EGU24-539, updated on 20 May 2024
https://doi.org/10.5194/egusphere-egu24-539
EGU General Assembly 2024
© Author(s) 2024. This work is distributed under
the Creative Commons Attribution 4.0 License.

Revisiting the common approaches for hydrological model calibration with high-dimensional parameters and objectives

Songjun Wu[1], Doerthe Tetzlaff[1,2], Keith Beven[3], and Chris Soulsby[4,1]
[1]Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, Germany
[2]Department of Geography, Humboldt University Berlin, Berlin, Germany
[3]Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YW, UK
[4]School of Geosciences, Northern Rivers Institute, University of Aberdeen, Aberdeen, UK

Successful calibration of distributed hydrological models is often hindered by complex model structures, incommensurability between observed and modeled variables, and the complex nature of many hydrological processes. Many approaches have been proposed and compared for calibration, but the comparisons were generally based on parsimonious models with limited objectives. The conclusions could change when more parameters are to be calibrated with multiple objectives and increasing data availability. In this study four different approaches (random sampling, DREAM, NSGA-II, GLUE Limits of acceptability) were tested for a complex application - to calibrate 58 parameters of a hydrological model against 24 objectives (soil moisture and isotopes at 3 depths under vegetation covers). By comparing the simulation performance of parameter sets selected from different approaches, we concluded that random sampling is still usable in high-dimensional parameter space, providing comparable performance to other approaches despite of the poor parameter identifiability. DREAM provided better simulation performance and parameter convergence with informal likelihood functions; however, the difficulty in describing model residual distribution could possibly result in inappropriate formal likelihood functions and thus the poor simulations. Multi-criteria calibration, taking NSGA-II as an example, gave ideal model performance/parameter identifiability and explicitly unravelled the trade-offs between objectives after aggregating them (into 2 or 4); but calibrating against all 24 objectives was hindered by the “curse of dimensionality”, as the increasing dimension exponentially expanded the Pareto front and increased the difficulty to differentiate parameter sets. Finally, Limits of acceptability also provided comparable simulations; moreover, it can be regarded as a learning tool because detailed information about model failures is available for each objective at each timestep. However, the limitation is the insufficient exploration of high-dimensional parameter space due to the use of Latin-Hypercube sampling.

Overall, all approaches showed benefits and limitations, and a general approach to be easily used for such complex calibration cases without trial-and-error is still lacking. By comparing those common approaches, we realised the difficulty to define a proper objective function for many-objective optimisation, either for aggregated scalar function (due to the difficulty of assigning weights or assuming a form for the residual distribution) or the vector function (due to the

expansion of the Pareto front). In this context, the Limits of Acceptability approach provided a more flexible way to define the "objective function" for each timestep, though it introduces extra demands in understanding data uncertainties and deciding on what should be considered acceptable. Moreover, in such many-objective optimisation, it is possible that not a single parameter set can capture all the objectives satisfactorily (not in 8 million run in this study). The non-existence of any global optimal in the sample suggests that the concept of equifinality should be embraced in using an ensemble of comparable parameters to represent such complex systems.