Hydrology and Earth System Sciences

Reply to J. Vrugt's comment on "How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?"

P. Reed, Y. Tang, and T. Wagener

Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, Pennsylvania, USA

We would like to thank Jasper Vrugt for his comment on our recent paper Tang et al. (2006) in which we compare the Strength Pareto Evolutionary Algorithm 2 (SPEA2), the Multi-objective Shuffled Complex Evolution Metropolis algorithm (MOSCEM-UA), and the Epsilon Dominance Non-dominated Sorted Genetic Algorithm II (ε-NSGAII) using a statistical metrics-based approach. To frame our response, we will provide a brief synopsis of the issues of concern discussed in the comment on our paper. Issue 1: Vrugt contends that the exclusion of the recommendation of Vrugt et al. (2003) that a single objective methodology should be used to first find the endpoints of the Pareto set to precondition search for MOSCEM-UA and that our use of initial uniform random sampling for the three algorithms did not accurately portray the performance of MOSCEM-UA. Issue 2: Vrugt contends that our approach in attaining the reference Pareto front in Fig. 5 using the 15 000 000 model simulations from all of the runs from all of the algorithms (i.e., 3 algorithms * 50 random seed trials/algorithm * 100 000 model simulations/random seed trail) is inefficient relative to his assertion that MOSCEM-UA would reliably identify the true reference front in approximately 22,000 model evaluations if we had first used a single objective algorithm to pre-condition MOSCEM-UA's search. We will address each of these issues individually and then provide some brief concluding remarks.

1 Response to Issue 1

In our comparison of the three algorithms, our computational experiment was designed to maximize the performances of MOSCEM-UA and SPEA2 relative to our own algorithm the ε-NSGAII. As detailed in Sect. 4.1 of Tang et al. (2006), multiple configurations of both SPEA2 and MOSCEM-UA

were tested to maximize the algorithms' performances. As noted in Sect. 6.1, SPEA2's performance is highly sensitive to an appropriately sized archive and we maximized the algorithm's performance by providing an archive size from the ε-NSGAII's results where epsilon dominance (ED) archiving dynamically sizes the archive without user input. For all three algorithms our goal was to test their abilities as independent multiobjective solvers using options and recommendations available in their source codes. In the case of MOSCEM-UA, the source code provided by Jasper Vrugt did not provide users with an option for first using single objective optimization to identify the end points of the Pareto front and therefore this recommendation requires another single objective algorithm to augment the search. It was not the goal of our study to develop dual single objective and multiobjective tests of SPEA2, MOSCEM-UA, and ε-NSGAII. The study highlights the strengths and weaknesses of all the algorithms in the context of multiobjective search. For all three algorithms, a uniform initial distribution is a standard starting point for applications and provides a full characterization of their runtime search dynamics.

We do agree that pre-conditioning of search does dramatically enhance the efficiency and reliability of evolutionary multiobjective optimization methods. This is in fact, a contribution from the work used to develop and test the ε-NSGAII (Kollat and Reed, 2006; Tang et al., 2006, 2007; Kollat and Reed, 2007). As described in Sect. 2.2.1, the ε-NSGAII exploits dynamic ED archiving and solution injection in a series of "connected runs" where initial small populations precondition multiobjective search and speed convergence. In ε-NSGAII, ED archiving and solution injection are the key defining properties of the algorithm that have been implemented in the algorithm's source code structure to minimize user inputs. Epsilon dominance archiving and solution injection are used to dynamically size the search population and represent a search enhancement termed *time continuation* (Goldberg, 2002). In the algorithm, after an initial small pop-

---

Correspondence to: P. Reed
(preed@engr.psu.edu)

ulation searches and identifies an initial approximation of the Pareto optimal set using a minimal number of model simulations, the initial ED archive solutions are then injected into a new search population where they represent 25-percent of the new population and the remaining 75-percent of the population's members are generated randomly. The random solutions ensure population diversity and allow the ε-NSGAII to "continue" search without premature convergence. Our recent work highlights that simple parallelization strategies, ED archiving, and time continuation can dramatically enhance the computational scaling, efficiency, and reliability of multiobjective search (Kollat and Reed, 2007; Tang et al., 2007).

2 Response to Issue 2

The development and meaning of Fig. 5 should be clarified in the context of Vrugt's comment. In the comment, it is proposed that the dual use of single-objective and multiobjective optimization would allow MOSCEM-UA to be far more efficient than our results (i.e., 22 000 model simulations versus 15 000 000). This is not an accurate description of the results in Fig. 5 since the figure simply provides an illustration of the best known reference set for the Leaf River test case as described in Sect. 5.1 of Tang et al. (2006). The reference set was developed using all search results for all random seeds for all of the tested algorithms to ensure that the best known reference set was being used in metric calculations. The key results that should be interpreted from Fig. 5 are that the Leaf River test case has large false fronts and that no single algorithm contributed the full reference set (ε-NSGAII found 58% of the reference set, SPEA2 found 42% of the reference set, and MOSCEM-UA did not contribute any exact reference solutions). Although none of the algorithms found the full exact set, all of them found approximations that ranged in quality as quantified using the hypervolume and epsilon indicator metrics. Figure 6 and Table 5 highlight that all of the algorithms had some random seed trials that failed to closely approximate the reference set.

A better portrayal of the potential search efficiencies of the algorithms for the Leaf River test case is shown in Fig. 7 which shows the runtime performance of the best performing seeds from each algorithm. There are several relevant observations that can be made from Fig. 7. First, all of the algorithms rapidly found approximations to the Leaf River reference set (i.e., hypervolume metrics <0.625 and epsilon indicator metrics <0.3125). The ε-NSGAII exceeded these thresholds in approximately 2000 model simulations whereas MOSCEM-UA and SPEA2 required approximately 12 500 and 20 000 model simulations, respectively. Also Fig. 7 shows that after 15 000 model simulations, MOSCEM-UA failed to maintain search whereas SPEA2 and ε-NSGAII continued to improve their approximation sets.

A key question that readers should consider when interpreting "best run" results for any evolutionary algorithm such as those shown in Fig. 7 is: "How reliably can the algorithms perform this efficiently?". Tang et al. (2006) show in Table 5 that SPEA2 performed the most reliably on the Leaf River test case followed by ε-NSGAII., which was followed by MOSCEM-UA. Augmenting any of the algorithms by using results from another single objective algorithm would likely enhance their search, but this does not represent a stand alone test of their search capabilities and failures which was the focus and intent of Tang et al. (2006). Additionally, there are several possible ways to address search failures as shown in our recent work (Tang et al., 2007) in which a very simple parallelization scheme exploits the ε-NSGAII's use of time continuation to dramatically enhance search efficiency and reliability for the Leaf River test case while maintaining algorithmic and parametric simplicity.

3 Concluding remarks

We appreciate the time and effort spent in developing the comment on our paper. Multiobjective optimization is garnering interest across a broad suite of water resources systems applications. The goal of this reply is to clarify that our use and testing of MOSCEM-UA as an independent multiobjective solver tried to maximize the algorithm's performance and provide a rigorous statistical test of its search dynamics. We appreciate Dr. Vrugt's continued contributions to multiobjective optimization in the context of his new developments. Since Tang et al. (2006) tested MOSCEM-UA, we did not comment on the new approach since it does not appear to be relevant to the focus of our study. Our research has progressed as well (e.g., see Kollat and Reed, 2007, and Tang et al., 2007) and we look forward to future interactions.

Edited by: H. H. G. Savenije

References

Goldberg, D. E.: The Design of Innovation: Lessons from and for Competent Genetic Algorithms, Kluwer Academic Publishers, Norwell, MA, 2002.

Kollat, J. B. and Reed, P.: Comparing State-of-the-Art Evolutionary Multi-Objective Algorithms for Long-Term Groundwater Monitoring Design, Adv. Water Resour., 29(6), 792–807, 2006.

Kollat, J. B. and Reed, P.: A Computational Scaling Analysis of Multiobjective Evolutionary Algorithms in Long-Term Groundwater Monitoring Applications, Adv. Water Resour., 30(3), 408–419, 2007.

Tang, Y., Reed, P., and Kollat, J. B.: Parallelization Strategies for Rapid and Robust Evolutionary Multiobjective Optimization in Water Resources Applications, Adv. Water Resour., 30(3), 335–353, 2007.

Tang, Y., Reed, P., and Wagener, T.: How efficient and effective are evolutionary multiobjective algorithms at hydrologic model

calibration, Hydrol. Earth Syst. Sci., 10, 289–307, 2006, http://www.hydrol-earth-syst-sci.net/10/289/2006/.

Vrugt, J., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39(8), 1214, doi:1210.1029/2002WR001746, 2003.