

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

**R. Kumar (Referee)**

rohini.kumar@ufz.de

Received and published: 7 October 2016

This paper by Boer-Euser et al describes the modeling inter-comparison results undertaken by a group of modelers over the Ourthe catchment in Belgium. They follow a fixed set of protocols for their model set-up; and the results shows the benefit of using event-specific metrics for understanding intermodal differences. The paper is quite well written and in my opinion it fits quite wells within the scope of the HESS journal. In the following, I provide some general/specific comments, which the authors may consider while revising the manuscript.

1. Missing references (relevant to this study) - specifically of the distributed model inter-comparison studies (phase 1 and 2), which have also analyzed model skill using

C1

event specific metrics for specifying inter-model differences.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D. J., & Participants, D. M. I. P. (2004). Overall distributed model intercomparison project results. *Journal of Hydrology*, 298(1), 27-60.

Smith, M. B., Koren, V., Zhang, Z., Zhang, Y., Reed, S. M., Cui, Z., ... & Anderson, E. A. (2012). Results of the DMIP 2 Oklahoma experiments. *Journal of Hydrology*, 418, 17-48.

2. I would suggest the authors to illustrate the modeling protocol in a step-wise manner or a flow chart to better follow the content.

3. It was bit difficult for me to follow the results - switching from one catchment to another. The rational/reason behind presenting results in this way should be clear to the readers.

4. I would also suggest the authors to comprehensively present their results for different catchments in a tabular form.

5. Blind validation in space: How different are the selected catchments in terms of their (dynamic) hydrologic behaviors? - for example in terms of correlation metrics of daily/hourly stream flows -

6. Among all the selected models, the GR4H model appears to have the least (or almost no) uncertainty in model outputs. Please explain?

7. Section 4.2 “Modelling the highest peaks” could be revised as “Modelling the flood peaks”

Good luck with the Revision

Rohini Kumar