Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2015-532-RC1, 2016

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale prealpine basin" *by* Lieke Melsen et al.

Anonymous Referee #1

Received and published: 14 February 2016

Review of "Representation of spatial and temporal variability in large-domain hydrological models: Case study for a mesoscale prealpine basin" by Melsen et al., HESS 2016

This study focuses on the transferability of parameter sets over different temporal and spatial resolutions, using the VIC model on a basin in Switzerland as a case study. The authors use a hierarchical latin hypercube sample to identify behavioral parameter sets at different temporal (hourly, daily, monthly) and spatial (1km, 5km, 10km) resolutions. The study considers the overlap in behavioral parameter sets between different resolutions to be an indicator of transferability, which may indicate a poor model representation of either temporal or spatial variability.

The study is well-designed and executed. The results are presented clearly and address a relevant practical need, as large-scale hydrologic models are often calibrated at a coarser scale than that at which they are later applied, out of concern for computation time. The paper will be a strong contribution to HESS, subject to the clarification of a few minor issues.

(1) The percentage of overlapping behavioral parameter sets across temporal and spatial scales is used as a measure of transferability. The number of behavioral parameter sets in each case is 32. Are there sample size issues here? Should confidence intervals for proportions be reported for the transferability metric? Understandably, the amount of computation time is limited by the most finely resolved scale, so this is not to say that the authors should perform additional model runs. But some discussion of sample size seems warranted. Because the behavioral cutoff is 1% of samples, it might be interesting to see what the results look like with a more lenient restriction (2-5%) to allow a larger sample size in the estimate of the proportion.

(2) Near L95 the GLUE approach is mentioned, i.e., comparing the uninformative prior and behavioral posterior distributions of parameters. Would it have been possible to use this comparison to derive the transferability metric? For example the distance between PDFs or CDFs. It's not clear that this would have been "better" than a proportion metric, but perhaps a bit less reliant on the particular parameter sets that were sampled. Again this is only a matter for clarification.

(3) One methodological point that deserves more explanation is the use of spatially lumped parameters, even when spatial resolutions are increased. Thus the use of different spatial resolutions is really only a matter of distributed forcing data, not the parameter fields themselves. This is explored to some extent in Section 5.3 with the transfer of parameters to sub-basins, which proves difficult. The authors mention on L520 that spatial resolution does not affect the distribution of behavioral parameters very much. This makes sense as distributed forcing (especially considering the heterogeneity of precipitation over the model period as shown in Figure 3) will allow hopefully

for a more accurate time of concentration. The authors claim that spatial variability is underestimated, but how much of this finding is due to the fact that parameters are lumped? The setup and results of the spatial transferability experiment seem to contradict the hypothesis near L80 that parameters should hardly be transferable over spatial scales.

(4) Before the HLHS sample is performed, the authors find that parameter sensitivity (using the DELSA method) does not change much across scales. Is this similar or different to the finding that behavioral parameter sets (values) DO change across scales? Is there an interpretation of this result that can be discussed? Many readers may find parameter sensitivity, and its transferability across scales, equally interesting as the model performance itself.

(5) According to Figure 6, the behavioral parameter sets at finer temporal resolutions (hourly, daily) are not so good at reproducing observed streamflow (NSE $\sim$ 0.5-0.6). This may warrant further discussion. The selection of behavioral parameter sets is based on the top 1% of model runs, not the performance metrics like NSE, KGE, etc. But using those criteria, it may be that none of the model runs are "behavioral". Are there any implications of this?

Again these are largely clarifying points related to the experimental setup. Overall the experiment is well-designed and the paper well-written, with very nice figures to complement the text, and it will make a nice contribution to HESS.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2015-532, 2016.