Hydrology and
Earth System
Sciences
Discussions

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

Tanja de Boer-Euser et al.

t.euser@tudelft.nl

Dear Rohini Kumar,

Thank you for your positive evaluation of our manuscript. We value the comments and suggestions you have made and would like to respond to them below.

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 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.

Thank you for pointing out these two related studies, we will consult them and include them in the revised version of the manuscript.

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.

We agree with you that a flow chart with the modelling protocol would make it easier to follow for the reader. In our reply to the first reviewer, we proposed a possible schematisation for this chart.

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.

We understand that currently the selection of presented results might seem a bit arbitrary. However, in each section of the results, we selected a catchment that illustrated most clearly the message we wanted to convey. We will make sure that the reason for choosing one catchment over the other is clearly stated in the text. The results for all catchments are shown in the supplement to support the selection made for the manuscript itself.

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

We agree with you that in certain cases a table can be very informative to present results. However, we do not think that a table with NSE/NSElog values for all catchments will be very informative, as we also pointed out in our reply to the first reviewer. We could try to make a table which qualitatively summarises the results we describe in the

different paragraphs of section 4 for all catchments, but this table has as well a high risk of being very large and therefore difficult to interpret.

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 -

Thank you for asking this question. In general, the hydrological behaviour of the catchments is similar, we will provide some more details about the daily/hourly stream flows in the revised version of the manuscript; either in the text or in Table 1.

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

The band width in the 20 model realisations of GR4H is indeed very narrow. This is due to the calibration algorithm applied, which followed a stepwise calibration to an optimum after pre-filtering of the parameter space (see Table 2). This aspect of the results was discussed in section 5.3, however we will point this out in the result section as well.

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

Thank you for pointing this out, we will change the title of this section.

Good luck with the Revision Rohini Kumar

Thank you!

––––––––––––––––––––––––––––––