Automatic Model Calibration with Expert Knowledge

Jonathan Quebbeman, Shaun Carny, Gerald Day, John Park, Paul Micheletty

August 26, 2018

Abstract

Automatic model calibration using multiple objective algorithms has been in practice for decades, but several challenges include the development of pareto optimal solution sets with no physical meaning, inconsistent parameter convergence between trials, and increased parameter variance in the non-dominated population. Large sets of parameters can lead to *compensatory calibration* with similar fitness but widely varied solution sets. To help guide the multi-objective evolutionary algorithm solution process, expert knownledge can be infused into an objective function leading to pareto solutions adhering to physically meaningful ranges, maintaining relationships between parameters, and creating increased regional consistency. This approach was applied to a case-study for calibrating the Animas River basin in Colorado using a the Research Distributed Hydrologic model and the Sacramento Soil Moisture Accounting and Snow-17 algorithms using the NSGA-II evolutionary algorithm. As a result of this process, fitness equal to or exceeding manual calibration was achieved, in addition to the development of realistic parameter set solutions.

1 Introduction

Genetic algorithms can be used for automatic model calibration using multiple objectives, which leads to a population of non-dominated solutions. Although the fitness of these solutions may be acceptable for each of the objectives, the corresponding parameter sets may not be physically realistic or meaningful. Manual review of the solutions by experts or third-parties may question the models given parameters on the extremes of acceptable ranges, or values may be contradicting each other.

This can occur in over-determined problems with one parameter compensating for another, otherwise known as *compensatory calibration*. Further, this creates population solutiions with increased variance. Since genetic algorithms include random perturbations, parameter convergence may occur in different regions with another independent trial if a random seed is not fixed. With multiple basins calibrated over a region, the increased diversity of solutions can create problems with parameter regionalization[5].

These are not challenges when using a manual calibration process where the expert can utilize experience to retain expected ranges of parameters and maintain a regional consistency in the manual process. For regions with a large number of basins to be calibrated, or if the data is regularly updated leading to new calibrations, the manual process may not be feasible or cost-effective. Genetic algorithms are efficient

1.1 A subsection

2 Incorporation of Expert Knowledge

$Fitness = f(Obs\,Q,\ Model\,Q,\ Parameter\,Set)$

3 Case Study: Animas River Basin

The Animas River Basin is a widely studied and monitored basin in the mountains of Colorado, USA, and is critically important for spring snow melt and water supply. Development of reliable estimates of yield from basins in this region is critical in support of basin water management. As part of a larger study, a Research Distributed Hydrologic Model (RDHM) model[4] was developed using the Sacramento Soil Moisture Accounting (SAC-SMA) model [National Weather Service], and snow accumulation and ablation was estimated with Snow-17 [1].

A Non-Dominated Sorting Genetic Algorithm (NSGA-II) [2] was wrapped around the hydrologic model using three objective fitness functions. Each objective should be *minimized*, therefore negative values were uses where needed.

3.1 Fitness Functions

Fitness-1: Negative Kling-Gupta   The Kling-Gupta[3] fitness includes three components of bias, variance, and correlation, described as follows.

$Kling - Gupta = 1 - \sqrt{(R^2 - 1.)^2 + (Var - 1)^2 + (MeanBias - 1)^2)}$

where,

$MeanBias = \mu_{model}/\mu_{obs}$

$Var = (\sigma_{model}/\mu_{model})/(\sigma_{obs}/\mu_{obs})$

$R^2 = Correlation(Obs, Mod)$

The *negative* Kling-Gupta value is used in the NSGA-II algorithm.

Fitness-2: Monthly Volume Bias   This estimates the mean difference between total monthly observed and calculated volumes and selects the maximum mean monthly difference.

$max\{|mean_{month}(Vol_{obs} - Vol_{model})|\}$

Fitness-3: Expert Knowledge   The expert knowledge oibjective is the sum of a collection of penalties based solely on the parameters of the given solution. With no parameter violations, the objective would be perfect and have a score of 0.

$\sum_{i=1}^{M} ParameterPenalty_i + \sum_{j=1}^{N} RelationPenalty_i$

Individual parameter penalty functions can be sharp, linearly gradated, or both. A sample figure of different parameter penalty functions is shown in Figure-2. Not for these examples, each parameter includes a region of no-penalty, a region on the higher end of values with a slowly gradated increase in penalty (the not-preferred zone), and a region on the lower end of relatively sudden increases in penalty.

Penalties are also used for the relationship between certain parameter sets as shown in Figure-3. In these cases, experts expressed that one variable should not be above a threshold if the other is below some level. These joint penalty functions are expressions of their regional knowledge - alternative forms are certainly possible, and could include non-linear relationships if users are aware of these functions and are able to express them mathematically.

3.2   Testing Approach

As additional objective functions are added, the non-dominated set may adversely affect other solution spaces with reduced fitness. Leaving the first two fitness functions in-place, would the addition of the expert knowledge objective (Fitness-3) result in degraded performance for Fitness-1 and Fitness-2? Further, can it be demonstrated that the algorithm results in decreased variance between trials, and decreased variance between solutions in a single pareto set?

Two parallel and identical model calibration trials were developed, with the exception of one trial including the third fitness expert knowledge fitness function (i.e. 2-Objective vs. 3-Objective).

A second test was developed using the same model configurations, which involved the repeated optimization process in 35 independent trials. This test is used to compare the ability for consistent convergence of parameters in certain ranges, relative to the 2-Objective model without expert knowledge.

3.3   Results

Individual members in the non-dominated pareto front produce very similar fitness in the hydrographs - Figure-4 shows a selection of hydrographs from five different members compared to both the observations and the manually calibrated values. Kling-Gupta values and monthly biases are very similar between all non-dominated solutions and additional pareto solution selection metrics would be needed to choose a single member (e.g. regionalization, expert preference or weighting methods for objectives, or a new post-processing objective). These figures show a successful use of the algorithm to develop calibrated parameters.

3

When compraing the 3-Objective to the 2-Objective approach, a similar level of fitness for the first two objectives is achievable. Figure-5 shows a) all solution member from all fronts, and b) the non-dominated solutions from both approaches. Although the shape of the field appears different when extending from 2 to 3-dimensions, the range on the non-dominated sets are very similar. This same information can also be viewed as histograms and is shown in Figure-6.

This process was repeated for 35 independent trials using the 2 or 3-Objective approaches. As shown in Figure-7 for the parameter $UADJ$, there is increased variance across the pareto front for each individual trial when expert knowledge is not incorporated. Further, the range of results extends across the entire acceptable space with each trial converging in different regions. Alternatively with the inclusion of 3-Objectives and the expert knowledge, the individiual trial variance is reduced, and the intra-trial variance is also reduced creating consistency between independent trials. The values destimated are within acceptable and believable ranges.

Although $UADJ$ demonstrated decreased variance, in some cases the variance increased.

4 Conclusions

Increased variance in certain parameters.

Recognize different regions and models will require different expert elicitation and development of penalties.

References

[1] Eric Anderson. Snow accumulation and ablation model - snow-17. 2006.

[2] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

[3] Hoshin V. Gupta, Harald Kling, Koray K. Yilmaz, and Guillermo F. Martinez. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377, 2009.

[4] Victor Koren, Seann Reed, Michael Smith, Ziya Zhang, and Dong-Jun Seo. Hydrology laboratory research modeling system (hl-rms) of the us national weather service. *Journal of Hydrology*, 291:297–318, 2004.

[5] Scott Steinschneider, Yi-Chen E. Yang, and Casey Brown. Combining regression and spatial proximity for catchment model regionalization: a comparative study. *Hydrological Sciences Journal*, 60(6):1026–1043, 2015.
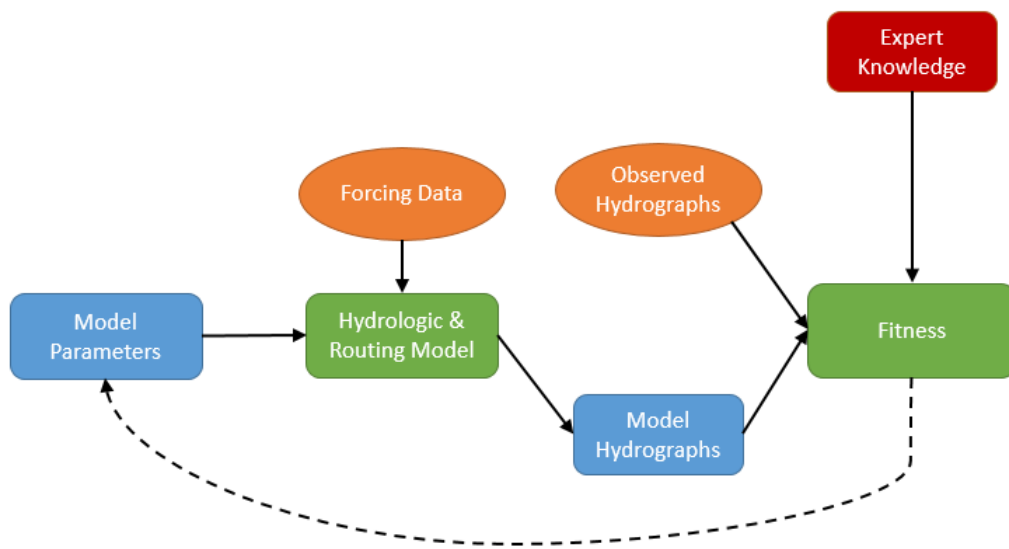
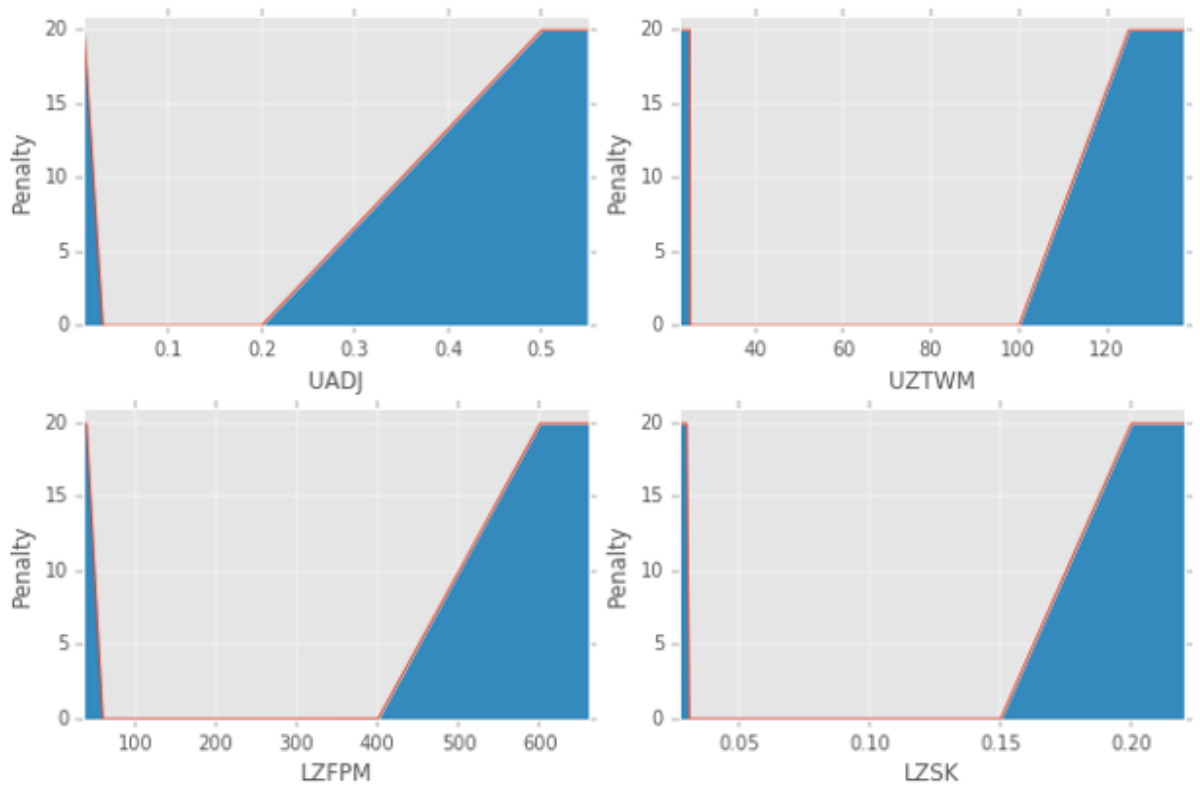Figure 1: Inclusion of Expert Knowledge in genetic algorithm

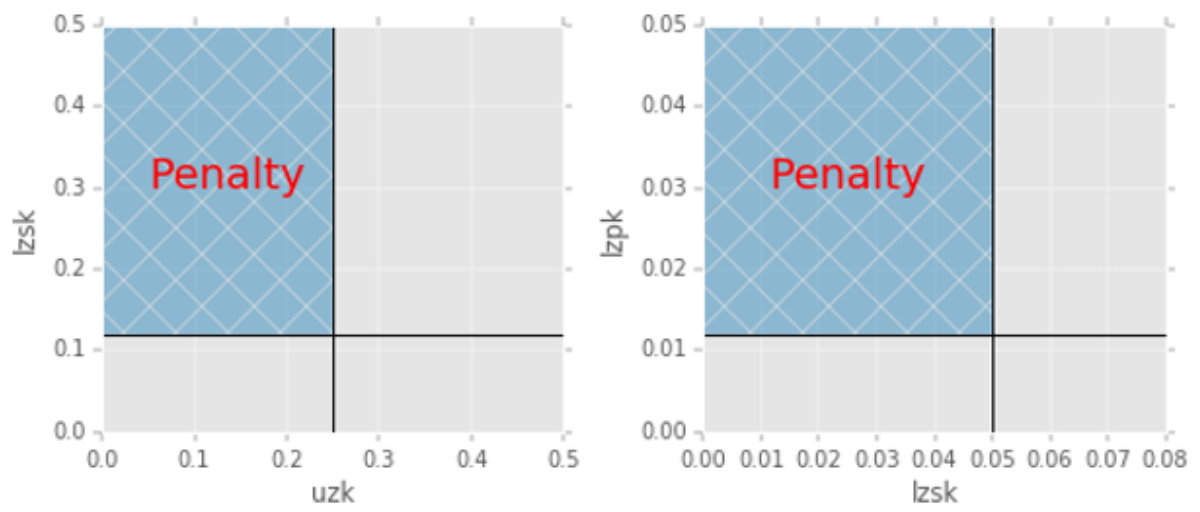Figure 2: Examples of parameter ranges and shoulder penalties

Figure 3: Example of correlated parameter penalties (LZSK and UZK, LZPK and LZSK)
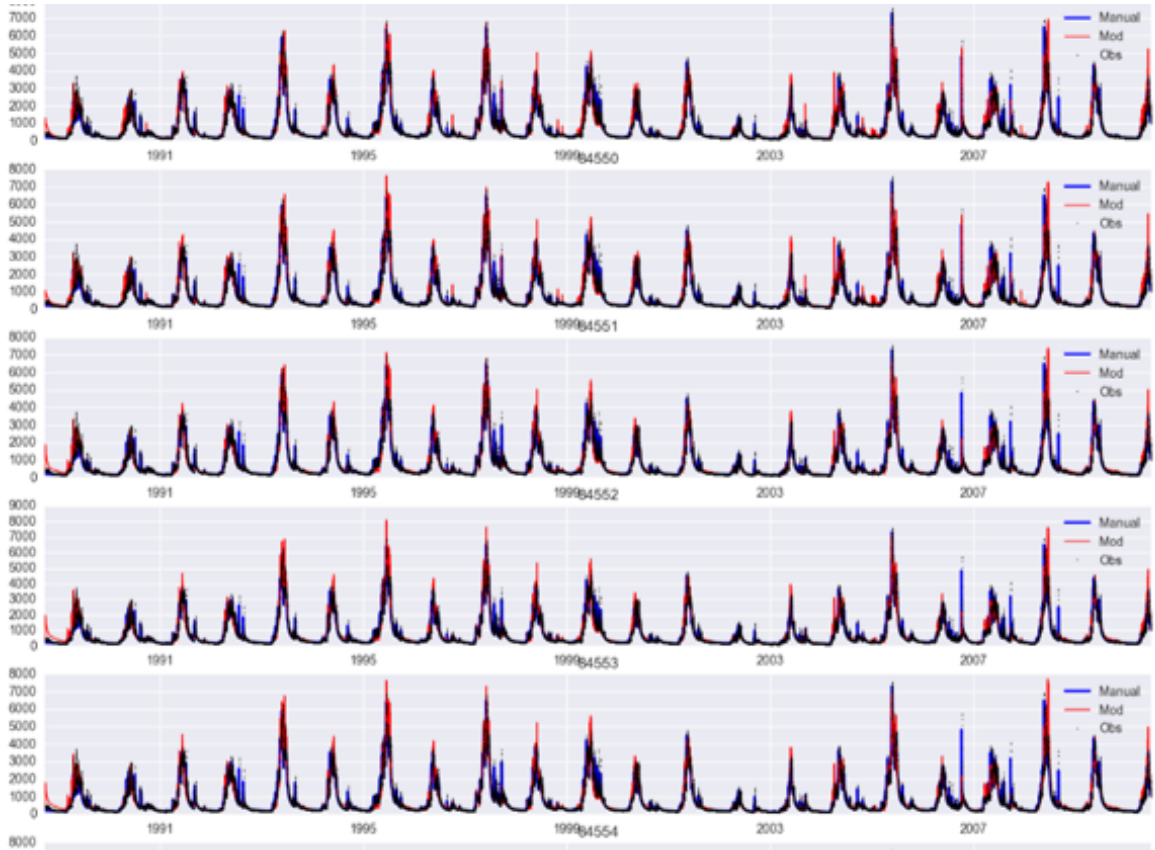
Figure 4: Comparison of calibrated hydrographs for five selections on the Pareto front against manual solution and observed data
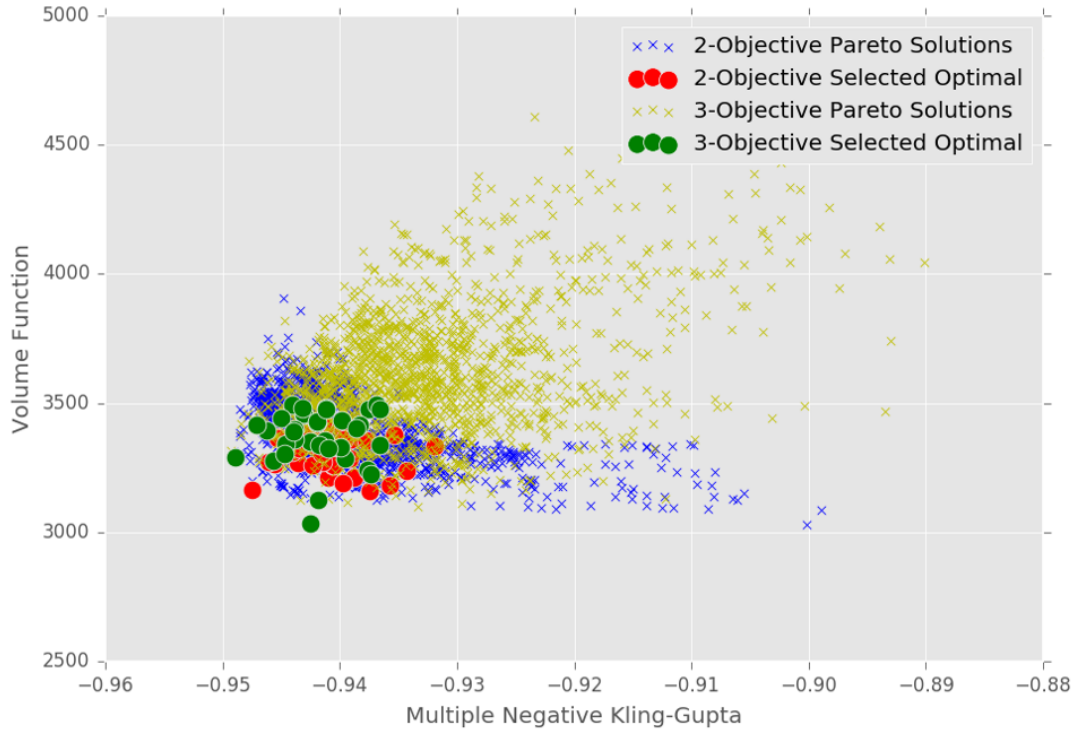
Figure 5: Population and Non-dominated solution objectives for both 2-objective and 3-objective approaches (inclusion of expert knowledge)
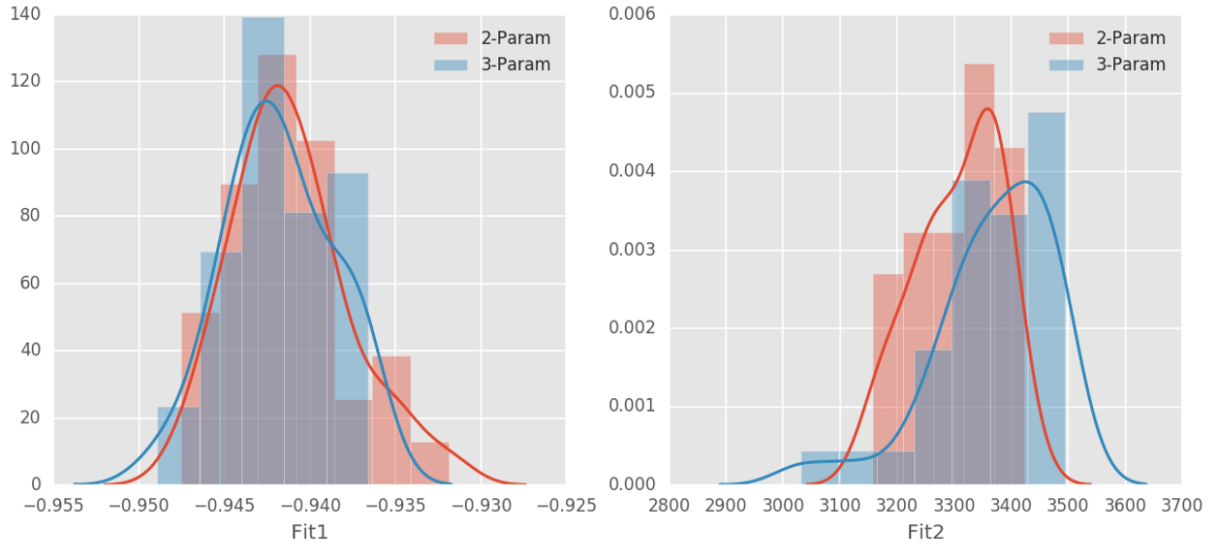
Figure 6: Comparison of first two fitness functions using either a) only those two objectives, or b) the additional parameter constraint objective.
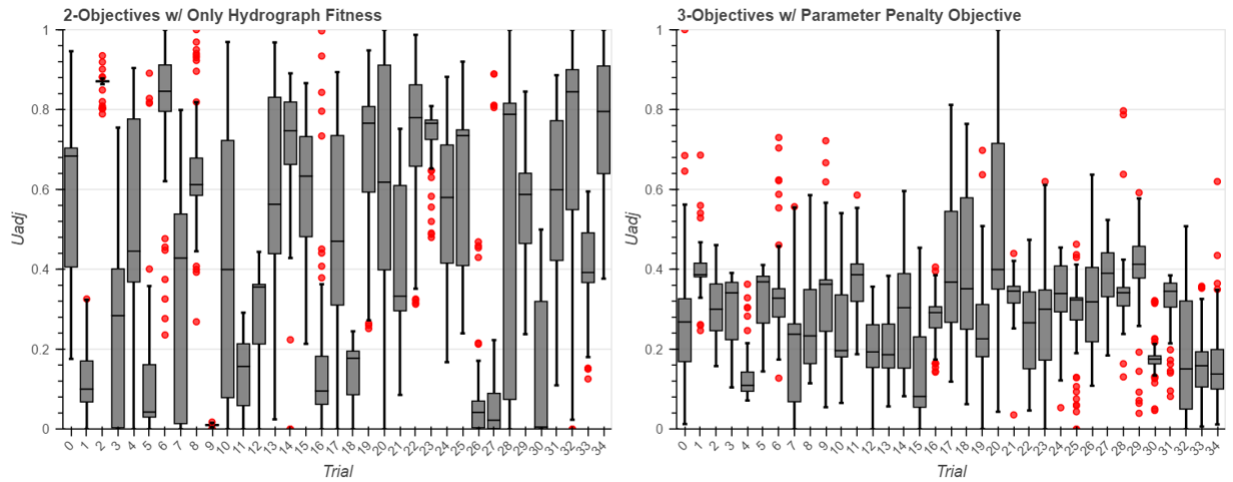

Figure 7: Inter-trial variance from independent calibration runs improving consistency and reproducibility