

Summary:

This manuscript provides detailed comparisons of multiple hydrologic response variables using a sophisticated integrated hydrology model and highly controlled experiment at the Landscape Evolution Observatory. The authors experiment with different levels of complexity within the model and demonstrate the importance of model heterogeneity if the goal of the model is to match spatially distributed points as opposed to integrated responses. Results also indicate the importance of considering more than just integrated hydrologic response variables when determining model parameters.

Recommendation:

Overall I find the paper to be well written. I think it provides an interesting comparison of a state of the art experiment with state of the art modeling that will be interesting to the hydrologic community and should be published in HESS. I find their scientific approach to be sound; however, I do think that some changes to the manuscript to better outline all of the test cases and highlight differences would make the discussion easier to follow. I also think that the manuscript would be of broader interest if the authors would devote some discussion the relevance of these findings to other commonly used or similar modeling approaches. I have provided detailed suggestions to this effect below.

Major Comments:

1. The introduction is focused on the need for multi objective parameter optimization. This is a good motivator for this work, but also the study is not really presenting advances for parameter optimization. Rather it's evaluating the impact of different parameterizations on model response. Therefore, I think it would be helpful to provide more background on heterogeneity and variably saturated flow processes and the state of the practices for both modeling and observations. I think this would provide a better context for where both the modeling and observations used here compare to previous work.
2. I would appreciate more details on why the observational experiments were setup the way they were. For example, how were the rainfall rates and timing determined?
3. It can be hard to keep all of the different simulations setups straight throughout the paper. I think this could be addressed by expanding on Figure 2 to better label different aspects of the domain that are discussed in the model setup and creating a new table or conceptual model that summarizes all of the runs in one place.
4. The discussion of differences between basins is mostly qualitative. I think some additional figures that plot differences between scenarios for key metrics and discussion points would strengthen the conclusions.
5. This study uses the CATHY model, but it is focuses on addressing larger questions in model uncertainty and parameterizations. Given this goal I think some additional discussion on the degree to which these results are specific to the model you are using or would be universal to other integrated flow and transport models would be quite helpful.

Specific Comments:

1. Page 2, line 8: Please expand on this point. What do you mean by 'an important example of this complexity'? Are you saying that parameter estimation has been particularly challenging for mass transport?
2. Page 3, line 6: Clarify, "infrastructure" for what?
3. Page 3, line 10: From this description it sounds like a simple sloping slab but from Figure 2 it appears that it is actually a tilted v sloping to the center of the domain. Please clarify. Also you could annotate the slopes on Figure 2 to make this even more clear.
4. Page 4, line 2: You should clarify that you are talking about just the rain from the first event here not 'all the rain water'
5. Page 4, line 2: Also here you switch from using the term 'irrigation' to 'rain'. It will be easier to follow if you pick one term and stay consistent.
6. Page 7, line 14: Please expand here to clarify how you decided on this lateral resolution.
7. Page 7, line 25: This is a very dense and long sentence. In my opinion it would easier to follow and refer back to if this information were provided in the form of a table. Also, if you keep this in paragraph form you should tie the three numbered experiments listed to simulations a-f in Table 3?
8. Page 8 line 16: How did you determine the 38cm depth for evaporation? This seems arbitrary.
9. Page 9 line 6: It would be helpful to have visual on your model figure for where the seepage face is occurring.
10. In my opinion the source sink terms listed in Tables 1 and 2 would be more easily interpreted graphically. Alternatively, I'm not sure that this information is necessary for the interpretation of the results as long as you describe how you got these terms so potentially these tables could also be deleted.
11. Table 3: Why is simulation e repeated twice in this table
12. Table 4 is difficult to follow. I think you need a separate table describing the setup of runs g-l and then report only the output metrics in this table. Also, it might help to just focus on runs g-l here and add the information for simulations a-f to Table 3.
13. Figure 3: Please describe what 'simulated, preceding case' means in the caption.
14. Figures 4, 6 and 8: I think the diamonds for the measured values should be smaller so that they are not overlapping each other or the axes so much.