

Interactive comment on “Benchmarking hydrological models for low-flow simulation and forecasting on French catchments” by P. Nicolle et al.

P. Nicolle et al.

pierre.nicolle@irstea.fr

Received and published: 21 February 2014

We thank Reviewer 1 for his careful reading and evaluation of our manuscript and his detailed suggestions, which will help improving the manuscript. In the following, we explain how we will account for his comments. Each time, the comment is repeated and our reply is given.

Reviewer’s comment (RC): The authors present the results of a large experiment on low flow simulation and forecast. They compare different hydrological models (with different complexities) for their performance and try to answer interesting

C7963

research questions. The very recent low flow literature is included and referred in an appropriate way. Overall the article is well written and quite clear for the reader although there is room for some improvements.

Authors’ reply (AR): We thank Reviewer 1 for his positive feedbacks and useful comments on our manuscript.

RC: 1) Section 2.3.3: Instead of using real forecast inputs, long term meteorological archive was used. The justification of long-term archive is somewhat surprising. Was long term data necessary? It would be nice to have a short test period but with real forecast meteorological ensemble forcings (e.g. the period of 2002-2005 as in Demirel 2013b) to see the effect of input uncertainty due to the different ensembles. Could you explain/justify (a bit more) the link between possible future conditions based on the historical dataset?

AR: We agree with the reviewer that testing the models with actual series of past meteorological ensemble forecasts would have been better to account for the actual uncertainty linked to meteorological forecasts, especially for short lead times. However, there were several reasons for running the models using archives of past observations instead of actual ensemble meteorological forecasts in the context of the PREMHYCE project:

- First we wanted to test models on long series to get general results, i.e. including a few key drought events that occurred in France in the past decades, that date back to the 1970s. Such long archives of past forecasts do not exist to our knowledge.
- Second, the lead times targeted in the project were up to a few weeks, i.e. much longer than the medium-range forecasts of about two weeks that are available today. Running models up to a few weeks therefore means that medium-range ensemble forecasts should have been extended with other information, basically

C7964

based on climatic series. Since the objective of the project was not to build scenarios but rather to concentrate on hydrological models, this was not an option we considered. We think that using past observed series provides a representative ensemble of likely conditions for the period of the year, even though the ensemble is probably too large for the short lead times. However, since the target is on low flows, the catchment response to meteorological inputs is much more smoothed than in high flow conditions, which makes this problem probably less essential.

- Third, the use of past observed series is one option that was chosen to run one of the tested models in operational conditions, and which provides interesting results.

For these reasons, it was chosen not to use actual meteorological forecasts. It will be difficult to include results with actual forecasts in the article. Indeed, we think that building scenarios combining medium-range forecasts and climatic archives to reach the targeted lead times may correspond to various options that should be considered. Actually, in a separate ongoing work at Irstea (PhD of Louise Crochemore), we are doing tests to investigate this issue and we intend to report it shortly. So, to answer the reviewer's comment, we propose to extend the discussion explaining why we used this option in the context of this work and better acknowledge the possible limitations for short lead times.

RC: 2) Section 2.3.3: Using historical SAFRAN data is more straightforward than downscaling the ECMWF forecast data. I find it an interesting, pragmatic and sound approach. This approach also avoids different errors due to downscaling. But representativeness of historical data for future scenarios should be better described. This can be in a subbasin for a short period of data, just to see if the two input dataset (51/39 ECMWF ensembles and 51 SAFRAN ensemble) are compatible.

AR: As explained above, this aspect was a bit out of the scope of the PREMHYCE
C7965

project. The preliminary tests we did in a separate work to compare the use of ECMWF forecasts with the SAFRAN archive option or even combined versions of these two sources of information showed that very little information is brought by the medium-range forecast in terms of reduction of uncertainty for low-flow forecasts. Our interpretation is that the smoothing effect of the catchment is much stronger than in high flow conditions. We plan to finalize these analyses and publish this work in due course. We will add a paragraph to better explain our choice of meteorological scenarios (see previous answer).

RC: 3) Section 4.1 concludes as “a better model in simulation does not systematically mean a better model in forecasting“. The reader can be curious why? May be it is the model sophistication handling the input uncertainty (behavior during wetter or dryer inputs)? Is there a similar situation in Demirel 2013b to support this result? For example, in Demirel 2013b while GR4J (NSlow: 0.65) outperforms HBV (NSlow: 0.52) for calibration period, the model output uncertainty of the HBV (the grey range in Fig 3) was lower than GR4J.

AR: The differences in relative performance between simulation and forecasting modes can have several origins. We think that one key aspect is the way models assimilate observed flows and/or use post-processing techniques in forecasting mode. We will better analyze the added value of this part by including new tests in which models will be run in forecasting mode but without any access to observed flows. We think that this will provide insights on this key aspect.

RC: 4) The second part of the sentence “... which strengthens the need for an evaluation relative to specific modeling objectives.“ is unclear to me. What do you mean? There was a specific modelling objective in this study i.e. low flows. What else?

AR: This is indeed not fully clear. By specific modeling objectives, we meant simulation or forecasting, which are used for different operational applications (e.g. low-flow

estimation for simulation, operational real-time hydrological drought management for forecasting). This will be clarified in the revised manuscript.

RC: 5) Another unclear sentence: “These differences in performance in simulation and forecasting can be explained by the specific tools used in forecasting, which assimilate streamflow and/or correct model outputs (see Table 3).“ What kind of specific tools?

AR: By specific tools, we mean the different methods used by modelers to improve the forecasts quality, i.e. streamflow assimilation or post-processing methods. A better model performance in forecasting mode can result from these methods instead of the model himself. As mentioned in our answer to reviewer's comment 3, we will provide additional insights on this aspect and therefore we will clarify this sentence.

RC: 6) Another unclear sentence: “However, given the variety of assimilation and correction methods applied in this study, it is difficult to conclude on the relative advantages of each of them and more systematic tests would be needed.“ ..the relative advantages of each of them (of What?) Please can you explain?

AR: Linked to comments 3 and 5 above, this will be clarified. Here, we refer to the assimilation and correction methods and meant that testing the reliability of these methods would require systematically applying them to each hydrological model and comparing the performance. Here the spirit of the project was to consider modelling tools, i.e. hydrological models and the associated assimilation and/or post-processing methods selected by the modellers. Therefore, we did not distinguish the two aspects and did not investigate the sensitivity of results to each of them. However, with the additional tests we will introduce (see answer to comment 3), the added value of these methods in issuing forecasts will be commented. This sentence will be rephrased.

RC: 7) Section 4.3: variable-weight average forecast model seems similar to Bayesian model averaging. If so I would recommend the authors to include relevant references from bayesian model averaging literature e.g. Parrish, M., H.

C7967

Moradkhani, C.M. DeChant (2012), Toward Reduction of Model Uncertainty: Integration of Bayesian Model Averaging and Data Assimilation, Water Resources Research,48, W03519, doi:10.1029/2011WR011116.

AR: Here, the method's principle looks similar to the BMA from Parrish et al., but is different because we do not use the probability density of forecast for each model to combine them. As for the other methods, each member of the multi-model corresponds to the weighted mean of the forecasts issued by the five models using the same meteorological scenario. We will clarify this point and better explain the differences with the method cited by the Reviewer.

RC: 8) The authors' effort on presenting the catchment characteristics to explain the relations to model performance is very much appreciated although the strength of relations was not significant to reveal a pattern.

AR: We thank the Reviewer for this comment since the choice to include or not these results was a bit difficult, given the lack of clear relationship. But we agree that it is something important since such relationships could be expected.

RC: 9) Page 14004, line 4 “The relative gain compared to the benchmark (daily average streamflow) is very high and showed the usefulness of hydrological simulation for low flows.“ What do you mean by relative gain?

AR: The term relative may be confusing here. Actually, we meant the performance gain relatively to the benchmark. The term relative will be removed and the sentence will be rephrased.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 13979, 2013.

C7968