

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

Anonymous Referee #1

Received and published: 3 January 2014

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 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.

Specific comments

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

C7031

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?

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.

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.

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?

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?

6) Another unclear sentence: “However, given the variety of assimilation and correction

C7032

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?

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. 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.

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.

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?

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