: “... estimation of the value of flood forecasts for decision-making.”

Page 1372, Line 16: The statement is made that “The differences between the scale ... deserve more development than the direct use of the forecast data.” I agree that this is an important issue, but some additional explanation is needed for clarification, since it appears that forecast data were used directly.

Page 1373, Line 16: Suggested wording: “An ensemble streamflow forecast ...”

Page 1375, Line 10: Suggested wording: “The methods are mainly from the field ...”

Page 1375, Line 23: Suggested wording: “... consists of the totality ...”

Page 1376, Line 7: Suggested wording: “Two different factorizations of the joint distribution into conditional and marginal distribution conveniently relate accuracy ...”

Page 1377, Line 21: Suggested wording: “... formulation comes from Richardson (2000).”

Page 1378, Line 8: Math error: “act if $C + \mu(L - L_1) < \mu L$, that is if $C/L_1 < \mu$, and ...”

Page 1378, Line 14: Suggested wording: “... forecasts may allow a decision maker to adapt ...”

Page 1378, Line 21: Suggested wording: “The four possible combinations of deterministic forecasts and event outcomes are defined in Fig. 4.”

Page 1379, Line 23: The statement that “For most of the cost-loss situations, ...” is confusing, because it is before any results on this have been presented. Please find another transition phrase to use to motivate the two-stage decision problem.

Page 1380, Line 12: Suggested wording: “... a forecast is issued, let’s say ...”

Page 1380, Line 14: Suggested wording: “... a second forecast is issued, based on

HESSD
3, S799–S804, 2006

EGU

which ..."

Page 1381, Line 6: Suggested wording: "The pair of thresholds ..."

Page 1381, Line 21: Suggested wording: "This skill results from the high resolution of the forecast (RRES), combined with their good reliability ..." Note that the RRES is not equivalent to good RREL.

Page 1383, Line 12: Suggested wording: "... with forecast lead time."

Page 1385, Line 4: Correction: "... forecast day D+3 to ..."

Page 1385, Line 17: Correction: "... estimated on the basis ..."

Page 1386, Line 9: Suggested wording: "... with fictitious costs ..."

Page 1386, Line 11: Suggested wording: "... exceeds the 95th percentile during the forecast horizon."

Page 1386, Line 13: Suggested wording: "... forecasts are issued ..."

Page 1386, Line 20: Suggested wording: "The values of the pair of thresholds ..."

Page 1388, Lines 20 and 21: Correction: Change "base" to "basis".

Page 1388, Line 25: Suggested wording: "... with fictitious costs ..."

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 3, 1369, 2006.