Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-339-RC2, 2016

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Interactive comment on "Looking beyond general metrics for model comparison – lessons from an international model intercomparison study" *by* Tanja de Boer-Euser et al.

Anonymous Referee #2

Received and published: 5 October 2016

Overall The paper summerises the outcome of a model intercomparison exercise, using 11 models in one catchment with performance evaluation both in time and space. The experiment seems well performed and the paper is well written. In general, model inter-comparisons should be encouraged to learn more about different model concepts, increase transparency in science and identify appropriate process descriptions. This paper is a good role-model for how such studies should be performed and documented. It is especially important as it considers a transboundary river, which is probably exposed to many model approaches resulting in different model results – this might be very confusing to decision-makers and lead to disputes across the borders.

The paper is based on state-of the-art modelling practices and presents a neat study

Printer-friendly version

Discussion paper

in a clear and concise way. It is ambitious in linking hydrological understanding to the differences in model performance. However, I think the authors should try to extract even more knowledge about processes controlling runoff at various stages of the hydrograph in the catchment as well as identify knowledge gaps, based on the ensemble modelling experiment. I also think it is important to highlight that even though the models showed on average good model performance, some parts of the hydrograph is still poorly described and should be identified. This is looking beyond what is available in general metrics! I suggest publication in HESS with only minor changes.

I have added a few detailed comments, which may improve the paper even further:

Abstract: I think the abstract can be a bit longer. I miss this sentence from the Methods section to describe the analysis made in the experiment “Three types of statistical analyses and comparisons of simulation results and observations were conducted: cumulative discharges, empirical extreme value distribution of the peak flows and flow duration curves for low flows.” I would also recommend highlighting in the abstract more findings when it comes to understanding controls of these parts of the hydrograph, related to differences in model performance, as well as, the fact that some parts of the hydrograph is poorly understood and here we can thus identify present knowledge gaps.

Introduction The first part is a bit too general and has many references to the literature, which are not necessary for the understanding of the paper and as a reader I miss guidance to how the citied papers actually contributed to the topics discussed. Often the references are lumped and support general statements that doesn’t help the reader much in deeper understanding of the literature and previous work, see example below.

Page 2, row 12: reading "Hydrological studies at different scales and under different climates have shown a large variety of hypotheses on hydrological functioning (e.g., McDonnell, 2013; Zehe et al., 2013; Fenicia et al., 2013; Clark et al., 2016; Seibert et al., 2016)." This is an example of a very general statement with no guidance to what

hypothesis the references refer to. Please, give examples and divide the lumped chain of references to explain to the author why these references are chosen – what variety did they show in hydrological understanding?

Page 2 row 22: I suggest starting the Introduction here – the paragraphs above don't contribute much but are common knowledge. This is where the study is motivated and from this point the text is much more interesting and straightforward. (Avoid reference dropping!)

Page 2 row 31: This sentence is difficult to understand: Ceola et al. (2015) concluded that deriving the causes of performance differences between various model structures is not trivial, mainly due to the considerable differences in model structures which disturbs the identification of model features that increase model performance. Are you trying to explain the problem of equifinality here? (i.e. that many different parameter settings may result in similar model performance, due to compensating processes in the model description?) Please, rephrase!

Page 4 row 15: I guess the notation of 'local time' can be removed in this decadal context.

Method The experiment is very straight forward and described with relevant level of details. I like the schematic pictures of model structures of Fig 2 and the descriptions in Table 2. Very useful for understanding!

Results I suggest including statistical metrics for model performance seen in Fig 3-9. Each graph can be assigned with a (few) metric(s) for the data it is showing, to support the analysis.

Section 4.2 Modelling the highest peaks: please, elaborate on potential causes why some models are more successful than other to capture the peaks. Are there processes they did describe that others did not? Where they more carefully calibrated? (did they apply other considerations for parameter choices?)

Fig 3: it would be more interesting to see relative error of annual flow instead of absolute; at least for a reader who is not familiar with the catchment but with general model performances.

Section 4.5 Transition from low to high flows – what can be learnt from these results of poor model performance? Interesting that a snow routine was not necessary, but could be compensated for... how?

Discussion Section 5.1: With the title 'Findings about the Meuse basin' I suggest you add some findings about which processes that seem to control flow of high peaks, low flow, etc. based on your results from model performance using different process descriptions/structures. Alternatively, you should change the title to "Model performance in the Meuse basin" but this will make the paper less appealing in my view. I would rather like to see the hydrological interpretation from the model results (i.e. change perspective to describe nature instead of models).

Page 11, row 10: I suggest to rephrase "We therefore hypothesize..." to We therefore suggest...or rather: " The results thus indicate...". Even if results may raise new questions, I think you should take the opportunity to make a statement from your study here.

Section 5.2 Benefits of an intercomparison study: Please, specify more clearly what direct benefit you could draw from using a model ensemble in this catchment. It is a costly and time-consuming process, so what are the proofs from your results claiming that it is worthwhile?

Page 11, row 28: I suggest acknowledging the importance of following a firm protocol (as you seem to have done here) in the collaboration, to ensure transparency for the result analysis and reproducibility of computational experiments. See for instance: Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C. and Arheimer, B., 2016. Most computational hydrology is not reproducible, so is it really science?. Water Resour. Res.. Accepted Author Manuscript. doi:10.1002/2016WR019285

Printer-friendly version

Discussion paper

Page 12 row 6: Note that it is quite common that catchment models underestimate extremes even though they do reasonable good on mean values, also in a multi-basin analysis. See for instance: Donnelly, C, Andersson, J.C.M. and Arheimer, B., 2016. Using flow signatures and catchment similarities to evaluate a multi-basin model (E-HYPE) across Europe. Hydr. Sciences Journal 61(2):255-273, doi: 10.1080/02626667.2015.1027710

Conclusion I would recommend highlighting your findings in understanding controls of the various parts of the hydrograph, related to differences in model performance. Was it only events during dry conditions you could refer to improved processes understanding? Check if there is more from your work to extract here! For 'Transition from low to high flows' as well as extreme flows I think you should acknowledge that we see knowledge gaps in process understanding and identify weaknesses in the model descriptions.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-339, 2016.

Printer-friendly version

Discussion paper