Advances in Water Resources 145 (2020) 103718

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Can modern multi-objective evolutionary algorithms discover high-dimensional financial risk portfolio tradeoffs for snow-dominated water-energy systems?

Rohini S. Gupta[a,*], Andrew L. Hamilton[b,c], Patrick M. Reed[a], Gregory W. Characklis[b,c]

[a] Department of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA
[b] Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[c] Center on Financial Risk in Environmental Systems, Gillings School of Global Public Health, UNC Institute for the Environment, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

ARTICLE INFO

Keywords:
Financial risk management
Direct policy search
Many objective optimization
Evolutionary algorithms
Hydropower
Algorithm benchmarking

ABSTRACT

Hydropower generation in the Hetch Hetchy Power System is strongly tied to snowmelt dynamics in the central Sierra Nevada and consequently is particularly financially vulnerable to changes in snowpack availability and timing. This study explores the Hetchy Hetchy Power System as a representative example from the broader class of financial risk management problems that hold promise in helping utilities such as SFPUC to understand the tradeoffs across portfolios of risk mitigation instruments given uncertainties in snowmelt dynamics. An evolutionary multi-objective direct policy search (EMODPS) framework is implemented to identify time adaptive stochastic rules that map utility state information and exogenous inputs to optimal annual financial decisions. The resulting financial risk mitigation portfolio planning problem is mathematically difficult due to its high dimensionality and mixture of nonlinear, nonconvex, and discrete objectives. These features add to the difficulty of the problem by yielding a Pareto front of solutions that has a highly disjoint and complex geometry. In this study, we contribute a diagnostic assessment of state-of-the-art multi-objective evolutionary algorithms' (MOEAs') abilities to support a DPS framework for managing financial risk. We perform comprehensive diagnostics on five algorithms: the Borg multi-objective evolutionary algorithm, Non-dominated Sorting Genetic Algorithm II (NSGA-II), Non-dominated Sorting Genetic Algorithm III (NSGA-III), Reference Vector Guided Evolutionary Algorithm (RVEA), and the Multi-objective Evolutionary Algorithm Based on Decomposition (MOEA/D). The MOEAs are evaluated to characterize their controllability (ease-of-use), reliability (probability of success), efficiency (minimizing model evaluations), and effectiveness (high quality tradeoff representations). Our results show that newer decomposition, reference point, and reference vector algorithms are highly sensitive to their parameterizations (difficult to use), suffer from search deterioration (losing solutions), and have a strong likelihood of misrepresenting key tradeoffs. The results emphasize the importance of using MOEAs with archiving and adaptive search capabilities in order to solve complex financial risk portfolio problems in snow-dependent water-energy systems.

1. Introduction

The Western United States (US) is strongly dependent on a complex and highly interdependent suite of water and energy infrastructure systems (Voisin et al., 2018; O'Connell et al., 2019; Liu et al., 2019). The hydrologic variability of the region coupled with growing pressures from increasingly variable climate extremes poses risks to the financial stability of these systems, which must find ways to cope with the variable revenues associated with lost hydropower generation as well as the potential need to supplement hydropower with more expensive thermal generation (Foster et al., 2015; Clarke et al., 2018). Snow-dependent hydropower generation within California (CA) provides an excellent example case. There is a growing need for improved decision support frameworks that are capable of helping utilities to discover and navigate the complex tradeoffs that are emerging as they confront hydro-climatic extremes that strongly impact power generation and their financial stability. In CA, hydropower primarily comes from snowpack runoff stored in high-altitude reservoirs in the Sierra Nevada mountain range. The historic 2012–2016 drought, brought on by subsequent years of low precipitation and high temperatures, drastically reduced water availability for hydropower production and uncharacteristically warm winters in 2014 and 2015 led to the lowest snowpack on record in CA (Gonzalez et al., 2018). Low snowpack corresponded to less runoff and consequently, the worst years of the drought reduced

* Corresponding author.
E-mail address: rg727@cornell.edu (R.S. Gupta).

CA's average hydropower production from 13% to 5% of the state's total energy mix (Lund et al., 2018). However, in the winters of 2016 and 2017, the Sierra Nevada saw historical snowfall from a series of strong atmospheric rivers (Gonzalez et al., 2018) that ended the persistent drought and brought the percentage of hydropower production in 2017 to 14.7% of California's total generation mix (California Energy Commission, 2018). This degree of inter-annual snowpack and hydro-climatic variability poses a significant management challenge for hydropower-reliant utilities Gonzalez et al. (2018). Variability in hydropower generation due to fluctuations in snowpack can affect the financial stability of hydropower-dependent utilities which often receive revenue proportional to electricity generated while incurring fixed costs every year, such as payments related to debt service, operations and maintenance, and personnel. In dry periods, decreases in net revenue can lead to an inability to make these payments, ultimately resulting in credit rating downgrades or even bankruptcy (Kern et al., 2015; Foster et al., 2015).

In an effort to mitigate financial volatility, hydropower-dependent utilities can manage their hydrological risks by investing in financial risk instruments such as hedging contracts to supplement revenue with payouts in dry years (Foster et al., 2015). It is common for a utility to also maintain a reserve fund to be used as a buffer in years that result in unanticipated losses. Hamilton et al. (2020) introduces a financial hedging instrument based on an index derived from a weighted average of February 1st and April 1st snow water equivalent depth (SWE) observations in the Sierra Nevada. The contract, termed a contract for differences (CFD), provides the buyer of the contract with payouts in years of low SWE, when revenue shortfalls are likely. In return, the buyer makes payments to the contract seller in years of high SWE, when the utility expects to have ample revenue. The study shows that the CFD, especially when combined with a reserve fund, is an effective tool for managing financial risk associated with variable hydropower revenues. Hamilton et al. (2020) apply these tools in the context of a stochastic simulation model for the hydropower generation and revenues of San Francisco Public Utilities Commission's (SFPUC) Power Enterprise Division. SFPUC supplies electricity, primarily produced from hydropower, to San Francisco's International Airport, municipal buildings, and a variety of customers classes (San Francisco Public Utilities Commission, 2016).

SFPUC and power utilities in general, have a growing interest in designing and optimizing portfolios of financial risk management tools. Bolton et al. (2011) suggests that the key to effective risk management requires consideration of both hedging tools and financial liquidity management in a state-aware and dynamic context. Traditionally, dynamic risk management has been framed as a stochastic dynamic programming problem (Mulvey and Shetty, 2004) but more recent studies suggest that simulation-optimization approaches hold promise to better capture the complex nature of real-world financial applications requiring frequent and adaptive decisions that are robust to a wide variety of potential futures (Better et al., 2008). The challenge with these simulation-optimization financial portfolio formulations, beyond being stochastic, is that their resulting mathematical formulations are typically severely nonlinear and must consider a broad array of objectives that encompass the financial interests of the institution, such as maximizing annual and minimum cash flow and minimizing portfolio complexity as well as a reserve fund's size. Moreover, their decision structure is analogous to high dimensional control problems (Powell, 2019). These mathematical properties have motivated a transition towards using heuristic global search algorithms such as multi-objective evolutionary algorithms (MOEAs) to discover optimal tradeoff solutions.

MOEAs are population-based stochastic search tools that use mating, mutation, and selection operators to evolve a candidate population

a portfolio of financial assets that minimize risk and maximize return (Markowitz, 1952; Steinbach, 2001). However, the Markowitz model is characterized by simplistic assumptions and has brought to light the necessity of increasing the complexity of financial models to better represent the dynamic conditions that real-world decision-makers face by moving towards formulations that are highly adaptive and state-aware (Ponsich et al., 2013). Direct Policy Search (DPS), first introduced by Rosenstein and Barto in the robotic control theory literature, is classified as a simulation-based policy approximation control formulation (Rosenstein and Barto, 2001; Powell, 2019). In short, time sequences of decisions are abstracted as parameterized policies using universal approximators such as radial basis functions or neural networks. The parameters of these policies are then optimized through simulation-optimization to meet system objectives. The DPS approach has a significant history in the water resources literature where it is also known as parameterization-simulation-optimization and was first used to model single-objective reservoir operations (Koutsoyiannis and Economou, 2003).

The DPS methodology has been extended to the multi-objective context with the Evolutionary Multi-Objective Direct Policy Search (EMODPS) framework formalized by Giuliani et al. (2014) and Giuliani et al. (2016). The EMODPS framework has, since its inception, garnered a broad array of applications including multi-sector reservoir operations (Desreumaux et al., 2018; Biglarbeigi et al., 2018; Quinn et al., 2019) and energy systems (Giudici et al., 2019). Building off of these successes, a 4-objective stochastic application of EMODPS is formulated to create dynamic and adaptive financial risk management strategies for SFPUC's snowpack-dependent hydropower generation. The financial portfolio planning policies abstract SFPUC as a state-aware agent planning a 20-year horizon of decisions for the optimal value of hedging contracts that the utility should buy each year and the amount of money that should be withdrawn or deposited into the utility's reserve fund. A 20-year time horizon is utilized to capture both a typical planning period for a utility and system dynamics in response to dry and wet years that tend to persist beyond single years. The resulting policies take key portfolio planning input concerns and recommend a balanced set of actions. Policy performance can be monitored and re-optimized if performance substantially deviates from what was projected.

While offering many advantages, the success of the EMODPS approach for addressing SFPUC's financial risks is highly contingent on the ability of the chosen MOEA to solve the highly challenging 4-objective stochastic formulation. Over the past decade of theoretical developments, MOEAs can now be classified into four categories of methods: (1) Pareto dominance techniques, (2) decomposition-based population search, (3) reference vector and reference point directed search, and (4) hyper-heuristics. Pareto dominance-based MOEAs such as the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002), sort a population into a sequence of fronts that are ranked to determine the next generation population. This class of algorithms faces challenges when solving multi-objective problems that have many objectives. As the number of objectives grow, so does the potential for solutions to be non-dominated with respect to one another which results in a lack of selection pressure when attempting to drive the search towards convergence (Palakonda and Mallipeddi, 2017). Decomposition-based approaches, such as a multi-objective evolutionary algorithm based on decomposition (MOEA/D) (Zhang and Li, 2007) use an approach such as a Tchebycheff decomposition to break down a multi-objective problem into populations of single-objective subproblems. Decomposition algorithms do not utilize a dominance-based approach and therefore tend to scale more effectively with increasing objective counts. However, as the number of objectives increases, the number of corresponding subproblems grows

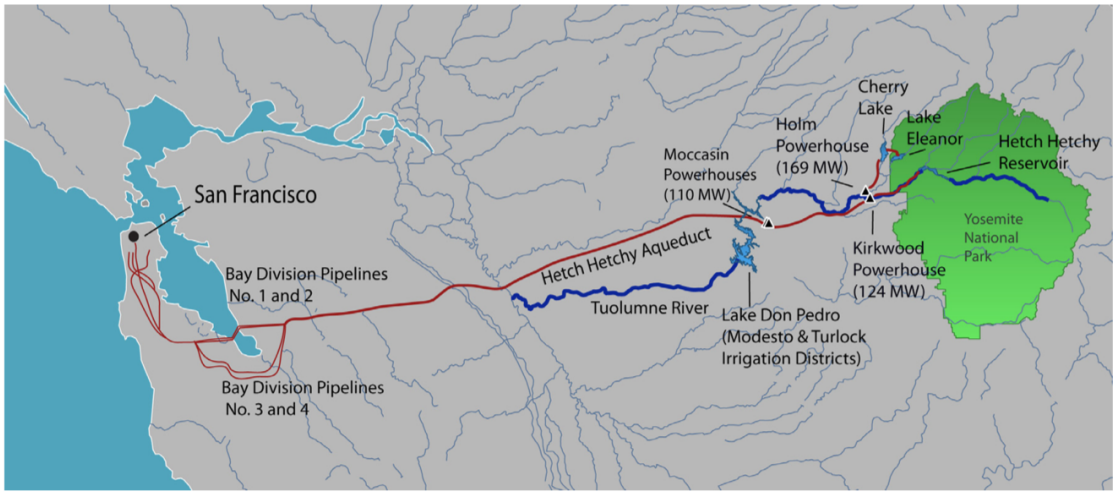


Fig. 1. The high-altitude reservoirs of Hetch Hetchy, Cherry Lake, and Lake Eleanor drive the hydropower turbines at the Holm, Kirkwood, Moccasin, and Moccasin Low-Head Powerhouses to generate electricity for the city of San Francisco and other retail customers. Some surplus power is sold to irrigation districts, and SFPUC also buys and sells power on the wholesale market (Map adapted from San Francisco Public Utilities Commission (2018)).

erence point or reference vector based-algorithms. These algorithms such as the Non-dominated Sorting Genetic Algorithm III (NSGA-III) (Deb and Jain, 2014) and the Reference Vector-Guided Evolutionary Algorithm (RVEA) (Cheng et al., 2016), target search to a reduced finite set representation of problems' tradeoffs or, if appropriate, a specific sub-region of focus in the objective space to reduce computational overhead. All of the aforementioned algorithms are non-adaptive and exploit largely the same suite of search operators which can cause them to be limited in generalizing to new classes of problems due to their utilization of fixed population sizes and static operators (Burke et al., 2013). Hyper-heuristic approaches and frameworks were created to automatically generate cooperative combinations of alternative search heuristics or operators as a means of extending their applicability across a wide variety of problems. The Borg MOEA (Hadka and Reed, 2013) implements an adaptive population sizing strategy to escape local optima and maintain diversity and can adaptively adjust its utilization of recombination operators to favor those that maximize its progress during search.

Studies to date have benchmarked MOEAs' abilities to approximate the tradeoffs of suites of highly challenging mathematical test functions and water resources applications (Hadka and Reed, 2012; Reed et al., 2013; Ward et al., 2015; Zatarain Salazar et al., 2016), but have not comprehensively considered the latest innovations in MOEAs. Moreover, there remains a dearth of studies focused on addressing the ability of state-of-the-art MOEAs to solve high dimensional financial risk management problems, especially for complex western US water and energy systems. The SFPUC benchmarking application therefore is a valuable test case to provide insight into advancing our understanding of the capabilities of state-of-the-art MOEAs to represent the tradeoffs for these complex systems. The SFPUC test case is characterized by a high-dimensional decision space (36 decision variables) as well as a combination of risk-neutral (mean-focused) and risk-averse stochastic objectives. The objectives are nonlinear, non-convex, and discretely discontinuous, yielding a Pareto front that has a severely disjoint geometry that has not been represented in previous water resources benchmarking studies. The policy representations for the financial decisions are inherently complex due to the utilization of constraints and multiple sets of informational inputs to inform the annual decisions that comprise the overall optimized policy. Thus, this study broadens the suite of MOEAs tested relative to previous benchmarking studies and introduces a new application of algorithms: NSGA-II, MOEA/D, the Borg MOEA, NSGA-III, and RVEA. More broadly, this study highlights inherent mathematical challenges posed in balancing the tradeoffs in coupled water and energy financial risk management problems.

2. San Francisco Public Utilities Commission benchmark

As discussed in the introduction, the purpose of this study is to contribute a comprehensive diagnostic benchmarking study to assess the ability of modern MOEAs to solve challenging financial risk mitigation problems that are emerging for coupled water-energy systems given growing hydroclimatic uncertainties. The assessment is centered around optimizing a portfolio of annual financial instruments for Hetch Hetchy Power Enterprise, the electricity division of SFPUC, represented in Fig. 1. The Hetch Hetchy Power enterprise operates three high altitude reservoirs: Hetch Hetchy Reservoir, Cherry Lake, and Lake Eleanor, in the headwaters of the Tuolumne River, that are fed by snowmelt from the central Sierra Nevada mountain range. Water from these reservoirs drives hydropower turbines at the Holm, Kirkwood, Moccasin, and Moccasin Low-Head Powerhouses. The generated power is sold to customers such as the San Francisco International Airport, municipal buildings in San Francisco, and a small number of other retail customer classes. Surplus power is often sold to irrigation district customers at a lower fixed rate and on the wholesale market (Western Systems Power Pool) at prevailing market rates. When hydropower production is insufficient to meet firm contractual obligations to retail customers, SFPUC must purchase the deficit on the wholesale market San Francisco Public Utilities Commission (2016).

2.1. Overview of the Hetch Hetchy financial risk model

The financial stability of SFPUC's Hetch Hetchy Power Division is tied to the variability of snowpack in the Sierra Nevada mountain range. Therefore, the utility could benefit from implementing financial risk management tools that help to hedge against this hydrologic uncertainty and stabilize inter-annual hydropower revenue. Hamilton et al. (2020) introduce a snow water equivalent (SWE) based index contract, termed a contract for differences (CFD), as a financial

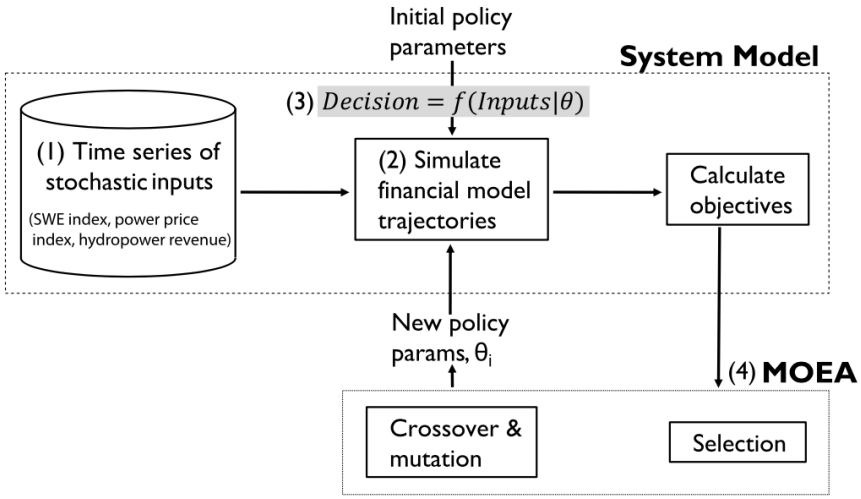


Fig. 2. Schematic of the evolutionary multi-objective direct policy search (EMODPS) approach. Stochastic inputs feed into the system model that simulates annual utility operations. The utility's financial decisions are represented as policies whose parameters are optimized by an MOEA. Figure adapted from Giuliani et al., 2016.

financial risk, especially when used in conjunction with an appropriately sized reserve fund.

SFPUC's use of CFD contracts in tandem with a reserve fund is optimized using an evolutionary multi-objective direct policy search (EMODPS) framework that utilizes closed-loop feedback and state information to inform multi-year sequences of optimal financial portfolio decisions. Fig. 2 shows a schematic of the stochastic EMODPS framework for the SFPUC benchmarking test case. The framework has four main components: (1) stochastic scenario sampling for uncertain simulation model inputs, (2) a coupled hydropower and financial risk simulation model, (3) candidate parameterized policies (or rules that guide state-aware financial decision sequences, and (4) an external MOEA that searches the space of candidate policies for Pareto-optimal solutions that characterize the financial risk tradeoffs for the SFPUC system.

The stochastic scenario sampling supports Monte Carlo simulations that account for three primary sources of uncertainty: snow water equivalent (SWE), power price indices, and hydropower revenue. Each sample of these factors are input into the system model which simulates annual utility operations and estimates revenue dynamics. The EMODPS formulation of the SFPUC system informs the utility's two major financial decisions every water year that dictate their annual cash flow. The first decision entails determining how much money the utility should deposit or withdraw from the reserve fund at the end of the year, after revenues and contract payouts or payments are observed. Then, the utility must determine the value of the CFD contract that it should enter into for the next water year. As illustrated in Fig. 2, these decisions are formulated as decision policies whose parameters, θ, represent the decision variables of the optimization problem. The MOEA, shown below the system model in Fig. 2, optimizes the policy parameters to approximate the Pareto front (or tradeoffs) across four financial objectives. In this study, we carefully benchmark how well state-of-the-art MOEAs solve the EMODPS formulation of the SFPUC test case, which is representative of the class of financial risk management problems that snow-dominated hydropower utilities are facing.

Each of the core components of the Hetch Hetchy EMODPS problem formulation are presented in greater detail in the remainder of Section 2.

2.2. Stochastic scenarios

For the SFPUC test case, the stochastically generated exogenous inputs that characterize each state of the world (SOW) are a SWE index, data to model key trends while better capturing rare extreme conditions that may not be observed in historical records. Considering only historical SWE, for instance, would severely underestimate the impacts of hydrologic variability and extremes on SFPUC operations and lead to myopic solutions that do not perform well if conditions arise that have not been observed historically.

Table 1
Stochastic inputs, ϵ_t, into the simulation model.

Stochastic Input	Variable	Units
SWE Index	ϵ_t^S	m
Power Price Index	ϵ_t^P	$/MWh
Hydropower Revenue	ϵ_t^R	$

Synthetic SWE, ϵ_t^S, is the amount of water that is stored in snowpack that the utility can use to produce hydropower. It is generated by first fitting historic February 1st and April 1st SWE measurements, which are available for all years from 1952 to 2016, excluding 1963, to gamma distributions. Then, a Gaussian copula is fit to capture the correlation between the months and used to generate synthetic February 1st and April 1st SWE measurements. A weighted average of the two measurements is used to develop an index that accounts for the relative importance of each of the months to capturing the timing of hydropower production. The SWE index is used to determine the net CFD payout. The payout function, h, illustrated in Fig. 3, dictates a contract payout or payment, termed c. If SWE measurements are below a specified threshold ($\epsilon_t^S < 0.63$ m), the utility receives compensation ($c_t = h(\epsilon_t^S) > 0$). However, for this compensation, the utility will make payments to the contract seller ($c_t = h(\epsilon_t^S) < 0$) in years when SWE measurements are above the threshold ($\epsilon_t^S > 0.63$ m).

The power price index, ϵ_t^P, represents the utility's best guess of the generation-weighted power price for the upcoming water year, $t+1$. The generation-weighted average power price for a water year, t, is shown in Eq. (1).

$$\bar{P}_t^{gen-wt} = \frac{1}{12}\frac{\sum_{m=1}^{12}\bar{G}_m P_{m,t}}{\sum_{m=1}^{12}\bar{G}_m} \quad (1)$$

In Eq. (1), $\bar{G}_m$ represents the average excess generation sold to the wholesale market for a given month m and $P_{m,t}$ represents the wholesale power price for month m in water year t, generated synthetically accord-

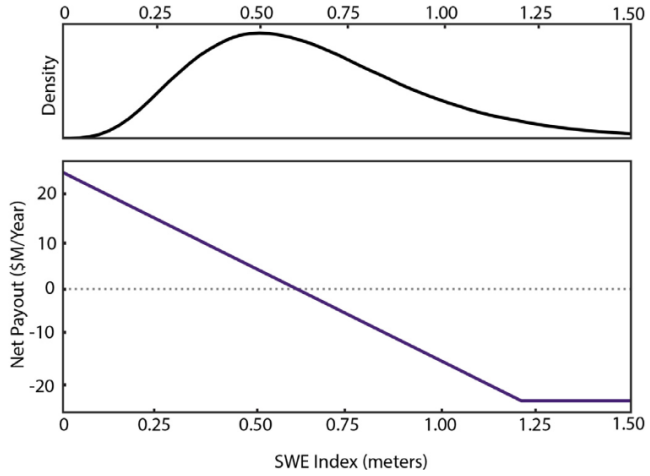


Fig. 3. (top) Probability density for the SWE index using a weighted average of February 1, and April 1, observations. (bottom) Contract payout function, *h*. When stochastically generated SWE is below 0.63 m, the utility receives a payout. If SWE is above 0.63 m, the utility makes payments to the contract seller.

t can then be defined in Eq. (2)

$$\epsilon_t^P = \tilde{P}_{t+1}^{gen-wt} = \beta_0 + \beta_1 \tilde{P}_t^{gen-wt} + \beta_2 P_{12,t} \quad (2)$$

where $\tilde{P}_{t+1}^{gen-wt}$ is the generation-weighted average power price for the upcoming water year, as predicted by linear regression where the predictors are $\tilde{P}_t^{gen-wt}$, the current year's generation-weighted average power price, and the power price in September of the current water year, $P_{12,t}$. The β_i are the coefficients of the linear regression.

The last stochastic input is yearly hydropower revenue, ϵ_t^R, which is calculated from synthetic hydropower generation. Historical hydropower generation is first fit to a series of linear models that capture dependence on snowpack based on the month of the year. An autoregressive (AR) model is fit to the de-seasonalized residuals of the linear models. The residuals of the AR model resemble a normal distribution. Thus, a synthetic time series of hydropower generation can then be generated by first sampling points from a normal distribution of the same variance and running these points through the AR model to obtain synthetic residuals. These residuals are then fed into the linear models in order to produce a hydropower generation time series that maintains correlation with SWE. The synthetic revenue, ϵ_t^R, is determined by combining the synthetic hydropower generation with a financial model. First, SFPUC must satisfy the demand of their retail customer base, which includes the San Francisco International Airport and government buildings. Then, if hydropower generation is in excess of the retail demand, a portion of the power is sold to irrigation district customers. Any additional surplus generation is sold into the Western Systems Power Pool (wholesale power market) at prevailing prices, which vary stochastically. In months where hydropower generation is insufficient to meet retail electricity demand, the utility must purchase the balance on the wholesale power market.

For the SFPUC test case, 1000-member ensembles of 20-year sequences of the exogenous drivers are generated. The hydropower and SWE inputs are generated concurrently in order to preserve historical correlation while the power price input is generated independently using a seasonal autoregressive-moving average (SARMA) model. A 20-year simulation period is chosen because it represents a typical planning period for a utility and can capture system dynamics in response to dry and wet years which tend to persist beyond single years. stochastic inputs can be found in Figure S1 of the Supporting Information to give readers an indication of the statistical characteristics of each input. Some limitations of this approach include that the independent generation of power price neglects to take into consideration the potential correlation between hydrology and power price. Furthermore, the stochastic inputs are generated as stationary stochastic processes informed by limited time series of historical data. In reality, longer records capturing more extreme events would potentially add more stochastic variability and pose an even larger challenge to modern MOEAs.

2.3. Defining financial flows and decision policies

The next stage of the EMODPS framework uses the synthetically generated stochastic samples and current state information to inform the annual financial decision policies for the utility. These candidate decisions policies are parameterized and direct policy search (DPS) is employed to discover the solutions that comprise SFPUC's optimal financial tradeoffs (i.e, the Pareto front). More formally, the policies are generally represented by a family of functions (e.g. linear, piecewise linear, radial basis functions, artificial neural networks) whose parameters, θ, are then optimized with respect to an objective vector, J. In the SFPUC test case, the component policies are formulated using Gaussian radial basis functions (RBFs) to provide the flexibility necessary to represent complex financial portfolio decision dynamics while also providing flexibility for optimizing heterogeneous performance objectives.

SFPUC's cash flow for a given year, y_t, can be broken down into three stages in Eq. (3), represented by superscripts: s_1, s_2, and s_3 respectively. At the end of a water year, September 30, the utility enters a hedging contract and proceeds through the next year, generating hydropower. The evaluation of the cash flow and all financial decisions takes place at the end of the year. The first stage of the cash flow, $y_t^{s_1}$, is the value of the revenue received by the utility from the year's hydropower generation, represented by the stochastic input, ϵ_t^R. The second stage of the cash flow, $y_t^{s_2}$, adds in the contract payout received for the year. The payout is represented as the value of the hedging contract, u_{t-1}^H, determined at the end of the previous water year, multiplied by the contract payout for that year, c_t. The final stage of the utility's cash flow, $y_t^{s_3}$ or u_t^F is the utility's first decision for the year. It is determined by either adding a withdrawal to the second stage of the cash flow or subtracting a deposit. A withdrawal or a deposit is represented by v_t. A value of $v_t < 0$ represents that a deposit was made from the current cash flow into the reserve fund while a value of $v_t > 0$ represents that a withdrawal was made from the reserve fund to be added into the cash flow. The utility's second and final decision of the year is to determine the value of the hedging contract, u_t^H, to enter into for the upcoming water year. The process is repeated for a T year simulation period.

$$\mathbf{y}_t = \begin{bmatrix} y_t^{s_1} \\ y_t^{s_2} \\ y_t^{s_3} or(u_t^F) \end{bmatrix} = \begin{bmatrix} \epsilon_t^R \\ y_t^{s_1} + u_{t-1}^H c_t \\ y_t^{s_2} + v_t \end{bmatrix} \quad (3)$$

The two decisions that the utility must make during the year are represented by the control variables, u_t^H and u_t^F, whose superscripts, H and F designate a Hedge and Final Cash Flow policy respectively. The parameters that define each policy will carry the corresponding superscript henceforth and a further summary of notation can be found in Appendix A. The control decisions are outputs of the overall policy $\mathcal{P}$ which is a function of some informational inputs (state variables) and of the parameter vectors θ_H and θ_F as shown in Eq. (4).

$$\mathcal{P}(u_t^H(\theta_H), (u_t^F(\theta_F)) \quad (4)$$

Eq. (5).

$$u_t^H = \phi^{HC}\left(\phi^{HN}\left(a^H + \sum_{i=1}^{n} w_i^H \exp\left(-\sum_{j=1}^{m^H}\left(\frac{(x_t^H)_j - c_{i,j}}{b_{i,j}}\right)^2\right)\right)\right) \quad (5)$$

Eq. (5) can be decomposed into an intermediary normalized contract slope, $\widetilde{u}_t^H$, before the normalization and constraint functions, ϕ^{HN} and ϕ^{HC}, are applied to obtain the unnormalized contract value, $u_t{}^H$.

$$\widetilde{u}_t^H = a^H + \sum_{i=1}^{n} w_i^H \exp\left(-\sum_{j=1}^{m^H}\left(\frac{(x_t^H)_j - c_{i,j}}{b_{i,j}}\right)^2\right) \quad (6a)$$

$$u_t^{H*} = \phi^{HN}(\widetilde{u}_t^H) \quad (6b)$$

$$u_t^H = \phi^{HC}(u_t^{H*}) \quad (6c)$$

In the intermediary Eq. (6a), a^H is an applied constant shift, and w_i^H, $c_{i,j}$, and $b_{i,j}$ are the weights, centers, and radii of n RBFs that represent the hedging contract slope policy, and $(x_t^H)_j$ is the value of the j^{th} of m^H input characteristics at time t. The $m^H = 1$ informational input into this policy is the current normalized balance in the reserve fund, denoted by $\widetilde{f}_t$. Therefore, for a given water year t, $x_t^H = [\widetilde{f}_t, 0, 0]$. Because there is only one input to this policy, but three inputs to the final cash flow policy, x_t^F, the number of inputs, m, must be three. Therefore, two of the inputs are inactive and set to two zeros at the end of x_t^H.

Then, shown in Eqs. (6b), and (6c) the ϕ^{HN} function unnormalizes the intermediary policy and ϕ^{HC} applies a constraint transformation. The unnormalization function, ϕ^{HN}, scales the contract value to the range $[0, 2k^H]$ as shown in Eq. (7).

$$\phi^{HN}(\widetilde{u}_t^H) = k^H \widetilde{u}_t^H \quad (7)$$

The functional form of ϕ^{HC} is represented in Eq. (8).

$$\phi^{HC}\left(u_t^{H*}\right) = \begin{cases} u_t{}^{H*}, & \text{if } u_t{}^{H*} \geq k^H d^H \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

This constraint dictates that the contract slope before constraints are applied, u_t^{H*}, must be greater than or equal to some threshold constant, $k^H d^H$, otherwise the contract is not purchased. The threshold variable, d^H, is set during optimization.

Given how the formulation implements RBFs and utilizes a different subset of inputs for each decision, the hedging policy is implicitly influenced by the centers and radii of the full set of inputs even after they are set to zero. This is an important consideration when trying to interpret how an input into the hedging policy will map to the policy's output. The current representation of the RBF network does effectively mitigate financial risks dynamically and adaptively while yielding a challenging benchmarking problem that can lead to different families of solutions in the hedging objective space. Furthermore, an alternative formulation that utilizes separate RBF networks for each decision was tested and the similarity of the resulting Pareto front of solutions between the two formulations is displayed in Figure S5 of the Supporting Information.

The last step in the utility's annual cash flow is to decide if money should be withdrawn or deposited into the reserve fund. This value is determined implicitly by first fitting a policy that determines the utility's final cash flow, u_t^F. This decision is represented in Eq. (9).

$$u_t^F = \phi^{FO}\left(\phi^{FI}\left(\phi^{FN}\left(a^F + \sum_{i=1}^{n} w_i^F \exp\left(-\sum_{j=1}^{m^F}\left(\frac{(x_t^F)_j - c_{i,j}}{b_{i,j}}\right)^2\right)\right)\right)\right) \quad (9)$$

Eq. (9) can be decomposed into an intermediary normalized final cash flow, $\widetilde{u}_t^F$, before a normalization function ϕ^{FN}, inner constraint ϕ^{FI}, and an outer constraint ϕ^{FO} are applied.

$$\widetilde{u}_t^F = a^F + \sum_{i=1}^{n} w_i^F \exp\left(-\sum_{j=1}^{m^F}\left(\frac{(x_t^F)_j - c_{i,j}}{b_{i,j}}\right)^2\right) \quad (10a)$$

In Eq. (10a), a^F is an applied constant shift, and w_i^F, $c_{i,j}$, and $b_{i,j}$ are the weights, centers, and radii of n RBFs that represent the annual cash flow policy and $(x_t^F)_j$ is the value of the j^{th} of m^F inputs at time t. The informational inputs corresponding to this decision are the normalized current power price index, $\widetilde{\varepsilon}_t^P$, the most recent normalized reserve fund balance, $\widetilde{f}_{t-1}$, and the current normalized cash flow at this point in the year, $\widetilde{y}_t^{s_2}$. Since the decisions at this point are being made prior to updating the reserve fund, the balance from the year $t-1$ is used. Therefore, the $m^F = 3$ policy inputs for a given year, t, are $x_t^F = \left[\widetilde{f}_{t-1}, \widetilde{\varepsilon}_t^P, \widetilde{y}_t^{s_2}\right]$.

The normalized final cash flow, $\widetilde{u}_t^F$, is unnormalized to the scale $[0, 2k^R]$, using a constant unit value for revenue, k^R, as shown in Eq. (11).

$$\phi^{FN}(\widetilde{u}_t^F) = k^R \widetilde{u}_t^F \quad (11)$$

We apply both an inner constraint, ϕ^{FI}, and an outer constraint ϕ^{FO}, to ensure that the resulting policy is feasible. The function, ϕ^{FO}, defined in Eq. (12), ensures that the reserve fund never exceeds the maximum allotted size represented by $k^F d^F$, where d^F is a decision variable that is set during optimization. That is, a deposit cannot be larger than the available space left in the reserve fund.

$$\phi^{FO}(u_t^{F*}) = \begin{cases} y_t^{s_2} + f_{t-1} - k^F d^F, & \text{if } (y_t^{s_2} + f_{t-1} - u_t^{F*}) > k^F d^F \\ u_t^{F*}, & \text{otherwise} \end{cases} \quad (12)$$

The inner function, ϕ^{FI}, constrains the amount of money that can be withdrawn from or deposited into the fund balance and is defined in Eq. (13).

$$\phi^{FI}(u_t^{F*}) = \begin{cases} \min\left(u_t^{F*}, y_t^{s_2} + f_{t-1}\right) & \text{if } v_t \geq 0 \\ \max\left(u_t^{F*}, y_t^{s_2} - \max\left(y_t^{s_2}, 0\right)\right) & \text{if } v_t < 0 \end{cases} \quad (13)$$

The first case of Eq. (13) represents the constraint that if any withdrawal, represented as $v_t \geq 0$, is made, it cannot be greater than the balance in the reserve fund. The second case enforces that deposits can only occur when the incoming cash flow is positive and that the deposit cannot be larger than the cash flow.

Finally, the amount of money that has been withdrawn or deposited into the fund, v_t, can be determined by subtracting the unnormalized cash flow before the withdrawal or deposit was made from the final cash flow, as shown in Eq. (14).

$$v_t = u_t^F - y_t^{s_2} \quad (14)$$

The decision variables for each of the decisions are represented by the parameter vectors θ_H and θ_F. Each vector is composed of a constant shift parameter, the RBF centers, radii, and weights, and a variable denoting either a minimum contract value threshold, d^H, or a maximum reserve fund balance threshold, d^F. These parameter vectors are shown in Eqs. (15) and (16).

$$\theta_H = \left[a^H, w_i^H, c_{i,j}, b_{i,j}, d^H\right] \quad (15)$$

$$\theta_F = \left[a^F, w_i^F, c_{i,j}, b_{i,j}, d^F\right] \quad (16)$$

The weights are constrained to be positive ($w_i^H \geq 0$ and $w_i^F \geq 0$) and sum to unity ($\sum_{i=1}^{n} w_i^H = 1$ and $\sum_{i=1}^{n} w_i^F = 1$). The centers and radii of each RBF are constrained between -1 and 1 and 0 and 1 respectively ($-1 \leq c_{i,j} \leq 1$, $0 \leq b_{i,j} \leq 1$) and are shared across the two decisions. Combining the policy parameters from the two decisions in Eqs. (15) and (16) results in the overall vector of decision variables Θ.

Four RBFs were chosen to represent each decision. Following the recommendations from prior EMODPS work, we set the number of RBFs equal to one more than the sum of the number of inputs and outputs (Giuliani et al., 2016). In our study, the first decision requires one input for one output, which corresponds to three RBFs. The second decision requires three inputs for one output, which corresponds to five

to $2n(m+1)+4=36$ decision variables for the overall policy vector Θ. Here, $m = 3$ and is set to $max(m^H, m^F)$, the maximum number of policy inputs between the two decisions. The optimal policy parameters, Θ^*, in Eq. (17), are solved for with respect to the objectives defined in Section 2.4.

$$\Theta(\theta_H, \theta_F) = [(c_{1,1}, \ldots, c_{n,m}), (b_{1,1}, \ldots, b_{n,m}), (w_1^H, \ldots, w_n^H, w_1^F, \ldots, w_n^F), (a^H, a^F), (d^H, d^F)] \tag{17}$$

2.4. Summary of financial performance objectives

The SFPUC benchmarking problem explores tradeoffs between four objectives that capture the utility's financial interests and stability. Each evaluation of the system's objectives are modeled over a simulation time horizon, T, of 20 years and across 1000 Monte Carlo simulations of the key uncertainties vector: $[\boldsymbol{\varepsilon}_t^S, \boldsymbol{\varepsilon}_t^P, \boldsymbol{\varepsilon}_t^R]$. The resulting matrix of 1000 samples of the vector of key uncertainties is denoted $\boldsymbol{\varepsilon}$.

Maximize expected annualized final cash flow (AnnualizedCash)

SFPUC wants to maximize their annualized final cash flow to meet their annual fixed costs which consume, on average, over 90% of their yearly revenue (Hamilton et al., 2020). As formulated in Eq. (18) below, this objective maximizes the expected value of the final cash flow, u_t^F, experienced in a given year, t, over a T-year period. This objective maximizes the sum of all discounted cash flow over the T years and the present value of the reserve fund, f_T, at the beginning of year $T+1$. This sum is divided by an annualization factor, using a discount rate of $r^A = 0.96$. The expectation operator, E_ε, calculates the expectation of the objective over the 1000 Monte Carlo simulations of the key uncertainties vector, $\boldsymbol{\varepsilon}$.

$$J^{AnnualizedCash}\left(u^F_{t\in(1,\ldots,T)}, f_T\right) = E_\varepsilon\left[\frac{1}{\sum_{t=1}^T (r^A)^t}\left(\sum_{t=1}^T \left((r^A)^t u_t^F\right) + (r^A)^{T+1}\left(f_T\right)\right)\right] \tag{18}$$

Maximize expected minimum final cash flow (MinCash)

SFPUC also may be interested in maximizing the worst-case final cash flow that they could receive in any given year to further hedge against a situation where they cannot pay their fixed costs. This objective is formulated to maximize the expected value of the minimum final cash flow, u_t^F, attained over the T years and the expectation of the objective is calculated over the 1000 Monte Carlo simulations of the key uncertainties vector, $\boldsymbol{\varepsilon}$.

$$J^{MinCash}\left(u^F_{t\in(1,\ldots,T)}\right) = E_\varepsilon\left[\min_{t\in(1,\ldots,T)}\left[u_t^F\right]\right] \tag{19}$$

Minimize expected maximum hedging frequency (Hedge)

SFPUC and utilities would prefer to limit buying into portfolios that utilize frequent hedging in order to limit transaction costs and additional fees associated with writing contracts. Therefore, this objective, formulated in Eq. (20) is implemented to minimize the expected hedging frequency across the 1000 Monte Carlo simulations of the key uncertainties vector, $\boldsymbol{\varepsilon}$. For any given stochastic sample, the indicator function, $\mathbf{1}_{u_t^H}$, defined in Eq. (21), returns a 1 if a non-zero hedging contract was used at any point over the T years, and zero otherwise.

$$J^{Hedge}\left(u^H_{t\in(0,\ldots,T-1)}\right) = E_\varepsilon\left[\max_{t\in(0,\ldots,T-1)}\left[\mathbf{1}_{u_t^H>0}\right]\right] \tag{20}$$

$$\mathbf{1}_{u_t^H} = \begin{cases} 1, & \text{if } u_t^H > 0 \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

This objective is, therefore, a measure of the likelihood that a utility will

tinuous decisions has the potential for generating similar features (e.g., see the disjoint water portfolios tradeoffs in Kasprzyk et al. (2009)).

Minimize expected maximum reserve fund balance (Reserve)

As power utilities in the United States are typically regulated by elected or politically appointed Public Utilities Commissions, it may be difficult to gain approval to hold a very a large sum of money in their reserve fund. Therefore, this objective is formulated to minimize the maximum reserve fund balance, f_t, seen over the T years and across the 1000 Monte Carlo simulations of the key uncertainties vector, $\boldsymbol{\varepsilon}$.

$$J^{Reserve}\left(f_{t\in(0,\ldots,T)}\right) = E_\varepsilon\left[\max_{t\in(0,\ldots,T)}\left[f_t\right]\right] \tag{22}$$

The optimal parameters of the financial policies, Θ^*, can now be determined by solving the following multi-objective problem:

$$\boldsymbol{\Theta}^* = \arg\min_{\boldsymbol{\Theta}}\left[-J^{AnnualizedCash}(\boldsymbol{\Theta}), -J^{MinCash}(\boldsymbol{\Theta}), J^{Hedge}(\boldsymbol{\Theta}), J^{Reserve}(\boldsymbol{\Theta})\right] \tag{23}$$

All maximization objectives are multiplied by negative one to convert the optimization into a minimization problem.

3. Methods

3.1. Diagnostic framework

This study contributes a comprehensive diagnostic assessment of the ability of current state-of-the-art MOEAs to solve complex water-energy financial risk portfolio planning problems as represented by the SFPUC benchmarking test case. The diagnostic assessment framework utilized in this study is illustrated in Fig. 4. MOEAs are heuristic search tools that use different types of parameterized search operators to mimic the natural processes of mating, mutation, and selection in order to evolve and improve a population of solutions (Coello et al., 2007). The default parameters of an MOEA tend to be set to values that give the best performance for a specific test instance. However, these values are not necessarily generalizable to other problems. Ideally, MOEAs should perform well under a wide range of parameterizations (Goldberg, 2002). Therefore, the diagnostic framework illustrated in Fig. 4 tests the sensitivity of the MOEAs to their parameterizations by sampling their full feasible parameter spaces using a Latin hypercube sample (LHS). Each sample, signified as a point in the parameter block of Fig. 4, represents one fully specified parameter instance of an MOEA. In this benchmarking study, each instance is used to solve the SFPUC optimization problem defined in Eq. (23) to obtain a Pareto approximate set of solutions. Given that MOEAs are stochastic global search tools that can be strongly sensitive to their pseudo-random sequences of initial random population generation and probabilistic search operators, each sampled parameterization in the benchmarking framework is run for 25 replicate random seed trials to account for these effects. The quality of the Pareto approximate set is assessed by comparison to a reference set that is obtained by merging the best solutions found across all algorithms. Performance metrics are then calculated with respect to this reference set. These performance metrics can be visualized to assess an MOEA's effectiveness, reliability, and controllability. "Effectiveness" refers to an MOEA's ability to attain high quality approximation sets. An MOEA is considered "reliable" if it can attain these sets with minimal variability across parametric and/or random seed trials. "Controllability" is a measure of the sensitivity of an MOEA to its parameterization. A highly controllable

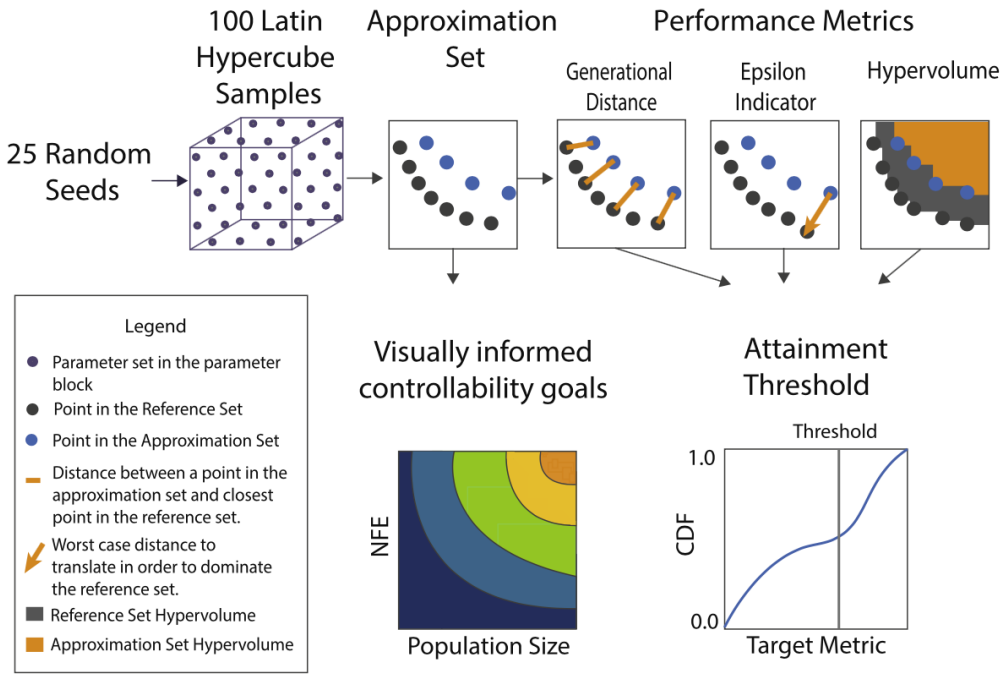


Fig. 4. Diagnostic assessment framework used to evaluate the performance of each MOEA tested in the study (adapted from Zatarain Salazar et al. (2016)). The parameters for each MOEA are sampled across their full ranges using Latin Hypercube Sampling. The approximation sets for each MOEA parameterization are assessed through metrics measuring the convergence, consistency, and diversity of approximation sets. Each MOEA parameter set is run for 25 random seed trials to account for the random initialization of populations and random components in the MOEAs' search.

3.2. Performance metrics

The metrics calculated in this study are generational distance, hypervolume, and additive epsilon indicator. These metrics give an indication of the convergence and diversity of the Pareto approximate set. Convergence indicates how close the approximate set is to the reference set. Diversity indicates how well the approximate set spans across the reference set (Coello et al., 2007).

Generational Distance: Generational distance is a measure of convergence of the approximate set to the best-known reference set. In order to calculate generational distance, the Euclidean distance between each test point and the closest point in the reference set is calculated. Then the average distance is calculated considering all generated Pareto points. Therefore, the goal is to minimize this metric. Generational distance is considered the most basic metric to meet as a near perfect value can be achieved if the reference set contains only one solution that is close to the best approximate reference set. Therefore, this metric is meaningful primarily to identify that when an algorithm cannot meet this metric, it has exhibited poor performance (Van Veldhuizen and Lamont, 1998b; 1998a).

Additive Epsilon Indicator: Additive epsilon indicator is a measure of the worst-case distance that the approximation set has to be translated to dominate the reference set. Thus, the goal is to minimize this metric value. If an approximate set has many gaps, then solutions from farther regions must be translated a large distance to dominate the best known approximation. Thus, this will lead to a high additive epsilon indicator value. This metric is a good measure of an MOEA's consistency, or ability to capture all parts of the Pareto front (Zitzler et al., 2003; Hadka and Reed, 2012; Reed et al., 2013).

Hypervolume: Hypervolume measures the volume of the objective space that is dominated by the approximation set. Hypervolume is normalized relative to the reference set hypervolume when metric results are reported.

3.3. Multi-objective evolutionary algorithms

MOEAs are popular tools for multi-objective optimization of complex, non-linear problems because their population-based approach requires less knowledge of the topology of a problem than their deterministic counterparts. Therefore, they are particularly suited towards and successful in water resources applications which tend to be characterized by non-convexity, stochasticity, nonlinearity, and high dimensionality (Nicklow et al., 2010; Maier et al., 2014). The capabilities of MOEAs have expanded over time to accommodate increasingly relevant and challenging water application problems, a review of which is provided in Reed et al. (2013). While there have been many new MOEA search advances over the last decade, many of the more modern algorithms have not been rigorously tested on their ability to solve challenging real-world problems such as the SFPUC financial risk test case. Consequently, this study contributes a benchmark of five state-of-the-art evolutionary algorithms on the highly non-convex and discrete, four objective financial decision test case. These algorithms are summarized below.

NSGA-II: First proposed by Deb et al. (2002), the NSGA-II is an elitist algorithm that dramatically advanced the capabilities of MOEAs to address challenging problems through three key innovations: elitism, efficient non-domination sorting, and incorporation of a diversity maintenance that does not require a user-specified parameter. The NSGA-II was one of the first algorithms to use the Pareto dominance relationship to search for an entire Pareto front in a single run (Coello et al., 2007). As an elitist algorithm, it first sorts a population, composed of an equal number of parents and children, into a sequence of fronts. Mem-

in the next generation, it employs a crowding distance operator that maximizes diversity by giving priority to solutions in sparser regions of the objective space. NSGA-II's incorporation of elitism helps to prevent non-dominated solutions from being lost through the search, but when used in conjunction with a fixed population size, this greatly limits the algorithm's ability to incorporate new solutions into the population. Furthermore, NSGA-II does not have an archive to store non-dominated solutions and while innovative at the time, its crowding distance operator was later found to have limitations beyond two objectives (Deb et al., 2002). In cases where NSGA-II only finds one front, selection is only based on this crowding operator. Thus, valid solutions may be dropped, resulting in deterioration. Deterioration occurs when an MOEA's solution set contains one or more solutions dominated by another member solution. In the extreme, deterioration can cause an MOEA to diverge away from the Pareto front (Hadka and Reed, 2012). Still, NSGA-II remains the most popular algorithm used today and is an appropriate historical baseline algorithm to include in the diagnostic study.

MOEA/D: Zhang and Li introduced MOEA/D, a decomposition based multi-objective evolutionary algorithm, in 2007. MOEA/D reformulates the multi-objective optimization problem into N single-objective optimization subproblems that are solved simultaneously. The decomposition is performed using methods such a weighted sum, Tchebycheff decomposition, or a Boundary Intersection approach to formulate each of the N sub-problems as a linear or non-linear aggregation of the problem objectives. Each sub-problem is given a different weighting vector to maximize diversity of search that results in N solutions. The next generation population is determined by mating each of the N population members with other members that reside within a pre-defined neighborhood around the point. Therefore, this algorithm solves each optimization problem using information from neighboring subproblems (Zhang and Li, 2007). However, one limitation of this algorithm is the necessity of the user to specify the size of the neighborhood around each solution, which can be subjective. Nevertheless, MOEA/D was chosen as a representative of an emerging class of decomposition-based algorithms (Giagkiozis and Fleming, 2014) and won the 2009 IEEE Congress on Evolutionary Computation (CEC 2009) competition (Zhang and Suganthan, 2009). This study implements the winning version of MOEA/D that utilizes a Tchebycheff decomposition approach and proposes a strategy for allocating the computational resources to different subproblems (Zhang et al., 2009).

The Borg MOEA: The Borg MOEA is a unified optimization framework that represents a class of self-adaptive algorithms whose variational operators are adaptively selected through search based on the problem's local topology. The Borg MOEA contains many novel components that build off of its parent algorithm, ϵ-MOEA (Deb et al., 2005), including implementation of epsilon dominance archiving to maintain non-dominated solutions during the search process and an adaptive population size. It also utilizes adaptive time continuation through epsilon progress, a metric used to monitor for stagnation in the search process and to escape local optima. If the algorithm fails to make progress discovering solutions that dominate members of the archive, it will implement a randomized restart to inject more diversity into a population. The population is emptied and repopulated with all archived solutions and a uniform mutation is applied to archive solutions to fill any remaining spots (Hadka and Reed, 2013). The Borg MOEA's self-adaptive characteristics allow it to be to less sensitive to underlying parameterizations than algorithms without these capabilities and thus has been shown to have applicability across a wide set of problem classes (Hadka and Reed, 2012; Reed et al., 2013).

NSGA-III: In 2014, Deb proposed NSGA-III to bridge the frameworks of NSGA-II and MOEA/D. NSGA-III utilizes the same non-dominated sorting as NSGA-II but implements a different niching strategy that in-

tive. A hyperplane is determined from extreme points and then reference points are evenly spaced across an $(M - 1)$ - dimensional simplex. Reference lines are specified for each reference point and population members are assigned to the closest reference line. If no new population members are associated with the reference vector, the former population member with the closest perpendicular distance to the reference vector is chosen to be added to the new population. If a prospective population member is associated with a vector that already has a member specified, then a random member is picked to move into the next generation population, P_{t+1}. After P_{t+1} is formed, it is then used to create an offspring population, Q_{t+1}, with usual crossover and mutation operators. NSGA-III suffers from the inability to preserve non-dominated solutions due to the lack of an archive, like its NSGA-II counterpart (Deb and Jain, 2014).

RVEA: RVEA is a reference vector-based algorithm similar to NSGA-III and motivated by decomposition-based approaches like MOEA/D. The reference vectors not only can be used to decompose the multi-objective problem into single-objective subproblems but also can also be tuned to target search in a user-preferred region of the Pareto front. RVEA adopts an elitism strategy similar to NSGA-II where a parent population is combined with an offspring population that is generated using traditional crossover and mutation operations. The prospective population is split into N subpopulations by associating each population member with one of N reference vectors. The main new contribution proposed is the implementation of an Angle-Penalized Distance (APD) metric to select which solution member associated with each reference vector will pass on to the next generation of the population. The metric seeks to balance convergence and diversity by taking into account both the distance between a solution and the reference vector along with measuring the acute angle the solution makes with its reference direction. The metric formulation prioritizes convergence early in the search and diversity is emphasized in the later stages of the search (Cheng et al., 2016). As the most modern algorithm in the suite, RVEA has been minimally benchmarked on variety of applications.

4. Computational experiment

As described in the Section 2.3, we introduce a DPS framework for abstracting SFPUC's yearly financial decisions. Gaussian radial basis functions are used to represent well-informed policies that map current utility state information and exogenous inputs to optimal contract values and end-of-the-year cash flow. Each radial basis function has three parameters: a radius, center, and weight. Four radial basis functions are used for this test case that is characterized by 36 decision variables. The candidate policies that inform yearly financial decisions over a 20-year time period are optimized with respect to these decision variables. Five state-of-the-art MOEAs are used to discover optimal policies with respect to the four objectives outlined in Section 2.4. The diagnostic assessment is performed using MOEA Framework, a free and open source Java library that allows users to design, execute, and assess the performance of a variety of popular MOEAs. The following subsections discuss the computational experiments that are executed through MOEA Framework to conduct the diagnostic assessment. First, Section 4.1 elaborates on the procedure for sampling of algorithm parameterizations. Then Section 4.2 describes the process of generating and verifying reference sets.

4.1. Sampling of algorithm parameterizations

The experimental setup for the diagnostic assessment requires testing MOEAs in both their default parameterizations and across 100 Latin hypercube samples from their full feasible suite of parameter ranges (displayed in Table 2). The default instance of each MOEA is first used to

Table 2
Latin hypercube sampling of MOEAs' operators and their associated parameter ranges as well as the MOEAs' default parameterizations. D corresponds to the number of decision variables (36).

	Parameter	LHS range	Default	Algorithms
Crossover	SBX rate	0 -1	1.0	Borg, NSGA-II, RVEA, NSGA-III
	SBX distribution index	0–100	15	Borg
			30	NSGA-II RVEA, NSGA-III
	DE crossover rate	0–1	0.1	All algorithms
	DE step size	0–1	0.5	Borg, 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/D	All algorithms
	PM distribution index	0–100	20	All algorithms
	UM rate	0–1	1/D	Borg
Selection	Neighborhood Size	0-0.2	0.2	MOEA/D
	Delta	0-1	0.9	MOEA/D
	Eta	0-0.02	0.02	MOEA/D
	Injection Rate	0.1-1	0.25	Borg
Population Size		10-250	100	Borg, NSGA-II MOEA/D
Divisions		4-9	8	NSGA-III, RVEA
NFE			200,000	All algorithms

Latin hypercube samples are taken from the range of acceptable parameter values for each MOEA. Each sample is an instance of the MOEA and represented by a single point in the parameter block in Fig. 4. Each MOEA instance is replicated for 25 random seeds and evaluated over 200,000 function evaluations. Archive output and runtime dynamics are reported every 5000 function evaluations to understand how algorithm performance evolves through the search. Performance metrics are calculated from the results of the optimization and then visualized in figures that serve to demonstrate algorithm behavior across search time and parameterization.

4.2. Generation and verification of reference sets

Five MOEAs outlined in Section 3.3 are used to discover optimal policies with respect to the four objectives outlined in Section 2.4. The best solutions that each MOEA finds individually comprise the algorithm's individual reference set. In order to compare algorithms, performance is assessed relative to the best known reference set for the SFPUC test case which is found by merging the best solutions across the algorithms using consistent epsilon sorting. An epsilon precision that dictates numerical precision for each objective must be specified for algorithms that utilize epsilon box dominance. For NSGA-II, this requires transforming point dominance to epsilon-box dominance to ensure consistent comparisons across MOEAs (Hadka and Reed, 2012). The epsilon values are 0.05 for the expected annualized final cash flow ($M/year) and the expected hedging frequency objectives (unitless) and 0.1 for the expected minimum final cash flow ($M) and the expected maximum fund balance objective ($M).

As stated in Section 2.2, the optimized objective values that define each solution in the reference set are averaged across 1000 simulations of 20-year periods defined by three different stochastic inputs: a SWE wider variety of potential worlds that would not be possible if using a fixed ensemble. The validity of the approximate sampling approach is confirmed by re-evaluating the solutions on a larger set of 100,000 independent stochastic samples and verifying that solution performance is stable. The re-evaluated reference sets can be found in the Supporting Information Figure S2. Re-calculating runtime dynamics with respect to the new verified reference set is computationally intractable. Therefore, Sections 5.1–5.4 display results and metrics with respect to the reference set determined from the optimization. Verified overall and default reference sets are only used in Section 5.5 of the results. As further empirical support for the selected sample size, Figures S3 and S4 in the Supporting Information show the resulting Pareto fronts from optimizing the SFPUC test case using the default parameterization of each algorithm and a 500 and 10,000-member sample, respectively. Given the equivalent performance, a 1,000-member sample is chosen for its sufficient convergence and computational efficiency to facilitate the demands of our diagnostic study. Prior studies provide more background on evolutionary optimization under uncertainty (e.g. Smalley et al., 2000, Gopalakrishnan et al., 2003, Chan Hilton and Culver, 2005, Beyer and Sendhoff, 2007, Deb et al., 2009, Kasprzyk et al., 2009, and Kasprzyk et al., 2012).

5. Results

5.1. Analysis of reference sets

Fig. 5 shows the overall best known reference set of Pareto approximate solutions that represent the financial risk tradeoffs for the SFPUC test case. As noted in Section 4.2, these solutions are the result of epsilon non-domination sorting across all trial runs performed for all of the tested algorithms. The theoretical ideal solution is represented by a black star. The SFPUC financial risk problem poses a challenge to

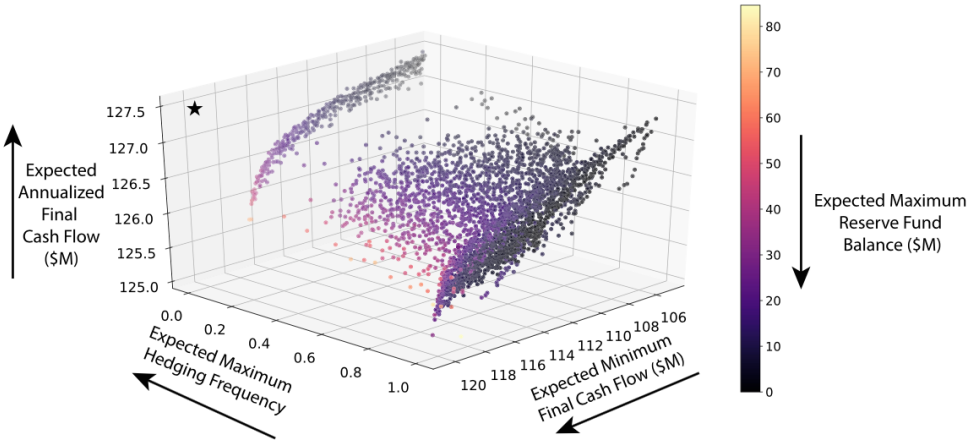


Fig. 5. The overall best known reference set and resulting tradeoffs for the SFPUC test case, attained by merging the best solutions found across all MOEA runs. The ideal point is shown as a black star. The fourth objective, minimizing the expected maximum reserve fund balance, is represented by the color scale (darker is preferred).

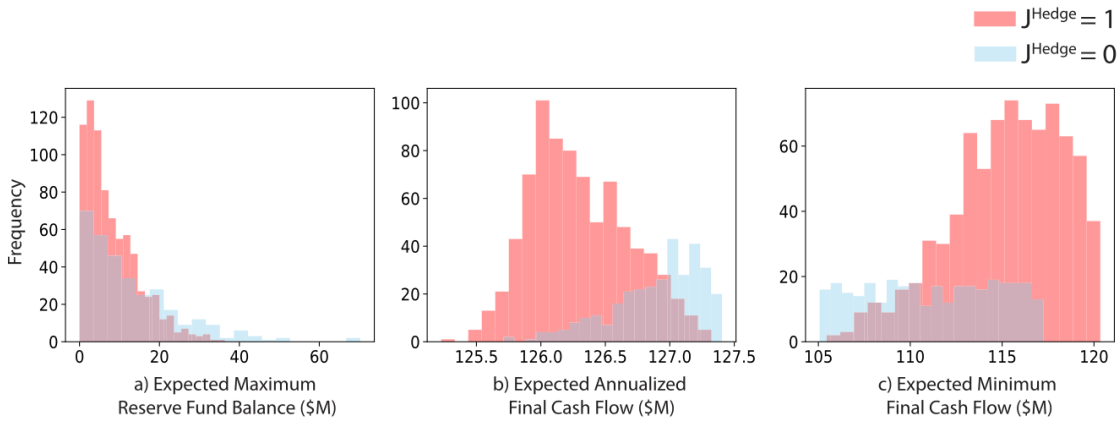


Fig. 6. Histograms of the range of the expected maximum fund balance, expected annualized final cash flow, and expected minimum final cash flow objective values associated with reference set solutions that have either a minimum or maximum value of the expected hedge frequency objective. The purple color indicates where the two histograms overlap. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

expected minimum final cash flow objective and the expected annualized final cash flow objective. The color gradient with lighter colored points in the lower front portion of the plot highlights that increasing expected minimum final cash flow usually necessitates the presence of a larger reserve fund balance. However, the lack of lighter points in the rightmost highest hedging frequency lobe of solutions suggests that a larger reserve fund balance is not necessary if a portfolio structure that utilizes more hedging is implemented.

Formalizing these relationships, the paneled histogram in Fig. 6 shows the range of values for the expected annualized final cash flow, expected minimum final cash flow, and the expected maximum reserve fund balance objectives for the reference set solutions that utilize a minimum and maximum hedging frequency. As seen in Fig. 6a, the minimum hedging frequency solutions implement a wider range of reserve fund balances while solutions that utilize the greatest hedging frequency do not require a reserve fund larger than $40 million. The maximum hedging contract structures reduce the potential self-insurance opportunity costs that SFPUC face when they have to fix a large amount of their money in a reserve fund. The hedging objective is a measure of the likelihood that a utility will enter into at least 1 contract over the 20-year period for any given realization of states of the world. This objective can that there are zones of highly concentrated solutions corresponding to extremes of the hedging frequency objective (i.e., probability equal to 0 or 1). An intermediate objective value between 0 and 1 indicates the fraction of the 1000 statistical 20-year replicate samples in which the solution implemented at least one contract during the planning period. Fig. 6b and c indicate that solutions that utilize a lower hedging frequency on average results in a larger expected annualized final cash flow whereas solutions that utilize a higher hedging frequency see a higher expected minimum final cash flow.

Fig. 7 builds on Fig. 5 by quantifying the reference set contributions for each of the tested MOEAs. The percentages displayed in the bar graph take into consideration solutions that are identified by multiple algorithms as well as solutions identified uniquely by specific algorithms. As seen in Fig. 7, the Borg MOEA contributed 81% of the reference set solutions, while RVEA contributed only 0.8% of reference set solutions. Notably, the reference vector and decomposition MOEAs contributed the least to the reference set. The NSGA-II had the second highest rate of contribution at 8.8% of the reference set solutions. The Borg MOEA and NSGA-II found a similar number of solutions. The Borg MOEA found 2700 non-dominated solutions while NSGA-II found 2000. During the sort to produce the overall reference set, many of NSGA-II's

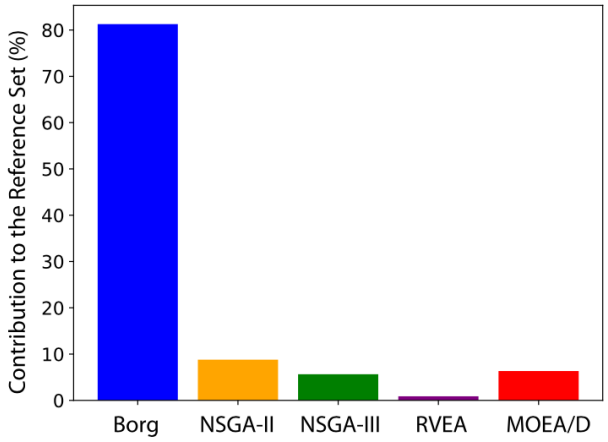


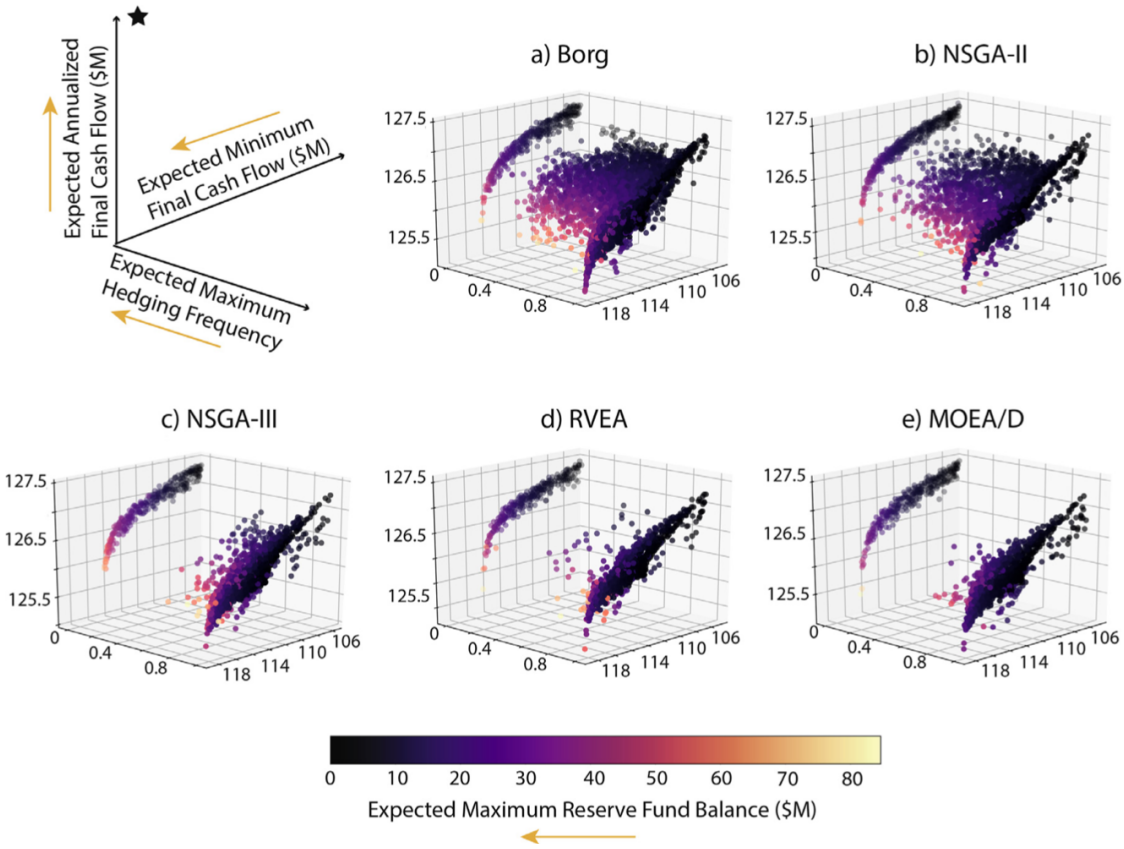
Fig. 7. The percent of the reference set contributed by each algorithm. The Borg MOEA single-handedly produced over 80% of the reference set while the newer algorithms struggled to discover solutions that were not dominated by Borg or NSGA-II.

pressure, and variational operators as more of the problem's decision space is explored through the search (Hadka and Reed, 2013). These characteristics allowed the Borg MOEA to discover more non-dominated solutions in more diverse areas than the other algorithms.

The five panels in Fig. 8 show the best attained reference sets obtained by each algorithm after completing 200,000 function evaluations. Each algorithm's reference set was attained across all of its trial runs. The Borg MOEA (Fig. 8a) and NSGA-II (Fig. 8b) are the only algorithms that obtain reference solution sets that closely replicate the geometry of the overall best known reference set in Fig. 5. The decomposition strategy of MOEA/D (Fig. 8e) as well as the reference point-based search of NSGA-III (Fig. 8c) and RVEA (Fig. 8d) all yield significantly fewer solutions that are biased toward the two extreme lobes of the hedging objective. This can be attributed to the underlying strategies employed by reference point, reference vector, and decomposition approaches to maintain solution diversity. MOEA/D's decomposition-based approach assigns uniform weighting to sub-problems and NSGA-III and RVEA implement uniformly distributed reference vectors and points to aid in search. Both approaches assume that the targeted Pareto front is smooth and continuous. This is not the case for optimization problems with a discontinuous Pareto front such as the SFPUC test case. Hence, if any reference points or vectors cannot locate a new population member, they are disregarded, reducing the density of solutions that can be discovered (Cheng et al., 2016). Potential fixes for the reference point techniques have been proposed in Cheng et al. (2016) and Deb and Jain (2014) that suggest adaptively regenerating or re-locating reference points and vectors rather than removing them completely. A key challenge for these proposed fixes remains: a generalized open-source accessible version of the algorithm codes that is scalable to real-world problems with more than three objectives does not exist at present.

5.2. Algorithmic effectiveness and reliability

Successful tradeoff analyses in decision support applications require that MOEAs are able to effectively discover high quality approximations of the Pareto front. Moreover, they should do so reliably across their



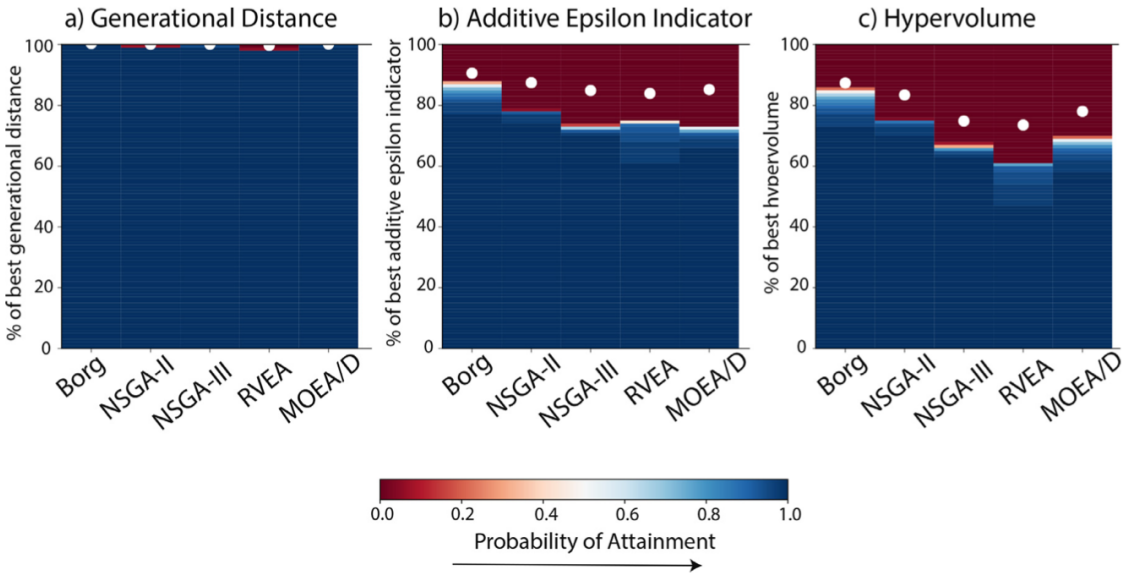


Fig. 9. Attainment plots that capture the best overall metric values achieved by each MOEA (white circles) as well as their success probabilities in attaining increasingly higher quality metric values. The red-to-blue shading indicates the probability that a single run of an MOEA reaches a given percentage of the best possible metric value for a) generational distance b) additive epsilon indicator and c) hypervolume. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

candidate parameterizations and random seed trials. That is, an MOEA should be both effective and reliable for any given run. The attainment plots in Fig. 9 provide a probabilistic assessment of MOEA performance. Each MOEA's best single trial run's overall performance in hypervolume, generational distance, and additive epsilon indicator is designated by a white dot in Fig. 9. For each of the metrics, the probability of attainment is defined as the percentage of an MOEA's trial runs (across all parameterizations and random seed trials) that attain a given level of the best single run metric value. Ideal performance would be indicated by a dark blue bar with a white dot at 100% indicating that the algorithm can attain ideal performance 100% of the time. A white dot below the 100% mark indicates that the algorithm was unable to attain ideal performance for the metric (i.e. achieving the reference set hypervolume or a value of zero for generational distance and additive epsilon indicator).

From Fig. 9a, it is clear that all of the algorithms were able to obtain high levels of performance for generational distance consistently. Generational distance is the easiest of the three metrics to satisfy, as a near perfect value can be achieved if the reference set contains only one solution that is close to the best approximate reference set. Therefore, poor performance in this metric would denote abject failure of an algorithm. Additive epsilon indicator is a more challenging metric, measuring the worst-case distance that the reference set has to be translated to dominate a given approximation set. All of the MOEAs show a degradation in their attainment performance for the additive epsilon indicator. This is not surprising given that additive epsilon indicator is particularly sensitive to gaps where the approximation set is missing solutions that are present in the best known reference set. The visualizations of the best reference sets for each of the MOEAs shown Fig. 8 highlight that several algorithms never identify solutions in the intermediate hedging objective compromise region. The worst case translation distances grow very rapidly given these gaps (see the discussions in Hadka and Reed, 2012; Hadka and Reed, 2013). Fig. 9b indicates that the Borg MOEA achieves the highest level of attainment for the additive epsilon indicator performance in both the additive epsilon indicator and hypervolume. In Fig. 9b and c, it is apparent that MOEA/D, NSGA-II, NSGA-III, and RVEA are able to obtain acceptable metric values for their single best performing trial run but are not likely to obtain these values consistently. The reference point (NSGA-III and RVEA) and decomposition (MOEA/D) algorithms also have the steepest decline in their single trial run attainments overall. In a practical context, the attainment results in Fig. 9 highlight that the SFPUC benchmarking problem is difficult and that all of the algorithms would struggle to reliably attain ideal additive epsilon indicator and hypervolume results with single random seed trials.

5.3. Algorithmic controllability and efficiency

MOEAs should provide consistent performance across any of their candidate parameterizations (i.e., ease-of-use). Often in algorithmic studies, the capabilities of MOEAs are reported after a trial-and-error analysis establishes a single best performing parameterization, especially when default parameterizations struggle (see Deb and Jain, 2014; Cheng et al., 2016; Qi et al., 2019). However, the single best parameterization is typically highly test-case dependent and under emphasizes the difficulty for users in trying to parameterize the algorithms while maintaining high levels of performance.

The control maps, shown in Fig. 10, are used to assess an algorithm's "sweet spot" or how sensitive that algorithm is to its parameterization. Each map shows the hypervolume attained as a function of NFE and proxies of population size, since these parameters consistently have dominant effects on MOEA performance and computational demand (Hadka and Reed, 2012; Reed et al., 2013). The color legend indicates the percent of the expected hypervolume that each algorithm attained by averaging across 25 random seed trials used to evaluate each parameterization. In short, this represents what would be expected from a single trial run of one of the algorithms solving the SFPUC test case. Ideal performance would be indicated by a

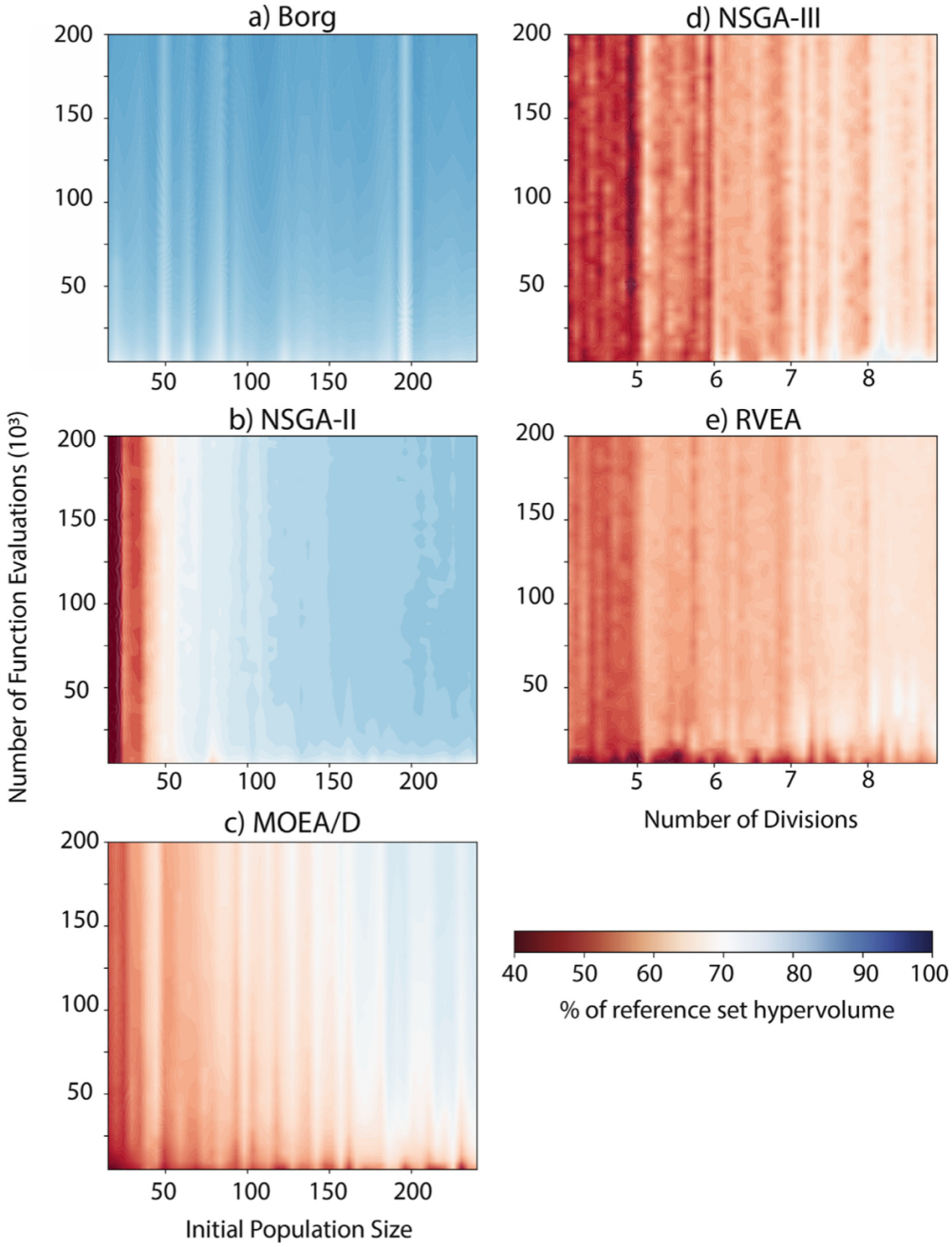


Fig. 10. Hypervolume performance control maps for the SFPUC test case capturing the controllability and efficiency of each MOEA. The color scale represents the percent of the best known global reference set's hypervolume captured by each local 25-seed reference approximation set for each tested parameterization. Ideal performance is shown in zones of dark blue and poor performance is designated by dark red. The control maps are a sub-space based on the full set of Latin hypercube samples of the parameters for each algorithm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The uniform blue shading in the Borg MOEA control map (Fig. 10a) indicates that the algorithm is less sensitive to its parameterization. The Borg MOEA also typically requires less than 25,000 function evaluations to achieve 80% of the best known hypervolume. The control map for NSGA-II (Fig. 10b) displays a high sensitivity to its population size, shown by the abrupt threshold for hypervolume performance, fore have a reduced ability to incorporate newer non-dominated solutions in every generation. Peak performance for NSGA-II is contingent on the user specifying a population size greater than 200, which is not the default specification for the algorithm or apparent to any user in advance. The control map for MOEA/D (Fig. 10c) exhibits a nonlinear sensitivity to both population size and search duration. Al-

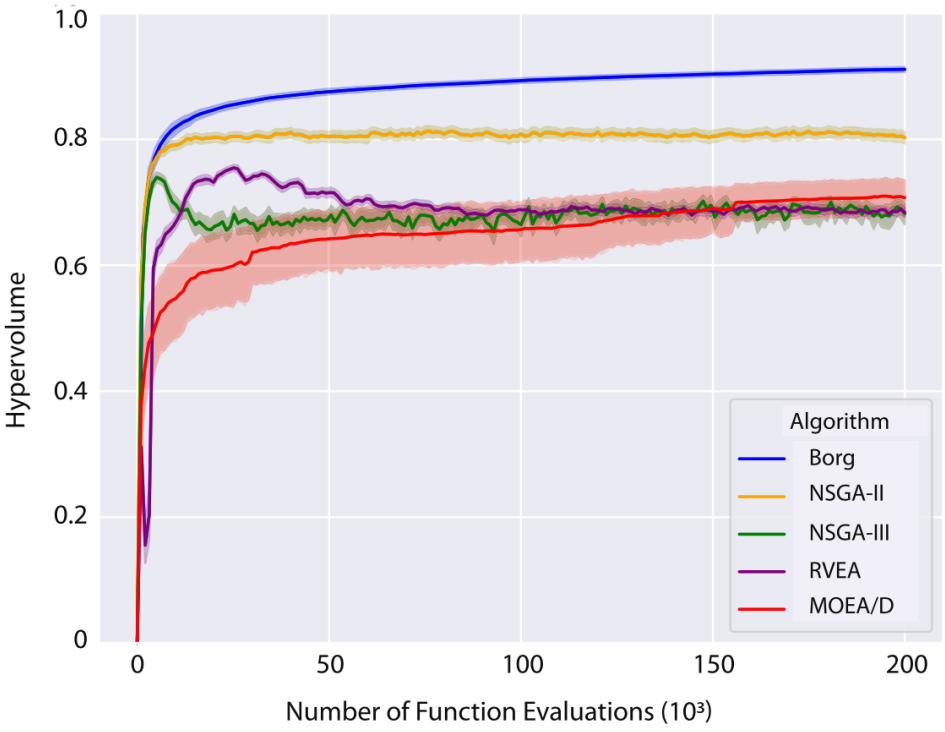


Fig. 11. Runtime dynamics of hypervolume performance across 50 seeds of each algorithm's default parameterization. The solid line represents the mean hypervolume achieved through the search process and the shading bounds the 5th and 95th percentile confidence interval. A hypervolume of 1 with thin shading is preferred (i.e., high performance reliability).

While MOEA/D has been shown to be successful in solving test functions, its sensitivity to the relative scaling of objective functions makes it more difficult to predict if it will be successful for other water resources applications. Moreover, MOEA/D's algorithmic computational time grows very rapidly with population size due to its neighborhood decomposition. Surprisingly, these computational demands can grow to an extent that they are no longer negligible relative to the normal demands for the function evaluations required in search (Hadka and Reed, 2012).

The control maps for NSGA-III and RVEA (Fig. 10d and e) show significant failure to attain an acceptable hypervolume for any range of parameterizations. Furthermore, deterioration in the algorithms is apparent by the color fluctuations along any given vertical segment of the control map. These color variations signify a non-monotonic variation in hypervolume as the number of function evaluations increases (i.e., solutions that are important to hypervolume progress have been lost). Overall, the Borg MOEA displays the strongest performance by consistently achieving high hypervolume over the suite of its parameterizations, highlighting that it would be difficult to make the algorithm fail. This lack of sensitivity to parameterization is due to the Borg MOEA's adaptive search techniques which have shown to be successful for a variety of benchmarking problems (Hadka and Reed, 2012; Reed et al., 2013; Ward et al., 2015; Zatarain Salazar et al., 2016). The Borg MOEA implements operators that can adaptively adjust through the evolutionary process and adaptive population sizing helps the algorithm to be insensitive to initial population size. The non-adaptive algorithms' lack of these mechanisms and consequent strong sensitivity to their parameterizations makes them difficult to use in real-world applications. The control maps emphasize that more effort should be directed towards de-

well for specific parameterizations that are challenging to discover do not help to facilitate this goal.

5.4. Default parameterization runtime and operator dynamics

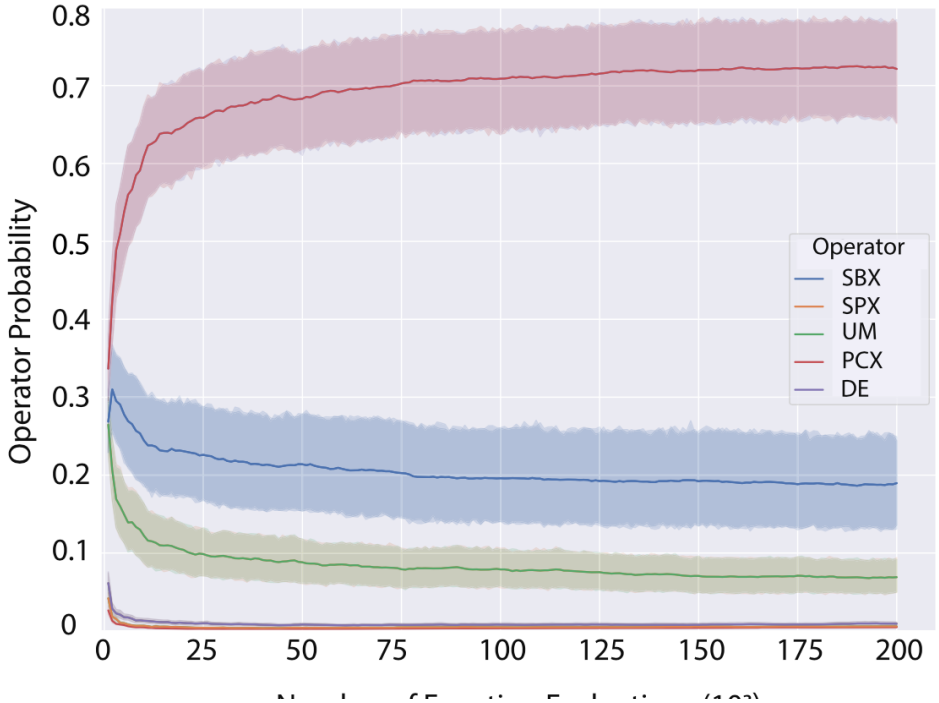
The default runtime analyses in Fig. 11 capture the expected performance of each of the MOEAs using their author-recommended default parameterizations summarized in Table 2. The hypervolume runtime dynamics allows the user to visualize how quickly and reliably each of the algorithms achieve hypervolume performance. Reliability is assessed by running 50 random seed replicate trials for each of the algorithm's default parameterizations. In Fig. 11, the solid lines represent the mean hypervolume achieved over the 50 trial runs for each algorithm. The shading bounds in the figure designate the 5th and 95th percentile confidence interval. The default runtime dynamics show that most of the algorithms achieve their best hypervolume performance after 10,000 function evaluations. For NSGA-III, RVEA, and MOEA/D, that found relatively few solutions, these results suggest that the algorithms converged on the two extreme disjoint regions of the objective space very quickly but were unable to discover the interior solutions (see Fig. 8). MOEA/D exhibits the largest variability across its random seed trial runs, which corroborates the results displayed in the attainment plots in Fig. 9. All of the algorithms aside from the Borg MOEA exhibit clear deterioration, which was first identified in the control maps in Section 5.3. In Fig. 11 deterioration (or loss of solutions) is shown with the fluctuating, non-monotonic hypervolume dynamics. This deterioration is especially prominent in RVEA and NSGA-III. Both algorithms achieve high hypervolume early in the search, but then non-linearly decrease in their hypervolume performance.

could be attributed to how they exploit their reference points and reference vectors. NSGA-III normalizes objectives in every generation, which means that solutions that are associated with a reference point in one generation may become located farther away from that reference point in a subsequent generation. Thus, while still a valid solution, the point may be excluded from the population if a new population member becomes associated with the reference point. For smooth Pareto fronts, this will likely not be an issue because normalization will be consistent across generations. However, if part of a disjoint set is encountered that has not been found yet, this can drastically change the normalization scheme and cause divergent deterioration. RVEA implements a vector adaptation strategy that, unlike NSGA-III's objective normalization, is not performed during every generation with the intent of stabilizing convergence (Cheng et al., 2016). However, as seen in Fig. 11, even with this solution strategy in place, RVEA still suffers from deterioration.

The smooth convergence of the Borg MOEA's hypervolume shown in Fig. 11 can be attributed to its employment of a closed-loop feedback where the effectiveness in generating new solutions informs auto-adaptive multi-operator search dynamics. While the other algorithms benchmarked in this study utilize a static parameterization and a single pre-specified suite of exploratory operators, the Borg MOEA's operators as well as key parameters are adaptively modified without any user input. The Borg MOEA's operators are initially equally likely to be used and then these operator likelihoods are adjusted through the search progress feedbacks to favor operators that contribute more solutions to the archive (Hadka and Reed, 2012). Fig. 12 shows how the operator dynamics of the Borg MOEA's six operators evolve across 50 random seed trial runs of the algorithm using its default parameterization and for 200,000 function evaluations. The solid lines represent the mean operator probabilities achieved over the 50 seeds and the shading bounds the 5th and 95th percentile confidence interval. Fig. 12 demonstrates how the Borg MOEA utilizes multiple exploratory operators, but primarily cooperatively utilizes simulated binary crossover (SBX), parent-centric crossover (PCX), and uniform mutation (UM). PCX emerges as the dominant operator and is employed upwards of 70% of the time in the latter half of the search. While SBX is known to work best for mathematically separable (i.e., independent) decision variables, it is interesting to note that in dynamic collaboration, SBX, PCX, and UM generate an ensemble of new exploration strategies (a non-specified hyper-variational search operator) that does not require strict separability. PCX is a rotationally invariant operator, meaning that it does not require strict decision variable independence and can generate translational moves in any direction. Fig. 12 clearly highlights that the SFPUC benchmarking test case is characterized by strong decision variable interdependence and non-separability (Deb et al., 2002; Hadka and Reed, 2012). Even though some of the operators have a smaller probability of use compared to PCX or SBX, it is the cooperative use of all operators that contribute to the overall success of the search. The significant use of the highly randomized UM operator signals a designed feedback within the Borg MOEA that allows the algorithm to escape local optima or attractors. In this study, the extreme lobes in the SFPUC test case's Pareto front clearly trap most modern MOEAs and the UM-driven exploratory search of the Borg MOEA aids the algorithm in discovering solutions in the sparse compromise region of the space. This cooperative behavior is demonstrated in previous studies (Hadka and Reed, 2012; Vrugt and Robinson, 2007; Vrugt et al., 2009).

5.5. Decision making implications of algorithmic choice

The prior results diagnose the algorithm's search performance. It is also interesting to explore how a typical user would perceive the SFPUC system's financial risk tradeoffs. Fig. 13 simulates a hypothetical interactive decision support exploration in which SFPUC stakeholders specify a performance criteria that the expected minimum final cash flow should be no less than 90% of the expected annualized final cash flow. This criterion represents reasonable expectations that could be set by the utility in order to be in an advantageous position to meet fixed



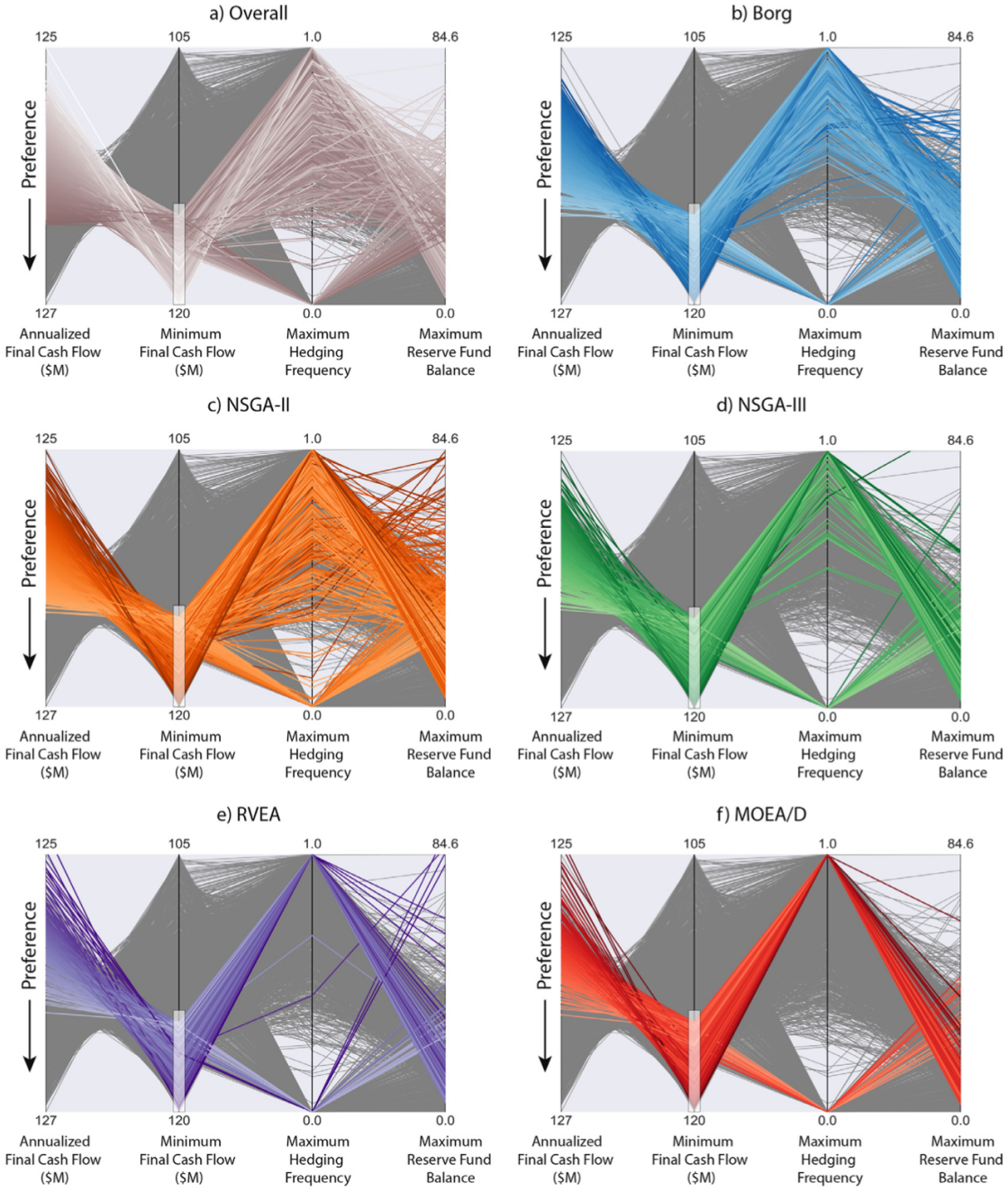


Fig. 13. Parallel axis plots highlighting the tradeoff reference sets that would be attained across the 50 random seed trial runs for the algorithms' default parameterizations. Each vertical axis in the panels represent an objective where the preferred direction is down. Each line in the figure represents a candidate solution whose re-evaluated performance meets the specified performance criteria. The background dark lines represent the best known overall reference set from all runs of all algorithms. The color gradient ranging from light to dark corresponds to high and low expected annualized final cash flow respectively.

annual costs. MOEAs facilitate this type of interactive *a posteriori* tradeoff analysis where decision makers view the full suite of possible solutions and then brush, or eliminate, solutions that don't meet the specified performance requirements. Less attention in literature is spent recognizing that the choice of algorithm can distort the decision maker's perception of their tradeoffs and the suite of solutions available to them. This issue is illustrated in Fig. 13 by showing how the perception of tradeoffs

every axis. Each line in the figure represents a candidate solution whose performance meets the specified performance criterion. The overall best known reference set is included in gray to provide context to what is the actual best set of possible solutions. Fig. 13a shows that when the criterion is imposed, a wide variety of solutions that utilize varying hedging frequencies remain. In focusing on the individual algorithms, it becomes clear that when the performance requirement is imposed,

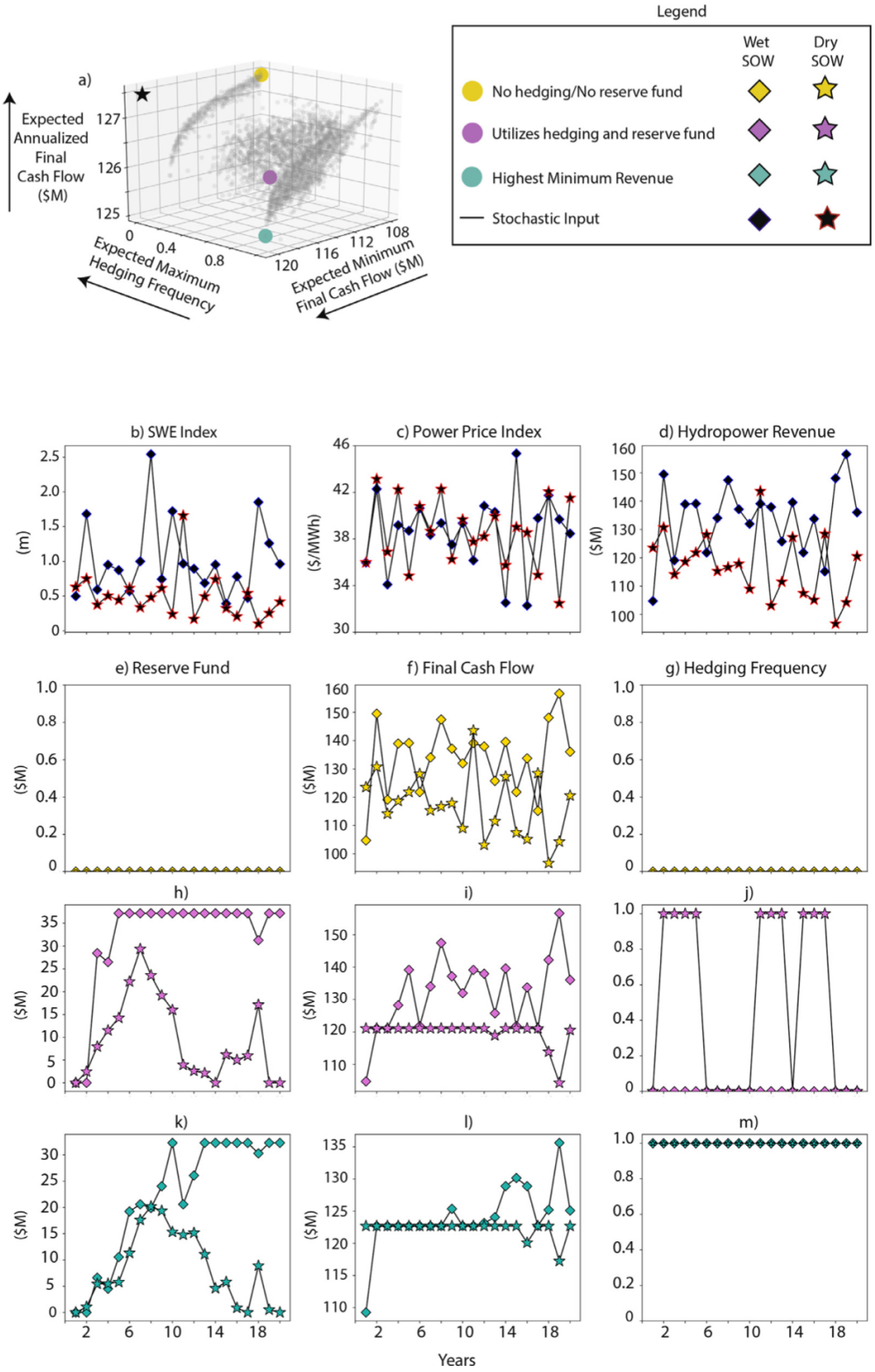


Fig. 14. Three example polices, portrayed in yellow, purple, and blue, were chosen from the overall reference set and re-evaluated in a characteristically wet and dry state of the world. Diamonds represent the time series of stochastic inputs or system states across the 20-year wet period, while stars represent the time series in the 20-year dry period. The first row of the figure shows the stochastic inputs across the 20 years in the wet and dry states. The subsequent rows show the performance of

Fig. 13c shows that NSGA-II finds solutions that most closely resemble the set of solutions remaining in the overall reference set in Fig. 13a. Shown in Fig. 13d, NSGA-III finds significantly fewer solutions in the interior space between the two lobes. Furthermore, most of these solutions are spaced quite closely together, offering close to equivalent performance on the expected annualized final cash flow and expected maximum fund balance objectives. Therefore, these solutions likely would not be deemed substantially different to the decision maker. Fig. 13e shows that RVEA can only locate two solutions in the interior region of the reference set, one of which leads to the worst performance in the expected maximum reserve fund balance objective. Fig. 13f demonstrates that MOEA/D is unable to preserve any interior solutions.

From a decision-making standpoint, these results suggest that that the choice of algorithm can strongly limit the perception of both the number and variety of solutions available. For instance, a decision maker that uses the default parameterization of MOEA/D may erroneously assume that the only contract structures which meet the minimum final cash flow requirement either utilize a minimum or maximum hedging frequency. However, the rest of the algorithms suggest otherwise. Using MOEA/D in this instance would severely constrain the number of options available to the decision maker while algorithms such as the Borg MOEA and NSGA-II would offer a more flexible set of solutions. Moreover, although NSGA-II happens to perform well for the SFPUC test case here using its default parameterization, this is not a generalized expectation as several studies have shown that non-adaptive MOEAs can yield drastically different behavior even for modest changes in water resources problems (Ward et al., 2015). Fig. 13 ultimately highlights that depending on performance requirements that are imposed, choice of algorithm can severely constrain the utility's perception of the choices that are available to them.

5.6. Stochastic input implications on policy behavior

When SFPUC establishes their preference and ultimately chooses a policy to implement, they will likely be interested in understanding how the policy will operate in a variety of potential SOWs. Fig. 14 explores how three policies representing alternative tradeoff preferences operate in a representative dry and wet 20-year simulation realization. Fig. 14a shows the three highlighted solutions in the context of SFPUC's tradeoffs. The yellow solution represents a scenario in which the utility receives a high expected annualized final cash flow and does not use a reserve fund or snow-driven hedging. The purple solution applies hedging in 70% of the SOWs and utilizes a reserve fund. The blue point represents a solution that provides the highest minimum final cash flow in any given year, which the utility might implement to hedge against situations where they cannot meet fixed annual costs. Each of these policy solutions are then evaluated in relatively dry and wet 20-year simulation periods. These periods were selected by identifying the 20-year period from the independent 100,000-member sample of stochastic inputs that had the least and most total SWE accumulation, respectively. Fig. 14b–d show the time series of stochastic inputs: yearly SWE index, power price index, and hydropower revenue. Blue diamonds and red stars distinguish the wet and dry time series, respectively. As expected, the wet period is characterized by greater SWE accumulation and consequently higher hydropower revenue. In these states of the world, the wet and dry periods have similar power price indices.

The three solutions, when evaluated in the wet and dry scenarios, show strong behavioral differences in how their reserve fund, final cash flows, and hedging evolve over the respective 20-year periods being simulated. The three bottom rows of Fig. 14 show the objective performance for the yellow (Fig. 14e and g), purple (Fig. 14h and j) and blue solutions (Fig. 14k and m), respectively. In these figure panels, the Fig. 14f is, therefore, equivalent to the hydropower revenue generated for the specific scenario inputs. Thus, this solution illustrates the difficult position that utilities are currently experiencing where their revenues are problematically tied to the potentially large swings in annual snowpack. The purple compromise solution uses both the reserve fund and hedging to yield very different behavior and adapts according to the SOW. As shown in Figures 14h and 14i, in the wet SOW, SFPUC maintains a stable and high reserve fund balance and cash flow from the hydropower generation. Hedging is not implemented in this wet period due to the consistently high reserve fund. However, in the dry period, hedging is applied in 10 of the 20 years where the utility keeps a low reserve fund. In Figure 14i, through careful use of hedging and supplements from the reserve fund, the revenue is generally stabilized. There are a few extreme years characterized by particularly low hydropower revenue and power prices. It should be noted that the final cash flow in these years is still higher than if no financial instruments were being used. The blue highest minimum cash flow solution (Figures 14k-14m), unlike the purple compromise solution, requires hedging in every year of both the wet and dry simulation periods. Through very active use of the reserve fund in Figure 14k, the solution has a higher stabilized final cash flow in the drier period relative to the purple compromise solution (Figure 14i), and d also a less erratic cash flow in the wet period.

Fig. 14 emphasizes the strong diversity of SFPUC's capacity to implement hedging and use a reserve fund to navigate the complex dynamics of their snow-dominated hydro-climatology. These results also further emphasize the importance of the interior solutions, which use state-aware closed loop feedbacks to adjust hedging behavior appropriately depending on the SOW being experienced. While the blue solution may be ideal with respect to the minimum final cash flow objective, it requires the utility to engage in the higher transactional costs and institutional challenges associated with hedging every year while accepting a more variable reserve fund, which may not be preferable. If an algorithm is unable to discover these compromise solutions, this once again could present more expensive and limiting options for the utility.

6. Conclusions

The volatile dynamics of snow-dependent hydrology in the Western US poses a significant financial risk management challenge for hydropower utilities. There is a growing need to develop risk mitigation policies that can incorporate the complex dynamics of these systems in tandem with emerging solution tools that can represent the resulting tradeoffs across alternative financial management strategies. The state-aware adaptive actions that are enabled by the EMODPS framework hold significant promise in addressing these challenges. However, the resulting financial risk portfolio simulation-optimization problems pose significant challenges to modern evolutionary multi-objective optimization tools. Our capability to access the advantages of EMODPS is highly contingent on the ability of MOEAs to solve for the resulting stochastic financial risk portfolios, including allowing users the flexibility to rapidly iterate across alternative formulations if stakeholders want to adapt their formulations (models, uncertainties, financial instruments) as they gain insights. However, highly flexible decision framing and support assumes that the underlying MOEA has four key properties: reliability across random trials and parameterizations, effectiveness in attaining high quality approximations of tradeoff solution sets, efficiency in minimizing computational demands, and high controllability (i.e., insensitive to algorithmic parameters).

Our study focuses on a highly challenging test case of using EMODPS to optimize policies that represent the major financial decisions for San

by a highly disjoint Pareto front of tradeoffs. Overall, the Borg MOEA was the only algorithm to display consistently high levels of performance across all assessments. Through the use of adaptive variational operators and population size, the Borg MOEA was able to represent the extent of the geometry of the overall reference set and contribute the most solutions relative to the other algorithms. Furthermore, it reliably attained high levels of generational distance, additive epsilon indicator, hypervolume performance, and demonstrated controllability, or ease of use, across all tested parameterizations. The rest of the suite of algorithms were unable to consistently achieve high levels of performance. While NSGA-II was also able to discover acceptable representations of the SFPUC application's tradeoffs, many of its solutions were ultimately dominated by other algorithms and it displayed a strong sensitivity to initial population size. The more modern NSGA-III, RVEA, and MOEA/D algorithms proved to struggle with solving the SFPUC test case. These reference point and decomposition techniques pose problem-specific challenges if the tradeoff solution sets are not uniform and convex. Given the complex disjoint, non-convex solution set for the SFPUC benchmarking test case, NSGA-III, RVEA, and MOEA were unable to locate interior compromise points. Furthermore, while achieving high best possible metric values, all the algorithms struggled to attain these values reliably. The control maps highlight MOEA/D's poor controllability and the abject failure of NSGA-III and RVEA to attain an acceptable hypervolume across any of their tested feasible parameterizations.

Most concerning is that the more modern algorithms displayed clear deterioration, or the tendency to lose solutions during the search. Deterioration can result due to a combination of algorithm characteristics: (1) the algorithm's lack of an archive to preserve non-dominated solutions and (2) the implementation of a normalization scheme in every generation which leads to instability in convergence towards the Pareto front. These characteristics along with a utilization of a static set of search operators renders these algorithms ineffective for solving the SFPUC test case. The success of the Borg MOEA lies in its adaptive search operators which allows it to adjust parameterizations to favor those that advance search progress and an adaptive population size operator to help maintain diversity and escape local optima. Furthermore, its implementation of an epsilon-dominance archive ensures both diversity and preservation of strong non-dominated solutions. The results from this benchmarking study suggest that not all algorithms can effectively and reliably approximate the tradeoffs of the SFPUC test case, and by extension, the classes of optimization problems that are characterized by complex and disjoint tradeoffs. A complex solution topology can result due to the presence of discrete objectives, as in the SFPUC test case, or discrete decision variables. For example, this behavior could manifest in capital investment infrastructure optimization problems that introduce capital costs within the simulation period. After a disaster, a resource-limited water resource system may seek temporary capacity expansion solutions that maximize safe water supply in a minimal time frame. A renewable energy company could aim to optimize placement of wind turbines across a section of farmland. Only the Borg MOEA has shown the ability to effectively approximate a complex and disjoint solution space. It should be stated that the formulation of optimization problems in financial risk or other areas can be as simple as a two-objective formulation to minimize risk and maximize return to a more complex EMODPS framework as demonstrated in this study. Furthermore, the specific benchmarked DPS formulation chosen for the study can be extended to consider many alternative formulations that feature different objectives, information inputs, and configurations of policy basis functions. A hyper-heuristic algorithm such as the Borg MOEA can adaptively adjust parameterization to accommodate diverse problem framings (Quinn et al., 2017), and shows the ability to dis-

has been shown to be false in this and prior controllability benchmarking efforts. This work highlights that potential future extensions need to be considered carefully and jointly with sufficiently adaptive solution tools that do not show the failure mechanisms highlighted in this work.

While the test case focuses on SFPUC and the Tuolumne Basin, many other snow-dominated hydropower systems are characterized by strong annual and interdecadal snowpack variability, from the Andes (Masiokas et al., 2006) to the Atlas Mountains (Boudhar et al., 2016). Utilities in these areas could benefit from an EMODPS framework for financial risk management and policies that utilize financial instruments to stabilize cash flow in hydroclimatic extremes. In order to find these solutions that offer a variety of hedging options and tend to lie in the interior of the space, it is important to utilize algorithms that are able to discover these solutions. Only the Borg MOEA has shown the ability to discover these solutions reliably, effectively, efficiently, and across broad ranges of its parameterization space. In order to realize the potential for using an EMODPS framework for financial risk management and other applications outside of the financial sector that require state-adaptive decision making across a range of alternative problem formulation hypotheses, more attention must be paid to the development of self-adaptive MOEAs. Future search tools need to effectively generalize well across a variety of applications and problem formulations. Therefore, focus should be spent developing new self-adaptive hyper-heuristic algorithms, like the Borg MOEA, that perform stably, require little user interaction, and therefore will facilitate easier and more effective decision support. Advancements in algorithm capabilities coupled with parallel and cloud computing to increase efficiency and size of experimentation, visual analytics to enhance interpretation of solutions, and effective use of state and exogenous information to inform state-aware management actions will ultimately provide the best means to approach the growing complexity of confronting snow-dominated hydropower systems and other complex water resources systems problems in the coming decades.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Rohini S. Gupta: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Visualization, Funding acquisition. **Andrew L. Hamilton:** Conceptualization, Methodology, Data curation, Writing - original draft, Writing - review & editing. **Patrick M. Reed:** Conceptualization, Methodology, Writing - review & editing, Funding acquisition. **Gregory W. Characklis:** Conceptualization, Writing - review & editing, Funding acquisition.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650441. Partial funding for this work was also provided by the National Science Foundation (NSF), Innovations at the Nexus of Food-Energy-Water Systems, Track 2 (Award 1639268). The authors would like to thank Alexis Dufour and Darryl Dunn of the San Francisco Public Utilities Commission for providing the underlying data. The views expressed in this work represent those of the authors and do

Appendix A. Notation Guide

Variable	Definition
$t/i/j$	index value for a given year/RBF/information input
y_t	cash flow for year t
$y_t^{s_1}$ / $y_t^{s_2}$/$y_t^{s_3}$	stage 1/2/3 cash flow for year t
v_t	deposit or withdrawal for year t
f_t	reserve fund balance for year t
c_t	net contract payout for year t
w_i^H/w_i^F	weight of RBF (hedge/final cash flow policy)
$c_{i,j}$	center of RBF i for input j
$b_{i,j}$	radii of RBF i for input j
$(x_t^H)_j/(x_t^F)_j$	information input j for hedge/final cash flow policy in year t
u_t^H/u_t^F	hedge contract slope/final cash flow for year t
$\widetilde{u}_t^H/\widetilde{u}_t^F$	policy values before normalization in year t
u_t^{H*}/u_t^{F*}	policy values before constraint application in year t
ϕ^{HN}/ϕ^{FN}	hedge/final cash flow normalization
ϕ^{HC}	hedge constraint
ϕ^{FI}/ϕ^{F0}	inner/outer final cash flow constraint
d^H/d^F	hedge/final cash flow constant threshold variable
a^H/a^F	hedge/final cash flow constant shift variable
θ^H/θ^F	hedge/final cash flow parameter vector
Θ	overall policy vector
$\mathcal{P}$	overall policy representation
r^A	discount rate=0.96
k^H	Normalization for hedge contract slope=$4 million/inch (0.025 m)
k^R	Normalization for revenues and cash flows=$250 million
k^F	Normalization for reserve fund=$150 million

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2020.103718.

References

Better, M., Glover, F., Kochenberger, G., Wang, H., 2008. Simulation optimization: applications in risk management. In: International Journal of Information Technology and Decision Making, vol. 7, pp. 571–587. https://doi.org/10.1142/S0219622008003137.

Beyer, H.G., Sendhoff, B., 2007. Robust optimization - a comprehensive survey. Comput. Methods Appl. Mech. Eng. 196 (33–34), 3190–3218. https://doi.org/10.1016/j.cma.2007.03.003.

Biglarbeigi, P., Giuliani, M., Castelletti, A., 2018. Partitioning the impacts of streamflow and evaporation uncertainty on the operations of multipurpose reservoirs in arid regions. J. Water Resour. Plann. Manage. 144 (7). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000945.

Bolton, P., Chen, H., Wang, N., 2011. A unified theory of Tobin's q, corporate investment, financing, and risk management. J. Finance 66 (5), 1545–1578. https://doi.org/10.1111/j.1540-6261.2011.01681.x.

Boudhar, A., Boulet, G., Hanich, L., Sicart, J.E., Chehbouni, A., 2016. Energy fluxes and melt rate of a seasonal snow cover in the Moroccan High Atlas. Hydrol. Sci. J. 61 (5), 1–13. https://doi.org/10.1080/02626667.2014.965173.

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. doi:10.1057/jors.2013.71

California Energy Commission, 2018. Toward A Clean Energy Future, 2018 Integrated Energy Policy Report Update Volume I. Technical Report. California Energy Commission.

Chan Hilton, A.B., Culver, T.B., 2005. Groundwater remediation design under uncertainty using genetic algorithms. J. Water Resour. Plann. Manage. 131 (1), 25–34. https://doi.org/10.1061/(ASCE)0733-9496(2005)131:1(25).

Cheng, R., Jin, Y., Olhofer, M., Sendhoff, B., 2016. A reference vector guided evolutionary algorithm for many-objective optimization. IEEE Trans. Evol. Comput. 20 (5), 773–791. https://doi.org/10.1109/TEVC.2016.2519378.

Clarke, L., Nichols, L., Vallario, R., Hejazi, M., Horing, J., Janetos, A.C., Mach, K., Mastrandrea, M., Orr, M., Preston, B.L., Reed, P., Sands, R.D., White, D.D., 2018. Sector interactions, multiple stressors, and complex systems. In: Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, vol. II, pp. 638–668. https://doi.org/10.7930/NCA4.2018.CH17.

Coello, C.A.C., Lamont, G.B., Veldhuizen, D.A.V., Goldberg, D.E., Koza, J.R., 2007. Evolutionary algorithms for solving multi-objective problems second edition genetic and evolutionary computation series series editors selected titles from this series. https://doi.org/10.1046/j.1365-2672.2000.00969.x.

Deb, K., Jain, H., 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: solving problems with box constraints. IEEE Trans. Evol. Comput. 18 (4), 577–601. https://doi.org/10.1109/TEVC.2013.2281535.

Deb, K., Mohan, M., Mishra, S., 2005. Evaluating the ϵ-domination based multi-objective evolutionary algorithm for a quick computation of Pareto-optimal solutions. Evol. Comput. 13 (4), 501–525. https://doi.org/10.1162/106365605774666895.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. Technical Report 2.

Desreumaux, Q., Côté, P., Leconte, R., 2018. Comparing model-based and model-free streamflow simulation approaches to improve hydropower reservoir operations. J. Water Resour. Plann. Manage. 144 (3), 05018002. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000860.

Foster, B.T., Kern, J.D., Characklis, G.W., 2015. Mitigating hydrologic financial risk in hydropower generation using index-based financial instruments. Water Resour. Econ. https://doi.org/10.1016/j.wre.2015.04.001.

Giagkiozis, I., Fleming, P.J., 2014. Pareto front estimation for decision making. Evol. Comput. 22 (4), 651–678. https://doi.org/10.1162/EVCO_a_00128.

Giudici, F., Castelletti, A., Garofalo, E., Giuliani, M., Maier, H.R., 2019. Dynamic, multi-objective optimal design and operation of water-energy systems for small, off-grid islands. Appl. Energy 605–616. https://doi.org/10.1016/j.apenergy.2019.05.084.

Giuliani, M., Castelletti, A., Pianosi, F., Mason, E., Reed, P.M., 2016. Curses, tradeoffs, and scalable management: advancing evolutionary multiobjective direct policy search to improve water reservoir operations. J. Water Resour. Plann. Manage. 142 (2), 04015050. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000570.

Giuliani, M., Herman, J.D., Castelletti, A., Reed, P., 2014. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. Water Resour. Res. 50 (4), 3355–3377. https://doi.org/10.1002/2013WR014700.

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

Gonzalez, P., Garfin, G., Breshears, D., Brooks, K., Brown, H., Elias, E., Gunasekara, A., Huntly, N., Maldonado, J., Mantua, N., Margolis, H., McAfee, S., Middleton, B., Udall, B., 2018. Southwest. In Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, vol. II. USGCRP, p. 11011184. https://doi.org/10.7930/NCA4.2018.CH25.

Gopalakrishnan, G., Minsker, B.S., Goldberg, D.E., 2003. Optimal sampling in a noisy genetic algorithm for risk-based remediation design. J. Hydroinf. 5 (1), 11–25. https://doi.org/10.2166/hydro.2003.0002.

Hadka, D., Reed, P., 2012. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. Evol. Comput. 20 (3), 423–452. https://doi.org/10.1162/EVCO_a_00053.

Hadka, D., Reed, P., 2013. Borg: an auto-adaptive many-objective evolutionary computing framework. Evol. Comput. 21 (2), 231–259. https://doi.org/10.1162/EVCO_a_00075.

Hamilton, A. L., Characklis, G. W., Reed, P. M., 2020. Managing financial risk tradeoffs for hydropower generation using snowpack-based index contracts. Preprint on https://essoar.org (2020). doi:10.1002/essoar.10502068.1.

Kasprzyk, J.R., Reed, P.M., Characklis, G.W., Kirsch, B.R., 2012. Many-objective de Novo water supply portfolio planning under deep uncertainty. Environ. Modell. Softw. 34, 87–104. https://doi.org/10.1016/j.envsoft.2011.04.003.

Kasprzyk, J.R., Reed, P.M., Kirsch, B.R., Characklis, G.W., 2009. Managing population and drought risks using many-objective water portfolio planning under uncertainty. Water Resour. Res. 45 (12). https://doi.org/10.1029/2009WR008121.

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 Resour. Res. 51, 2412–2427. https://doi.org/10.1002/2014WR016533.

Koutsoyiannis, D., Economou, A., 2003. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. Water Resour. Res. 39 (6). https://doi.org/10.1029/2003WR002148.

Liu, L., Hejazi, M., Iyer, G., Forman, B.A., 2019. Implications of water constraints on electricity capacity expansion in the united states. Nat. Sustain. 2 (3), 206–213. https://doi.org/10.1038/s41893-019-0235-0.

Lund, J., Medellin-Azuara, J., Durand, J., Stone, K., 2018. Lessons from California's 2012–2016 drought. J. Water Resour. Plann. Manage. 144 (10). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000984.

Maier, Holger R., Jain, A., Dandy, G.C., Sudheer, K.P., 2014. Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. Elsevier.

Markowitz, H., 1952. PORTFOLIO SELECTION*. J. Finance 7 (1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

Masiokas, M.H., Villalba, R., Luckman, B.H., Le Quesne, C., Aravena, J.C., 2006. Snowpack variations in the Central Andes of Argentina and Chile, 1951–2005: large-scale atmospheric influences and implications for water resources in the region. J. Clim. 19 (24), 6334–6352. https://doi.org/10.1175/JCLI3969.1.

Mohammadi, A., Omidvar, M.N., Li, X., 2012. Reference point based multi-objective optimization through decomposition. 2012 IEEE Congress on Evolutionary Computation, CEC 2012. https://doi.org/10.1109/CEC.2012.6256486.

Mulvey, J.M., Shetty, B., 2004. Financial planning via multi-stage stochastic optimization. Comput. Oper. Res. 31 (1), 1–20. https://doi.org/10.1016/S0305-0548(02)00141-7.

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. J. Water Resour. Plann. Manage. 136 (4), 412–432.

Palakonda, V., Mallipeddi, R., 2017. Pareto dominance-based algorithms with ranking methods for many-objective optimization. IEEE Access 5, 11043–11053. https://doi.org/10.1109/ACCESS.2017.2716779.

Ponsich, A., Jaimes, A.L., Coello, C.A., 2013. A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. IEEE Trans. Evol. Comput. 17 (3), 321–344. https://doi.org/10.1109/TEVC.2012.2196800.

Powell, W.B., 2019. A unified framework for stochastic optimization. Eur. J. Oper. Res. 275, 795–821. https://doi.org/10.1016/j.ejor.2018.07.014.

Qi, Y., Li, X., Yu, J., Miao, Q., 2019. User-preference based decomposition in MOEA/D without using an ideal point. Swarm Evol. Comput. 44, 597–611. https://doi.org/10.1016/j.swevo.2018.08.002.

Quinn, J.D., Reed, P.M., Giuliani, M., Castelletti, A., 2019. What is controlling our control rules? Opening the black box of multireservoir operating policies using time-varying sensitivity analysis. Water Resour. Res. 55 (7), 5962–5984. https://doi.org/10.1029/2018wr024177.

Quinn, J.D., Reed, P.M., Keller, K., 2017. Direct policy search for robust multi-objective management of deeply uncertain socio-ecological tipping points. Environ. Modell. Softw. 92, 125–141. https://doi.org/10.1016/j.envsoft.2017.02.017.

Reed, P., Minsker, B.S., Goldberg, D.E., 2003. Simplifying multiobjective optimization: an automated design methodology for the nondominated sorted genetic algorithm-II. Water Resour. Res. 39 (7). https://doi.org/10.1029/2002WR001483.

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. In: iEMSs 2012 - Managing Resources of a Limited Planet: Proceedings of the 6th Biennial Meeting of the International Environmental Modelling and Software Society https://doi.org/10.1016/j.advwatres.2012.01.005.

Rosenstein, M.T., Barto, A.G., 2001. Robot weightlifting by direct policy search. Int. Joint Conf. Artif.Intell. 17, 839–846.

San Francisco Public Utilities Commission, 2016. Comprehensive Annual Financial Report, Fiscal Years Ended June 30, 2016 and 2015. Technical Report.

Smalley, J.B., Minsker, B.S., Goldberg, D.E., 2000. Risk-based in situ bioremediation design using a noisy genetic algorithm. Water Resour. Res. 36 (10), 3043–3052. https://doi.org/10.1029/2000WR900191.

Steinbach, M.C., 2001. Markowitz revisited: mean-variance models in financial portfolio analysis. SIAM Rev. 43 (1), 31–85. https://doi.org/10.1137/S0036144500376650.

Van Veldhuizen, D.A., Lamont, G.B., 1998. Evolutionary Computation and Convergence to a Pareto Front. Technical Report.

Van Veldhuizen, D.A., Lamont, G.B., 1998. Multiobjective Evolutionary Algorithm Research: A History and Analysis. Technical Report.

Voisin, N., Kintner-Meyer, M., Wu, D., Skaggs, R., Fu, T., Zhou, T., Nguyen, T., Kraucunas, I., 2018. Opportunities for joint water-energy management: sensitivity of the 2010 western u.s. electricity grid operations to climate oscillations. Bull. Am. Meteorol. Soc. 99 (2), 299–312. https://doi.org/10.1175/BAMS-D-16-0253.1.

Vrugt, J.A., Robinson, B.A., 2007. Improved evolutionary optimization from genetically adaptive multimethod search. Proc. Natl. Acad. Sci. U.S.A. 104 (3), 708–711. https://doi.org/10.1073/pnas.0610471104.

Vrugt, J.A., Robinson, B.A., Hyman, J.M., 2009. Self-adaptive multimethod search for global optimization in real-parameter spaces. IEEE Trans. Evol. Comput. 13 (2), 243. https://doi.org/10.1109/TEVC.2008.924428.

Ward, V.L., Singh, R., Reed, P.M., Keller, K., 2015. Confronting tipping points: can multi-objective evolutionary algorithms discover pollution control tradeoffs given environmental thresholds? Environ. Modell. Softw. 73, 27–43. https://doi.org/10.1016/j.envsoft.2015.07.020.

Zatarain Salazar, J., Reed, P.M., Herman, J.D., Giuliani, M., Castelletti, A., 2016. A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. Adv Water Resour. https://doi.org/10.1016/j.advwatres.2016.04.006.

Zhang, Q., Li, H., 2007. MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans. Evol. Comput. 11 (6), 712–731. https://doi.org/10.1109/TEVC.2007.892759.

Zhang, Q., Liu, W., Li, H., 2009. The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances. In: 2009 IEEE Congress on Evolutionary Computation, CEC 2009, pp. 203–208. https://doi.org/10.1109/CEC.2009.4982949.

Zhang, Q., Suganthan, P.N., 2009. Final Report on CEC'09 MOEA Competition. Technical Report.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., Da Fonseca, V. G., 2003. Performance assessment of multiobjective optimizers: an analysis and review. doi:10.1109/TEVC.2003.810758.