Hydrol. Earth Syst. Sci. Discuss., 9, C1997–C2001, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C1997/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Multi-criteria parameter estimation for the unified land model" *by* B. Livneh and D. P. Lettenmaier

Anonymous Referee #2

Received and published: 3 June 2012

Livneh and Lettenmaier examined the benefits the multi-criteria calibration and multiple sources of observations for a land surface model for the Conterminous United States (CONUS). The model is a combination of the Noah Land Surface Model and the Sacramento Soil Accounting Model and includes components for vegetation, evapotranspiration (ET), snow dynamics, frozen soils, surface heat, radiative fluxes, soil moisture, runoff generation and infiltration. The authors calibrate the model for different scales using available information from streamflow gages, satellite-based ET, regional reanalysis and gravity-based terrestrial water storage (TWSC). The authors perform the multi-objective calibration using error components of the Nash-Sutcliffe-Efficiency (NSE) for streamflow, ET and TWCS. Based on their results, the authors evaluate tradeoffs in performance and usefulness of different sources of information used to

evaluate the hypotheses embedded in land surface models. This manuscript reflects a great effort on combining multiple sources of information to increase the fidelity in the representation of water partitioning by land surface model for a large number of catchments in CONUS. The authors have explained the steps followed in the analysis; however, additional details on the methodology and analysis of the results are required to fully understand their methodology and contribution.

Major comments:

1) One of the underlying messages of the manuscript is the importance of obtaining enough information to evaluate the formulations and parameter of complex models. In this regard, the initial discussion on equifinality can be extended to include methods to increase the power of model evaluation. Literature on how hydrologic signatures and diagnostic evaluation of hydrologic models would help to explain the selection and design of performance measures used in analysis. Right now, the manuscript does not explain properly how/why the criteria were selected for the evaluation. The selection of performance measures should be guided by the actual processes and hypothesis embedded in the model. Ideally, performance measures should relate directly to each of model component. i.e. A low flow statistic can be related to parameters related to depletion rate parameters of the lower storage, flow bias or water balance measures can be related to processes controlling water partitioning between streamflow and ET.

2) All major elements of the analysis seem sound. However, the calibration methodology and analysis require further explanation, and possible comparison to other methods. It is not clear how many parameters were optimized and what regularization strategies, if any, was used to make the problem tractable. Also, it is not properly explained the appropriateness of MOCOM-UA algorithm for the analysis, for the example one of the cited papers, Vrugt et al. (2003), concludes that MOCOM-UA "has the tendency to cluster the Pareto solutions in the most compromised region among the objectives". Are the single-criterion end points included in the estimated Paretos? The authors should explain if their results are similar if a more robust multi-objective

optimization approach is used.

3) Part of the initial intent of Nash and Sutcliffe (1970) paper was to introduce the concept of comparing models of increasing complexity. The regular NSE statistics compares to a model equal to the long-term mean of the observations. Given the larger scale and complexity of this work, the analysis could benefit from adding additional model benchmarks to the analysis. For example, the results could identify locations were the addition of increasing complexity provides a better estimate of Q and ET fluxes. Q can be compared to daily/monthly means, while the ET component can be compared to empirical ET equations. Similarly, a comparison to previous hydrologic simulations for the CONUS domain can be useful to evaluate the appropriateness of model structure and data.

4) The note related to future work to understand the impacts of structural error is taken. However, it is important that structural error in the analysis does not lead to misinterpretation of the results. From your results, it is clear that parameters can compensate for structural deficiencies of the model in locations such as Colorado basin. What is your approach to deal with structural errors? Is there any other region in CONUS where you suspect large structural errors? Is one single structure able to represent all the catchments included in the analysis?

Other comments:

4420, 21. Introduce nomenclature for Sacramento (Sac) model.

4420, 23-24. Why these sets of observations were selected for evaluation? The authors should explain which observations are necessary to calibrate properly the model, and which ones are missing, if any. It is not clear in the remaining part of the text, which parameters were not calibrated because of lack of observations or relative insensitivity.

4421, 21. Indicate the source of the naturalized flows data. Are the methods and assumptions the same for each major basin? Which natural flow characteristics are

preserved or lost?

4423, 24. Include a brief description of the methods and assumptions for the generation of gridded precipitation. For example, it is not clear if the model parametrization is artificially compensating for precipitation bias or if there is a multiplier parameter used to account for this.

4425, 10. Is there any information available to identify the contributions to TWSC by snow and groundwater? The assumptions to calculate TWSC from the model are not clear. Are these assumptions valid for all locations?

4426, 13. Table 2 can be extended to include all the parameters used in ULM and the ones that were selected for calibration. That way it would be easier to understand which model components and processes were evaluated.

4427, 9. Gupta et al. (2009) explain how MSE based calibrations can result in the underestimation of flow variability because of the interaction of error components. Were you able to test and compare MOCOM to single criteria such as KGE statistic?. This approach can decrease the dimension of the Pareto space and provide similar results with less computational effort.

4427, 18. Provide details on calibration strategy. It is not clear from the text if NSE/MSE error components from each criteria were used for the multi-objective calibration. For a combined Q, ET and TWCS calibration, this would result in nine objective functions used simultaneously by the optimization algorithm. If this is the case, did you have to adjust the calibration algorithm? What was the success rate of MOCOM during calibration?

4428, 13. I disagree with this line. I would imagine that applicability to flood forecasting would require detailed hydraulics component, and measures related to timing of the events.

4430, 9. Clarify units used to normalized ETsat.

4430, 21. Clarify units used for NDVI index since it is supposed to vary between -1 and 1.

4432, 18. I would rather mention “less skill that the long-term mean”

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 4417, 2012.