Evolutionary Multiobjective Optimization in Water Resources: The Past, Present, and Future

Patrick M. Reed, David Hadka, Jonathan D. Herman, <u>Joseph R. Kasprzyk</u>, Joshua B. Kollat
The Pennsylvania State University (preed@engr.psu.edu, dmh309@psu.edu, jdh33@psu.edu, jrk301@psu.edu, juk124@psu.edu)

Abstract: This study contributes a rigorous diagnostic assessment of state-of-the-art multiobjective evolutionary algorithms (MOEAs) and highlights key advances that the water resources field can exploit to better discover the critical tradeoffs constraining our systems. This study provides the most comprehensive diagnostic assessment of MOEAs for water resources to date, exploiting more than 100,000 MOEA runs and trillions of design evaluations. The diagnostic assessment measures the effectiveness, efficiency, reliability, and controllability of ten benchmark MOEAs for a representative suite of water resources applications addressing rainfall-runoff calibration, long-term groundwater monitoring (LTM), and risk-based water supply portfolio planning. The suite of problems encompasses a range of challenging problem properties including (1) many-objective formulations with four or more objectives, (2) multi-modality (or false optima), (3) nonlinearity, (4) discreteness, (5) severe constraints, (6) stochastic objectives, and (7) non-separability (also called epistasis). The applications are representative of the dominant problem classes that have shaped the history of MOEAs in water resources and that will be dominant foci in the future. Recommendations are provided for which modern MOEAs should serve as tools and benchmarks in the future water resources literature.

Keywords: many-objective optimization, evolutionary algorithms, water supply, model calibration, long-term groundwater monitoring, interactive visual analytics

1   INTRODUCTION

There is a clear demonstrated need for multiobjective problem formulations in water resources, as evidenced by Haimes and Hall [1977]. When designing a water project, decision makers and stakeholders want to minimize the cost, but also maximize the system's reliability, environmental quality, and so forth. The decision objectives are often calculated using complex simulation models that exhibit mathematical properties that preclude using classical optimization, including nonlinearity, stochasticity, discreteness, high dimension decisions, severe combinatorial growth rates, and uncertainties. These properties have motivated research into multiobjective evolutionary algorithms (MOEAs) that can effectively solve this problem (see the review by Coello Coello [2007]). MOEAs are a class of *a posteriori* decision support tools, which means that the decision makers are presented with an explicit representation of their tradeoffs before they express their preference for one or more selected solutions. No weighting or commensuration is done before the search, which allows the analyst to discover surprising problem properties during the search process. Additionally, a complex simulation model is fully embedded within the MOEA, which means that the mapping from design to objective outcome is calculated using a trusted method.

Cohon and Marks [1975] defined the multiobjective optimization problem as "vector optimization", where the objective function is a vector comprised of a number of scalar objectives, f(x), subject to a set of constraints. Decision variables comprising the vector x can be real-valued, integer, or binary. In the absence of additional preference information, there is no single optimal solution when conflicts exist between one or more objective functions. Pareto optimal solutions have performance that is not exceeded with respect to all objectives by any other feasible solution in the search space. Since it is infeasible to enumerate the entire search space for nontrivial problems, MOEAs find a "Pareto approximate set" which is the best known approximation to the set. The set of tradeoff solutions is often called the Pareto front. Nicklow et al. [2010] provides a review of water resources applications using MOEAs.

This study provides the most comprehensive diagnostic assessment for MOEAs ever attempted. The results provide a careful assessment of effectiveness, efficiency, reliability, and controllability of ten benchmark MOEAs for a suite of water resources applications. The applications span a mathematical test function, rainfall-runoff calibration, long-term groundwater monitoring, and risk-based water supply portfolio planning. Interested readers can refer to the full analysis for further detail [Reed et al. In-Press]. Section 2 outlines recent innovations in multiobjective optimization and the tested algorithms. Section 3 describes the testing framework, and section 4 introduces the test problems. Section 5 provides summary results, with section 6 giving conclusions.

2. MULTIOBJECTIVE EVOLUTIONARY ALGORITHMS
2.1 Recent Innovations

As reviewed by Nicklow et al. [2010], early MOEAs were used to solve problems in several important areas in the water literature [e.g., Meyer et al. 1994; Gupta et al. 1998]. However, these tools were limited in their ability to sustain convergence (evolving to the Pareto front) while maintaining diversity (spreading solutions on the full extent of the Pareto front). One important innovation is the proof of convergence and diversity [Rudolph 1998; Rudolph 2000; Laumanns et al. 2002], or the theoretical assertion that given unlimited search time the algorithm has a non-zero probability of generating a well-spread set of Pareto optimal solutions. Laumanns et al. [2002] introduced epsilon dominance archiving, in which point-based nondomination sorting is replaced by epsilon "blocks" of user-defined precision. Kollat and Reed [2007a] demonstrated that epsilon dominance archiving improves the scalability of MOEAs, in terms of their ability to solve large problems with a reasonable amount of increasing computational demand. Section 4.1 will introduce a mathematical problem in this study that tests the scalability of algorithms for high objective counts.

Recent research has also explored *many-objective* problems of four or more objectives simultaneously [e.g. Reed and Minsker 2004]. Brill et al. [1990] showed that for complex planning problems, solutions from low-dimensional problem formulations can be proven inferior when new objectives are added to the analysis. In contrast, many-objective formulations can overcome two key cognitive challenges. Cognitive myopia [Hogarth 1981] refers when stakeholders ignore alternatives that influence their decision preferences (i.e., focusing on a narrowly-defined problem formulation). Cognitive hysteresis [Gettys and Fisher 1979] is the idea that low-dimensional highly constrained problems reinforce initial preconceptions and biases (i.e., not being able to form new hypotheses). While many-objective problems expand the complexity of problems that can be considered, MOEAs should also avoid *dominance resistance*, where the rapid growth of non-dominated solutions overwhelms the ability of MOEA selection to distinguish high-quality solutions [Purshouse and Fleming 2007].

Interactive visual analytics is a third innovation that aids the process of using MOEAs for many-objective problems. First introduced in Thomas and Cook [2005], visual analytics refers to the rapid analysis of large data sets using interactive

software that enables multiple connected views of planning problems. Kollat and Reed [2007b] demonstrated how visual analytics can be used to explore high-dimensional tradeoff sets to discover and negotiate high-quality solutions. Watching the iterative solution sets can also be used during the search to determine suitable stopping criteria, and afterwards to "brush" away solutions that do not meet user-defined performance thresholds.

Table 1. MOEAs tested in this study.

Group	Algorithm Names	Notes
Historical Benchmarks	NSGAII, SPEA2	Spawned creation of more sophisticated MOEAS, widely used
Pareto Approximation	ε-MOEA, ε-NSGAII, OMOPSO	Uses epsilon archiving to avoid search deterioration
Indicator-Based Methods	IBEA	Uses hypervolume instead of non-domination to measure quality
Aggregate Functions	MOEA/D	Decomposes multiobjective problem into single-objective aggregations
Differential Evolution	GDE3	Utilizes a rotationally indifferent operator for non-decomposable problems
Adaptive Operator Selection	AMALGAM, BORG	Adapts different evolutionary operators during search commensurate on search progress

2.2   Tested Algorithms

Table 1 presents the ten algorithms tested in this study. The algorithms are grouped based on their properties as shown in the first column. For more information on each algorithm, please consult Hadka and Reed [In-Press] and Reed et al. [In-Press].

3.   TESTING FRAMEWORK

An important consideration for implementing the ten parameters of Table 1 is finding values for parameters such as crossover and mutation that result in good performance for each algorithm. Consequentially, a robust algorithm will perform well across large regions of the parameter space, meaning that the choice of parameter value is not important for determining performance. The experimental design of this study follows Hadka and Reed [In-Press] and represents the most comprehensive test of MOEAs ever performed in the water literature.

First, a Latin Hypercube Sample (LHS) of values for all algorithm parameters is generated. Since algorithm performance can be affected by random numbers used within the initial population and operators, 50 replicate random number generator seeds are used for each LHS sample point. We generate a best-known approximation to the Pareto optimal set for each water resources problem across all runs and random seeds. This set is termed the "reference set".

Each algorithm run is tested relative to the reference set using three metrics: generational distance [Van Veldhuizen 1998], additive epsilon indicator [Zitzler et al. 2003], and hypervolume [Zitzler et al. 2003]. Generational distance measures the Euclidian distance of points in the approximation set relative to the nearest

corresponding points in the reference set. This is the easiest measure to satisfy since a single approximation solution being close to the reference set satisfies this indicator. The additive epsilon indicator measures the worst case distance required to translate an approximation set solution to dominate its nearest neighbor. The additive epsilon indicator is a good measure of diversity, since it focuses on the worst case distance and can show whether or not the approximation set has "gaps" in its tradeoff solution set. Hypervolume, the third metric, is the most challenging to satisfy. It measures the volume of objective space dominated by an approximation set. According to Knowles and Corne [2002], the hypervolume indicator measures both convergence and diversity of an approximation set.

Two tests are used with each of the three indicators. The first test is the attainment threshold, which measures the probability of success across the LHS samples. Our use of attainment thresholds in this context is the first study to statistically characterize the "controllability" of algorithms, or the likelihood that a random parameterization of an algorithm will perform well on an application problem [Goldberg 2002]. The second test visualizes "control maps" of the algorithm performance, showing which *values* in the parameter space perform well across the three indicators.

4.   APPLICATION TEST CASES

4.1   DTLZ2

Although the majority of MOEA studies to date have focused on two or three objective problems, many real-world applications have four or more objectives to consider. We test the scalability of the MOEAs to problems with higher numbers of objectives by using a scalable version of the DTLZ2 problem first introduced in Deb et al. [2002]. There are M objective functions (where M ranges from 2 to 16), and the number of decision variables for each problem is L, where $L = M + 9$.

4.2   Rainfall-Runoff Model Calibration (HBV)

Here, we test the ability for MOEAs to calibrate conceptual rainfall-runoff models that transform input rainfall into runoff at an outlet point. The HBV model is a widely used lumped conceptual rainfall-runoff model [Bergstrom 1995]. The model utilizes 14 real-valued parameters that control three routines (a degree-day snow model, soil moisture accounting, and outflow routing). The Williams River, West Virginia, US (USGS Gauge 03186500) was selected for calibration based on a screening analysis that showed the difficulty of calibration. Four objectives are used: the Nash-Sutcliffe Efficiency (focusing on high-flows); the Box-Cox transform Root Mean Squared Error (focusing on low-flows); the runoff coefficient error (focusing on the long-term water balance); and the slope of the flow duration curve error (focusing on the long-term variability of flows). The problem, hereafter referred to as HBV, is nonlinear and multimodal, in that the objective space contains false optima.

4.3   Groundwater Monitoring Design (LTM)

The groundwater monitoring design test case tries to eliminate redundancies in sampling a contamination plume in the subsurface. The problem class is termed Long Term Monitoring (LTM), since the wells sample and monitor the development of the plume over long periods of time. The sampling domain is a hypothetical plume of perchloroeythlene drawn from an existing study site at Lawrence Livermore National Lab. An existing network of 58 sampling points exists at the site, with the decision variables determining whether or not to sample at a well (i.e., a discrete space of size 2^{58} yes/no decisions). Four objectives are used: minimize the cost of the system, minimize concentration error, minimize concentration uncertainty, and minimize the error in estimating total contamination mass. Quantile Kriging is used to test the performance of each sampling plan relative to the full case of sampling all 58 ports. A constraint enforces that each sampling plan

must provide a sufficient spatial coverage of the full extent of the plume such that an interpolation estimate can be made across each spatial grid location. The problem is constrained and discrete with a very large decision space. For more information see Kollat et al. [2008].

4.4   Water Supply Portfolio Planning (LRGV)

The water supply portfolio planning test case seeks to help a single city in the Lower Rio Grande Valley (LRGV) of Texas, USA find the best combination of traditional supply and market-based leases and options for maintaining a reliable water supply. The portfolio is controlled by eight real-valued decision variables that control rights (one variable), options (three variables), and thresholds that determine market acquisitions using ratios of supply to demand (four variables). A Monte Carlo simulation is used to test the performance of each portfolio, which simulates 10 years of water supply decisions. Five objectives are used, which minimize system cost, maximize reliability, minimize surplus water, minimize dropped transfers, and minimize the number of leases of each portfolio. The problem is "noisy", in that each objective function calculation depends on random draws from the Monte Carlo simulation. It also exhibits a highly irregular Pareto front geometry partially due to its combination of integer and discrete decision variables. For more information see Kasprzyk et al. [2009].

5.   REPRESENTATIVE RESULTS

The study used more than 100,000 separate MOEA runs and trillions of function evaluations to evaluate the 10 MOEAs across four separate test case problems as summarized in section 4. Results for all runs are analyzed using the attainment function and control maps as previously discussed. The five best performing algorithms were identified as the Borg MOEA, GDE3, ε-NSGAII, AMALGAM, and OMOPSO. Table 2 provides summary notes about the performance characteristics of the five best algorithms, with the full analysis presented in Reed et al. [In-Press].

Table 2. Summary results for the top five performing algorithms.

Algorithm	Notes
Borg	Best scalability on DTLZ2 problem; dominant contributor to all water resource problem reference sets; superior performance at small populations and low number of function evaluations.
GDE3	One of the most parsimonious parameterizations; low scalability on DTLZ2; high success rates on HBV and LTM but low contributions to reference sets; requires careful parameterization.
ε-NSGAII	First algorithm to use adaptive population sizing and epsilon dominance archiving; good scaling performance and contribution to LTM reference set; performance expected to be exceeded by Borg.
AMALGAM	First auto-adaptive multi-operator MOEA; good performance for HBV and LTM but struggled on LRGV case; performance expected to be exceeded by Borg.
OMOPSO	Reduced scalability on DTLZ2; low contributions to reference sets; good performance on LRGV case in terms of attainment and controllability.

6.   CONCLUSIONS

Based on the suite of results, the Borg MOEA was the only algorithm to consistently perform well across all applications, also showing the highest controllability for the water resources problems. The five MOEAs in Table 2 should be the focus of future benchmarking studies, which should follow rigorous statistical designs such as the one presented in this study.

Complex water resources problems in the future will require the ability to design and use algorithms that are controllable, efficient, and reliable across different problem types. This is bolstered by the development of auto-adaptive algorithm frameworks that can modify the search procedure with minimal user input. Advances in additional domains including human-computer interaction, parallel cloud computing services, and visual analytics will also aid this effort, allowing discovery of new water resources innovations in a changing future.

ACKNOWLEDGMENTS

REFERENCES

Bergstrom, S. *The HBV Model*, Water Resource Publications, Highlands Ranch, 1995.

Brill, E.D., J. Flach, L. Hopkins, S. Ranjithan, MGA: a decision support system for complex, incompletely defined problems, *IEEE Trans. Syst. Man Cybernet.* 20(4): 745-757, 1990.

Coello Coello, C., G.B. Lamont, and D.A. Van Veldhuizen (eds) *Evolutionary Algorithms for Solving Multi-Objective Problems.* Springer, New York, 2 ed, 2007.

Cohon, J. and D. Marks, A review and evaluation of multiobjective programming techniques, *Water Resources Research* 11(2): p. 208-20, 1975.

Deb, K., L. Thiele, M. Laumanns, and E. Zitzler, Scalable multi-objective optimization test problems. In: *Proceedings of the congress on evolutionary computation*, p. 825-830, 2002.

Gettys, C and S. Fisher, Hypothesis plausibility and hypothesis generation, *Organ. Behav. Human Perform.* 24(1): 197-217, 1979.

Goldberg, D., *The design of iinnovation: lessons from and for competent genetic algorithms*, Kluwer Academic Publishers, Norwell, 2002.

Gupta, H., S. Sorooshian, and P. Yapo, Toward improved calibration of hydrologic models: multiple and noncommensurable meausures of information, *Water Resources Research* 34, p. 751-763, 1998.

Hadka, D. and P. Reed, Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization, *Evol. Comput.,* In-Press.

Haimes, Y. and A. Hall, Sensitivity, responsivity, stability and irreversibility as multiple objectives in civil systems, *Advances in Water Resources* 1(2): p. 71-81, 1977.

Hogarth, R., Beyond discrete biases: functional and dysfunctional aspects of judgemental heuristics, *Psychol. Bull.* 90(1): 197-217, 1981.

Kasprzyk, J.R., P.M. Reed, B.R. Kirsch, and G.W. Characklis, Managing population and drought risks using many-objective water portfolio planning under uncertainty, *Water Resources Research,* 45, 2009.

Kollat, J., and P. Reed, A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications, *Advances in Water Resources* 30(3): p. 408-419, 2007a.

Kollat, J., and P. Reed, A framework for visually interactive decision-making and design using evolutionary multiobjective optimization (VIDEO), *Environmental Modelling and Software*, 22(12): 1691-1704, 2007b.

Kollat, J., and P. Reed, Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design, *Advances in Water Resources*, 29(6): 792-807, 2006.

Kollat, J., P. Reed, and J. Kasprzyk, A new epsilon-dominance hierarchical Bayesian optimization algorithm for large multi-objective monitoring network design problems, *Advances in Water Resources* 31(5): 828-845, 2008.

Knowles, J. and D. Corne, On metrics for comparing non-dominated sets. In: *Proceedings of the 2002 world congress on computational intelligence*, p. 711-716, 2002.

Laumanns, M, L. Thiele, K. Deb, and E. Zitzler, Combining convergence and diversity in evolutionary multiobjective optimization, *Evol. Comput.,* 10(3): p. 263-282, 2002.

Meyer, P., A. Valocchi, and J. Eheart, Monitoring network design to provide initial detection of groundwater contamination, *Water Resources Research,* 1994.

Nicklow, J. et al. State of the art for genetic algorithms and beyond in water resources planning and management, *Journal of Water Resources Planning and Management*, 136, p. 412-432, 2010.

Purshouse, R.C. and P. Fleming, On the evolutionary optimization of many conflicting objectives, *IEEE Trans. Evol. Comput.* 11(6): 770-784, 2007.

Reed, P.M., D. Hadka, J. Herman, J. Kasprzyk, and J. Kollat, Evolutionary multiobjective optimization in water resources: The past, present, and future, *Advances in Water Resources*, In-Press.

Reed, P. and B. Minsker, Striking the balance: long-term groundwater monitoring design for conflicting objectives, *Journal of Water Resources Planning and Management* 130(2): p. 140-149, 2004.

Rudolph, G., Evolutionary search for minimal elements in partially ordered sets, In: *Evolutionary programming VII: 7th annual conference on evolutionary programming* p. 345-353, 1998.

Rudolph, G., and A. Agapie, Convergence properties of some multi-objective evolutionary algorithms, In: *Congress on evolutionary computation*, vol 2, p. 1010-1016, 2000.

Thomas, J., and K. Cook, *Illuminating the path: the research and development agenda for visual analytics*, IEEE Computer Society Press, Los Alamitos, 2005.

Van Veldhuizen, D., and G. Lamont, Evolutionary computation and convergence to a Pareto front, In: *Genetic Programming 1998: Proceedings of the 3rd annual conference*, p. 22-25, 1998.

Zitzler, E., L. Theile, M. Laumanns, C. Fonseca, Grunert da Fonseca, V., Performance assessment of multiobjective optimizers: an analysis and review, *IEEE Trans. Evol. Comput.* 7(3): 117-132, 2003.