Recent Developments and Applications of the OSTRICH Calibration Toolkit

Drew Allan Loney, PhD PE, Bureau of Reclamation, Lakewood, CO, dloney@usbr.gov
Douglas Woolridge, PE, Bureau of Reclamation, Lakewood, CO, dwoolridge@usbr.gov
Kristin Mikkelson, PhD PE, Bureau of Reclamation, Lakewood, CO, kmikkelson@usbr.gov
L. Shawn Matott, PhD, University of Buffalo, Buffalo, New York, lsmatott@buffalo.edu
Juliane Mai, PhD, University of Waterloo, Waterloo, Ontario, juliane.mai@uwaterloo.ca

Introduction

Calibration is a critical step in developing a successful hydrologic model when observation data is available adjust physical process representations. One of the largest sources of uncertainty in a hydrologic model comes from error in estimating parameters that are input into the model. The model's input parameter set is adjusted during calibration to reduce or minimize the model error with respect to the observed data. Manual calibration is a time-consuming and laborious task in which parameters are adjusted to improve model performance, and a new model simulation is initiated with the adjusted parameters. This process is repeated many times until the model achieves sufficient performance metric values, and the model is deemed successful. The manual calibration process can take days or weeks depending on the number of parameters and the complexity of the model. In addition to the substantial time investment, manual calibration likely does not result in parameter sets that fully optimize the model results and is unrealistic for many of the highly parametrized hydrologic models used today.

Many parameter optimization tools already exist that address the issues with manual calibration. Optimization algorithms can be scripted into software and run automatically to find an optimal parameter set. The tools can be customized for each hydrologic model by making modifications to the tool to read the different input or outputs specific to that model. In addition to optimizing parameter sets, the tools can be used to provide parameter uncertainty estimates. This is especially useful when working in a probabilistic framework, which is often required to quantify a system's risk and reliability. Implementation of optimization techniques save substantial time and budget from hydrologic model development.

This work highlights recent improvements to and applications of the Optimization Software Toolkit for Research Involving Computational Heuristics (OSTRICH) toolkit. Maintained as a collaboration between the Bureau of Reclamation (Reclamation), University of Waterloo, and University of Buffalo, OSTRICH is intended to be an approachable, flexible, and scalable optimization toolkit. OSTRICH utilizes an input structure that allows rapid development and debugging of optimization workflows. Additionally, OSTRICH provides multiple optimization algorithms to match the objective function topology as well as for rapid transition among the algorithms. Several recent deployments of OSTRICH within Reclamation will be highlighted as use cases.

Algorithm Modernization

OSTRICH, which is written primarily in C++, was developed by the University of Buffalo, the University of Waterloo, and Reclamation's Technical Service Center (TSC). Reclamation TSC has recently assumed the role of hosting and maintaining the OSTRICH code, which is available on GitHub. To address new use cases, improve code performance, and enhance software maintainability, Reclamation initiated a process to modernize the OSTRICH architecture.

OSTRICH includes numerous optimization algorithms from which the user can select the appropriate approach. The algorithms can be broadly classified into local and global algorithms, those that look in the region near to the initial condition and those that broadly search the parameter space, respectively. The utilized algorithm should be selected based on the problem, the model, and the confidence in the initial condition. Oftentimes, multiple algorithms may be combined in sequence to explore the full parameter space prior to refining the local solution. OSTRICH facilitates readily switching among algorithms by replacing the algorithm block within the input file. The focus of the modernization began with the genetic and Gauss-Marquart-Levenberg solvers with the intention to continue to other algorithms as part of future work. The OSTRICH repository currently has a main branch ("main") and a development branch ("dev"). The configuration files developed for this effort are set up to support the updated architecture which, at the time of this effort, is available on the development branch of the project.

OSTRICH can be run in either a Linux-based or Windows environment. It can be run in parallel mode to have multiple model simulations occurring simultaneously, which is essential for models such as Precipitation Runoff Modeling System (PRMS; Markstrom et al., 2016), which has thousands of spatially-distributed parameters to calibrate, or Soil Water Balance Model (SWB; (Westenbroek et al., 2018), where each model run can take a fairly long time. Calibrations that were performed on a Linux-based operating system used the USGS High Performance Computing (HPC) center's Denali HPC (Falgout, Gordon, and Davis 2019). To run OSTRICH in parallel on Windows, Microsoft Message Passing Interface (MPI) must be installed. OSTRICH simulations are executed from the command line once all files are in the working directory and the configuration file is set up. For more details on installation, see OSTRICH repository and documentation (usbr/ostrich, 2023).

Template Development

An OSTRICH calibration typically requires three items – an input file, a template file, and model postprocessing logic – in addition to other files required by a model. OSTRICH needs an input file that is used to specify basic model configurations such as paths to model files and associated files, optimization algorithm and parameters, objective function, hydrologic parameter information, and performance metric information. The input file must be placed in the same directory as the OSTRICH executable. A parameter template file must also be included in the directory. The template file is the hydrologic model's input parameter file with parameter placeholder values that are linked to names in the configuration file.

Finally, the user typically must include model output post-processing code to calculate performance metrics for comparison with those specified in the configuration file. This output of the performance metric code can be considered the objective function that OSTRICH seeks to optimize. The model output post-processing code and the objective function code are typically written in R, Python, or some such coding structure. While OSTRICH contains the ability to read directly from some model output formats, post-processing operations are often more complex than can be readily addressed directly by OSTRICH.

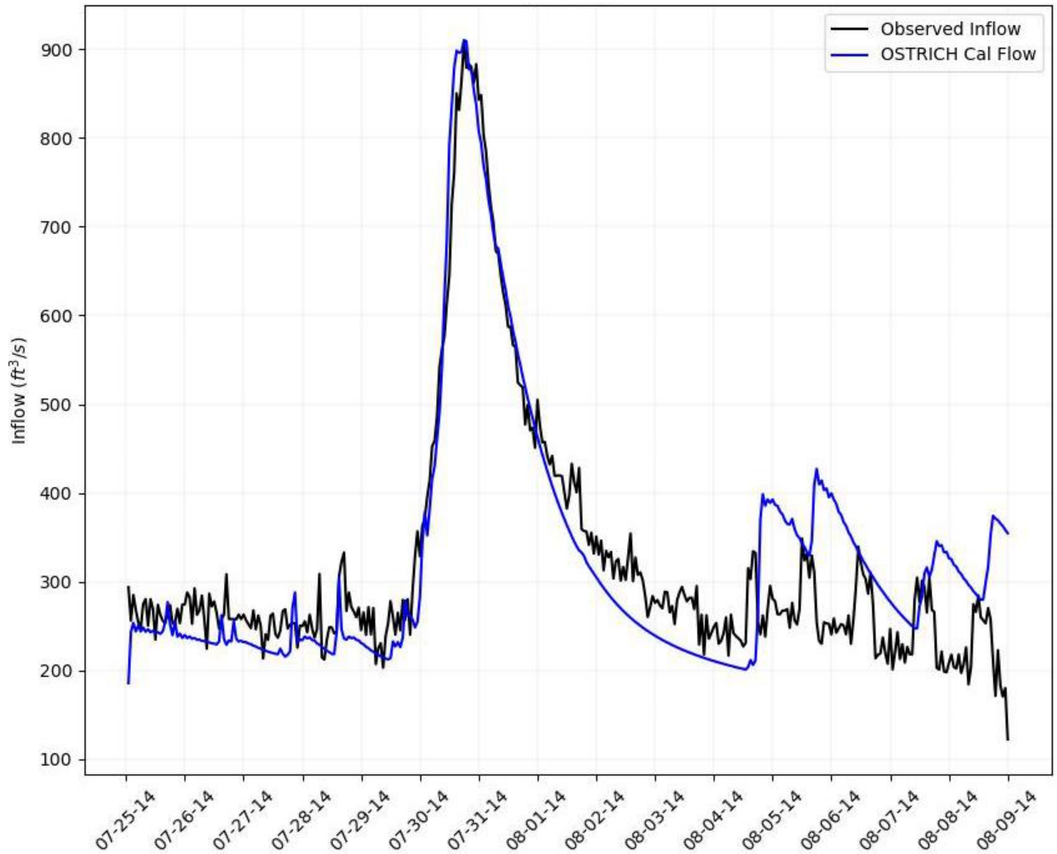

Figure 1. OSTRICH calibration results for a Reclamation SEFM model

The algorithms used for this study include the Genetic Algorithm (GA; Yoon and Shoemaker 2001) and Particle Swarm Optimization (PSO; Kennedy and Eberhart 1995), from the development and main branches, respectively. The GA is a global optimization algorithm that efficiently solves complex, nonlinear problems. The GA operates by repeatedly mutating and shuffling parameters over multiple generations to evolve the population of alternatives toward the global optimum. The mutation operations ensure a global exploration of the parameter space while the parameter shuffle acts to converge the population. PSO is a highly parallel, global optimization algorithm which treats alternatives as particles within high-dimensional parameter space. The algorithm iterates toward a solution by moving the particles through the space using a combination of inter-particle communication, local particle motion, and random particle motion. The combination of inter-particle communication and random particle motion allows the PSO algorithm to find an optimal solution within the parameter space regardless of how the parameter set is initialized. As the number of calibrated parameters grows, the number of particles (i.e., alternatives) must grow along with the number of iterations across the particle swarm. The objective function describes the function that OSTRICH seeks to optimize. The objective functions available in OSTRICH are weighted sum of squared error (WSSE) and General-purpose Constrained Optimization Platform (GCOP).

OSTRICH was implemented for five different hydrologic models for this effort: SWB (SWB; Westenbroek et al., 2018), the PRMS (PRMS; Markstrom et al., 2015), Corps of Engineers Quality Width Averaged 2D (CE-QUAL-W2; Cole and Wells, 2015), the Stochastic Event Flood Model (SEFM; MGS Software, 2018), and the Sacramento Soil Moisture Account Model (SAC-SMA; Burnash et al., 1973).

Conclusions

OSTRICH is a flexible, readily approachable framework for hydrologic model calibration. The modernization done as part of the described effort provides a foundation for ongoing OSTRICH development. Additionally, the configuration files created for each of the models can be used on future projects to simplify implementation of OSTRICH. Reclamation remains committed to ongoing development and use of the OSTRICH calibration toolkit.

References

Burnash, R.J.C., R.L. Ferral, and R.A. McGuire. 1973. "A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers." Sacramento, CA.

Cole, T.M., and Wells, S.A. 2015. "CE-QUAL-W2: A Two-Dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 4.0." Portland, OR.

Falgout, J., J. Gordon, and M. Davis. 2019. "USGS Denali Supercomputer."

Kennedy, J., and R. Eberhart. 1995. "Particle Swarm Optimization." In *International Conference on Neural Networks 1995*, edited by IEEE. https://doi.org/10.1007/978-3-030-61111-8_2.

Markstrom, S.L., R.S. Regan, L.E. Hay, R.J. Viger, R.M.T. Webb, R.A. Payn, and J.H LaFontaine. 2015. "PRMS-IV, the Precipitation-Runoff Modeling System, Version 4." In *U.S. Geological Survey Techniques and Methods 6. Modeling Techniques*, 263–64.

MGS Software LLC. 2018. "Stochastic Event Flood Model (SEFM) Technical Support Manual." Olympia, WA.

usbr/ostrich. U.S. Bureau of Reclamation, 2020. Accessed: Feb. 27, 2023. [Online]. Available: https://github.com/usbr/ostrich

Westenbroek, Stephen M., John A. Engott, Victor A. Kelson, and Randall J. Hunt. 2018. "SWB Version 2.0--A Soil-Water-Balance Code for Estimating Net Infiltration and Other Water-Budget Components." *U.S. Geological Survey*, no. Techniques and Methods 6-A59: 118. https://pubs.usgs.gov/tm/06/a59/tm6a59.pdf.

Yoon, J., and C.A. Shoemaker. 2001. "Improved Real-Coded GA for Groundwater Bioremediation." *Journal of Computing in Civil Engineering* 15 (3).