

Response to Review of HESSD-11-5599-2014 by Hoshin Gupta

Note that original reviewer comments are in blue and author responses are in black throughout.

Overall Review Comments: I congratulate and commend the authors for an excellent job in compiling and reporting this “large-sample” data set of watersheds across the CONUS that can be used for hydrological and modeling investigations emphasizing spatial extensiveness (breadth) and generality of hydrological understanding (and therefore associated model performance). The paper is an excellent example of a relatively comprehensive study and report, and should function as a “model” or “template” for similar studies to be conducted for the other continents on our planet. In addition to reporting a carefully considered and reasoned approach to basin selection, they authors provide a benchmark assessment of simulation performance based on a “standard” lumped catchment model and calibration approach. I have only a very few suggestions for how the paper might be improved.

The authors thank Hoshin for his positive review and thoughtful consideration for areas of improvement. Regarding the specific comments, we have attempted to address them all in a thorough manner.

- 1) It might be nice to see (in Section 2) some more analysis of “basin characteristics” that would facilitate comparison/contrasting of the CONUS basins with ones from other continents. This analysis should probably include a) physical descriptors such as size, mean elevation distribution, shape (length to width ratio), and river characteristics such as distribution of stream order, dendritic pattern etc., and b) climatological descriptors including both annual values and monthly climatology.

We have added two new figures (Figs. 2 & 3) to address this comment. Figure 2 provides CDFs of basin physical descriptors while Figure 3 includes CDFs of annual climatological variables. There are two new paragraphs discussing these figures in Section 2.2. The input data will be provided with the basin set download as well.

- 2) While $NSE = 0.55$ is “OK” as a benchmark, it does reflect a relatively low level of model performance. Perhaps the authors could also slightly expand the discussion in the text to mention also the fraction of catchments exceeding (say) $NSE = 0.8$. This would help to set the tone for future studies by setting an “OK” level and a “Good level”.

This is a good idea and we have included this discussion in the text. We have included 0.55 and 0.8 NSE levels as “OK” and “Good.”

- 3) I think it would be good to more strongly emphasize the role of large-sample studies to help identify “outlier” catchments (and regions), along with the important function of “characterizing” the nature of the “outlier” (e.g., as being likely due to data errors, model inadequacy, calibration failure, etc).

We have added discussed outlier basins and their underlying causes throughout the text in a more direct manner (e.g. Fig. 4).

- 4) The issue of performance on basins with strong annual climatology does not come through very strongly in the discussion. I wonder if it would help to include a map showing where the “climatology” is strong and where it is relatively weak (e.g. strength of flow correlation (?)).

We have added a flow autocorrelation plot (Fig. 8d) and discussed how those spatial patterns relate to NSE - MNSE performance differences.

- 5) The authors report MSE decomposition components for bias and variability. For completeness, perhaps they could also report the obs-sim cross-correlation coefficient. We’ve added the correlation coefficient plot (Fig. 5b) and corresponding discussion to the text.

Minor Comments:

- 1) The sentence “Gupta et al. (2014) emphasize ...” beginning on page 5602 line 4 cites the paper twice (at beginning and end).

Fixed sentence

- 2) The phrase “well know” on page 5605 line 20 should be “well known”

Fixed statement

- 3) On page 5606 line 3, the phrase “necessitating a snow model is required,” either the word “necessitating” or the phrase “is required” should be removed.

Fixed statement

- 4) On page 5607 line 24 “is shown by” should be “as shown by”.

Fixed statement

- 5) On page 5609 line 23, the term “poorlyfollowing” should be “poorly following”.

Fixed statement

- 6) In Section 4.3 it might be interesting to compare spatial variability results with those reported by Martinez & Gupta (2010, 2011 WRR).

We have added some discussion of these papers. It is interesting to see similar spatial patterns in model performance across CONUS, even with a monthly water balance model.

- 7) On page 5615 line 8 should this be “which utilized 425 of the basins ...”?

Removed statement

Additional Suggestion: As the basis for a complementary study, I think it would be interesting to repeat the calibration-evaluation study while interactively removing the (say) 5% or 10% of the time-steps (used to compute the performance measure) that correspond to the largest simulation errors (and therefore strongly influence the selected “best” parameter set). While this might lead to interesting response surface artifacts during calibration (but none that SCE should not be able

to handle), I wonder if this would lead to more stable calibration results, when viewed for the evaluation period (functioning as sort of a fault-detection strategy)?

We agree, this would be an interesting study. We also agree that it may lead to more stable calibration results and may improve the validation period model performance. We do discuss the impact of large errors in the later figures in this text, but it would be very worthwhile to explore this in a rigorous manner. These are the types of studies we are envisioning coming from this dataset and we are pleased to see ideas for future work come quickly.