

Hydrol. Earth Syst. Sci. Discuss., referee comment RC1  
<https://doi.org/10.5194/hess-2021-325-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.



## Review Comment on hess-2021-325

Anonymous Referee #1

---

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-RC1>, 2021

---

This well-written manuscript compares three different non-likelihood based model calibration methods for a single case study with the aim of showing the relative value of each of the methods, given the same amount of model runs.

The paper is based on a case study presented in the work of Finger et al. 2011. It takes from this earlier study the case study, the (multi-objective) data, the hydrological model, the metrics and part of the model calibration methods. What is added is the use of the simulated annealing method and the Genetic Algorithm.

While the idea of comparing the relative value of different model calibration methods given a fixed amount of simulations is somewhat interesting, I cannot recommend the publication of this paper because all results are conditional on the single case study and, more importantly, on the chosen algorithmic parameters of the compared search algorithms (simulated annealing, Genetic algorithm). These algorithmic parameters are not discussed, they are simply fixed. Furthermore, GA is treated as a single method whereas there is a multitude of implementations with different performances.

Thus: The conclusion that GA outperforms the other two within a fixed amount of simulations is not really interesting: it certainly outperforms random search (MC); whether it outperforms or not simulated annealing depends on the problem at hand and on the algorithmic parameters (how the search algorithm is tuned).

All conclusions on the value of multi-data calibration for this case study re-iterate, re-inforce earlier conclusions.

The paper does not present new methods on how to compare the algorithm outputs nor on how to analyze the optimization outputs (metrics taken from Finger et al., 2011). Accordingly, the paper does not present new methods nor new transferable insights into existing methods (except into the exact algorithms used in this paper) nor new insights into hydrological processes.

As far as I see, this paper does thus not fit HESS.

Additional comment:

I do not understand how the paper can mix Pareto-optimality and multi-objective optimization via objective function weighing: if we optimize a weighted sum of objective functions, you cannot get the Pareto-frontier (or only if you explore different weightings). Either an algorithm looks specifically for solutions on the frontier or it does not. Judging a posteriori how many we found by chance seems a rather unfair criteria to compare different algorithms. But perhaps I misunderstood something here.