Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/hess-2021-325-RC2, 2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on hess-2021-325

Anonymous Referee #2

Referee comment on "Improving the Pareto Frontier in multi-dataset calibration of hydrological models using metaheuristics" by Silja Stefnisdóttir et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-325-RC2, 2021

The study analyses the performance of three frequently used automatic optimization algorithms including Monte Carlo simulations, simulated annealing (SA) and a genetic algorithm (GA) for a multivariable calibration in a small glaciated catchment considering streamflow, snow cover area and glacier mass balance. The results are evaluated based on the objective function values achieved by the best 100 and best 10 parameter sets as well as by uncertainty widths regarding prediction and parameter uncertainties. The authors conclude that the genetic algorithm outperformed the two other methods as it achieved better solutions and narrower confidence intervals than the other two methods.

The paper is generally well structured. The problem of model calibration is within the scope of HESS and the question of which search technique to select is a practical and relevant question that modelers have to decide upon.

However, I have several major concerns with this paper.

- The authors see the novelty of their work in "confronting for the first time these three metaheuristics most frequently applied in hydrology within a multi-output calibration framework to derive practical recommendations for further applications". This seems overstated. What about other comparison of optimization techniques? (See e.g. studies cited in Efstratiadis and Koutsoyiannis (2010)). Are Monte Carlo, GA and SA really the three most frequently applied optimization methods in hydrologic model calibration? What has been found by other studies that compare different optimization techniques and what are the research gaps that are addressed in this study?
- A fundamental problem of the current study is that the authors emphasize the multi-objective nature of the problem and aim at analyzing which of the three optimization methods provides the most balanced pareto front. However, as far as I understand, the authors did not apply optimization techniques that are designed for this task (e.g. multi-objective variants of GA). Instead it seems that the multi-objective problem was

summarized to a single-objective problem (using a weighted sum approach with fixed weightings) and SA and GA were applied in their single-objective forms (which is fine in principle but not if the aim is to study the pareto front).

- For analyzing which method performs best in representing the pareto front, the study focuses on objective function values and the number of non-dominated solutions, concluding that GA performs best. In a multi-objective setting, one should additionally consider the diversity of the solutions, i.e. how well they are spread along the pareto front.
- Some of the conclusions cannot be drawn from the results of this study. The study concludes that the results demonstrated the value of multi-dataset calibration for realistically simulating different runoff components. However, while this might have been a finding from a previous study I cannot see how this can be concluded based on results from the current study. The study also states that "it appears to be essential to give equal weights to all modelled runoff components". However, only one weight configuration has been tested so that this statement cannot be derived from the presented results.