Hydrol. Earth Syst. Sci. Discuss., 8, C1471–C1476, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C1471/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "

Simplifying a hydrological ensemble prediction system with a backward greedy selection of members – Part 1: Optimization criteria" *by* D. Brochero et al.

G. THIREL (Referee)

guillaume.thirel@jrc.ec.europa.eu

Received and published: 16 May 2011

This article presents a method for selecting a reduced set of members from a grand ensemble, using the ECMWF EPS and 16 hydrological members. The number of members can be reduced from 800 to 100 or even less with no deterioration of the scores or even a noteworthy improvement. The methodology used for the selection, which

is quite complex and not commonly used in hydrology, is well explained. However the article would benefit from improving the explanations about the reasons why this selection is done. Finally, an important methodological issue has to be clarified. Maybe I did not understand correctly the method, but the ECMWF EPS members are completely independent from one forecast to another so it is not correct to select them as done in this article.

This paper (and its companion) belongs to the Special Issue: "Latest advances and developments in data assimilation for operational hydrologic forecasting and water resources management". I think it should be removed from this special issue, since the articles are not about data assimilation.

Even if the authors explain very carefully how to obtain a reduced (and better) set of ensembles from a grand ensemble, they do not explain what the goal of this study is and what the possible applications could be. What I mean is that after running a grand ensemble, and performing the backward greedy technique, they obtain a slightly better ensemble. However, the computation time of the backward greedy technique is quite important. One could wonder if it is worth to perform this. Indeed, the real-time forecasts (that are the objective of this work, I guess) require quasi immediate availability of the forecasts. Do the authors think that the reduced ensemble given by the backward greedy technique over the 17-month period of this study can be used for any future real-time forecasts? My question is not about the local time (in fact the prediction lead-time) or regional space generalization (both of them are justified in the companion paper), but about a generalization for future forecasts. Maybe the application of this study expected by the authors is totally different, but I think it should be discussed in this paper.

My major concern about this article is that ECMWF EPS members are totally independent from one forecast to another future or past one. To put it in other words, the member number 1 of the forecast issued on day N (i.e. for days N to N+9) is totally independent from the member number 1 of the forecast issued on day N+1 (i.e. for days

N+1 to N+10) and from member number 1 of any future forecasts. What I understood is that on some periods (several consecutive or not forecasts), the best members are selected, and are then proven to be still better than the initial 800 members on different and not correlated periods. How can it be that some members selected because they were giving the best performance during a testing period, can be good for other periods if their only common point is their "number"? Maybe I misunderstood the methodology used in this article, and in this case a major revision has to be done on this point for justifying the methods. If this is not the case I see here an important failure in what the article relies on and it seems difficult to publish this work. Please consider this important issue and if I am wrong justify why so that future readers will not make the same conclusions that I did.

General Comments:

- If the ECMWF EPS is quite consistent during the 2005-2006 period, it evolves often, modifying its characteristics (probabilistic, spatial, etc...). Moreover the hydrological model can also evolve. How much reliable can thus the results of this study be after modification of this features?

- One of my concerns is about the study period. This period is rather short, only 17 months, which is already a low limit for drawing conclusions about HEPS and their scores because of the small number of extreme (i.e. interesting) events during this period. The authors are aware of this problem but do not discuss it in the article. In the study, this period is split in training, validation and test periods, which reduces a lot their length. It seems to me important to extend the study period, even if I know that the CPU cost is enormous.

- The 10 catchments used in this study and presented in Tab. 1 are quite small. Could you discuss how an 8x8km meteorological analysis (SAFRAN) and the even coarser ECMWF EPS are used (any kind of interpolation / disaggregation?) and can represent the processes at this small scale?

- The main characteristics of these 10 catchments are described and the authors state that it represents "a large range of hydro-climatic conditions". Thus it would be interesting to try to link the results to these hydro-climatic conditions. Are the improvements higher for smaller/larger basins? For rainy/dry basins? Etc...

- The backward greedy selection technique is applied through removing one by one the members that decrease the error the most. The justification of this choice is missing and has to be provided in this paper before publication. Furthermore, it should be explained what is "the error" in this case (i.e. that it is one of the given statistical scores, or the CC).

- I also think that a discussion is needed on the impact of selecting some members of some ensembles and using them together. All the members of a given ensemble are equally likely. Can we consider when we use 5 members from 1 ensemble, plus 26 members from another one and finally 19 from a last one (for example), that all the members of the newly created ensemble are still equally likely? If not, can we still use the rank diagram that relies on this specific assumption?

- Did you compute scores for a selection of n (with n=30, 50, 100,...) members randomly chosen from the 800? Are these scores really worse than those coming from the backward greedy technique? I think this would be a good justification of the fact that the improvement comes from the selection method and not from any statistical artifact based on the number of members for example.

Specific Comments:

- Page 2741 lines 22-23: please add "EPS" in "of the European Center for Medium-range Weather Forecasts (ECMWF) EPS". Moreover the correct spelling is "Centre", not "Center".

- Section 2: could you please discuss why you chose these 5 scores and not other ones?

- CRPS description (section 2.1): the authors should add what is the best value for this score, and what is a bad value. This could help the readers not knowing this score very well.

- The IGNS (section 2.2) is not a classical score for hydrologists, thus I think that more efforts have to be put in order to introduce it in this paper. Like it is now, it is not clear what this score shows, what a good score is, what it brings more than the other scores. Please try to improve this part.

- Section 2.4 line 10: it seems to me that Sc is not the rank of the observation, but the population of each interval.

- Section 2.4 line 16-17: Delta is the deviation, not the flatness, please put the delta symbol earlier in this sentence.

- The MDCV (section 2.5); please discuss what you consider to be the best MDCV value (from eq. (7) it seems to be 1) and why.

- Section 2.6: I think that the coefficients w1,. . ., 5 are not necessarily useful because only w3 is different to 1 in Eq. (7).

- Could you explain the sentence “Preliminary analysis showed. . .” (lines 5-6 section 2.6)? What do you mean by “covering all scenarios”? Isn’t this combined criterion supposed to give the best forecast when CC is minimal?

- The Velazquez et al. (2010) reference used in part 3 to refer to the description of the HEPS is not enough. Consider replacing this EGU abstract reference with the Velazquez et al. (2011) paper published in Advances in Geosciences, which is much more complete.

- Section 3: please add this reference for SAFRAN: Analysis of Near Surface Atmospheric Variables: Validation of the SAFRAN analysis over France, P. Quintana-Seguí, P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C. Canellas, L. Franchisteguy, S. Morel, Journal of Applied Meteorology and Climatology, 47, 92-107, 2008.

<doi:10.1175/2007JAMC1636.1>. If the 50 year reanalysis has been used, please add: Vidal, J.-P.; Martin, E.; Franchistéguy, L.; Baillon, M. & Soubeyroux, J.-M. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. International Journal of Climatology, 2010, 30, 1627-1644

- Section 3.1: from 10-day MEPS forecasts you obtain 9-day HEPS. Are you using the EPS forecasts issued at 12:00? If yes, this information is missing.

- Section 3.3: “The hydrological models were calibrated with 29 years as mean length”. Could you explain where this difference comes from, since you use the same input?

- Section 4: it is not clear for me how the training and validation subsets are used. If I understand right, the training subset is used to find the best members with the backward greedy technique, the test subset is used for verification (scores discussed in Part 6), but what is the validation subset used for? Could you explain it better?

- Page 2757, line 27: why unit weights are used here? The RD weight was 2, why did you modify it?

- Conclusions page 2763 lines 12-14: the 100 members as optimal number is not proved in this paper. This sentence should thus be less categorical.

- References: please indicate the link for downloading Talagrand et al. (1997), because it is not so easy to find.

- Fig. 1: please draw the area of the 10 basins.

- Fig. 3: it is difficult to see anything on the time series plots of CRPS and IGNS, please consider improving them.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 2739, 2011.