Hydrol. Earth Syst. Sci. Discuss., 8, C6078–C6081, 2012
www.hydrol-earth-syst-sci-discuss.net/8/C6078/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions" *by* G. Seiller et al.

Olga Semenova (Referee)

omakarieva@gmail.com

Received and published: 11 February 2012

The paper is addressing the issue of hydrological predictions in changing environment which remain the greatest challenge for the science. It aims to propose the methodology for assessment of model suitability and transposability in time within the task of hydrological projections in climate change context. The approach implies the application of calibration/validation procedure and analysis of models performance by several criteria as for single ones so for different combinations of twenty models. The differential split sample approach is used for testing the performance of the models outside of the conditions range in which they were calibrated.

The general idea to analyze models performance in contrasted conditions is very use-

ful. It may give a lot of information on the invalidity of model structure encouraging for further model development. At the same time it is not clear what the terms of "model suitability and transposability" do specifically mean in the context of this paper. I would suggest that the authors give clear definition of those expressions. While mimicking observed hydrographs given meteorological input basically has become rather trivial task (wide range of methods is available for calibration of both model structures and model parameters), hydrological modelling in ungauged basins, especially in extreme natural conditions, such as arid and cold regions, where the processes interactions are more nonlinear and complicated, is still embarrassing, let alone predictions for future. In this connection I would suggest that the proposed approach of evaluating hydrological model behaviour under contrasted conditions is considered and applied but AFTER attentive assessment of physical adequacy of models structure and estimated models parameters. May be some complementary explanation should be added to the paper in this sense?

In the paper the authors propose some general conclusions concerning suitability of the application of a single model or their ensemble. I can not agree that all of them are justified. First of all, let's have a look at the choice of the models. What are the fundamental differences between the models? Do they more differ than resemble each other? I would suggest that the authors give more detailed overview of them. At the page 10902 6 – 8 we can find "Only two structures, M01 and M05 do not include a slow routing storage (often considered as the groundwater storage). They compensate with overland ïňĆow routing storage…" It does imply that the model M05 does not have groundwater component but compensates it by OVERLAND flow routing. Can it be true at the mountainous woody watersheds? At the same time the model M05 is chosen as one of the models performing better then many others. So the authors themselves give us an example that good performance does not mean physical adequacy. Can it be that we rely on those models in climate change projections based on their ability to mimic observed hydrographs even in contrasted conditions?

In this sense the conclusion that a single model performance always loses in front of the ensemble is not correct. First of all it definitely depends on the choice of the models. The second reason why it is wrong is that formal evaluation of the model performance should not be based just on two criteria. At Fig. 8 – 9 one can find that the difference between the performance of the model-ensemble and the best model does not exceed 0.03 of NSE. So in general it may happen that there is a single model performing at some appropriate level in all cases (which may be a bit lower than the best) but because of formal evaluation this model is rejected. It would be useful to see the combination of NSE and PB values for all models for the comparison and drawing the conclusions. Also it would be interesting to compare the results of validation presented in the paper with the case if the models would be calibrated in ordinary conditions and validated at different contrasted years. Would be the difference in performance meaningful and how much? In this way the conclusions of the section 3.2 (10908, 10 – 18) that "Nevertheless, this methodology allows identifying best-compromise individual models for each catchment based on results illustrated in Figs. 5 and 6" is not justified. The proposed methodology does not allow such identifying but certainly may help it in some way.

I would agree that "Multimodel aims at extracting as much information as possible from the existing models" (section 1.2, 10899, 12 – 16) but it may happen only in that case if the ensemble contains the models that have really different conceptualizations of unknown processes which do not contradict at least to observations (remember the overland flow of M05 model). If initially we know that a model is wrong then how it can compensate the errors of other models' structures. It is not clear why the combination of several models with different wrong concepts will certainly compensate each other instead of summing and exaggerating their own deficiencies.

My general conclusion is that the paper proposes very useful approach for evaluating models performance but it should not be the only method. I would suggest that it may be an effective complement to many others aiming the analysis of model structures

on their adequacy to physical processes. Therefore I suggest that the statements and conclusions of the paper should be changed to less categorical.

Technical notes and corrections

1. I would suggest the use of modified PB criterion. Introducing absolute value of $|Q_{sim,i} – Q_{obs,i}|$ would give the opportunity to evaluate the performance of the model in a much more effective way. Otherwise positive and negative errors compensate each other and we get some very simple average value of water balance discrepancy.

2. 10902 – 12 at the a

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 10895, 2011.