

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

B. Schaepli (Editor)

bettina.schaepli@epfl.ch

Received and published: 5 June 2012

Both reviewers agree that this manuscript represents an interesting study and should be published in HESS after moderate to major revisions. They both point out that the used methodology is not presented in sufficient detail. I would like to ask the authors to reply to all these comments in the public discussion.

I agree with the reviewers that especially the calibration method needs a better description. And I have a number of additional comments:

- The manuscript states that multi-criteria calibration might help to select a single best parameter set (p. 4419). But how does this apply in the case where different
C2057

criteria show a trade-off (i.e. where the result is a set of parameters on the Pareto-front)? I suggest making clear whether your method uses the different criteria in a sequential manner (to converge to a single set) or in a proper multi-objective setting leading to Pareto optimality and a set of Pareto-optimal parameter sets (at the moment, the presented method is a mixture of both).

- Since you mention the GLUE methodology, you should justify why you actually would like to find a single best parameter set and how you can use your methodology to assess uncertainty (see reviewers' comments)
- TWSC is a state variable change (change of storage); Q and ET are fluxes; this fundamental difference should be mentioned in the discussion of the results; furthermore, Q and TWSC integrate the entire hydrological response whereas ET only a part. The inherent difference in information content of the three signals should be discussed.
- P. 4427: an important part of the trade-off between the criteria stems from observational uncertainties and from problems of commensurability, not only from model structure.
- In general, the NSE for different signals and from different case studies cannot be compared since they refer to a very different benchmark depending on the seasonality of the reference signal (see Schaepli and Gupta, 2007). This should be discussed in more detail especially because this might cast doubt on the usefulness of the NSE value comparison among catchments and signals (and on the usefulness of fig. 12) (see also reviewer 1 comments).
- I do not understand the following paragraph p. 4433: “Within each calibration set, a single optimal solution was selected that represents a tradeoff between optimizing its respective objective functions, giving equal weight to each. For example, the calibration labeled Q, ETAWB produced a set of simulations that

minimized the objective functions for each of these quantities (Q and ETAWB), creating an envelope of similarly scoring simulations (a Pareto front).” Did you identify the Pareto-front? If yes, how did you select a single best set (a concept which does not apply in the case of Pareto-optimality)? Did you select the best set by optimizing a weighted combined objective function? In this case, what role plays the Pareto-front?

- The current manuscript version suggests at different instances that using more than one reference signal for model calibration could improve the model calibration in terms of discharge (e.g. abstract, p. 4418, line 18, p. 4435, line 7). This seems to be a bit misleading: for a given calibration period, a multi-criteria optimization cannot outperform the single best solution with respect to one of the criteria, provided that the single best solution is found (i.e. that the end point of the Pareto front is found, see also reviewers' comments). The methodology and the discussion of the results needs to be more concise with this respect. In fact, an increase in discharge simulation performance can result from a multi-criteria optimization but only if this performance is assessed for different periods than the calibration period or with different criteria than the reference NSE value.
- P. 4433: you assume that the poor cross-performance (for other criteria) of parameter sets calibrated with ET and TWSC results from observational uncertainties; I suggest to also discuss how this result is simply related to the fact that these signals do not contain enough information to constrain the model parameters (e.g. you can only poorly constrain snowmelt parameters with observed evaporation); for this, more detail on the calibrated parameters is of course required (see further reviewers' comments).
- Conclusion: I have the impression that the manuscript does not give evidence for conclusion point 6 (travel times not mentioned earlier); can you give an outlook?

Table 3 and 4: improvement with respect to what?

C2059

Reference

Schaefli, B., and Gupta, H.: Do Nash values have value?, *Hydrological Processes*, 21, 2075-2080, 10.1002/hyp.6825, 2007.

-

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