"Fuzzy committees of specialised rainfall-runoff models: further enhancements" by N. Kayastha et al.

RESPONSE TO Referee 2 (Interactive comment)

The authors are grateful to the reviewer the valuable comments and advice. We tried to address all these comments in this answer.

1. General comments

The manuscript presents further developments and an application of the so-called 'fuzzy committee models' approach on three catchments using the HBV hydrological model. Compared to previous studies on similar issue, the originality of the paper lies in the comparative assessment in verification mode of the committee models with multi-objective solutions given by the NSGAII algorithm. The issue is clearly relevant, the methods are original and the objectives of the study clearly stated. Still, there are many methodological details that are missing and to my opinion, the results are not deeply analysed as they could be

2. Specific comments

I have three main comments on the manuscripts that shall be discussed by the authors

2.1. It does not appear clear to me what scheme is preconized at the end of the paper. A number of weighting and membership functions are tested. Some are prescribed (the weighting function parameters), other are optimized (the membership function parameters). These tests are clearly interesting to study the sensitivity of the approach to these functions but there is a lack of methodological guidelines for further studies. For instance, it is not clear at what stage the membership function parameters are optimized (and how are they optimized?).

ANSWER. The combination scheme is optimized by the exhaustive search for the best γ and δ ensuring the lowest RMSE. We have added the clarifying statements about it in Section 2 and 3.

Are they calibrated after the calibration of the specialised models? On the calibration period? These elements are missing or not clearly stated and the methodological choices could be better justified.

ANSWER. We agree we were not fully clear when describing the methodology. Indeed the membership functions are optimized using the calibration data after the specialized models are calibrated. In the new version we are explaining this process clearer in Section 2.

A related issue is the possible interactions between the parameters of the membership function and the parameter of the weighting function, a point that is not discussed at all.

ANSWER. We agree, this issue was not discussed. In this paper however we had to limit ourselves to a particular set of experiments, so we added a corresponding recommendation for the future work.

2.2. I am concerned by the robustness of the conclusions and the developments of the approach since only three catchments are tested in the paper (and the results on the Alzette catchment are not fully analyzed), with very mixed results.

ANSWER. We recognize our formulations in the Conclusions were not fully accurate and did not correctly reflect the results reported in Table 3. In fact, in Table 3 one can see that for all three case studies the best committee model has higher performance than the best single model (on verification set). In this paper we presented the method to find the best set of parameters for this committee model by optimization. In the new version of the manuscript we reformulated our conclusions to make them clearer. In Conclusions we now state:

"• In calibration a committee model is always better than the single model, independent of the values of parameters MFtype and N (however we have to optimize δ and γ).

• When tested on verification data, the best committee model (identified by calibration) outperforms the best single model (identified by calibration) on all case studies."

We think the three cases studies reported are enough to demonstrate the applicability and strength of this approach. For the Leaf catchment we tested 26 different parameterizations, however, indeed on the other two catchments the number of experiments was smaller, but in all of them we demonstrated the performance of the proposed model on both calibration and test data sets. We plan to do more work on other cases studies in our future studies.

Besides, I wonder if the record periods are long enough to draw robust conclusions, given the number of free parameters of the hydrological model and the additional calibration of the membership function. This may lead to objective functions that are driven by only one flood event and thus potentially less robust inferred parameters. At the view of the surface responses on Figure 4, this probably occurred on the Bagmati catchment. To me, the main advantage of the committee model is to increase the flexibility of a given model structure without increasing its degree of freedom. This advantage is likely decreased in the proposed study by increasing the number of choices in the membership functions and the weighting functions. The few catchments studied reinforce this impression.

ANSWER. We are not sure we understand the concerns about small record periods. Table 1 presents the data set sizes: 17720 (hourly), 3717 (daily), 2922 (daily). Unfortunately this was all data we had and we had no opportunity to collect more data, but we think even the data we used covers multi-year periods, all seasons and multiple peak flows reasonably well. However we agree that in was not clear from the text of the paper, so we added a sentence to reflect this (beginning of Section 3). Unfortunately size of the paper does not allow us to add all hydrographs for all case studies, so we provided a link to a web page with these figures.

We understand the expressed concerns of having the need to choose more parameters. However more flexibility does not come for free: we had to increase the number of parameters (degrees of freedom) in the overall model. Please note that these three or four parameters (*WStype, MFtype, δ, γ*,) are automatically identified by optimization process and there is no need for the user to select them.
We fully agree that the committee model can be sensitive to the choice of these parameters and indeed using more robust optimization methods for their identification (instead of the ACCO, GA and NSGA-II that do not take robustness into account) would bring advantages – this is now recommended for future work.

Note: We found that enough projections of the model parameterizations of Leaf catchment for calibration and verification in updated Figure 3 and the plots for Bagmati and Alzette are similar as Leaf. Adding more graphics would not be helpful to improve clarity this work and the reasons of limited space we decided not to present in this paper.

2.3. The results section is particularly short. Here are some points that could be discussed in more details (the list is clearly non exhaustive). One may expect a deeper analysis on the comparison of the committee models and the multi-objective calibration framework, e.g. is there a solution from the Pareto front that performed better than the committee model in verification mode? Table 3 is not really discussed in the paper.

ANSWER. This is a valid observation. In fact we performed such analysis but due to limitations on size did not include all plots in the paper. In the updated version we add the necessary plot showing the results of such comparison: they show that none of the single models from the Pareto set outperform the committee model (both on calibration and test sets).

It may be interesting to determine if there exists a generic solution for the three catchments concerning the parameters of the weighting and the membership functions. Is there any conceptual reason why these parameters shall differ from one catchment to another?

ANSWER. We do not think there is a generic solution, and we agree we were not clear about this. We updated the Conclusions to reflect this; we now state:
"We cannot suggest the "universal" best set of parameters MFtype and WStype applicable for any case study: in calibration all of them were good, and in verification their optimal values were different for different cases."

The differences of the ACCO and GA algorithms are not discussed. Consequently, why using two algorithms?

ANSWER. We agree it was not clearly explained, and in the updated manuscript we provide an explanation in Section 3:
"We used two different algorithms for calibration due to the following reasons: 1) the initially used GA appeared to be quite slow (in terms of the required model runs), so

we also employed the faster ACCO algorithm; 2) to cross-check one by another since they both use randomization and this may affect the results."

The best solutions in calibration of the weighting and membership functions do not provide the best solution in verification. How the authors interpret this?

ANSWER. Nobody can guarantee that a particular parameterization of some model which is best in calibration will be also best in verification – simply because data sets are different, contain noise, etc. Yet another reason could be that the calibrated models overfit the data, so in verification the models are not very accurate – but we have not tested this assumption. In the new version of the manuscript we added a recommendation for the future work to perform cross-validation during calibration (for its early stopping) to prevent overfitting.

3. Technical corrections

There are many typos in the text that should be corrected.

p.678,l.11-13 the sentence does not make sense.

ANSWER. It is now written: *We can be built several sub-models instead of using only one single model to better characterize the various regimes which represent the catchment hydrological behaviour. The sub-models are also called "specialised models".*

Section 2.2 is not very clear from the first reading. The authors should state more clearly that only four configurations of parameters alpha and N are tested in the paper, and refer to Figure 1.

ANSWER. We have added the clarifying statements in Section 2

p.680 l.14 viva -> vice ???

ANSWER. Corrected.

Section 2.3: As in section 2.2, state clearly in that paragraph how and at which stage gamma, delta and N are optimized.

ANSWER. We have added the clarifying statements in Section 2 and 3.

p.683 l.17-19: Please, state that this conclusion concerns the calibration mode, even if it is explained in the next paragraph.

ANSWER. Corrected.

p.683-684 l.24-: This result is very interesting and I wonder why it appears only in conclusion and not in the results section.

ANSWER. Corrected.

p.684 l.7: I am not very clear if the committee models approach is beneficial or detrimental for hydrological simulation under Climate Change. Could the authors discuss a little more this topic?

ANSWER. Our aim was to demonstrate that the committee approach improves the accuracy of conceptual hydrological modeling. So, if a (single) model A is used for CC studies (or any other studies) and a (committee) model B is better, we suggest using model B.
(However we are not sure that the model A and B calibrated for the current state of the natural system will be really representing reality of 20-50-100 years from now when the CC effect would be manifested. The catchment will probably change so much (forestation, urbanization, biochemical changes in soil etc.) that the new models have to be built based on the assumed scenario of change.)

Homogenize significant digits of Table 3 and Table 4.

ANSWER. Corrected