A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control

Jazmin Zatarain Salazar[a,*], Patrick M. Reed[a,*], Jonathan D. Herman[b], Matteo Giuliani[c], Andrea Castelletti[c]

[a]*School of Civil & Environmental Engineering, Cornell University, Ithaca, NY, USA*
[b]*Department of Civil & Environmental Engineering, University of California, Davis, USA*
[c]*Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milano, Italy*

Abstract

Globally, the pressures of expanding populations, climate change, and increased energy demands are motivating significant investments in re-operationalizing existing reservoirs or designing operating policies for new ones. These challenges require an understanding of the tradeoffs that emerge across the complex suite of multi-sector demands in river basin systems. This study benchmarks our current capabilities to use Evolutionary Multi-Objective Direct Policy Search (EMODPS), a decision analytic framework in which reservoirs' candidate operating policies are represented using parameterized global approximators (e.g., radial basis functions) then those parameterized functions are optimized using multi-objective evolutionary algorithms to discover the Pareto approximate operating policies. We contribute a comprehensive diagnostic assessment of modern MOEAs' abilities to support EMODPS using the Conowingo reservoir in the Lower Susquehanna River Basin, Pennsylvania, USA. Our diagnostic results highlight that EMODPS can be very challenging for some modern MOEAs and that epsilon dominance, time-continuation, and auto-adaptive search are helpful for attaining high levels of performance. The ϵ-MOEA, the auto-adaptive Borg MOEA, and ϵ-NSGAII all yielded superior results for the six-objective Lower

---

[*]Corresponding Authors
Email addresses: `jz583@cornell.edu` (Jazmin Zatarain Salazar), `pmr82@cornell.edu` (Patrick M. Reed)

Susquehanna benchmarking test case. The top algorithms show low sensitivity to different MOEA parameterization choices and high algorithmic reliability in attaining consistent results for different random MOEA trials. Overall, EMODPS poses a promising method for discovering key reservoir management tradeoffs; however algorithmic choice remains a key concern for problems of increasing complexity.

1. Introduction

Operational water management within river basins worldwide is confronting a challenging combination of growing population pressures, evolving multi-sector demands, and climate change [19]. These challenges are pressing existing and
5 planned hydropower operations to adopt integrated water resources management that takes into account a broad range of social, economic, and environmental issues [66]. Efficient multi-purpose reservoir management strategies are critical given the growing risks for flood and drought shocks as well as the need to meet evolving water allocation demands across a complex set of users
10 (e.g., balancing the variability of renewables or flow maintenance for ecosystem services [3, 9, 36]). However; identifying efficient and balanced reservoir management strategies that meet energy needs while maintaining other key river basin services remains a severe challenge for actual operations.

Reservoir policies need to realistically consider the complex dynamics that
15 typify river basin systems. Consequently, the optimization techniques used in their design need to avoid simplifications that widely discourage their application in real reservoir contexts [44]. Popular operational water management frameworks ranging from classical tools (e.g., dynamic programming (DP) or linear programming(LP) family of methods) to single-objective heuristics are
20 limited in the breadth of multi-objective formulations that they can resolve [6, 4, 24]. Traditionally, these approaches were developed for single objective

formulations and only recently have they extended to multi-objective formulations. Yet, they are still limited in their scalability and are not applied to many-objective formulations (with more than four objectives) [22]. The weighting schemes used in traditional multi-criterion implementations of single-objective methods are strongly sensitive to the convexity as well as the separability of the resulting aggregate management objectives [6, 7]. These issues pose important limits for formulations with heterogeneous objective functions. For instance, a minimax reliability objective and an expected cost objective may encounter difficulties when integrated into a single weighted function when using a DP framework. The classical approach for appropriately aggregating conflicting objectives requires an *a-priori*, well-specified set of weights [20]. Using the terminology of Cohon and Marks [11], DP-based solution strategies can also be used as generating methods, where a suite of optimization runs are executed as the weights for different objectives are varied to attain Pareto optimal solutions [57]. The Pareto optimal set represents the suite of solutions whose performance in a single objective cannot be improved without degrading their performance in one or more other objectives. Plotting this Pareto optimal set of solutions in a problem's objective space yields the Pareto front, or the geometric representation of the optimal tradeoffs. This scalarization process requires one optimization run for each point that defines a trade-off curve, which is computationally very demanding and often results in poor representations of the Pareto frontier [5]. These limitations make it important to understand the value of algorithms capable of approximating the Pareto front in a single run (e.g., Vamplew et al. [62], Castelletti et al. [5], Reed et al. [52]). Among these methods, multi-objective evolutionary algorithms (MOEAs) have been demonstrated to be capable of discovering high quality representations of complex tradeoffs [49, 52, 46, 25].

Evolutionary Multobjective Direct Policy Search (EMODPS) provides a flexible framework for employing MOEAs in complex multi-purpose reservoir systems. Giuliani et al. [24] formalized this approach, which features reservoir policy identification, multi-objective evolutionary optimization and visual ana-

3

lytics to characterize the baseline operations and discover the key operational tradeoffs to provide operators with guidance on balancing a reservoir system's competing demands. Rosenstein and Barto [55] first introduced direct policy search (DPS) in the general control theory literature. DPS is also known as parameterization-simulation-optimization in the water resources literature [42] with earlier water resources applications found in Guariso et al. [28] and Oliveira and Loucks [50]. EMODPS provides users with flexibility in how to formulate and solve multi-objective reservoir control problems. EMODPS benefits from (1) the simultaneous consideration of heterogeneous forms of objective functions (e.g., minimax and expected value) [23], (2) the potential use of exogenous information to condition control decisions, [26] and (3) simulation-based treatment of uncertainties in system dynamics or performance [25]. EMODPS copes with high dimensionality reservoir's operational decisions by instead optimizing the parameters of a control policy. This is a parsimonious approach that broadens analysis of complex reservoir systems; the systems do not need to be simplified as the methodology can accommodate more objectives and uncertainties without increasing substantially a problem's difficulty.

Despite these practical advantages, the success of EMODPS is highly dependent on appropriately representing the space of possible operating policies as well as the MOEA's capability to optimize them. The flexibility and accuracy of global approximators to represent alternative operating policies has been assessed in Giuliani et al. [24]. Although there are a growing number of studies exploring the EMODPS framework, at present no rigorous algorithmic assessments have been completed. The key contribution and focus of this study is to diagnose the difficulty of using MOEAs to support the EMODPS framework using the six-objective Lower Susquehanna test case, and analyze which MOEAs are more suitable for finding the best Pareto approximate set. The Lower Susquehanna test case is challenging due to its large number of conflicting multi-sector demands and the time resolution of the analysis, which is linked to the rapidly changing energy prices. Key system demands include hydropower production, urban water supply, recreation and environmental requirements.

4

2. Lower Susquehanna River Basin Benchmark

The Susquehanna River is the largest river in the eastern United States, contributing 50% of the inflows to the Chesapeake Bay. The basin drains over a 71,000 km^2 watershed and provides public water supply for a population of 4.1 million people. In the Lower Susquehanna River Basin, the Conowingo dam plays a key role in balancing the multi-sector water demands within the region, representing one of the largest non-federal hydroelectric dams in the U.S. (see Figure 1). The Conowingo Dam embodies a complex multi-objective system due to the competing demands between hydropower production, environmental flow requirements, cooling water for Peach Bottom Nuclear Power Plant, recreational use and water supply for Baltimore, MD and Chester, PA. (illustrated in Figure 2). To address these issues, the Susquehanna River Basin Commission has historically led computer-aided adaptive management [56] to mediate compromises across the system's multi-sector demands. More recently, Giuliani et al. [25] have contributed a more explicit analysis of the tradeoffs confronting the Lower Susquehanna, highlighting important potential conflicts between hydropower revenue, nuclear power cooling water, and environmental flow requirements. The Lower Susquehanna test case is representative of the management challenges faced in reservoir systems worldwide. A key question explored in this study is how capable of MOEAs are at capturing their tradeoffs. Building off of the initial contributions of Giuliani et al. [25], the Lower Susquehanna provides an excellent benchmarking test case to evaluate this question.

2.1. Susquehanna River Basin Model

The Lower Susquehanna simulation model used in this study is based on the historical formulation in Giuliani et al. [25], where a dynamic mass balance over a historical time series of inflows and evaporation rates as well as the Conowingo and Muddy Run Reservoirs' releases. Muddy Run Reservoir is a pumped hydropower operation which takes advantage of intra-daily cycles in energy prices. During off-peak hours, water is pumped uphill from Conowingo

Reservoir into Muddy Run Reservoir; this water is released during peak hours to maximize hydropower profit for the combined system.

The power house, located in Conowingo, MD, exploits the reduced pricing associated with excess grid capacity during off peak hours to pump water from the Conowingo Reservoir uphill into Muddy Run, the water then relies on gravity-based return flows to Conowingo to take advantage of peak power demand periods.

Direct rainfall over the reservoir surface can be negligible in relation to flow contributions from upstream contributing areas. Evaporation, in the other hand, is considered since this test case focuses on prolonged summer droughts were the losses are not negligible. These relationships are described in Equation 1:

$$
\begin{aligned}
s_{t+1}^{CO} &= s_t^{CO} + q_{t+1}^{CO} + q_{t+1}^{CO,L} - r_{t+1}^{CO} - E_{t+1}^{CO} - q_{t+1}^{p} + r_{t+1}^{MR} \\
s_{t+1}^{MR} &= s_t^{MR} + q_{t+1}^{MR} - r_{t+1}^{MR} - E_{t+1}^{MR} + q_{t+1}^{p}
\end{aligned}
\tag{1}
$$

where s_t^i are the volume of water stored at each reservoir (i=Conowingo Reservoir(CO), Muddy Run Reservoir(MR)), q_{t+1}^{CO} and $q_{t+1}^{CO,L}$, are the mainstem measured at the Marietta gauging station and lateral inflow to the Conowingo reservoir, respectively, q_{t+1}^{MR} is the inflow to Muddy Run, and q_{t+1}^{p} is the water pumped from Conowingo to Muddy Run. The volume released is given by the release function: $r_{t+1}^i = f(s_t^i, u_t^i, q_{t+1}^i, E_{t+1}^i)$, [57], which depends on the storage s_t^i, the release decision u_t^i, the inflow q_{t+1}^i, and the evaporation loss E_{t+1}^i. The release r_{t+1}^{CO} is actually a vector of 4 releases that supply water to the atomic power plant, Baltimore, Chester, and downstream through the Conowingo power plant. The time subscript of each variable represents the time instant at which it assumes a deterministic value. The reservoir storage is measured at time t, whereas inflow has subscript $t + 1$, denoting the inflow to the reservoir in the time interval $[t, t + 1)$. The decision time-step is set at 4 hours to balance the need to follow hourly energy prices and to have a time-step sufficiently long to

avoid impact by turbine operation mechanics.

The multi-stakeholder objectives for the dam are modeled over the simulation time horizon H of one year (see Figure 2). An annual simulation horizon is considered because of the minor regulation capacity of the reservoir system, and the negligible dependence to its initial state (i.e., reservoir levels at the beginning of the simulation).

Hydropower Revenue (to be maximized). Defined as the economic revenue obtained from hydropower production at the Conowingo hydropower plant. Revenue is a function of the hourly energy production (HP_t) given in MWh and the hourly energy price (ρ_t) in US \$/MWh, defined in Equation 2. The energy prices are defined by the seven hour moving average of the energy price trajectory in the Pennsylvania, New Jersey-Maryland (PJM) energy market [21].

$$J^{hyd} = \sum_{t=1}^{H}(HP_t \cdot \rho_t) \tag{2}$$

$$HP_t = \eta g \gamma_w \bar{h}_t q_t^{Turb} \cdot 10^{-6} \tag{3}$$

The hourly energy production (MWh) is defined by Equation 3, where η is the turbine efficiency, g is the gravitational acceleration (9.81 m/s^2), γ_w is the water density (1000 kg/m^3), $\bar{h}_t$ is the net hydraulic head in meters, namely the reservoir level minus tailwater level and q_t^{Turb} is the turbined flow in m^3/s.

Water Supply Reliability to Baltimore, Chester and the Atomic Power Plant (to be maximized). Each of the water demand objectives is measured as the daily average volumetric reliability defined accordingly to Hashimoto et al. [34] as:

$$J^{VR,i} = \frac{1}{H} \sum_{t=1}^{H} \frac{Y_t^i}{D_t^i} \tag{4}$$

where Y_t^i is the daily delivery in m^3, D_t^i is the corresponding daily demand in m^3, and subscript i represents the water supply to either Baltimore, Chester or to the Atomic Power Plant.

7

Recreation (to be maximized). Defined as the storage reliability (SR) in weekends of the touristic season. Given by the relationship between number of weekend days in the touristic season below the target level (n_F) and the total number of weekends in the touristic season (N_{we}). The target level is 32.5 meters (106.5 ft) to guarantee boating.

$$J^{SR} = 1 - \frac{n_F}{2N_{we}} \tag{5}$$

Environmental Shortage (to be minimized). Defined as the daily average shortage index (SI) relative to the FERC flow requirements [12]. The quadratic function in Equation 6 penalizes larger deficits while allowing small and more frequent shortages [33].

$$J^{SI} = \frac{1}{H} \sum_{t=1}^{H} \left(\frac{max(Z_t - Y_t, 0)}{Z_t} \right)^2 \tag{6}$$

where Y_t is the daily release and Z_t is the FERC flow requirement, both in m^3.

3. Methods

In the next sections, we first present a detailed overview of the EMODPS framework as formalized by Giuliani et al. [24]. Then, we summarize the diagnostic framework implemented in this study to benchmark the MOEAs. We conclude this section by describing the suite of algorithms selected for evaluation.

3.1. Evolutionary Multi-objective Direct Policy Search

Evolutionary Multi-objective Direct Policy Search (EMODPS) [24] is a simulation-based approach that combines direct policy search, nonlinear approximating networks, and multi-objective evolutionary optimization to discover Pareto approximate control policies for multi-purpose reservoir systems. The key steps in this framework are (1) the parameterization of candidate operating policies using non-linear approximators, and (2) the use of multi-objective evolutionary

8

search to identify parameterizations that yield Pareto approximate reservoir control policies that compose the tradeoffs across conflicting management objectives.

DPS refers to the explicit representation of a control policy (e.g., abstracting a water reservoir guide curve) which maps system states and other observed variables to an action (reservoir release) [59]. With DPS, the operating policy is first parameterized within a given family of functions (e.g., linear, piecewise linear, radial basis functions, etc.) and then the parameters that define those functions are optimized with respect to the operating objectives. Previous studies have introduced this concept to the water resources literature for single-objective reservoir optimization [50, 48, 42], assuming the shape of the operating rule empirically. EMODPS broadens direct policy search to a multi-objective context and generalizes our capability to abstract complex operating policies with nonlinear mathematical basis functions while accounting for mixtures of highly challenging, heterogeneous objectives (e.g., stochastic mixtures of minimax measures of multi-sector costs and risks). Non-linear approximators, such as artificial neural networks (ANN) and radial basis functions (RBFs) can be used to flexibly map the system's state onto decisions as time-varying, complex non-linear relationships which should well approximate the unknown optimal operating policies of complex multi-objective reservoir systems [61, 8, 2]. In particular, RBFs have been recently demonstrated to outperform ANNs in terms of their ease-of-use, parsimony, and their ability to generalize beyond initial training scenarios [24].

Reservoir operations are modeled by identifying the feedback policy p, defined by a temporal sequence of operating rules that give the release decision $u_t = p(t, x_t)$ at each time step t for the current system conditions x_t (i.e., the reservoir level). The day of year (time index t) implicitly considers energy prices, water demands and environmental flow constraints. With EMODPS, the operating policies are parameterized within a given class of functions, i.e. $u_t = p(t, x_t, \theta)$ where θ is a vector of time-varying parameters. The values of θ are determined through optimization of the policy parameters relative to the

9

reservoir's objective functions. Optimizing the policy parameters relative to the reservoir objectives seeks to yield policies that capture the potential preferences embodied in the tradeoffs across the dam's objectives.

In a multioutlet reservoir, the release decision associated to the kth outlet is defined as:

$$u_t^k = \sum_{i=1}^{n} w_i^k \varphi_i(z_t) \tag{7}$$

where n is the number of RBFs φ_i, w_i is the weight of the ith basis function, and z_t is the input vector whose elements are time t and state x_t (i.e. reservoir level). The weights sum to 1 and are nonnegative. Each RBF is defined as follows:

$$\varphi_i(z_t) = exp\left[-\sum_{j=1}^{m} \frac{(z_t - c_{j,i})^2}{b_{j,i}^2} \right] \tag{8}$$

where m is the number of input variables, z_t is either the time index or the reservoir level, c_i, b_i are the m-dimensional center and radius vectors of the ith RBF. The parameter vector θ is defined as $\theta = [c_{i,j}, b_{i,j}, w_i^k]$, with $i = 1, ..., n$, $j = 1, ..., m$ and $k = 1, ..., N_u$, where the center and radius are normalized (i.e. $c_{i,j} \in [-1, 1]$, and $b_{i,j} \in (0, 1]$). The inputs to the RBF policy (time index and reservoir level) are mapped into the outputs (downstream release decisions). Subsequently, their performance is evaluated in the objective space.

The policy parameters (θ) are determined by solving the following multi-objective problem:

$$f\theta^* = arg\min_{\theta} J(\theta) \tag{9}$$

The policy parameters θ, are the decision variables, the objective functions J are the reservoir's operating objectives, defined in Equations 2 through 6. The parameters are obtained by simulating the system over the time horizon H, under the policy $p = \{p(t, x_t, \theta); t = 0, ..., H - 1\}$. All the maximization objectives are multiplied by -1, to treat all the objectives as minimization functions.

Key benefits of the EMODPS framework include (1) multiple objectives can

10

be optimized simultaneously in a single run, (2) the approach does not require that the objective functions and the model constraints are time separable, (3) the framework can be combined with any simulation model (including Monte Carlo ensembles), and (4) the non-linear approximators used to represent reservoir policies can easily incorporate exogenous information (e.g., weather observations or forecasts) to condition the decision without significantly increasing computational challenges, enhancing the ease to evaluate further uncertainties. For a more detailed analysis of the resulting EMODPS policies and their implications for the Susquehanna River Basin, we direct the reader to [25].

3.2. Diagnostic Framework

This study implements a comprehensive diagnostic framework illustrated in Figure 3 [29, 52] to compare the performance of several well-known MOEAs. MOEAs are stochastic search tools that use different mating, mutation, selection and archiving parameters. Examples of such parameters can be found in Table 2. The default parameterizations for each MOEA are usually defined by finding parameter values that are highly tuned and perform well for specific applications or for test instances; however, as we deal with more complex problems, we encounter less predictable behavior and can no longer assume ideal parameters [29, 52]. The diagnostic framework used in this study removes this bias by sampling the full feasible parameter space for each evaluated MOEA using Latin Hypercube samples (LHS). Each point drawn from the Latin Hypercube sample in Figure 3 represents a full specification of an MOEA's parameters. Furthermore, each parameterization is benchmarked by running the MOEA using multiple randomly generated seeds to account for random effects when generating initial populations and guiding probabilistic search operators. Pareto approximate sets are then computed for each parameter sample. The fitness of the approximation sets is measured through performance metrics that evaluate the solution's convergence, diversity and consistency (discussed in more detail section 3.3). The Lower Susquehanna test case has an unknown true Pareto front, hence an approximation of the front was obtained by combining all the

11

non-dominated solutions attained across all runs of the seven MOEAs tested. To simplify our nomenclature, in this study the best known Pareto approxima-

₂₇₀ tion set will be referred to as the reference set, this set is used to calculate the performance metrics.

Our classification of the performance of an MOEA can be succinctly summarized through its effectiveness, reliability, efficiency, and controllability. Effectiveness measures if an MOEA attains high levels of performance. Reliability

₂₇₅ captures the variability in performance across parameterizations as well as random seed trials. Efficiency refers to attaining high levels of performance in a minimum number of function evaluations (NFE). Controllability measures the ease-of-use or sensitivity of MOEAs' to their parameterizations. If an algorithm is well suited for a problem, a wider range of its parameter combinations will

₂₈₀ yield good performance; conversely, if an algorithm is not well suited for a problem, or the problem is particularly difficult, it may be hard to find even a single successful algorithm parameterization. Ideally, an MOEA would generate a high quality Pareto approximation for any combination of its input parameters; this is referred to a large "sweet spot" in its parameter space [27].

₂₈₅ Although the MOEA diagnostic framework employed in this study globally samples the algorithms' full feasible parameterization spaces, Reed et al. [52] demonstrated that showing performance with respect to the number of function evaluations and population size is sufficient to visualize the ease-of-use and efficiency of the algorithms. Hence, the control maps depicted in Figure

₂₉₀ 3 summarize the full set of sampled parameterizations projected onto the two-dimensional subspace defined by the sampled ranges of NFE and population sizes.

3.3. Performance Metrics

The MOEA search performance metrics used in this study are generational

₂₉₅ distance, ϵ-indicator, and hypervolume. These metrics allow the comparison of the approximation sets by providing an appropriate quantification of proximity, consistency and diversity, respectively [37, 68, 10, 29]. The metrics were

calculated relative to the reference set as illustrated in Figure 3.

Generational distance [63, 64] is the most basic measure of proximity. It consists of the minimum point in the average Euclidean distance vector between solutions in an approximation set and its corresponding nearest solutions in the reference set in the objective space. This is the easiest metric to meet; it only requires a single point to be close to the reference set. The presence of a single point near the reference set will strongly influence the calculation of the minimum Euclidean distance between each of the objective vectors in the approximation set. Generational distance does not account for diversity and if our solution consists of few points that are close to the reference set, this solution may have good convergence but not diversity, meaning that a good representation of the full set of tradeoffs is not achieved. A large value for this metric implies that an MOEA has failed to find a single solution close to the reference set.

Additive epsilon-indicator [68] provides a measure of consistency, where consistency refers to Pareto approximate sets that capture all portions of tradeoffs. The metric is computed as the largest distance that an approximation set needs to be translated in order to dominate the reference set, thus making this metric very sensitive to gaps in tradeoffs. If a Pareto approximate set has gaps, then solutions must be translated a much further distance, dramatically increasing the additive ϵ-indicator metric value (see illustration of this effect in [29] or [52]). A low value for this metric is desired as it measures the worst-case distance from the reference set.

Hypervolume [68] provides a measure of diversity and proximity. It quantifies the volume of the objective space dominated by an approximation set; therefore, this metric is to be maximized. In this study, hypervolume was normalized relative to the reference set hypervolume; hence, a value of 1 would mean that the approximation set dominates the same volume as the reference set. Hypervolume is generally the most challenging and comprehensive metric providing insight of an algorithm's convergence and the diversity of their representation of tradeoffs.

13

3.4. Multi-objective Evolutionary Algorithms

Multi-objective evolutionary algorithms are general purpose stochastic search methods simulating natural selection and biological evolution. MOEAs use operators that imitate the processes of mating, mutation and selection to solve multi-objective problem formulations. They start from an initial population of randomly generated solutions, and then seek to iteratively improve this set of solutions using selection, mutation and mating operators. MOEAs population-basis poses important practical advantages that help them deal with challenging mathematical properties such as non-convexity, nonlinearity, stochasticity, and mixtures of continuous as well as discrete decisions. These advantages are highlighted by [15, 10, 18]. Reed et al. [52] provide a review of further innovations to MOEAs that dramatically enhanced their performance and usability. These innovations include epsilon-dominance archiving as introduced by Laumanns et al. [45] that archives the box non-dominated solutions within a grid with user specified resolution to guarantee convergence and diversity maintenance. Another important MOEA search innovation is termed time continuation, where MOEAs continuously check for search stagnation and introduce new diverse solutions to reinvigorate the search. Most recently, a new class of self-adaptive MOEAs that use feedbacks from their search progress to adapt their strategies have been developed. Self-adaptivity improves the effectiveness, efficiency, reliability, and controllability of the MOEAs [29, 30, 52], allowing the algorithm to have less dependence upon specific parameter configurations.

Encompassing the above reviewed MOEA innovations, this study benchmarks seven state-of-the- art algorithms that comprise the representative suite of modern tools described below and summarized in Table 1.

NSGAII. The Nondominated Sorted Genetic Algorithm II (NSGAII, Deb et al. [17]) represents a key historical MOEA benchmark and is the most widely used algorithm at the time of this study. It features a fast non-dominated sorting procedure that uses Pareto dominance relation to search for the entire Pareto front in a single run. Diversity is preserved using a crowding distance operator to measure how close an individual is to its neighbors. A large average

14

crowding distance results in better diversity. Finally, parents are selected from the population by using tournament selection based on the rank and crowding distance. Key advancements in NSGAII that have led to its wide use include elitist selection operators (i.e., the best parents are likely to survive) as well as its parameter free crowding distance operator to improve diversity maintenance.

ϵ-**MOEA**. The Epsilon Dominance Multi-objective Evolutionary Algorithm (ϵ-MOEA, Deb et al. [16]) is the first instance of an algorithm that actively exploits epsilon dominance archiving as a feedback to search, that is the archive of non-dominated solutions is iteratively updated in the presence of a new solution and the population size is also adapted based on the concept of epsilon-dominance. It also provides a theoretically guaranteed mechanism for maintaining convergence and diversity. In multi-dimensional problems convergence maintenance becomes challenging since the number of non-dominated solutions increases very quickly and it becomes difficult to discriminate between solutions; this is also known as dominance resistance. Epsilon-dominance avoids dominance resistance by providing stable and bounded archiving and enables the user to specify the desired precision for each objective allowing to maintain convergence. Epsilon-dominance also provides users with means of improving MOEA efficiency [41] by removing the computational burden of seeking Pareto approximate solutions that are not significantly different from a numerical precision perspective. Epsilon dominance requires the user to specify the desired level of precision to identify epsilon non-dominated solutions. ϵ-MOEA is a steady-state algorithm, meaning that only one solution in the population is replaced for each full mating, mutation and selection loop. ϵ- MOEA uses two co-evolving populations: a search population and an archive population. The epsilon-dominance archive is actively exploited in each evolutionary loop where one solution from the population and one from the archive are chosen for generating an offspring solution using the simulated binary crossover (SBX) recombination operator [13] and the polynomial mutation (PM) operator [14] to update the archive population.

ϵ-**NSGAII**. The epsilon dominance NSGAII (ϵ-NSGAII, Kollat [38], Kol-

lat and Reed [40]) extends the original NSGAII by including epsilon dominance archiving, adaptive population sizing, and time continuation as part of a limited degree of self-adaptive search. The ϵ-NSGAII uses a series of connected runs where small populations are exploited to precondition search with successively doubled population sizes. Pre-conditioning occurs by injecting current solutions within the epsilon-dominance archive into the initial generations of larger population runs. For example, when an initial smaller population evolves until it is no longer making significant progress, then the population size increases, where 25% of the new population is composed of archived solutions and the remaining 75% are randomly generated (i.e., time continuation of search).

MOEA/D. The Multi-objective Evolutionary Algorithm Based on Decomposition (MOEA/D, Zhang and Li [67]) exploits the decomposition strategy used in traditional multi-objective optimization. It decomposes a multi-objective optimization problem into several scalar optimization sub-problems and optimizes them simultaneously solving many single-objective Chebyshev decompositions in a single run. Each sub-problem is optimized by only using information from its neighboring sub-problems. The population is composed of the best solution found so far for each subproblem. Only the current solutions to its neighboring subproblem are exploited for optimizing a subproblem in MOEA/D. Since its introduction, the MOEA/D established itself as a benchmark for new MOEAs by winning the 2009 IEEE Congress on Evolutionary Computation (CEC 2009) competition [52].

OMOPSO. Optimized Multi-objective Particle Swarm Optimization Algorithm (OMOPSO, Reyes-Sierra and Coello [54]) contributed a popular multi-objective extension of the Particle Swarm algorithm, inspired by the movement of organisms in a bird flock or a fish school. OMOPSO incorporates epsilon dominance and uses a crowding factor for the selection of leaders. For each generation and for each particle, a leader is selected. Selection is made by binary tournament based on the crowding value of the leaders. This proposal uses two external archives: one for storing the leaders currently being used for performing the flight and another one for storing the final solutions. The crowding factor

is used to filter out the list of leaders whenever the maximum limit imposed on such list is exceeded. Only the leaders with the best crowding values are retained.

 $_{425}$ **GDE3**. The third evolution step of generalized differential evolution (GDE3, Kukkonen and Lampinen [43]) is the multi-objective extension of the differential evolution algorithm introduced by Storn and Price [58] for global optimization. GDE3 starts with an arbitrary number of objectives and constraints. It introduces an adaptive population and non-dominated sorting by pruning non dom-
 $_{430}$ inated solutions to decrease the population size at the end of each generation, aimed to improve solution diversity and to make it more stable to the population parameter. The diversity maintenance technique is based on a crowding estimation using the nearest neighbors of solutions in a Euclidean sense, and a nearest neighbor's search technique. GDE3 uses rotationally invariant op-
 $_{435}$ erators, producing offspring in any direction relative to the orientation of the fitness landscapes. This is an important characteristic for problems with high dependency among its decision variables [35, 30].

 Borg MOEA. Hadka and Reed [30] introduced the Borg MOEA, which is not a single algorithm; alternatively, it represents a hyper-heuristic framework
 $_{440}$ (for a review see Burke et al. [1]), whose search operators are adaptively selected based on the progress being made in solving a problem. The adaptive discovery of key operators is of particular importance for benchmarking how variation operators enhance search for complex many-objective problems. The Borg MOEA assimilates several design principles from existing MOEAs and introduces sev-
 $_{445}$ eral novel components. These components include: an ϵ-box dominance archive for maintaining convergence and diversity throughout search; ϵ-progress, which is a computationally efficient measure of search progression and stagnation; an adaptive population sizing operator based on ϵ-NSGAII's [39] use of time continuation to maintain search diversity and to facilitate escape from local optima;
 $_{450}$ multiple recombination operators to enhance search in a wide assortment of problem domains; and the steady-state, elitist model of ϵ-MOEA [16] which can be easily extended for use on parallel computing [51, 31].

17

4. Computational Experiment

As described in Section 3.1, the EMODPS framework abstracts the Conowingo reservoir's operations using RBF-based simulations of alternative operating policies. The experiments are based on the historical formulation in Giuliani et al. [25] which are run over the trajectories of inflows, evaporation, and energy prices of 1999 representing a dry, challenging year. The Pareto-approximate set of policies is composed of RBF functions mapping reservoir level and time into reservoir release decisions. Each radial basis function is defined by its center, radius and weight parameters. For the Lower Susquehanna test case, this problem has a total of 32 parameters that are used to generate four release decisions every four hours for a given operational year. These four release decisions are required for water supply to Baltimore, Chester and the Atomic Power Plant and for downstream release. The RBF output release decisions are a function of inputs for the time index and reservoir level as illustrated in Figure 4. The seven MOEAs evaluated in this study seek to identify Pareto approximate reservoir policies by finding the RBF's shape parameters that yield the best representation of the tradeoffs across the six objectives of the Conowingo reservoir. For those algorithms exploiting epsilon-box dominance, an epsilon precision must be specified to set the acceptable numerical precision to be used for each objective. The epsilon values are 0.5 for hydropower revenue, 0.05 for each volumetric reliability to Baltimore, Chester and the Atomic Power Plant, 0.05 for recreational storage reliability and 0.001 for the environmental shortage index as in Giuliani et al. [25]. The overall best known Pareto approximate set for the Lower Susquehanna test case was attained using consistent epsilon-dominance sorting across all algorithms. This reflects that it is always possible to transform point dominance results (e.g., those from NSGAII) into box-dominance results; the reverse is not true. It is critical to maintain consistent dominance relationships when benchmarking MOEAs [29].

The parameter space of each MOEA is sampled using 100 Latin Hypercube samples (LHS). Each point in the parameter space shown in Figure 3, represents

18

a full specification of the algorithm's crossover, selection, and mutation operators, as well as initial population sizes and NFEs. Table 2 provides a summary of the parameter ranges for each algorithm including Latin Hypercube samples and random seed replicates used in this study. Each point in the parameter block is replicated for 30 random seeds to account for effects on initial populations and probabilistic search operators. Since the Lower Susquehanna test case has an unknown true Pareto Front, the best known approximation, or reference set, is generated by sorting the non-dominated solutions found across 30 seed runs for 100 parameter samples with maximum number of function evaluations of 200,000 for the seven algorithms tested. The performance metrics described in section 3.3, are then computed relative to this reference set illustrated in Figure 3.

Runtime dynamics for each algorithm were captured in order to test the MOEAs' performance with their default parameterizations. The runtime dynamics provide snapshots of the hypervolume performance attained every one thousand funtion evaluations. The search was extended to 250,000 NFEs to explore if the MOEAs continue to improve their search under their typical use case. Each algorithm was run for 50 random seed trials with their default crossover, mutation and selection operators shown in Table 2.

5. Results and Discussion

5.1. Contributions to Best Known Reference Set

Figure 5 represents the reference set of Pareto approximate solutions that compose the tradeoffs for the Conowingo dam's potential operating policies. These tradeoffs were obtained across all runs for the seven algorithms tested in this study. The arrows in Figure 5 indicate the direction of preference for the environmental shortage index, recreation storage reliability and atomic power plant volumetric reliability. Although not visible, the ideal solution would be located in back lower corner of the box. Figure 5 also provides a measure of the percent of the reference set solutions that were captured by each MOEA

19

within the epsilon precisions specified in section 4. Color is used to visualize the portions of the reference Pareto approximate surface contributed by each algorithm. The Borg MOEA and ϵ-MOEA were the largest contributors to the
515 reference set with 48 and 43 percent respectively (shown with large purple and green areas in Figure 5). The remaining algorithms made very modest contributions to the reference set where ϵ-NSGAII contributed 4 % of the total solutions; both, NSGAII and MOEA/D contributed with 2%, GDE3 provided 1% contribution. OMOPSO failed to capture any of the reference set solutions. Overall,
520 ϵ-MOEA and the Borg MOEA alone capture 91% of the Lower Susquehanna test case's best known tradeoff solutions.

5.2. Effectiveness and Reliable Search

To gain a better understanding of the MOEAs effectiveness and reliability, Figure 6 provides attainment plots that quantify the probability that a given
525 MOEA attains a certain percent of the best possible metric value. Each of the columns in panels (a)-(c) in Figure 6 corresponds to an MOEA. The best overall metric attained in a single seed run for each algorithm is indicated by the black circle in each column. The vertical axis represents the percent of the best metric, and the gray shading indicates the probability of attaining a given
530 percent of the best metric value. Ideal performance would be indicated by a completely black bar with a black circle at the 100% level, designating that a single trial run of an algorithm is both perfectly reliable and effective. Reporting the attainment probability as opposed to only reporting the best overall run, provides a broader context of probabilistic search performance. If an algorithm
535 exhibits outstanding performance in its best single run, but fails to attain high performance consistently, then it has low value to users.

Figure 6a illustrates perfect generational distance for all of the MOEAs tested. This is the easiest metric to meet, since it only requires one solution to be close to the reference set. Failing to attain generational distance would
540 indicate that the algorithm failed to find a single solution near the reference set; this metric helps identify complete algorithmic failure. The fact that none

20

of the tested algorithms display abject failure when supporting direct policy search can provide intuition on the difficulty of the problem. Although all of the algorithms attained strong performance in the generational distance metric, this metric is generally only useful for extremely difficult problems where not even a single Pareto approximate solution is identified. However, the nature of a many objective problem requires the representation of the full set of tradeoffs, as well as capturing all the portions of the tradeoffs, which this metric fails to provide.

When transitioning to epsilon indicator in Figure 6b, all of the algorithms have degraded performance. This is a predictable result since this is a more challenging metric to satisfy and is very sensitive to gaps in the approximation set. Despite the more challenging metric, ϵ-MOEA, the Borg MOEA and ϵ-NSGAII, are highly effective. The ϵ-MOEA leads the attainment probability for high threshold levels, followed by ϵ-NSGAII and the Borg MOEA. This means that a single seed run of these MOEAs across the tested parameterizations would reliably approximate the Conowingo tradeoffs. It should be noted that in Figure 6b, MOEA/D and OMOPSO both show significant degradations in performance.

As expected, Figure 6c shows that hypervolume attainment is the most challenging test of performance across all of the algorithms. Hypervolume is typically challenging since it requires high performance for convergence and diversity; in other words, it requires a high quality representation of the full set of tradeoffs. The top performing algorithms are ϵ-MOEA, the Borg MOEA and ϵ-NSGAII. ϵ-MOEA and the Borg MOEA have the best overall metric value; however, their likelihood of attaining high threshold levels decreases. MOEA/D and OMOPSO have the worst performance. OMOPSO's best overall hypervolume achieves only 30% of the best metric value, while MOEA/D achieves 50%. MOEA/D is sensitive to heterogeneous scaling across an application's objectives due to its use of Chebyshev-based aggregations of objectives into a single weighted objective function.

The overall most effective and reliable algorithms based on their best per-

formance and high attainment probability are ϵ-MOEA, the Borg MOEA and ϵ-NSGAII. Any single run of these algorithms would have a good approximation to the best known Pareto policies. The overall results from Figure 6 highlight that although the 6-objective formulation of the Lower Susquehanna test represents a very challenging control problem, solving the EMODPS variant of the problem is not that challenging for some of the modern MOEAs.

5.3. Controllability and Efficiency

As discussed in section 3, we implemented visually informed controllability goals. This refers to the exploitation of visual analytics to determine what corresponding level of the hypervolume performance yields an acceptable representation of the tradeoffs if used in a decision support context. The hypervolume level that achieves a good representation of decision relevant compromises is then used as our target hypervolume. Figure 7 illustrates that the 75th percentile hypervolume provides a high quality representation of the reference set. Further refinements of hypervolume performance may not be warranted if they require significantly more computational effort, emphasizing the importance of visualization in MOEA benchmarking as recently noted by Reed and Kollat [53]. The control maps in Figure 8 present two-dimensional projections of the MOEAs performance sampled across their full feasible parameter spaces. They are constructed with population sizes ranging from 10 to 1000 and number of function evaluations ranging from 10,000 NFEs to 200,000 NFEs. These two parameters commonly have a very strong influence on algorithmic performance and computational demands. The color legend provides a measure of the percent of the target hypervolume (see Figure 7) captured across 30 random seeds for each of the MOEAs' tested parameterizations. In simpler terms, the results of Figure 8 differ from the attainment results of Figure 6 by assuming as is typically done in practice that for each of the MOEAs' parameterizations the resultant Pareto approximation set would be developed across 30 independent trial runs. Ideal performance would be represented by an entirely dark blue control map, indicating that an MOEA attained the target hypervolume across

22

all of its parameterizations (i.e., it is highly controllable). In other words, it indicates that it would be very difficult to make an MOEA fail as a result of how it is parameterized given 30 random seed trials. This would also mean that a good approximation to the Pareto front could be attained with minimum NFEs, providing insight of the algorithm's efficiency.

The ϵ-NSGAII and ϵ-MOEA control maps show that these algorithms have a very broad range of effective parameters, indicated by a large dark blue region in their control maps. These two algorithms are effective for this problem, attaining high levels of performance for a low number of function evaluations, and for a large range of population sizes. The Borg MOEA requires a larger investment of NFEs to maximize its success; it requires a minimum of 100,000 NFEs to consistently attain the target hypervolume level. Consistent with prior findings, the Borg MOEA is generally only sensitive to NFE [32, 52]. NSGAII, the classical benchmark used in this study, shows a broad zone of high performance; however, it does not attain 100% of the target hypervolume for any of the parameterizations tested. GDE3 requires larger population sizes and increased NFE to improve its performance, requiring more than 100,00 NFE and a population size larger than 400 to attain nearly 75% of the target hypervolume. OMOPSO has poor performance, achieving only 25% to 40% of the target hypervolume for the full range of tested parameters. MOEA/D fails to attain the target hypervolume for this problem. Although this algorithm has been shown to be very strong in solving test functions, MOEA/D is sensitive to scaling, this makes it less useful for many-objective direct policy search.

The results from Figure 8, emphasize the importance of controllable algorithms in water resources applications. Broadly, control maps provide a sense of the sweet spot for an MOEA's parameter space. In other words, they give insight on the sensitivity of an MOEA to its different parameter choices. As we transition to more challenging applications, we need to shift the focus from finding instances of MOEA parameterizations that work for specific applications to MOEAs that support the exploration of challenging problems, and that are capable of yielding high quality results regardless of the user-specified pa-

rameters. In this regard, the top performing algorithms are the Borg MOEA, ϵ-MOEA, and ϵ-NSGAII, which find broad zones of high performance. Namely, ϵ-MOEA shows high performance for broader parameter ranges than the rest of the MOEAs. The Borg MOEA's control map, for example, suggests that it is difficult to make the algorithm fail given sufficient NFE. A similar trend is shown for NSGAII and GDE3; these algorithms have improved results with larger NFE and population sizes. OMOPSO and MOEA/D show the worst overall performance. The overall results from Figures 6 and 8 are consistent in highlighting that the Lower Susquehanna's EMODPS formulation is readily solved by many current MOEAs. The challenge posed in these results is that it would have been very difficult to predict in advance the levels of failure for GDE3, OMOPSO and MOEA/D. When viewed in combination with other recent benchmarking efforts in water resources (Reed et al. [52] and Ward et al. [65]), only the Borg MOEA has consistently performed well across applications. All of the more traditional non-adaptive MOEAs have had mixed success on water resources applications. The benchmarking results for EMODPS in this study further highlight that algorithm choice remains a concern.

5.4. Runtime Dynamics

To represent the typical use case of running the MOEAs using their default parameterizations, as opposed to sampling the full feasible parameter space, each algorithm was run with their default parameters summarized in Table 2. Our typical use case evaluation exploits the MOEAs' absolute hypervolume performance; this metric is clearly demonstrated in Figure 6 as being the most challenging given the joint requirements of diversity maintenance and convergence. Figure 9 represents the average across 50 random seeds for each algorithm. Their search durations were extended to 250,000 NFE to check for continued progress. MOEA/D failed using its default parameterizations as shown by its horizontal line at 0 hypervolume for the entire run. The runtime dynamics show that MOEA/D's recommended parameterizations are not suited for this problem, even when sampling across a broader range of parameterizations, as shown

24

earlier by the control maps in Figure 8. NSGAII and GDE3 have moderate improvements within the first 50 thousand NFE; however, they cease to improve for the remainder of their runs. These algorithms are far less effective when using their default parameters. This suggests that a population size of 100 is not sufficient and a larger population may be required to improve their success (as shown in their control maps in Figure 8). ϵ-MOEA makes rapid progress within the first 100 thousand NFE; however, it exhibits very low improvements throughout the rest of the run. ϵ-NSGAII, outperformed the other algorithms, obtaining the highest hypervolume at the end of the run under default parameterizations, it also showed steep progress within few NFEs, outperforming ϵ-MOEA within 100,000 function evaluations. The Borg MOEA had the second largest hypervolume value at the end of the run; however, its average random seeds show a steady and continued progress throughout the run, indicating that Borg MOEA has potential to improve the search steadily with extended NFE. Both of the top performing algorithms, ϵ-NSGAII and the Borg MOEA, feature time continuation, which enables them to detect stagnation and reinvigorate the search injecting new random solutions. This enables them to continue to improve the search. Additionally, the Borg MOEA's adaptive multi-operator use makes it less sensitive to its parameters and predictably improves performance with increasing NFE. Table 3 provides a summary of hypervolume performance for the evaluated MOEAs. This metric is presented since it is the most challenging metric used in this study. The table presents the maximum and the mean hypervolume values along with their variance across all parameters and all seeds for each of the MOEAs tested. The percent contribution of each MOEA to the reference set is also provided.

6. Conclusions

This study has demonstrated that some modern MOEAs are capable of reliably and effectively finding control policies that balance the Lower Susquehanna test case's six-objective tradeoffs using Evolutionary Multi-objective Direct Pol-

25

icy Search (EMODPS). By acting directly in the policy space, EMODPS enables the evaluation of multiple objectives simultaneously while enhancing users' ability to exploit simulation models, include complex mixtures of different types of objectives (e.g., expected costs and minimax risks), and incorporate a broader array of system information in reservoir operating policies (e.g., model forecasts, upstream operations, etc.). The potential value of EMODPS increases if it facilitates the transition to complex reservoir control problems.

Understanding our ability to address more complex reservoir management applications using EMODPS requires rigorous assessments of the capabilities of modern MOEAs to support this framework. This study expands on the MOEA diagnostic study by Reed et al. [52] by benchmarking the relative efficiency, effectiveness, and reliability of ϵ-MOEA, Borg, ϵ-NSGAII, NSGAII, GDE3, OMOPSO and MOEA/D when solving a challenging EMODPS application. The diagnostic benchmarking results for these algorithms are based on a six-objective formulation for the management of the Conowingo reservoir located in the Lower Susquehanna River Basin. Overall the ϵ-MOEA, ϵ-NSGAII and the Borg MOEA demonstrated consistently high levels of performance. These three algorithms feature epsilon dominance-archiving which represents a diversity enhancement that also ensures stable and bounded archiving. Additionally, the Borg MOEA and ϵ-NSGAII feature time-continuation with adaptive population sizing to help enhance the search and guarantee continued progress with increased search time. These features help maintain search diversity and facilitate escape from local optima, observed by a continued and stable progress throughout their runtime. MOEA/D and OMOPSO exhibited poor performance. GDE3 and NSGAII also struggled to support direct policy search. This implies that algorithmic performance still remains a concern. Collectively, recent MOEA diagnostic studies demonstrate that only the Borg MOEA has performed consistently well across a wider water applications' suite [32, 29, 52, 65].

As we confront river basin systems with a higher number of reservoirs and increasingly uncertain tradeoffs impacting their operations (e.g., climate change, changes in energy markets, population pressures, ecosystem services, etc.), fu-

26

ture EMODPS solution strategies will require MOEAs that are highly scalable
and extensible to emerging parallel computing architectures. Future research
efforts should focus on extensions that consider broader envelopes of uncertainty to encompass societal challenges and climate change, and also effectively use of information to make policies more adaptive without significantly increasing the computational demands. Reservoirs are very complex and integrative coupled human-natural systems. Effective operations require an understanding of evolving hydro-climatology, conflicting demands, and risks. More research is also needed to bridge the longstanding gap between theoretical analyses of reservoirs and their actual operations.

7. Acknowledgements

Table 1: Multi-objective Evolutionary Algorithms (MOEAs) tested in this study.

Algorithm	Class	Reference
Borg MOEA	Adaptive multi-operator	Hadka and Reed, 2011
ϵ-NSGAII	Pareto front approximation	Kollat and Reed, 2006
ϵ-MOEA	Pareto front approximation	Deb et al, 2002
OMOPSO	Particle swarm optimization	Sierra and Coello Coello, 2005
GDE3	Differential Evolution	Kukkonen and Lampinen, 2005
MOEA/D	Aggregate functions	Zhang et al, 2009
NSGAII	Baseline	Deb et al, 2000

Table 2: Latin hypercube sampling of MOEAs' operators and their associated parameter ranges as well as the MOEAs' default parameterizations.

	Parameter	LHS range	Default	Algorithms
Crossover	SBX rate	0 -1	0.01	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII
	SBX distribution index	0-500	15	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII
	DE crossover rate	0-1	0.1	Borg, GDE3, MOEA/D, NSGAII
	DE step size	0-1	0.5	Borg, GDE3, MOEA/D
	PCX parents	2-10	3	Borg
	PCX offspring	1-10	2	Borg
	PCX eta	0 -1	0.1	Borg
	PCX zeta	0 -1	0.1	Borg
	UNDX parents	2 -10	3	Borg
	UNDX offspring	1 -10	2	Borg
	UNDX eta	0 -1	0.5	Borg
	UNDX zeta	0-1	0.35	Borg
	SPX parents	2-10	3	Borg
	SPX offspring	1-10	2	Borg
	SPX epsilon	0-1	0.5	Borg
Mutation	PM rate	0-1	1	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII, MOEA/D
	PM distribution index	0-500	20	Borg, ϵ-NSGAII, ϵ-MOEA, NSGAII, MOEA/D
	UM rate	0-1	1/L	Borg
	Perturbation Index	0-1	NA	OMOPSO
Selection	Neighborhood Size	0-0.2	0.1	MOEA/D
	Delta	0-1	0.9	MOEA/D
	Eta	0-0.02	0.01	MOEA/D
	Archive Size	10-1000	NA	OMOPSO
	Injection Rate	0.1-1	0.25	Borg, ϵ-NSGAII
Population Size		10-1000	100	All algorithms
NFE		10,000 - 200,000	250,000	All algorithms
LHS		100	NA	All algorithms

Table 3: Summary of hypervolume performance and set contribution. The maximum, mean and variance of the hypervolume is obtained across all seeds and parameterizations for each of the tested MOEAs. The set contribution refers to the percent of solutions from the best overall approximation obtained by each MOEA.

Algorithm	Maximum hypervolume	Mean hypervolume	Variance	Set Contribution
Borg MOEA	0.5063	0.1836	0.0155	48 %
ϵ-MOEA	0.4767	0.2354	0.0094	43 %
ϵ-NSGAII	0.4408	0.1950	0.0110	4 %
NSGAII	0.372	0.1301	0.0071	2 %
GDE3	0.3584	0.1028	0.0057	1 %
MOEA/D	0.2016	0.0096	0.00058	2 %
OMOPSO	0.1883	0.0413	0.00047	0 %

Figure 1: Map of the Susquehanna River Basin, highlighted in yellow is the Lower Susquehanna River Basin section. The Conowingo hydropower plant and the Muddy Run facility are represented by a large and a small triangle, respectively, located between Pennsylvania and Maryland. (Map adapted from [47]).
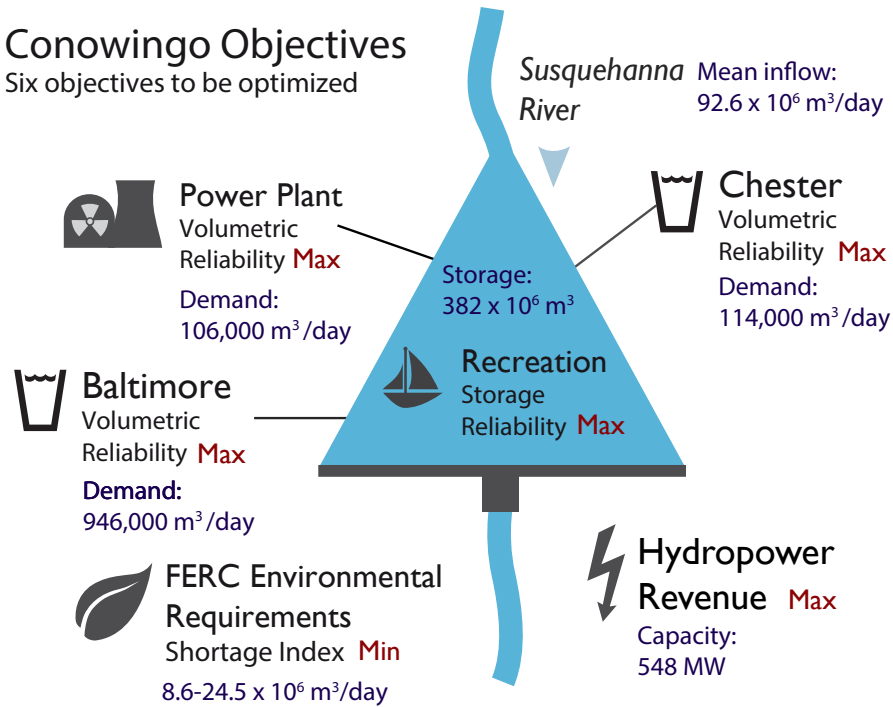
32

Figure 2: Illustration of the six objectives of the Conowingo reservoir. The reservoir has a storage capacity of 382×10^6 m^3/day (approximately 310,000 acre-feet) and hydropower capacity of 548 MW, with a mean annual inflow from the Susquehanna river of 92.6×10^6 m^3/day. Chester, PA and Baltimore, MD can withdraw a maximum of 114,000 and 946,000 m^3/day respectively, approved by the Susquehanna River Basin Commission. Peach Bottom Atomic Power Station uses water from the Conowingo reservoir for cooling purposes, evaporating up to 106,000 m^3/day. Additionally, FERC requires a minimum intermittent flow of 8.6×10^6 m^3/day and up to 25.5×10^6 m^3/day for the fish migratory period in the spring [60].
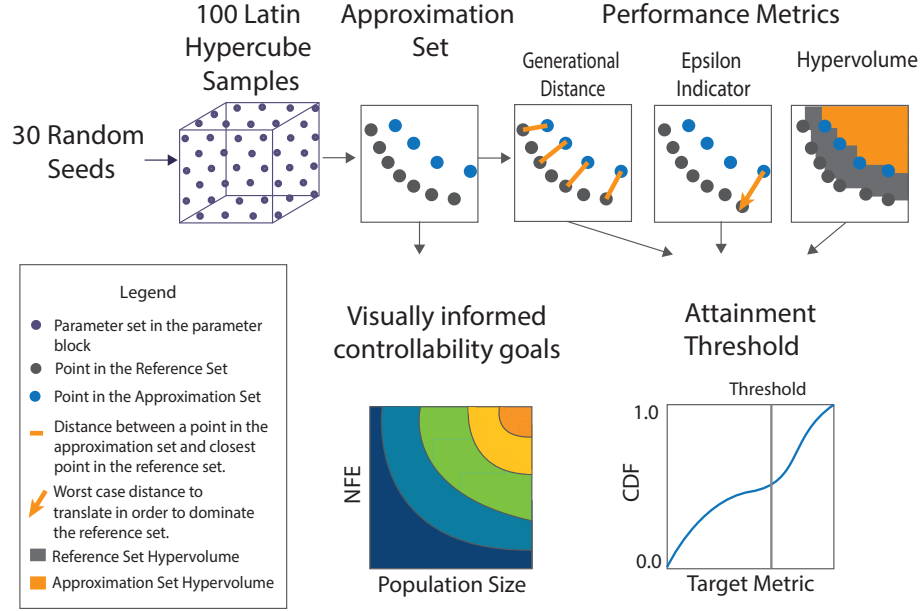
Figure 3: Diagnostic assessment framework used to evaluate the performance of each of the seven MOEAs tested in the study (adapted from Reed et al. [52]). The parameters for each MOEA are sampled across their full ranges using Latin Hypercube Sampling. Each MOEA parameter set is run in replicate for 30 random seeds to account for random seed effects. The approximation set for each MOEA parameterization is assessed through metrics measuring the convergence, consistency and diversity of approximation sets. Probabilistic assessments of metric attainment are then used to assess the effectiveness and reliability of each algorithm. Visually informed metric goals are then established to construct control maps that capture the efficiency and controllability of each algorithm.
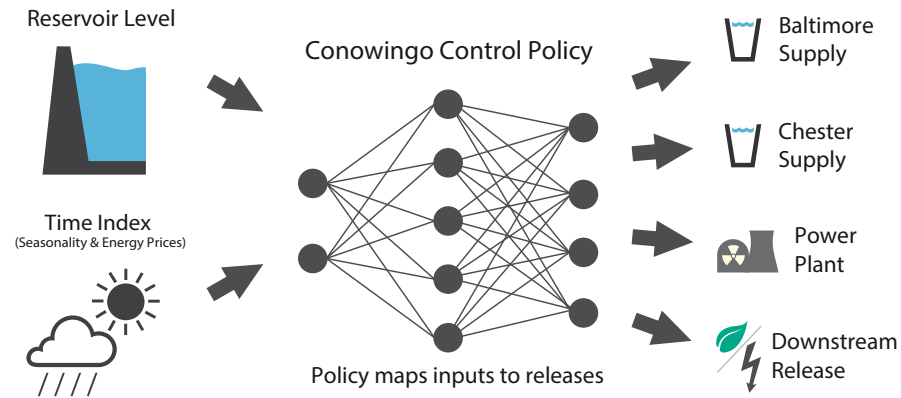
Figure 4: Illustration of the control policy represented using radial basis functions. Radial basis functions input reservoir level and time index to define the corresponding release decisions for water supply and downstream release for hydropower production and minimum environmental flow requirements.
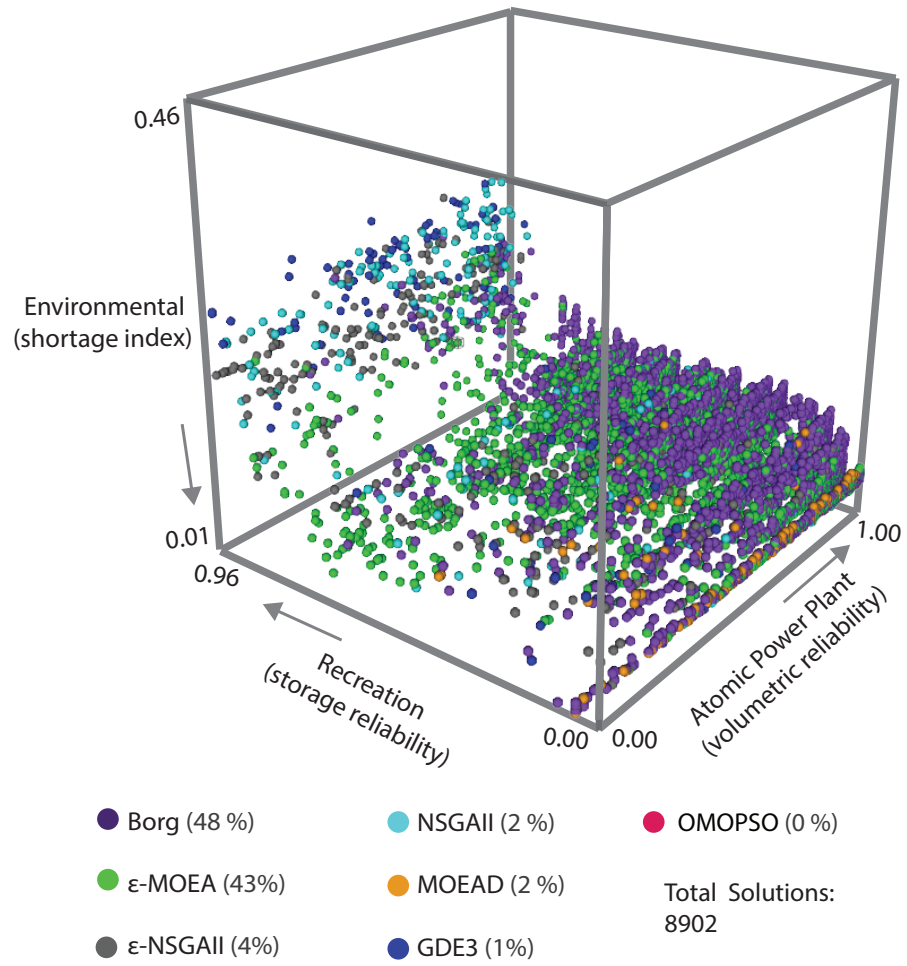
Figure 5: The reference set for the Lower Susquehanna test case attained across all MOEA runs. The arrows show the direction of preference for the plotted objectives. The ideal solution would be located in the lower-back corner. The colors represent contributions from each MOEA.
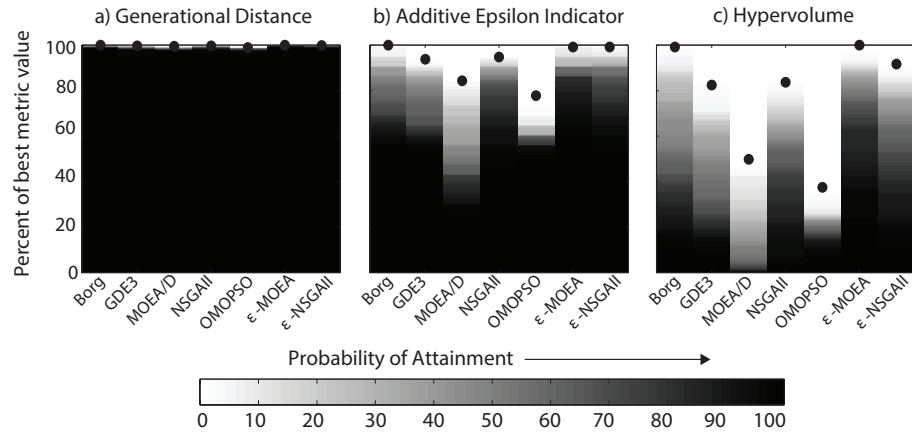
a) Generational Distance   b) Additive Epsilon Indicator   c) Hypervolume

Probability of Attainment

Figure 6: The best overall metric value achieved by each MOEA across all runs is designated using black circles. The grayscale shading shows the probability that a single random seed run of an MOEA reaches a given percentage of the best possible metric value for (a) generational distance, (b) additive epsilon indicator, and (c) hypervolume.
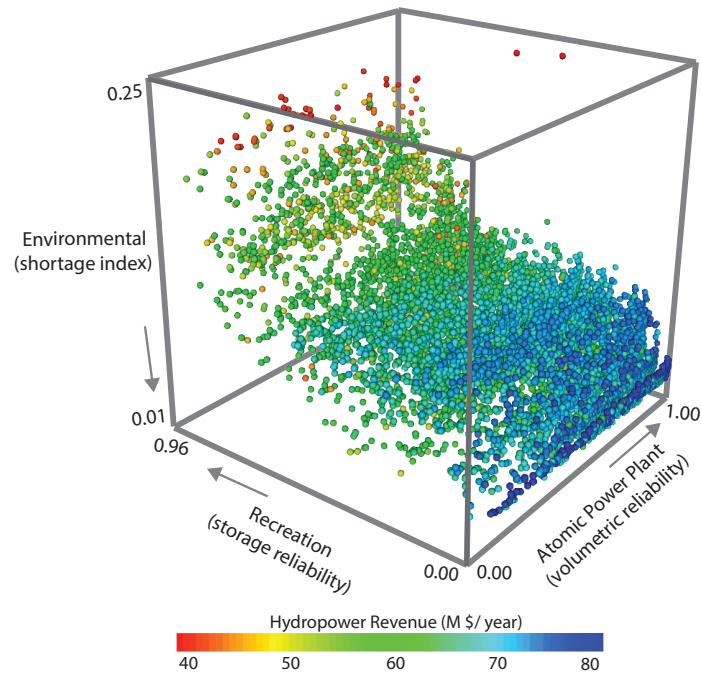
Figure 7: Approximation set corresponding to the 75th percentile hypervolume. This performance attainment goal provides a good representation of the Conowingo tradeoffs and defines the target hypervolume for the control maps.
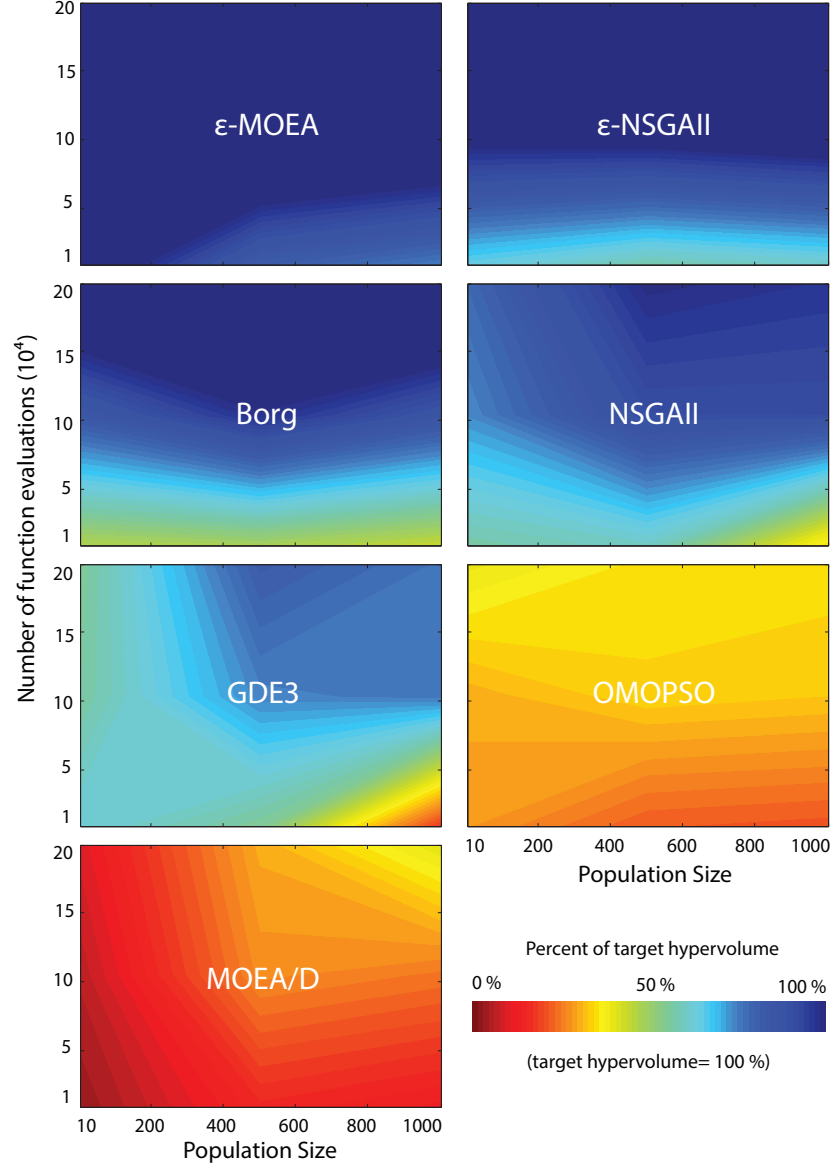
Figure 8: Hypervolume performance control maps capturing controllability and efficiency of each MOEA. The color scale represents the percent of the target (75th percentile hypervolume) captured by each local 30-seed approximation set from each tested MOEA parameterization. The control maps are subprojections of the Latin Hypercube samples for each MOEA's full feasible parameter ranges, focusing on the number of function evaluations (NFEs) and population sizes. Ideal performance is shown as dark blue shading indicating that 100% of the target hypervolume is captured by the MOEA, while dark red designates full algorithmic failure.
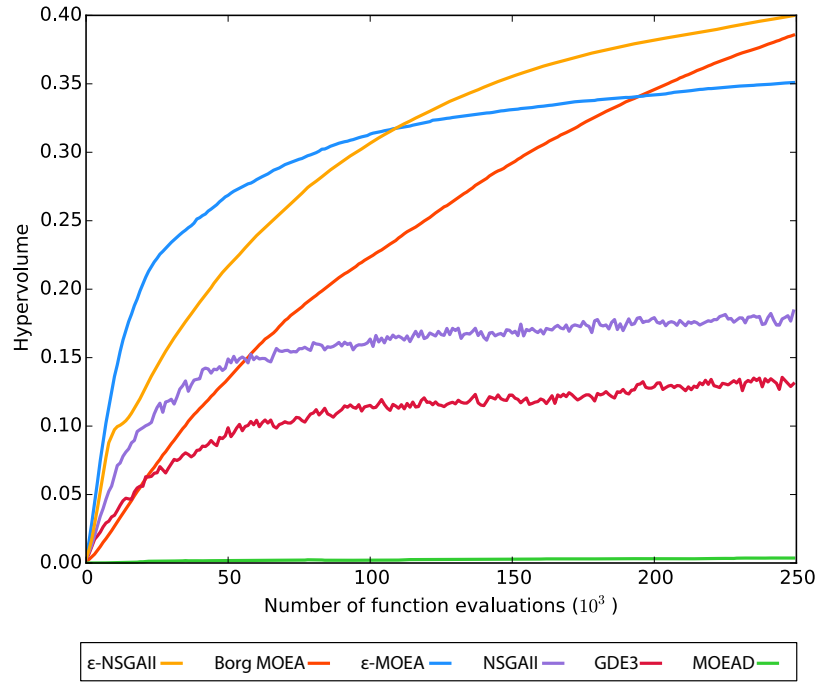
Figure 9: Random seed average search dynamics that result when each MOEA solves the Lower Susquehanna test case using their default parameterizations. Each line represents the average hypervolume attained as a function of the number of function evaluations across 50 random seed runs of each MOEA's default parameterizations.

References

[1] Burke, E.K., Gendreau, M., Hyde, M., Kendall, G., Ochoa, G., Özcan, E., Qu, R., 2013. Hyper-heuristics: A survey of the state of the art. Journal of the Operational Research Society 64, 1695–1724.

[2] Busoniu, L., Ernst, D., De Schutter, B., Babuska, R., 2011. Cross-entropy optimization of control policies with adaptive basis functions. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 41, 196–209.

[3] Castelletti, A., Antenucci, J., Limosani, D., Quach Thi, X., Soncini-Sessa, R., 2011. Interactive response surface approaches using computationally intensive models for multiobjective planning of lake water quality remediation. Water Resources Research 47.

[4] Castelletti, A., Galelli, S., Restelli, M., Soncini-Sessa, R., 2010. Tree-based reinforcement learning for optimal water reservoir operation. Water Resources Research 46.

[5] Castelletti, A., Pianosi, F., Restelli, M., 2013. A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run. Water Resources Research 49, 3476–3486.

[6] Castelletti, A., Pianosi, F., Soncini-Sessa, R., 2008. Water reservoir control under economic, social and environmental constraints. Automatica 44, 1595–1607.

[7] Castelletti, A., Pianosi, F., Soncini-Sessa, R., 2012. Stochastic and robust control of water resource systems: Concepts, methods and applications. System Identification, Environmental Modelling, and Control System Design , 383–401.

[8] Castelletti, A., de Rigo, D., Rizzoli, A., Soncini-Sessa, R., Weber, E., 2007. Neuro-dynamic programming for designing water reservoir network management policies. Control Engineering Practice 15, 1031–1038.

[9] Castelletti, A., Yajima, H., Giuliani, M., Soncini-Sessa, R., Weber, E., 2014. Planning the optimal operation of a multioutlet water reservoir with water quality and quantity targets. Journal of Water Resources Planning and Management 140, 496–510.

[10] Coello Coello, C.A., 2007. Evolutionary Algorithms for Solving Multi-Objective Problems. Genetic and Evolutionary Computation. 2 ed., Springer, New York.

[11] Cohon, J., Marks, D., 1975. A review and evaluation of multiobjective programming techniques. Water Resources Research 11, 208–220.

[12] Commission, F.E.R., et al., 1989. Order approving settlement agreement for the Conowingo hydroelectric project. Technical Report. Tech. Rep. Proj. 405-009, Docket EL80-38.

41

[13] Deb, K., Agrawal, R.B., 1994. Simulated binary crossover for continuous search space. Technical Report IITK/ME/SMD-94027. Indian Institute of Technology. Kanpur.

[14] Deb, K., Goyal, M., 1996. A combined genetic adaptive search (geneas) for engineering design. Computer Science and Informatics 26, 30–45.

[15] Deb, K., Gupta, H., 2006. Introducing robustness in multi-objective optimization. Evolutionary Computation 14, 463–494.

[16] Deb, K., Mohan, M., Mishra, S., 2003. A fast multi-objective evolutionary algorithm for finding well-spread pareto-optimal solutions. Technical Report 2003002. Kanpur Genetic Algorithms Laboratory.

[17] Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197.

[18] Deb, K., Sinha, A., 2009. Solving bilevel multi-objective optimization problems using evolutionary algorithms, in: Evolutionary Multi-Criterion Optimization. Springer Science $\mathplus$ Business Media, pp. 110–124.

[19] Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., et al., 2014. Ipcc, 2014: Climate change 2014: Mitigation of climate change. contribution of working group iii to the fifth assessment report of the intergovernmental panel on climate change. Transport .

[20] Efstratiadis, A., Koutsoyiannis, D., Xenos, D., 2004. Minimizing water cost in water resource management of athens. Urban Water Journal 1, 3–15.

[21] Exelon, 2010. Operations modeling calibration report, Technical Report, Addendum Conowingo hydroelectric project- RSP 3.11 (405) and Muddy Run Pumped Storage Project- RSP 3.2 (2355). Technical Report. Gomez and Sullivan Eng.

[22] Fleming, P.J., Purshouse, R.C., Lygoe, R.J., 2005. Many-objective optimization: An engineering design perspective, in: Evolutionary multi-criterion optimization, Springer. pp. 14–32.

[23] Giuliani, M., Castelletti, A., 2016. Is robustness really robust? how different definitions of robustness impact decision-making under climate change. Climatic Change , 1–16.

[24] Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., P.M., R., 2015a. Curses, tradeoffs, and scalable management: Advancing evolutionary multi-objective direct policy search to improve water reservoir operations. ASCE Journal of Water Resources Planning & Management .

[25] Giuliani, M., Herman, J., Castelletti, A., Reed, P., 2014. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. Water Resources Research 50, 3355–3377.

[26] Giuliani, M., Pianosi, F., Castelletti, A., 2015b. Making the most of data: an information selection and assessment framework to improve water systems operations. Water Resources Research 51, 9073–9093.

[27] Goldberg, D.E., 2002. The Design of Innovation: Lessons from and for Competant Genetic Algorithms. Kluwer Academic Publishers, Boston.

[28] Guariso, G., Rinaldi, S., Soncini-Sessa, R., 1986. The management of Lake Como: A multiobjective analysis. Water Resources Research 22, 109–120.

[29] Hadka, D., Reed, P., 2012. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. Evolutionary Computation 20, 423–452.

[30] Hadka, D., Reed, P., 2013. Borg: An auto-adaptive many-objective evolutionary computing framework. Evolutionary Computation 21, 231–259.

[31] Hadka, D., Reed, P., 2014. Large-scale parallelization of the borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems. Environmental Modelling & Software .

[32] Hadka, D., Reed, P.M., Simpson, T.W., 2012. Diagnostic assessment of the borg moea for many-objective product family design problems, in: Evolutionary Computation (CEC), 2012 IEEE Congress on, IEEE. pp. 1–10.

[33] Hashimoto, T., Loucks, D.P., Stedinger, J.R., 1982a. Robustness of water resources systems. Water Resources Research 18, 21–26.

[34] Hashimoto, T., Stedinger, J.R., Loucks, D.P., 1982b. Reliability, resiliency and vulnerability criteria for water resource system performance evaluation. Water Resources Research 18, 14–20.

[35] Iorio, A.W., Li, X., 2008. Improving the performance and scalability of differential evolution, in: Simulated Evolution and Learning. Springer, pp. 131–140.

[36] Kern, J.D., Characklis, G.W., Foster, B.T., 2015. Natural gas price uncertainty and the cost-effectiveness of hedging against low hydropower revenues caused by drought. Water Resources Research 51, 2412–2427.

[37] Knowles, J., Corne, D., 2002. On metrics for comparing nondominated sets, in: Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on, IEEE. pp. 711–716.

[38] Kollat, J.B., 2005. The Epsilon Non-Dominated Sorted Genetic Algorithm II: A Highly Effective Multi-Objective Evolutionary Algorithm for Water Resources Applications. Master's thesis. The Pennsylvania State University.

[39] Kollat, J.B., Reed, P.M., 2005. The value of online adaptive search: A performance comparison of NSGAII, epsilon-NSGAII and epsilon-MOEA, in: Coello Coello, C., Aguirre, A., Zitzler, E. (Eds.), EMO 2005: The Third International Conference On Evolutionary Multi-Criterion Optimization, Springer Verlag. pp. 386–398.

[40] Kollat, J.B., Reed, P.M., 2006. Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design. Advances in Water Resources 29, 792–807.

[41] Kollat, J.B., Reed, P.M., 2007. A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. Advances in Water Resources 30, 335–353.

[42] Koutsoyiannis, D., Economou, A., 2003. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. Water Resources Research 39.

[43] Kukkonen, S., Lampinen, J., 2005. Gde3: The third evolution step of generalized differential evolution, in: Evolutionary Computation, 2005. The 2005 IEEE Congress on, IEEE. pp. 443–450.

[44] Labadie, J.W., 2004. Optimal operation of multireservoir systems: State-of-the-art review. Journal of Water Resources Planning and Management 130, 93–111.

[45] Laumanns, M., Thiele, L., Deb, K., Zitzler, E., 2002. Combining convergence and diversity in evolutionary multiobjective optimization. Evolutionary computation 10, 263–282.

[46] Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., Dandy, G.C., Gibbs, M.S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D.P., Vrugt, J.A., Zecchin, A.C., Minsker, B.S., Barbour, E.J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M., Reed, P.M., 2014. Evolutionary algorithms and other meta-heuristics in water resources: Current status, research challenges and future directions. Environmental Modelling & Software 62, 271–299.

[47] Musser, K., 2007. Susquehanna river watershed with lower susquehanna river watershed highlited .

[48] Nalbantis, I., Koutsoyiannis, D., 1997. A parametric rule for planning and management of multiple-reservoir systems. Water Resources Research 33, 2165–2177.

[49] Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., Zechman, E., 2010. State of the art for genetic algorithms and beyond in water resources planning and management. Journal of Water Resources Planning and Management 136, 412–432.

[50] Oliveira, R., Loucks, D.P., 1997. Operating rules for multireservoir systems. Water Resources Research 33, 839–852.

[51] Reed, P.M., Hadka, D., 2014. Evolving many-objective water management to exploit exascale computing. Water Resources Research 50, 8367–8373.

[52] Reed, P.M., Hadka, D., Herman, J.D., Kasprzyk, J.R., Kollat, J.B., 2013. Evolutionary multiobjective optimization in water resources: The past, present and future. Advances in Water Resources 51, 438–456.

[53] Reed, P.M., Kollat, J.B., 2013. Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example. Advances in Water Resources 56, 1–13.

[54] Reyes-Sierra, M., Coello, C.C., 2006. Multi-objective particle swarm optimizers: A survey of the state-of-the-art. International journal of computational intelligence research 2, 287–308.

[55] Rosenstein, M.T., Barto, A.G., 2001. Robot weightlifting by direct policy search, in: International Joint Conference on Artificial Intelligence, Citeseer. pp. 839–846.

[56] Sheer, D.P., Dehoff, A., 2009. Science-based collaboration: Finding better ways to operate the conowingo pond. American Water Works Association. Journal 101, 20.

[57] Soncini-Sessa, R., Weber, E., Castelletti, A., 2007. Integrated and participatory water resources management-theory. volume 1. Elsevier.

[58] Storn, R., Price, K., 1997. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11, 341–359.

[59] Sutton, R.S., McAllester, D.A., Singh, S.P., Mansour, Y., et al., 1999. Policy gradient methods for reinforcement learning with function approximation., in: NIPS, Citeseer. pp. 1057–1063.

[60] Swartz, P., 2006. Conowingo pond management plan. Technical Report. Technical Report 242, Susquehanna River Basin Commission.

[61] Tikk, D., Kóczy, L.T., Gedeon, T.D., 2003. A survey on universal approximation and its limits in soft computing techniques. International Journal of Approximate Reasoning 33, 185–202.

[62] Vamplew, P., Dazeley, R., Berry, A., Issabekov, R., Dekker, E., 2011. Empirical evaluation methods for multiobjective reinforcement learning algorithms. Machine Learning 84, 51–80.

[63] Van Veldhuizen, D.A., Lamont, G.B., 1998a. Evolutionary computation and convergence to a pareto front, in: Late breaking papers at the genetic programming 1998 conference, Citeseer. pp. 221–228.

[64] Van Veldhuizen, D.A., Lamont, G.B., 1998b. Multiobjective evolutionary algorithm research: A history and analysis. Technical Report. Citeseer.

[65] Ward, V., Singh, R., Reed, P., Keller, K., 2015. (in-revision). confronting tipping points: How well can multi-objective evolutionary algorithms support the management of environmental thresholds? Environmental Modelling & Software .

[66] World Bank, W., 2009. World Development Report 2009: Reshaping Economic Geography. World Bank.

[67] Zhang, Q., Li, H., 2007. Moea/d: A multiobjective evolutionary algorithm based on decomposition. Evolutionary Computation, IEEE Transactions on 11, 712–731.

[68] Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Fonseca, V.G.d., 2003. Performance assessment of multiobjective optimizers: an analysis and review. IEEE Transactions on Evolutionary Computation 7, 117–132.