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 2 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 manuscript reports on a model intercomparison project (MIP) on low flow simulation and forecasting in a number of rivers in France. It is well written, and methods are generally clear, with some exceptions

(see below). The study is certainly relevant and could fill a niche as no such experiments have yet focused on low flows. Reading a paper with this topic I expected more specific conclusions on what needs to be done to better simulate low flow though – this information, i.e. the relation to of the results back to the actual model differences with respect to low flow modelling are still a bit weak. In order for the paper to make a convincing and useful contribution to wider hydrological sciences, therefore, it needs to be more focused, preferably shifted from a wide reporting of the (provocatively put: rather boring and not directly transferrable) very detailed results on all individual models and score measures to e.g. the (commendably: really interesting and useful) questions already now addressed in the discussion section. The manuscript would then have the potential for a much wider impact.

Authors' reply (AR): We thank Reviewer 2 for this positive general comment and for his proposal to extend the discussion section. As mentioned in our answer to Reviewer 1, we will add the results of model testing in forecasting mode without assimilation/post-processing methods, which will give useful insights on the added value of this key aspect of the forecasting methodology. Since the article is already quite long and we do not wish to make it much longer, we will also make the result section a bit shorter, even though we think that this section is essential and represent a valuable output of our work. We propose to remove the case study illustrative applications (section 3.3), which are probably less essential to understand results.

Major comments:

RC: 1) As said above, the interesting aspects are currently hidden in the discussion section. On the other hand a proper discussion relating the results to the literature is missing. The analysis done is largely sufficient for example to make these discussion questions into the main objectives of the manuscript and thus make it visibly more useful to a wider community than the French participants of this specific MIP.

AR: To answer the Reviewer's comment and better stress the key aspects in the discussion, we propose to (1) better emphasize these questions in the introduction section, (2) increase the discussion by referring to results already published in the literature to explain how our results corroborate or not past findings and (3) better stress in the conclusion the key directions we think useful to investigate in the future.

RC: 2) The frequently repeated conclusion that "model performance depends more on the catchment than on the choice of model", which is presented as the main outcome of the results as they are presented now, needs to be better supported. The analysis doesn't prove this consistently e.g. by testing the degree of systematic ranking of models versus the degree of systematic ranking of catchments with the same approach.

AR: This aspect is indeed not clearly shown in the article. The figure 1 shows the mean rank of the models for all the criteria and for each catchment. For each criterion, we ranked the performances of all catchments. Then, a mean rank is determined with the rank of the criteria. We see there is a general trend whatever the model, even though there is some large spread for some models. We will better discuss this aspect in the article.

RC: 3) Related to the previous point is the highlighted but unsupported not-found link to catchment characteristics. For example: p13995 lines 18/19 What is the basis of this statement "satisfactory streamflow simulation seems to depend more on catchment characteristics than on the model"? It contradicts the next sentence and the sentence in the abstract that "all attempts to relate model efficiency to catchment characteristics remained inconclusive". Two of the three predictors tested are not catchment characteristics, but streamflow characteristics. Only drainage density is a catchment characteristic. In any case what are the hypotheses that these three should be influencing model performance? If influences are tested statistically there should be a hypothesis. Besides characteristics, also processes not represented by the models will hopefully influence

the performance. What these might be could also be the basis for such tests and lead to more transferrable results. Here the approach needs to be clearer and more focussed.

AR: We agree that this is confusing. To avoid confusion, we will change these comments by referring to variability of performances between models and between catchments, using an objective index of variability (e.g. standard deviation) to better support these ideas.

RC: 4) One methodological aspect the functioning of which and the relevance for the results needs to be better explained and elucidated are the post-correction methods and the assimilation methods. As these are related to model application rather than to the models themselves they can distort results. It would have been useful to test these somehow separately. But at least their effect needs to be discussed in detail with reference to literature (see comment on lacking discussion in general).

AR: We agree with the Reviewer that this is a key aspect to interpret results and differences between models in forecasting mode. As explained in our answer to Reviewer 1, we will introduce results showing model performance in forecasting mode without assimilation/post-processing, to better emphasize the role of these methods on differences between models.

RC: 5) The approach of constructing scenarios appears not well suited to look at the two very severe events in forecasting mode, as this results in most 'spaghettis' above the actual low flow. The value of this sub-analysis should be reconsidered or it should be assessed more in terms of the models themselves. What do those that can model these extreme events well have that others don't?

We agree that the approach of scenario selection has limitations, as acknowledged in our answer to Reviewer 1. One scenario that could have been added is the "no rainfall" scenario, although it is very pessimistic and corresponds to very rare events. Actually,

the year 1976 in our record is close to this scenario, which is therefore accounted for to some extent. The illustration case studies we had introduced on severe low-flow events partly intended to illustrate the differences between models in such conditions. But the Reviewer is right in stressing that the differences in model performance could be related to model structures. Although it is difficult to draw conclusions on this aspect without a close investigation of model behavior and internal states, which was a bit beyond the objectives of the project, we will introduce tentative directions for further analysis on this issue.

Specific and technical comments:

RC: 6) The abbreviations and variable names of the skill scores are very variable with some being words, some three or four letter abbreviations, some one letter variables. This makes the manuscript difficult to read. Perhaps they could be homogenized in their presentation or less could be selected (see comment above on "too detailed")

AR: We will homogenize notations to clarify the text as suggested.

RC: 7) Catchment yield (Table 2), Runoff Yield p. 13994, line 4. Choose one term. "Runoff Ratio" may be the more common term, anyway.

AR: We will replace the term catchment yield in table 2 and the term runoff yield p.13994 by runoff ratio as suggested.

RC: 8) Last sentence of the abstract was unclear to me without having read the paper.

AR: We agree and the sentence will be rephrased/extended to improve clarity.

RC: 9) Last sentence of 1.1 is a bit disconnected and surprising and would require some further information of why this is mentioned at all.

AR: We had written this sentence to avoid confusion on the actual focus of the article.

But we agree that it is not well placed here. We will move it to the method section when discussing lead times.

RC: 10) p. 13983 line 8ff This sound a bit complicated. Why can it not be called ensemble low flow forecasting (similar to ensemble flood forecasting)?

AR: The sentence will be rephrased as suggested.

RC: 11) 1.3 The list of models or forecasting systems in France is of little use to the paper

AR: We think it is useful to explain the context of this research in France and acknowledge the previous efforts to develop forecasting tools since these efforts lead to the development of most of the models that were tested in the project. However, this link may be unclear and we will clarify it in this section.

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

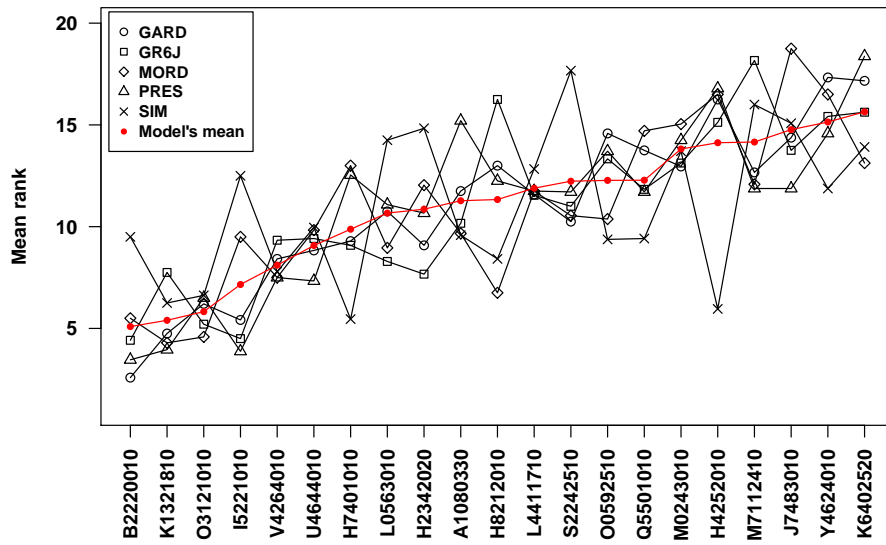**Fig. 1.** Models mean rank versus catchments

C7979