Man vs. Machine: An interactive poll to evaluate hydrological model performance of a manual and an automatic calibration

Johannes Wesemann, Reinhard Burgholzer, Mathew Herrnegger, and Karsten Schulz
Institute of Water Management, Hydrology and Hydraulic Engineering (IWHW), University of Natural Resources and Life Science, Vienna, Austria (johannes.wesemann@boku.ac.at)

In recent years, a lot of research in hydrological modelling has been invested to improve the automatic calibration of rainfall-runoff models. This includes for example (1) the implementation of new optimisation methods, (2) the incorporation of new and different objective criteria and signatures in the optimisation and (3) the usage of auxiliary data sets apart from runoff.

Nevertheless, in many applications manual calibration is still justifiable and frequently applied. The hydrologist performing the manual calibration, with his expert knowledge, is able to judge the hydrographs simultaneously concerning details but also in a holistic view. This integrated eye-ball verification procedure available to man can be difficult to formulate in objective criteria, even when using a multi-criteria approach.

Comparing the results of automatic and manual calibration is not straightforward. Automatic calibration often solely involves objective criteria such as Nash-Sutcliffe Efficiency Coefficient or the Kling-Gupta-Efficiency as a benchmark during the calibration. Consequently, a comparison based on such measures is intrinsically biased towards automatic calibration. Additionally, objective criteria do not cover all aspects of a hydrograph leaving questions concerning the quality of a simulation open.

This contribution therefore seeks to examine the quality of manually and automatically calibrated hydrographs by interactively involving expert knowledge in the evaluation. Simulations have been performed for the Mur catchment in Austria with the rainfall-runoff model COSERO using two parameter sets evolved from a manual and an automatic calibration. A subset of resulting hydrographs for observation and simulation, representing the typical flow conditions and events, will be evaluated in this study. In an interactive crowdsourcing approach experts attending the session can vote for their preferred simulated hydrograph without having information on the calibration method that produced the respective hydrograph. Therefore, the result of the poll can be seen as an additional quality criterion for the comparison of the two different approaches and help in the evaluation of the automatic calibration method.