CAN MODERN MULTI-OBJECTIVE EVOLUTIONARY ALGORITHMS DISCOVER HIGH-DIMENSIONAL FINANCIAL RISK PORTFOLIO TRADEOFFS FOR SNOW-DOMINATED WATER-ENERGY SYSTEMS?

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Rohini Gupta

December 2019

ABSTRACT

Hydropower-reliant power utilities are becoming increasingly vulnerable to hydrologic variability in states such as California, that have suffered from extensive droughts and reduced winter snowfall. One such utility is the San Francisco Public Utilities Commission (SFPUC), which operates the Hetch Hetchy Power System. SFPUC is strongly reliant on snowmelt from the Sierra Nevada to provide hydropower to San Francisco. Therefore, it is particularly financially vulnerable to changes in snowpack availability and timing, which translates to variability in yearly revenue. Evolutionary multi-objective direct policy search (EMODPS) can be used to identify time adaptive stochastic rules that inform optimal financial decisions based on state and exogenous information. However, the resulting financial risk mitigation portfolio planning problem is difficult to optimize due to its high dimensionality and mixture of nonlinear, nonconvex, and discrete objectives. We contribute a diagnostic assessment of state-of-the-art MOEAs' abilities to support an EMODPS 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 Multiobjective Evolutionary Algorithm Based on Decomposition (MOEA/D). MOEA performance is evaluated by analyzing controllability, reliability, efficiency, and effectiveness. 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.

BIOGRAPHICAL SKETCH

Rohini received a Bachelor of Science Degree from the University of Illinois at Urbana-Champaign in May 2017. She majored in Civil and Environmental Engineering with a concentration in Energy-Water-Environment Sustainability. She began her graduate studies at Cornell University in August 2017.

To my family and friends for all your love and support.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Patrick Reed for his invaluable guidance and support in helping to shape this thesis. I would also like to thank Dr. Scott Steinschneider for serving on my committee and for his helpful suggestions and comments during the review process.

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No.DGE-1650441. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the US National Science Foundation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CAN MODERN MULTI-OBJECTIVE EVOLUTIONARY ALGORITHMS DISCOVER HIGH-DIMENSIONAL FINANCIAL RISK PORTFLIO TRADEOFFS FOR SNOW-DOMINATED WATER-ENERGY SYSTEMS?

1.1 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). Growing pressures from increasingly variable climate extremes as well as broader financial pressures are already significantly stressing these systems in ways that are often hard to predict and manage (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 CA 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 to 14.7% of California's the total mix (California Energy Commission, 2018). This degree of inter-annual snowpack and hydro-climatic variability poses a significant management challenge for hydropower dependent utilities that is projected to potentially become more severe (Gonzalez et al., 2018).

In an effort to mitigate financial volatility, hydropower-dependent utilities often manage their hydrological risks by investing in financial risk instruments such as hedging contracts to supplement revenue with payouts in volatile years (Foster et al., 2015). It is not uncommon 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. (a)(in preparation) introduces a financial index insurance 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 dry years that result in revenue shortfalls. 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 to hedge against revenue volatility for San Francisco Public Utilities Commission's (SFPUC) Power Enterprise Division which supplies clean energy primarily to San Francisco's International Airport and municipal buildings (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 for better capturing the uncertainties involved in real-world financial applications (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. 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 of solutions that compose the tradeoffs of a system (Coello et al., 2007). To date, there are very few examples demonstrating the use of state-of-the-art MOEAs in financial applications. Early MOEA literature examples focused on the Markowitz model which seeks to assemble 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 2001 in

the robotic control theory literature, is classified as a simulation-based "policy approximation" control formulation (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 these successes, Hamilton et al. (b)(in preparation) formulates a 4-objective stochastic application of EMODPS 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 which 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 projected 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, or (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 exponentially, substantially increasing computational demands associated with the decomposition processes itself (Mohammadi et al., 2012). The limits in Pareto dominance and decomposition approaches motivated the emergence of another

class of algorithm, reference 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). Reed et al. (2013) comprehensively have not 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) and combination of risk-neutral (mean-focused) and risk-averse stochastic objectives. The objectives are nonlinear and non-convex and the inclusion of a discrete objective to represent portfolio complexity results in a Pareto front that is severely disjoint with a complex geometry that has not been represented in previous 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 domain as captured by the SFPUC financial risk portfolio problem. The benchmarking effort contributed here rigorously evaluates five state-of-the-art MOEAs that capture historical to emerging major classes of algorithms: the 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.

1.2 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 Figure 1.1. 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 the Modesto and Turlock Irrigation Districts at a lower fixed rate (San Francisco Public Utilities Commission, 2016).

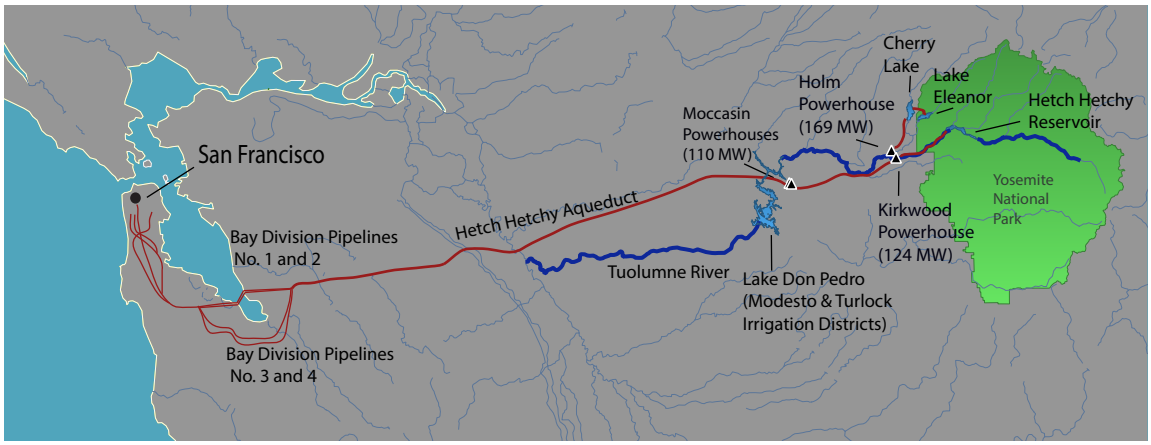

Figure 1.1: The high-altitude reservoirs of Hetch Hetchy, Cherry Lake, and Lake Eleanor drive the hydropower turbines at the Holm, Kirkwood, Moccasin, and Moccasin Low-Head Powerhouses to generate clean energy for the city of San Francisco and other retail customers. Surplus power is sold to irrigation districts (Map adapted from San Francisco Public Utilities Commission, 2018).

1.2.1 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. (a)(in preparation) introduce a snow water equivalent (SWE) based index contract, termed a contract for differences, as a financial risk management tool for SFPUC. The CFD allows the utility to make payments to the contract seller in wet years and receive payouts in dry years. The study shows that the SWE-based index insurance is highly correlated with hydropower and therefore annual revenues, and when used in conjunction with a reserve fund, the contract can help to stabilize the utility's annual revenue.

Hamilton et al. (b)(in preparation) optimizes SFPUC's use of CFD contracts in tandem with a reserve fund using an evolutionary multi-objective optimization direct policy search (EMODPS) framework that utilizes closed-loop feed-

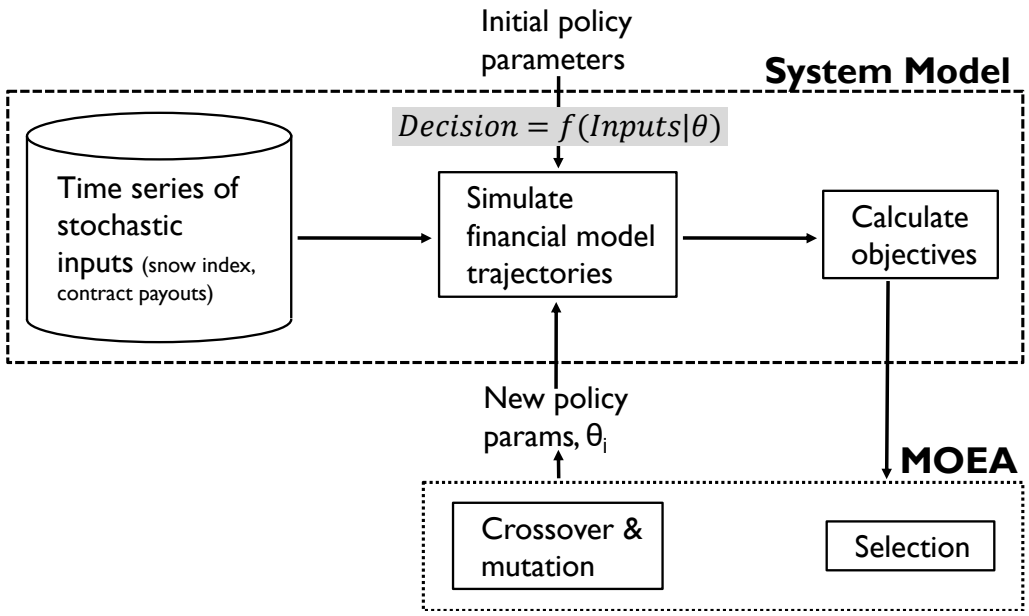

Figure 1.2: Schematic of the evolutionary multi-objective direct policy search (EMODPS) approach. Stochastic inputs feed into the system model that simulates annual utility operations. The utilities financial decisions as represented as policies whose parameters are optimized by an MOEA. Figure adapted from Giuliani et al., 2016.

back and state information to inform multi-year sequences of optimal financial portfolio decisions. Figure 1.2 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 to quantify optimal 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 year that dictate their annual cash flow. At the beginning of the water year, the utility must determine the value of the CFD contract that it will enter. Then at the end of the year, after revenues and contract payouts are observed, the utility determines how much money to deposit or withdraw from the reserve fund. As illustrated in Figure 1.2, 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 Figure 1.2, optimizes the policy parameters to approximate the Pareto frontier (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 1.2.

1.2.2 Stochastic Scenarios

For the SFPUC test case, the stochastic inputs that characterize each state of the world (SOW) are SWE, power price indices, and hydropower revenue, shown in Table 1.1. These stochastic inputs represent the key uncertainties that drive differences in the utility's inter-annual revenue dynamics. The Monte Carlo samples for these key uncertainties are synthetically generated from historical 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 change.

Table 1.1: Stochastic inputs, ε_t , into the simulation model.

Stochastic Input	**Variable**	**Units**
SWE Index	ε_t^S	Inches
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 measurements from 1952-2016 (excluding 1963) to gamma distributions. Then, a Gaussian copula is fit to capture the correlation between the months and used to generate synthetic SWE measurements. The

synthetically generated SWE is used to determine the CFD contract payout or payment. The payout function, h, illustrated in Figure 1.3, dictates a contract payout or payment, termed c. If $h(\varepsilon_t^S) > 0$ then SWE measurements are low and the utility receives compensation ($c_t > 0$). However, if $h(\varepsilon_t^S) < 0$, then SWE measurements are high and the utility makes payments to the contract seller ($c_t < 0$).

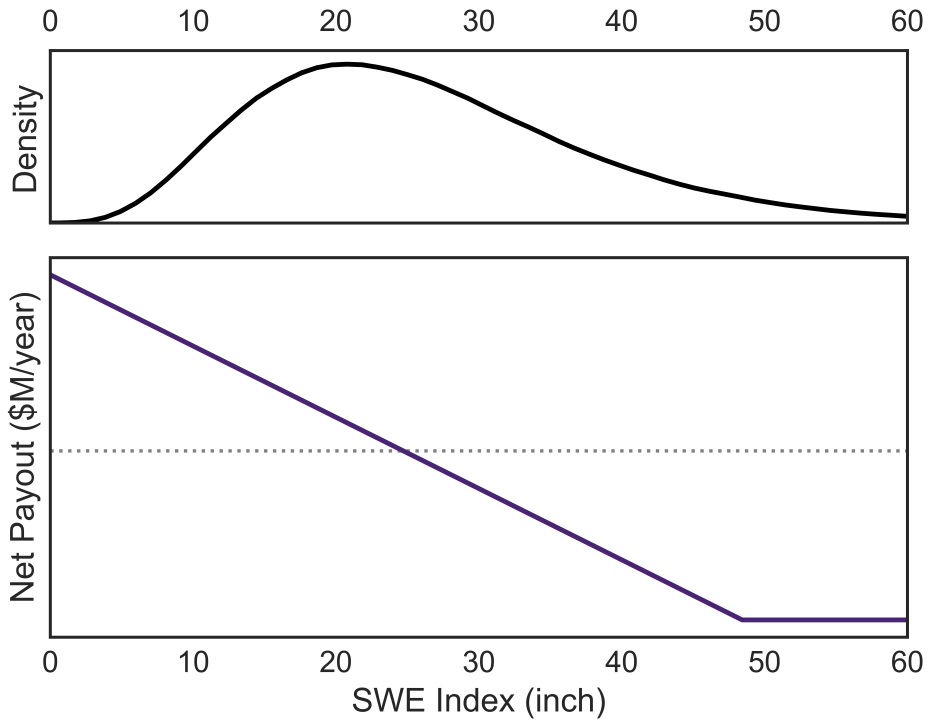

Figure 1.3: (top) Probability density for the SWE index using a weighted average of February 1 and April 1 observations. (bottom) Contract payout function, h. When stochastically generated SWE is below 25 inches, the utility receives a payout. If SWE is above 25 inches, the utility makes payments to the contract seller.

The power price index, ε_t^P, represents the utility's best guess for the upcoming year's power price and therefore their estimate of their ability to make revenue when selling excess hydropower. This index is determined by considering both the current power price and the generation-weighted power price over the

previous year as shown in equation (1.1).

$$\bar{P}_t^{gen-wt} = \frac{1}{12} \frac{\sum_{m=1}^{12} \bar{G}_m P_{m,t}}{\sum_{m=1}^{12} \bar{G}_m} \tag{1.1}$$

In equation (1.1), $\bar{G}_m$ represents the average excess generation sold to the wholesale market for a given month m and $P_{m,t}$ represents the power for month m in year t. In the months that hydropower production is insufficient and the utility must purchase power from the wholesale market, generation is negative. Therefore, a larger index indicates the belief that the upcoming year will yield higher, more beneficial prices, while a lower index indicates the opposite.

The last stochastic input is yearly 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 residuals of the linear models and used to generate a synthetic time series of hydropower generation. The sales of the synthetic hydropower to SFPUC's three major classes of customers along with respective fixed rates or wholesale prices determines ε_t^R. 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 the Modesto and Turlock Irrigation Districts (MTID). For the SFPUC test case, 1000, 20-year Monte Carlo ensembles of the three exogenous drivers are generated concurrently in order to preserve correlation between the inputs and across time. 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. More

detail on the generation of stochastic inputs can be found in Hamilton et al. (a)(in preparation).

1.2.3 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 those solutions that comprise SFPUC optimal financial tradeoffs (i.e, the Pareto frontier). 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 equation (1.2), represented by superscripts: s_1, s_2, and s_3 respectively. At the beginning of the water year, the first decision that the utility must make is to determine the value of the hedging contract, u_t^H, to enter. The utility then enters the contract and proceeds through the year, generating hydropower. The cash flow is evaluated 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, defined in Section 1.2.2. 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^H, multiplied by the contract payout for that year, c_t. The final stage of the utility's cash flow, $y_t^{s_3}$, is also referred to as the utility's final adjusted revenue, u_t^R. The final adjusted revenue is the utility's final financial 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.

$$\mathbf{y}_t = \begin{bmatrix} \mathrm{y}_t^{s_1} \\ \mathrm{y}_t^{s_2} \\ \mathrm{y}_t^{s_3} or(u_t^R) \end{bmatrix} = \begin{bmatrix} \varepsilon_t^R \\ \mathrm{y}_t^{s_1} + u_t^H c_t \\ \mathrm{y}_t^{s_2} + v_t \end{bmatrix} \tag{1.2}$$

The two decisions that the utility must make during the year are represented by the control variables, u_t^H and u_t^R, whose superscripts, H and R designate a "Hedge" and "Revenue" 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. These two decisions comprise the overall policy $\mathcal{P}$ as shown in equation (1.3). The overall policy is a function of the two decisions, which in turn are functions of respective parameter vectors θ_H and θ_R.

$$\mathcal{P}(u_t^H(\theta_H), (u_t^R(\theta_R)) \tag{1.3}$$

The first component of the policy, the slope of the hedge contract, u_t^H, is the

fraction of the stochastically generated contract payout that the utility should buy into for the year. The policy-prescribed slope, normalized on [0,1], is given by $\widetilde{u}_t^H$ and defined in equation (1.4).

$$\widetilde{u}_t^H = \phi^H \left(a^H + \sum_{i=1}^{n} w_i^H \exp\left(- \sum_{j=1}^{m} \left(\frac{(x_t^H)_j - c_{i,j}}{b_{i,j}} \right)^2 \right) \right) \tag{1.4}$$

Equation (1.4) can be decomposed into an intermediary normalized contract slope, $\widetilde{u}_t^{H*}$, before the constraint function, ϕ^H, is applied, and then the normalized contract value, $\widetilde{u}_t^H$, after the constraint is applied.

$$\widetilde{u}_t^{H*} = a^H + \sum_{i=1}^{n} w_i^H \exp\left(- \sum_{j=1}^{m} \left(\frac{(x_t^H)_j - c_{i,j}}{b_{i,j}} \right)^2 \right) \tag{1.5a}$$

$$\widetilde{u}_t^H = \phi^H(\widetilde{u}_t^{H*}) \tag{1.5b}$$

In the intermediary equation (1.5a), 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 state characteristics at time t. The $m = 1$ informational input into this policy is the current normalized balance in the reserve fund, denoted by $\widetilde{f}_t$. Therefore, for a given year t, $x_t^H = \left[\widetilde{f}_t\right]$. Then, shown in equation (1.5b), the function ϕ^H applies a constraint to the intermediary policy. This functional form of ϕ^H is represented in equation (1.6).

$$\phi^H(\widetilde{u}_t^{H*}) = \begin{cases} \widetilde{u}_t^{H*}, & \text{if } \widetilde{u}_t^{H*} \geq d^H \\ 0, & \text{otherwise} \end{cases} \tag{1.6}$$

This constraint dictates that the normalized contract slope before constraints are applied, $\widetilde{u}_t^{H*}$, must be greater than some threshold constant, d^H, otherwise the contract is not purchased. The threshold, d^H, is set during optimization. Then, the value of the contract slope, u_t^H, is determined by unnormalizing $\widetilde{u}_t^H$ with a constant unit value of the contract, k^H, as shown in equation (1.7).

$$u_t^H = k^H \widetilde{u}_t^H \tag{1.7}$$

The last step in the utility's annual cash flow is to decide if money should withdrawn or deposited into the reserve fund. This value is determined implicitly by first fitting a policy that determines the utility's final adjusted normalized revenue, $\widetilde{u}_t^R$. This decision is represented in equation (1.8).

$$\widetilde{u}_t^R = \phi^{RO}\left(\phi^{RI}\left(a^R + \sum_{i=1}^{n} w_i^R \exp\left(-\sum_{j=1}^{m}\left(\frac{(x_t^R)_j - c_{i,j}}{b_{i,j}}\right)^2\right)\right)\right) \tag{1.8}$$

Equation (1.8) can be decomposed into an intermediary normalized adjusted revenue, $\widetilde{u}_t^{R*}$, before an inner constraint, ϕ^{RI} , and an outer constraint ϕ^{RO} is applied.

$$\widetilde{u}_t^{R*} = a^R + \sum_{i=1}^{n} w_i^R \exp\left(-\sum_{j=1}^{m}\left(\frac{(x_t^R)_j - c_{i,j}}{b_{i,j}}\right)^2\right) \tag{1.9a}$$

$$\widetilde{u}_t^R = \phi^{RO}(\phi^{RI}(\widetilde{u}_t^{R*})) \tag{1.9b}$$

In equation (1.9a), a^R is an applied constant shift, and w_i^R, $c_{i,j}$, and $b_{i,j}$ are the weights, centers, and radii of n RBFs that represent the adjusted revenue policy and $(x_t^R)_j$ is the value of the j^{th} of m state characteristics 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 $m = 3$ policy inputs for a given year, t, are $x_t^R = \left[\widetilde{\varepsilon}_t^P, \widetilde{f}_{t-1}, \widetilde{y}_t^{s_2} \right]$.

We apply both an inner constraint, ϕ^{RI}, and an outer constraint ϕ^{RO}, to ensure that the resulting policy is feasible. The function, ϕ^{RO}, defined in equation (1.10), ensures that the reserve fund never exceeds the maximum allotted size represented by d^R, a decision variable that is set during optimization. That is, a deposit, (represented as $v_t < 0$), cannot be larger than the available space left in the reserve fund.

$$\phi^{RO}(\widetilde{u}_t^{R*}) = \begin{cases} \widetilde{y}_t^{s_2} + \widetilde{f}_{t-1} - d^R, & \text{if } v_t < 0 \\ \widetilde{u}_t^{R*}, & \text{otherwise} \end{cases} \tag{1.10}$$

The inner function, ϕ^{RI}, constrains the amount of money that can be withdrawn from or deposited into the fund balance and is defined in equation (1.11).

$$\phi^{RO}(\widetilde{u}_t^{R*}) = \begin{cases} \min\left(\widetilde{u}_t^{R*}, \widetilde{y}_t^{s_2} + \widetilde{f}_{t-1}\right) & \text{if } v_t \geq 0 \\ \max\left(\widetilde{u}_t^{R*}, \widetilde{y}_t^{s_2} - \max\left(\widetilde{y}_t^{s_2}, 0\right)\right. & \text{if } v_t < 0 \end{cases} \tag{1.11}$$

The first case of equation (1.11) 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 implies that deposits can only occur when the incoming cash flow is positive and that the deposit cannot be larger than the

cash flow.

The normalized adjusted revenue, $\widetilde{u}_t^R$, is unnormalized to the scale $\left[-k^R, k^R\right]$, as adjusted revenue can be positive or negative, using a constant unit value for revenue, k^R as shown in equation (1.12).

$$u_t^R = k^R\left(2\widetilde{u}_t^R - 1\right) \tag{1.12}$$

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 adjusted revenue, as shown in equation (1.13).

$$v_t = u_t^R - k^R\widetilde{y}_t^{s_2} \tag{1.13}$$

The decision variables for each of the decisions are represented by the parameter vectors θ_H and θ_R. 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^R. These parameter vectors are shown in equations (1.14) and (1.15).

$$\theta_H = \left[a^H, w_i^H, c_{i,j}, b_{i,j}, d^H\right] \tag{1.14}$$

$$\theta_R = \left[a^R, w_i^R, c_{i,j}, b_{i,j}, d^R\right] \tag{1.15}$$

The weights are constrained to be positive ($w_i^H \geq 0$ and $w_i^R \geq 0$) and sum to unity ($\sum_{i=1}^n w_i^H = 1$ and $\sum_{i=1}^n w_i^R = 1$). The centers and radii of each RBF are

constrained between -1 and 1 and 0 and 1 respectively ($-1 \leq c_{i,j} \leq 1,\ 0 \leq b_{i,j} \leq 1$). Combining the policy parameters from the two decisions in equation (1.14) and equation (1.15) results in the overall vector of decision variables Θ.

Four RBFs were chosen to represent each decision. Each RBF has two weights, one for each of the sub-policies, and m centers and radii. Adding in the shift and threshold variables amounts to $2n(m+1)+4 = 36$ decision variables for the overall policy vector Θ. The optimal policy parameters, Θ^*, are determined by solving the optimization problem in equation (1.16) with respect to the objectives defined in Section 1.2.4.

$$\Theta(\theta_H, \theta_R) = [(c_{1,1}, ..., c_{n,m}), (b_{1,1}, ..., b_{n,m}), (w_1^H, ..., w_n^H, w_1^W, ..., w_n^W), (a^H, a^R), (d^H, d^R)] \quad (1.16)$$

1.2.4 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: $[\varepsilon_t^S, \varepsilon_t^P, \varepsilon_t^R]$. The resulting matrix of 1000 samples of the vector of key uncertainties is denoted ε.

Maximize expected annualized adjusted revenue (AdjRev)

SFPUC wants to maximize their annualized adjusted revenue to meet their annual fixed costs which consume, on average, over 90% of their yearly revenue

(San Francisco Public Utilities Commission, 2016). As formulated in equation (17) below, this objective maximizes the expected value of the annualized adjusted revenue, u_t^R, experienced in a given year, t, over a T-year period. This objective maximizes the sum of all discounted adjusted revenue 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 a discount factor, $r^A = 0.96$, to convert it to an annualized value. The expectation operator, E_ε, calculates the expectation of the objective over the 1000 Monte Carlo simulations of the key uncertainties vector, ε.

$$J^{AdjRev}\left(u^R_{t\in(1,\ldots,T)}, f_T\right) = E_{\boldsymbol{\varepsilon}}\left[\frac{1}{\sum_{t=1}^{T}(r^A)^t}\left(\sum_{t=1}^{T}\left((r^A)^t u_t^R\right) + (r^A)^{T+1}(f_t)\right)\right] \tag{1.17}$$

Maximize expected minimum adjusted revenue (MinRev)

SFPUC also may be interested in maximizing the worst-case adjusted revenue that they could receive in any given year to further hedge against a situation where they cannot pay their fixed costs and must take on debt. This objective is formulated to maximize the expected value of the minimum adjusted revenue, u_t^R, attained over the T years and the expectation of the objective is calculated over the 1000 Monte Carlo simulations of the key uncertainties vector, ε.

$$J^{MinRev}\left(u^R_{t\in(1,\ldots,T)}\right) = E_{\boldsymbol{\varepsilon}}\left[\min_{t\in(1,\ldots,T)}\left[u_t^R\right]\right] \tag{1.18}$$

Minimize expected maximum hedge complexity (Complexity)

SFPUC and utilities would prefer to limit the complexity of the portfolios that they buy into in order to limit transaction costs and additional fees associated with writing contracts. Therefore, this objective, formulated in equation (1.19) is implemented to minimize the expected value of the maximum hedging com-

plexity attained over the T years and across the 1000 Monte Carlo simulations of the key uncertainties vector, ε. For any given stochastic sample, the indicator function, $\mathbf{1}_{u_t^H}$, defined in equation (1.20), returns a 1 if a non-zero hedging contract was used at any point over the T years, and zero otherwise.

$$J^{Complexity}\left(u^H_{t\in(0,\ldots,T-1)}\right) = E_{\boldsymbol{\varepsilon}}\left[\max_{t\in(0,\ldots,T-1)}\left[\mathbf{1}_{u_t^H>0}\right]\right] \tag{1.19}$$

$$\mathbf{1}_{u_t^H} = \begin{cases} 1, & \text{if } u_t^H > 0 \\ 0, & \text{otherwise} \end{cases} \tag{1.20}$$

This objective is, therefore, a measure of the likelihood that a utility will enter into at least one contract over the 20-year period in any of the 1000 Monte Carlo simulations.

Minimize expected maximum reserve fund balance (Reserve)

It is not viewed as politically favorable for a utility to hoard 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 ε.

$$J^{Reserve}\left(f_{t\in(0,\ldots,T)}\right) = E_{\boldsymbol{\varepsilon}}\left[\max_{t\in(0,\ldots,T)}\left[f_t\right]\right] \tag{1.21}$$

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

$$\boldsymbol{\Theta}^* = \arg\min_{\boldsymbol{\Theta}} \left[-J^{AdjRev}(\boldsymbol{\Theta}), -J^{MinRev}(\boldsymbol{\Theta}), J^{Complexity}(\boldsymbol{\Theta}), J^{Reserve}(\boldsymbol{\Theta}) \right] \tag{1.22}$$

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

1.3 Algorithmic Benchmarking Methodology

1.3.1 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 Figure 1.4.

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 more complex problems. Ideally, MOEAs should perform well under a wide range of parameterizations (Goldberg, 2002). Therefore, the diagnostic framework illustrated in Figure 1.4 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

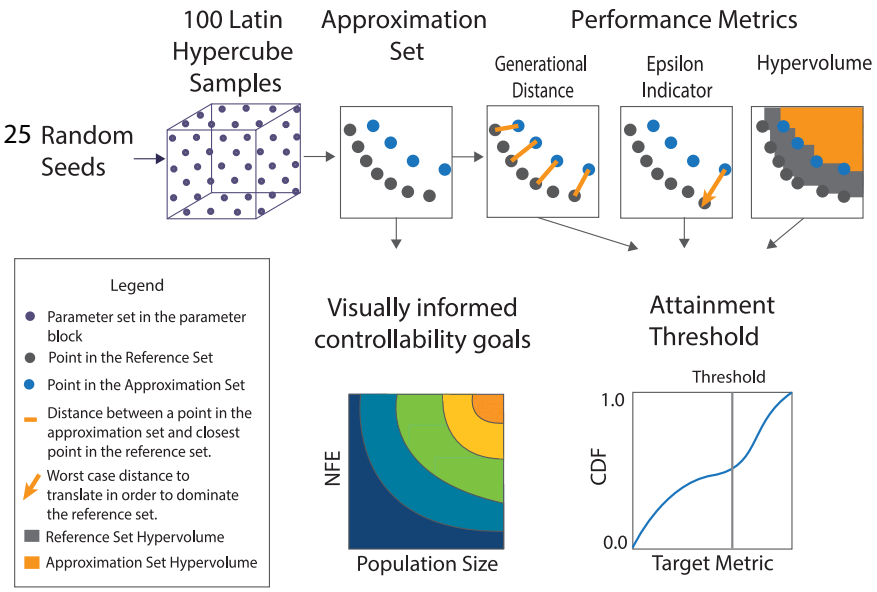

Figure 1.4: Diagnostic assessment framework used to evaluate the performance of each MOEA tested in the study (adapted from Zatarain Salazar et al. (2016)). The parameters for each MOEA are sampled across their full ranges using Latin Hypercube Sampling. The approximation set for each MOEA parameterization is assessed through metrics measuring the convergence, consistency and diversity of approximation sets. These metrics are used to construct visualization tools that capture the efficiency and controllability of each algorithm and assess the effectiveness and reliability of each algorithm. Each parameter set is further replicated for 25 random seed trials to account for the random initialization of populations.

point in the parameter block of Figure 1.4, represents one fully specified parameter instance for an MOEA. In this benchmarking study, each instance is used to solve the SFPUC optimization problem defined in equation (1.22) 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 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 MOEA is easy-to-use and not sensitive to its parameterization. Ideally, an MOEA should generate a high quality Pareto approximation for any parameterization; this is referred to a large "sweet spot" in its parameter space (Goldberg, 2002).

1.3.2 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 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, 1998a,b)

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 generally the most challenging and comprehensive metric to meet that can be used to assess the approximation set's convergence and diversity (Zitzler et al., 2003). In this study, hypervolume is calculated across normalized objective function values and is also normalized relative to the reference set hypervolume when metric results are reported.

1.3.3 Multi-objective Evolutionary Algorithms

MOEAs are popular tools for multi-objective optimization of complex, nonlinear problems because their population-based approach requires less knowledge of the topology of a problem than their deterministic counterparts. There-

fore, 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 nonconvex 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. Members in each front are non-dominated with respect to each other and are therefore given the same rank. The next generation of solutions is determined by incorporation of fronts that have the best ranks. When NSGA-II reaches a front with more members than the remaining slots 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 member 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 resource 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 monitor for stagnation in search process and 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 be 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 involves choosing population members with the greatest proximity to a set of pre-defined reference points to achieve diversity. After the Pareto front sorting procedure, the M problem objectives are normalized in each generation based on the maximum value achieved for each objective. A hyperplane is determined from extreme points and then reference points are evenly space 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 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 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.

1.4 Computational Experiment

As described in the Section 1.2.3, 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 utility state information and power prices to optimal contract values and end-of-the-year adjusted revenue. Each radial basis function has three parameters: a radius, center, and weight. Four radial basis functions are used for this test case and there are three informational parameters serving as inputs into the function to yield a total of 36 decision variables. The candidate optimal policies that inform yearly financial decisions over a 20-year time period are optimized with respect to these decisions variables. Five state-of-the-art MOEAs are used to discover optimal policies with respect to the four objectives outlined in Section 1.2.4. The diagnostic assessment is performed using MOEAFramework, a free and open source Java library that allows users to design, execute, and assess the performance of a variety of pop-

ular MOEAs. The following subsections discuss the computational experiments that are executed through MOEAFramework to conduct the diagnostic assessment. First, Section 1.4.1 elaborates on the procedure for sampling of algorithm parameterizations. Then section 1.4.2 describes the process of generating and verifying reference sets.

1.4.1 Sampling of Algorithm Parameterizations

The experimental setup for the diagnostic assessment requires testing MOEAs in their default parameterizations, displayed in Table 1.2. The default instance of the MOEAs is then used to solve the SFPUC test case. Each MOEA is allowed to search for 200,000 NFE and replicated for 50 random seeds to account for effects from the random initialization of the population. Then, in order to test MOEAs across their feasible range of parameter values, 100 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 Figure 1.4. Each MOEA instance is replicated for 25 random seeds and evaluated over 200,000 NFE. Archive output and runtime dynamics are reported every 5,000 NFE 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.

Table 1.2: 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

1.4.2 Generation and verification of reference sets

Five MOEAs outlined in Section 1.3.3 are used to discover optimal policies with respect to the four objectives outlined in Section 1.2.4. 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 using consistent epsilon sorting across the algorithms. 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 average annualized adjusted revenue and the average hedging complexity objectives and 0.1 for the minimum adjusted revenue and the fund balance objective.

As stated in Section 1.2.2, 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: snow water equivalent, power price index, and yearly revenue. In the SFPUC test case, objectives are optimized to a different ensemble of inputs in every functional evaluation, which allows exposure to a wider variety of potential worlds that would not be possible if using a fixed ensemble. This type of noisy sampling can serve to capture extreme events in the tails of the distributions of the stochastic inputs that would impact the utilities most severely. The sampling is also supported by a large body of work centered around evolutionary optimization under uncertainty (Smalley et al., 2000; Chan Hilton and Culver, 2005; Gopalakrishnan et al.,

2003; Kasprzyk et al., 2009). The stability of the approach is further confirmed by re-evaluating the solutions on a larger set of 100,000 independent of stochastic samples and verifying that solution performance is stable. The re-evaluated reference sets can be found in Appendix B. Re-calculating runtime dynamics with respect to the new verified reference set is computationally intractable. Therefore, Sections 1.5.1-1.5.4 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 1.5.5 of the results.

1.5 Results and Discussion

1.5.1 Analysis of Reference Sets

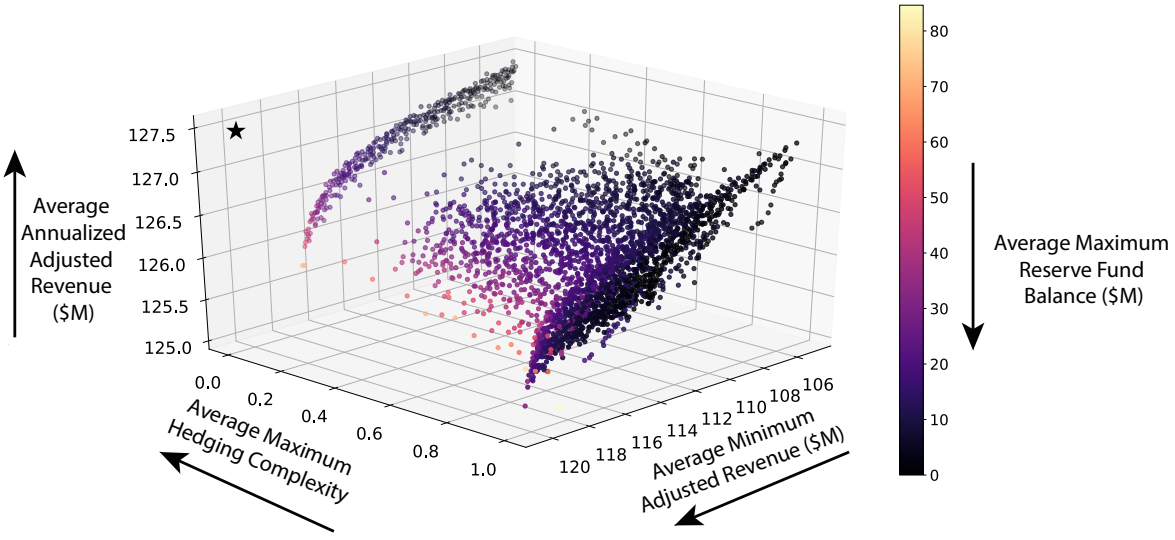

Figure 1.5: The overall best known tradeoffs for the SFPUC test case attained across all MOEA runs. The ideal point is shown as a black star. The fourth objective, minimizing the average maximum reserve fund balance is represented by the color scale (darker is preferred).

Figure 1.5 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 1.4.2, 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 MOEAs given that its best known Pareto front is characterized by a complex disjoint geometry that contains two distinct lobes with compromise solutions interspersed in the compromise region defined by the portfolio complexity objective. As evident in Figure 1.5, there is a strong tradeoff between the average minimum adjusted revenue objective and the average annualized adjusted revenue objective. The color gradient with lighter colored points in the lower front portion of the plot highlights that increasing average minimum adjusted revenue usually necessitates the presence of a larger reserve fund balance. However, the lack of lighter points in the rightmost highest complexity lobe of solutions suggests that a larger reserve fund balance is not necessary if a more complex portfolio structure is implemented.

Formalizing this relationship, the histogram in Figure 1.6 shows that the minimum complexity contracts implement a wider distribution of maximum reserve fund balances while the most complex financial portfolio solutions do not require a reserve fund larger than $40 million on average. The high complexity contract structures reduce the potential self-insurance opportunity costs that SFPUC face when they have to fix a large amount of their funds in a reserve account. The complexity objective is a measure of the likelihood that a utility will enter into at least 1 contract over the 20-year period for any state of the world. This objective can be seen as a proxy of transaction costs and minimizing it abstracts an aversion to complex financial instrument portfolios. Figure

1.5 shows that there are zones of highly concentrated solutions corresponding to extremes of the complexity objective (i.e., probability equal to 0 or 1). An intermediate complexity objective value between 0 and 1 indicates the fraction of the 1000 statistical 20-year replicate samples in which the solution implemented at least 1 contract during the planning period. Generally, lower complexity solutions are associated with a larger annualized average revenue whereas higher complexity solutions lead to a higher minimum revenue.

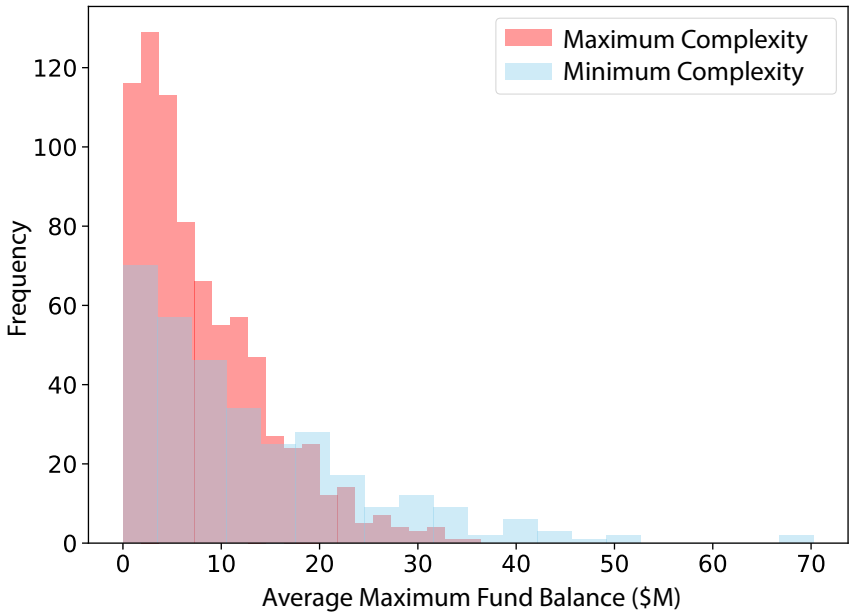

Figure 1.6: Histograms of the distributions of the maximum reserve fund balances associated with the most and least complex contract structures. The purple color indicates where the two histograms overlap.

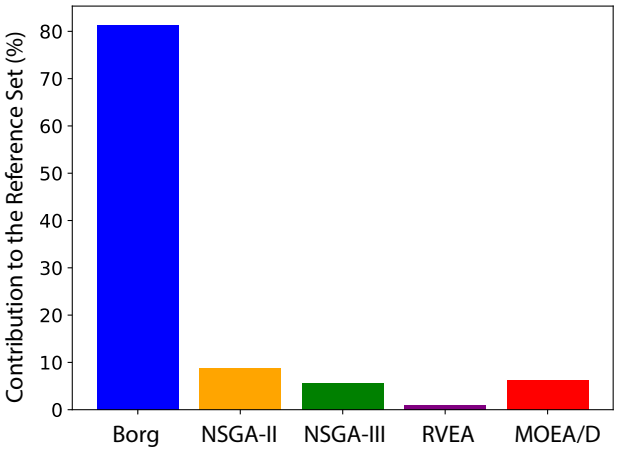

Figure 1.7: The percent of the reference set was contributed by each algorithm. The Borg MOEA single-handedly produced over 80% of the reference set while the newer algorithms struggled to discover solutions that were not dominated by Borg or NSGA-II.

Figure 1.7 builds on Figure 1.5 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 Figure 1.7, 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 solutions were ultimately dominated by the Borg MOEA solutions. The key distinguishing feature between the Borg MOEA and the rest of the algorithms lies in its ability to adaptively adjust population size, selection 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 and in more diverse areas than the other algorithms.

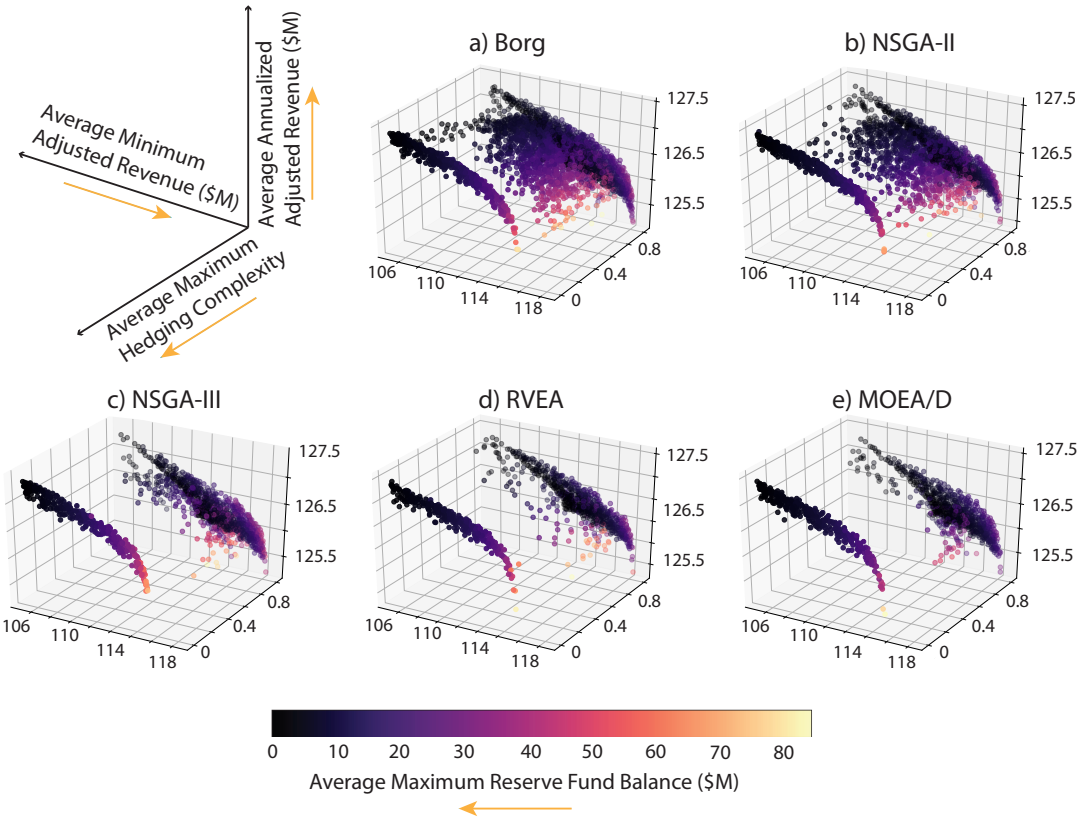

Figure 1.8: Each individual algorithm's best attained approximate reference sets discovered across all of its runs for all tested parameterizations.The yellow arrows indicated the preferred direction for the objectives.

The five panels in Figure 1.8 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 (Figure 1.8a) and NSGA-II (Figure 1.8b) are the only algorithms that obtain reference solution sets that closely replicate the geometry of the overall best known reference set in Figure 1.5. The decomposition strategy of MOEA/D (Figure 1.8e) as well as the reference point-based search of NSGA-III (Figure

1.8c) and RVEA (Figure 1.8d) all yield significantly fewer solutions that are biased toward the two extreme lobes of the complexity 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 subproblems 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.

1.5.2 Algorithmic Effectiveness and Reliability

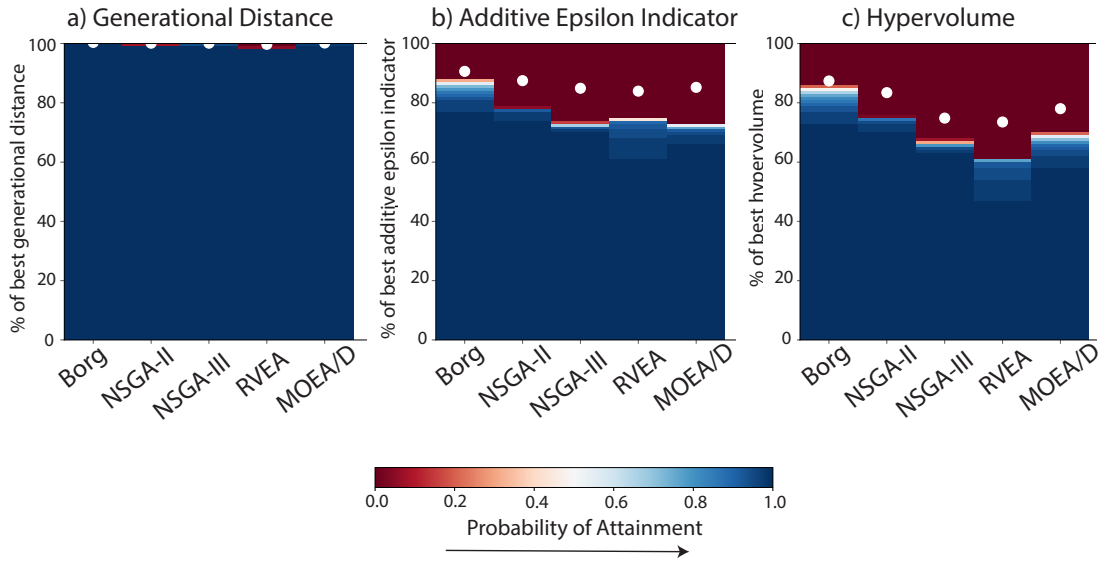

Figure 1.9: Attainment plots that capture the best overall metric values achieved by each MOEA (white circles) as well as their success probabilities in attaining increasingly higher quality metric values. The red-to-blue shading indicates the probability that a single run of an MOEA reaches a given percentage of the best possible metric value for a) generational distance b) additive epsilon indicator and c) hypervolume.

Successful tradeoff analyses in decision support applications require that MOEAs are able to effectively discover high quality approximations to the Pareto front. Moreover, they should do so reliably across their candidate parameterizations and random seed trials. That is, an MOEA should be both effective and reliable for any given run. The attainment plots in Figure 1.9 provide a probablistic 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 Figure 1.9. For each of the metrics, the probability of attainment is defined as the percentage of an MOEA's trial runs (across all parameterization 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 Figure 1.9a, it is clear that all of the algorithms were able to obtain a 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 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 senstive 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 Figure 1.8 highlight that several never identify solutions in the intermediate complexity compromise region. The worst case translation distances grow very rapidly given these gaps (see the discussions in Hadka and Reed, 2012, 2013). Figure 1.9b indicates that the Borg MOEA achieves the highest level of attainment for the additive epsilon indicator measure, consistently achieving 79% of the overall best value. The Borg MOEA similarly out performs the other MOEAs hypervolume attainment in Figure 1.9c, consistently achieving 75% of the best hypervolume metric value. Surprisingly, NSGA-II has the second best overall attainment performance in

both the additive epsilon indicator and hypervolume. In Figures 1.9b and 1.9c, 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 this value 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 Figure 1.9 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.

1.5.3 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.

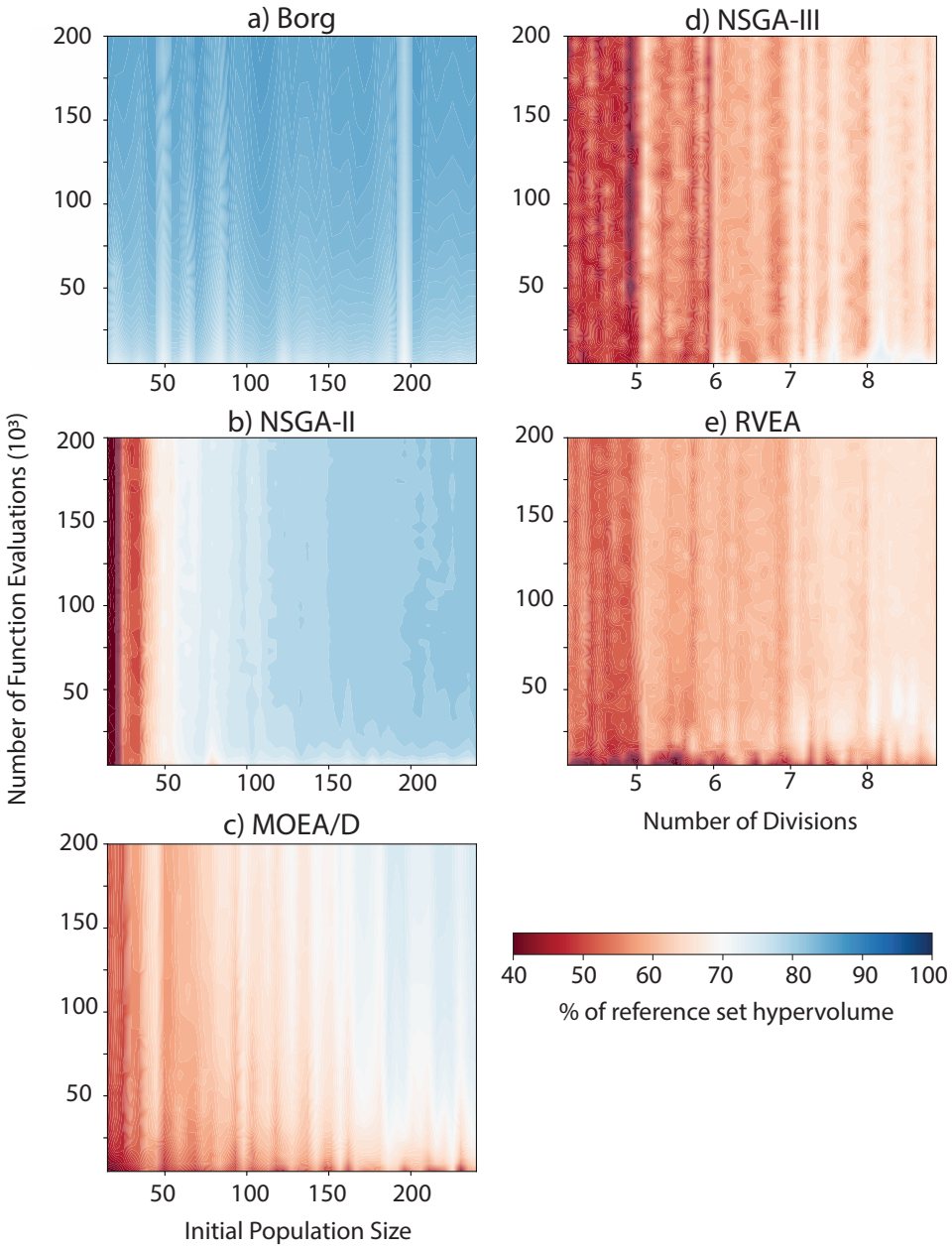

Figure 1.10: Hypervolume performance control maps for the SFPUC test case capturing the controllability and efficiency of each MOEA. The color scale represents the percent of the best known global reference set's hypervolume captured by each local 25-seed reference approximation set for each tested parameterization. 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.

The control maps, shown in Figure 1.10, 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 solid dark blue shading, which requires that the algorithm achieves high hypervolume (effectiveness) even if constrained by a limited number of function evaluations (efficiency) for all tested parameterizations (controllability).

The uniform blue shading in the Borg MOEA control map (Figure 1.10a) 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 (Figure 1.10b) displays a high sensitivity to its population size as has been widely documented in prior studies (Reed et al., 2003; Hadka and Reed, 2012; Reed et al., 2013), showing an abrupt threshold for hypervolume performance. This behavior is characteristic of elitist algorithms that utilize a fixed population size and therefore 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 (Figure 1.10c) exhibits a nonlinear sensitivity to both population size and search duration. Although MOEA/D exhibits an island of high hypervol-

ume performance, it would be difficult for users to specify successful parameter ranges a priori, meaning that this algorithm is not controllable, or "easy-to-use". 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 (Figure 1.10d and 1.10e) 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 sen-

sitivity to their parameterizations makes them difficult to use in real-world applications. The control maps emphasize that more effort should be directed towards developing and using MOEAs that can flexibly navigate the complex tradeoffs of real-world applications and whose endogenous self-adaptive nature require no user interaction. The ultimate goal should be to facilitate efficient and user-friendly decision-making; algorithms that only work well for specific parameterizations that are challenging to discover do not help to facilitate this goal.

1.5.4 Default Parameterization Runtime Dynamics

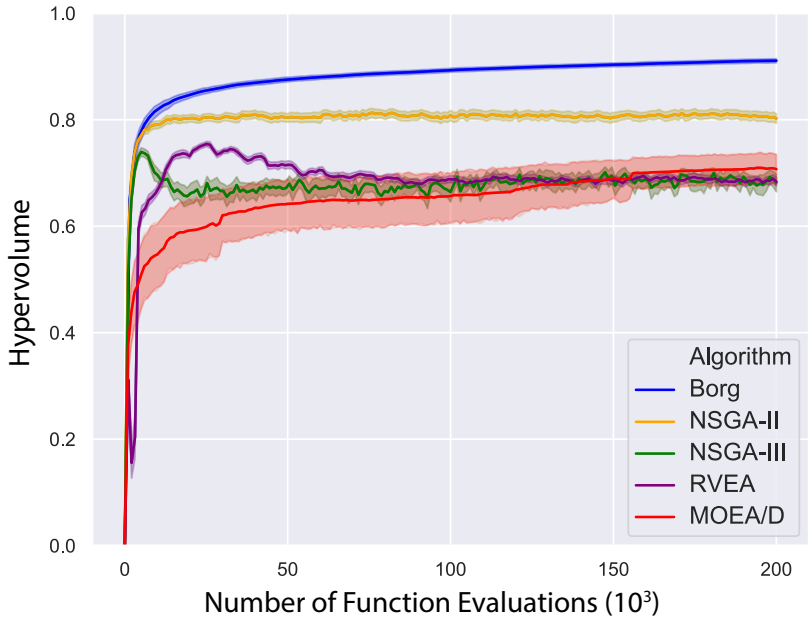

Figure 1.11: Runtime dynamics of hypervolume performance across 50 seeds of each algorithm's default parameterization The solid line represents the mean hypervolume achieved through the search process and the shading bounds the 5th and 95th percentile confidence interval. A hypervolume of 1 with thin shading is preferred (i.e., high performance reliability).

The default runtime analyses in Figure 1.11 capture the expected performance of each of the MOEAs using their author-recommended default parameterizations summarized in Table 1.2. 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 Figure 1.11, the solid lines represent the mean hypervolume achieved over the 50 trial runs for each algorithm. The shading bounds in the figure designate the 5^{th} and 95^{th} 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 most and least complex financial portfolio structures for SFPUC very quickly but were unable to discover the interior solutions (see Figure 1.8). MOEA/D exhibits the largest variability across its random seed trial runs, which corroborates the results displayed in the attainment plots in Figure 1.9. All of the algorithms aside from the Borg MOEA exhibit clear deterioration, which was first identified in the control maps in Section 1.5.4. In Figure 1.11 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.

A loss in hypervolume is common for algorithms that do not exploit formal solution archiving. However, if the lack of archive was the most prominent issue, then NSGA-II and MOEA/D would have exhibited more deterioration.

Therefore, the deterioration of the two reference point algorithms can be attributed to how they exploit their reference vector and reference points. 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 Figure 1.11, even with this solution strategy in place, RVEA still suffers from deterioration.

1.5.5 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. Figure 1.12 simulates an interactive decision support exploration in which SFPUC stakeholders specify a performance criteria that the expected minimum revenue should be no less than 90% of the average annual revenue. This criterion represents reasonable expectations that could be set by the utility in order to be in an advantageous position to meet fixed annual costs. MOEAs facilitate this type of interactive *a posteriori* tradeoff analyses

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 Figure 1.12 by showing how the perception of tradeoffs and candidate compromise solutions changes the algorithms. The vertical axes in each panel of Figure 1.12 represent each of the objectives where the preferred direction of performance is down. Therefore, the ideal solution would be a horizontal line that intersects the bottom of every axis. Each line in the figure represents a candidate SFPUC candidate tradeoff solution whose performance meets the specified performance criteria. The overall best known reference set is included in gray to provide context to what is the actual best set of possible solutions. Figure 1.12a shows that when the requirement is imposed, a wide variety of solutions of varying complexity remain. In focusing on the individual algorithms, it becomes clear that when the performance requirement is imposed, the algorithm's reference sets are not equivalent. All of algorithms tend to preserve solutions in the extreme complexity lobes of Figure 1.5. However, the Borg MOEA (Figure 1.12b) and NSGA-II (Figure 1.12c) reference sets preserve a larger number and a greater variety of solutions relative to the other MOEAs when the requirement is imposed. Figure 1.12c shows that NSGA-II finds solutions that most closely resemble the set of solutions remaining in the overall reference set in Figure 1.12a. Shown in Figure 1.12d, NSGA-III finds much fewer interior solutions. Furthermore, most of these solutions are spaced quite closely together, offering close to equivalent performance on the annualized adjusted revenue and maximum fund balance objectives. Therefore, these solutions likely would not be deemed substantially

different to the decision maker. Figure 1.12e shows that RVEA can only locate two interior solutions, one of which leads to the worst performance in the maximum fund balance objective. Figure 1.12f 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 complexity 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 revenue requirement either have a minimum or maximum complexity. 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). Figure 1.12 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.

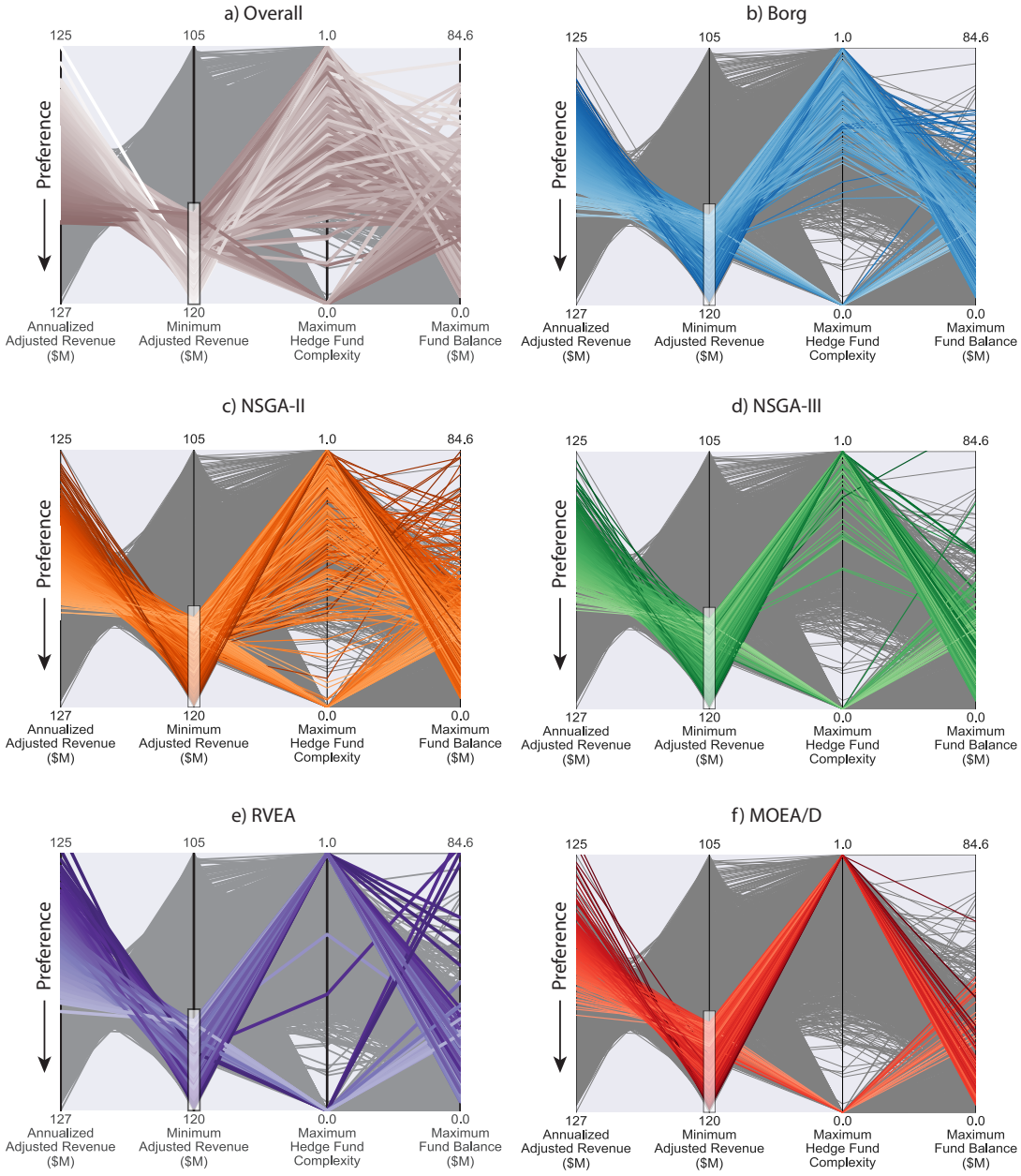

Figure 1.12: Parallel axis plots highlighting the tradeoff reference sets that would be attained across the 50 random seed trial runs for each of the algorithms' default parameterizations. Each vertical axis in the panels represent an objective where the preferred direction is down. Each line in the 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 adjusted revenue respectively.

1.6 Conclusion

The increasingly variable 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 evolving and 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 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 Francisco Public Utilities Commission. We contribute a rigorous assessment of the effectiveness, efficiency, reliability, and controllability of five state-of-the-art MOEAs (NSGA-II, MOEA/D, the Borg MOEA, NSGA-III, and RVEA) when solving the SFPUC test

case, which is characterized 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 algorithms' 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 financial problems that will face snow-dependent hydropower utilities. In order to access the potential for using an EMODPS framework for financial risk management, more attention must be paid to the development of adaptive MOEAs that can effectively discover optimal policies and generalize well across a variety of water resources applications. Only the Borg MOEA has shown the potential to do so. 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 policies will ultimately provide the best means to approach the growing complexity of financial challenges that Western US coupled water and energy systems will face in the coming decades.

APPENDIX A

NOTATION GUIDE

Indices	
$t/i/j$	index value for a given year/RBF/information input
Stochastic Inputs	
$\varepsilon_t^S/\varepsilon_t^P/\varepsilon_t^R$	stochastically generated SWE index/power index/revenue for year t
Financial Variables	
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	contract payout for year t
Policy Variables	
u_t^H/u_t^R	hedge contract slope/adjusted revenue for year t
$\widetilde{u}_t^{H*}$	hedge contract value before application of constraints in year t
$\widetilde{u}_t^{R*}$	annual adjusted revenue before application of constraints in year t
ϕ^H/ϕ^R	Hedge/Revenue Constraint
ϕ^{RI}/ϕ^{R0}	Inner/Outer Adjusted Revenue Constraint
d^H/d^R	Hedge Policy/Adjusted Revenue Constant Threshold
θ^H/θ^R	Hedge Policy/Adjusted Revenue Parameter Vector
Θ	Overall Policy Vector
$\mathcal{P}$	Overall Policy Representation
RBF Parameters	
w_i^H/w_i^R	weight of RBF (Hedge policy/Revenue policy)
$c_{i,j}$	center of RBF i for input j
$b_{i,j}$	center of RBF i for input j
$(x_t^R)_j/(x_t^H)_j$	information input j for Revenue Policy/Hedge Policy in year t
Constants/Miscellaneous	
r^A	discount rate=0.96
k^H	Normalization for hedge contract slope=$4 million/inch
k^R	Normalization for revenues and cash flows=$250 million
$\sim$	Refers to a normalized value

APPENDIX B

RE-EVALUATION OF REFERENCE SETS

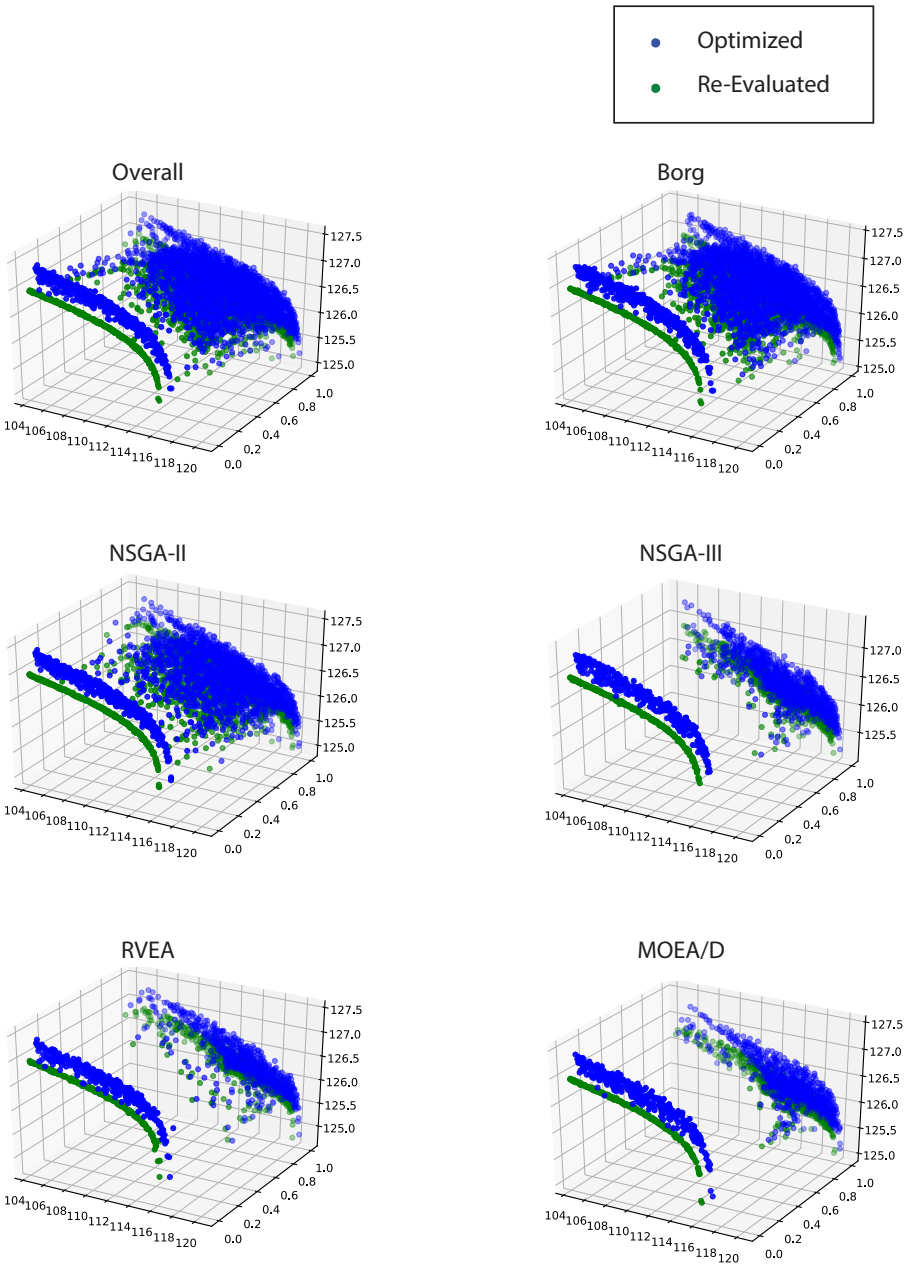

Figure B.1: Individual reference sets were re-evaluated with a new larger and independent set of 100,000 sets of stochastic samples. Blue points represent the original solution set and green points represent the re-evaluated sets

BIBLIOGRAPHY

M. Better, F. Glover, G. Kochenberger, and H. Wang. Simulation optimization: Applications in risk management. In *International Journal of Information Technology and Decision Making*, volume 7, pages 571–587, 12 2008. doi: 10.1142/S0219622008003137.

P. Biglarbeigi, M. Giuliani, and A. Castelletti. Partitioning the impacts of streamflow and evaporation uncertainty on the operations of multipurpose reservoirs in arid regions. *Journal of Water Resources Planning and Management*, 144 (7), 7 2018. ISSN 07339496. doi: 10.1061/(ASCE)WR.1943-5452.0000945.

P. Bolton, H. Chen, and N. Wang. A Unified Theory of Tobin's ¡i¿q¡/i¿ , Corporate Investment, Financing, and Risk Management. *The Journal of Finance*, 66(5):1545–1578, 10 2011. doi: 10.1111/j.1540-6261.2011.01681.x. URL `http://doi.wiley.com/10.1111/j.1540-6261.2011.01681.x`.

E. K. Burke, M. Gendreau, M. Hyde, G. Kendall, G. Ochoa, E. Özcan, and R. Qu. Hyper-heuristics: A survey of the state of the art, 12 2013. ISSN 01605682.

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

A. B. Chan Hilton and T. B. Culver. Groundwater Remediation Design under Uncertainty Using Genetic Algorithms. *Journal of Water Resources Planning and Management*, 131(1):25–34, 1 2005. ISSN 0733-9496. doi: 10.1061/(ASCE)0733-9496(2005)131:1(25).

R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. *IEEE Transactions*

on Evolutionary Computation, 20(5):773–791, 10 2016. ISSN 1089778X. doi: 10.1109/TEVC.2016.2519378.

L. Clarke, L. Nichols, R. Vallario, M. Hejazi, J. Horing, A. C. Janetos, K. Mach, M. Mastrandrea, M. Orr, B. L. Preston, P. Reed, R. D. Sands, and D. D. White. Sector Interactions, Multiple Stressors, and Complex Systems. In *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment*, volume II, chapter 17, pages 638–668. 2018. doi: 10.7930/NCA4.2018.CH17.

C. A. C. Coello, G. B. Lamont, D. A. V. Veldhuizen, D. E. Goldberg, and J. R. Koza. *Evolutionary Algorithms for Solving Multi-Objective Problems Second Edition Genetic and Evolutionary Computation Series Series Editors Selected titles from this series :*. 2007. ISBN 1852337877. doi: 10.1046/j.1365-2672.2000.00969.x.

K. Deb and H. Jain. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation*, 18(4):577–601, 2014. ISSN 1089778X. doi: 10.1109/TEVC.2013.2281535.

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

K. Deb, M. Mohan, and S. Mishra. Evaluating the ϵ-domination based multi-objective evolutionary algorithm for a quick computation of Pareto-optimal solutions. *Evolutionary Computation*, 13(4):501–525, 2005. ISSN 10636560. doi: 10.1162/106365605774666895.

Q. Desreumaux, P. Côté, and R. Leconte. Comparing Model-Based and Model-Free Streamflow Simulation Approaches to Improve Hydropower Reservoir Operations. *Journal of Water Resources Planning and Management*,

144(3):05018002, 3 2018. ISSN 0733-9496. doi: 10.1061/(ASCE)WR.1943-5452.0000860.

B. T. Foster, J. D. Kern, and G. W. Characklis. Mitigating hydrologic financial risk in hydropower generation using index-based financial instruments. *Water Resources and Economics*, 2015. ISSN 22124284. doi: 10.1016/j.wre.2015.04.001.

I. Giagkiozis and P. J. Fleming. Pareto front estimation for decision making. *Evolutionary Computation*, 22(4):651–678, 12 2014. ISSN 15309304. doi: 10.1162/EVCO_a_00128.

F. Giudici, A. Castelletti, E. Garofalo, M. Giuliani, and H. R. Maier. Dynamic, multi-objective optimal design and operation of water-energy systems for small, off-grid islands. *Applied Energy*, pages 605–616, 9 2019. ISSN 03062619. doi: 10.1016/j.apenergy.2019.05.084.

M. Giuliani, J. D. Herman, A. Castelletti, and P. Reed. Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management. *Water Resources Research*, 50(4):3355–3377, 2014. ISSN 19447973. doi: 10.1002/2013WR014700.

M. Giuliani, A. Castelletti, F. Pianosi, E. Mason, and P. M. Reed. Curses, Tradeoffs, and Scalable Management: Advancing Evolutionary Multiobjective Direct Policy Search to Improve Water Reservoir Operations. *Journal of Water Resources Planning and Management*, 142(2):04015050, 2 2016. ISSN 0733-9496. doi: 10.1061/(ASCE)WR.1943-5452.0000570.

D. E. Goldberg. *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*. Kluwer Academic Publishers, Boston, 2002. ISBN 1402070985.

P. Gonzalez, G. Garfin, D. Breshears, K. Brooks, H. Brown, E. Elias, A. Gunasekara, N. Huntly, J. Maldonado, N. Mantua, H. Margolis, S. McAfee, B. Middleton, and B. Udall. Southwest. In Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment. volume II, chapter 25, page 1101–1184. USGCRP, 2018. doi: 10.7930/NCA4.2018.CH25. URL https://nca2018.globalchange.gov/chapter/24/.

G. Gopalakrishnan, B. S. Minsker, and D. E. Goldberg. Optimal sampling in a noisy genetic algorithm for risk-based remediation design. *Journal of Hydroinformatics*, 5(1):11–25, 1 2003. ISSN 1464-7141. doi: 10.2166/hydro.2003.0002.

D. Hadka and P. Reed. Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization. *Evolutionary Computation*, 20(3):423–452, 2012. ISSN 10636560. doi: 10.1162/EVCO_a_00053.

D. Hadka and P. Reed. Borg: An auto-adaptive many-objective evolutionary computing framework. *Evolutionary Computation*, 21(2):231–259, 2013. ISSN 15309304. doi: 10.1162/EVCO_a_00075.

A. L. Hamilton, Characklis, G. W., Reed, and P. Reed. Managing financial risk tradeoffs for hydropower generation using snowpack-based index contracts. *In prep*, pages 1–47, a.

A. L. Hamilton, G. W. Characklis, and P. Reed. Dynamic Management of Environmental Financial Risks Using Evolutionary Multi-Objective Direct Policy Search. *In prep*, b.

J. R. Kasprzyk, P. M. Reed, B. R. Kirsch, and G. W. Characklis. Managing population and drought risks using many-objective water portfolio planning under uncertainty. *Water Resources Research*, 45(12),

12 2009. ISSN 00431397. doi: 10.1029/2009WR008121. URL http://doi.wiley.com/10.1029/2009WR008121.

D. Koutsoyiannis and A. Economou. Evaluation of the parameterization-simulation-optimization approach for the control of reservoir systems. *Water Resources Research*, 39(6), 6 2003. ISSN 00431397. doi: 10.1029/2003WR002148. URL http://doi.wiley.com/10.1029/2003WR002148.

L. Liu, M. Hejazi, G. Iyer, and B. A. Forman. Implications of water constraints on electricity capacity expansion in the United States. *Nature Sustainability*, 2 (3):206–213, 3 2019. ISSN 23989629. doi: 10.1038/s41893-019-0235-0.

J. Lund, J. Medellin-Azuara, J. Durand, and K. Stone. Lessons from California's 2012-2016 drought. *Journal of Water Resources Planning and Management*, 144 (10), 10 2018. ISSN 07339496. doi: 10.1061/(ASCE)WR.1943-5452.0000984.

H. Maier, A. Jain, G. Dandy, K. S. E. m. &, and u. 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Elsevier*, 2014.

H. Markowitz. PORTFOLIO SELECTION*. *The Journal of Finance*, 7(1):77–91, 3 1952. ISSN 00221082. doi: 10.1111/j.1540-6261.1952.tb01525.x. URL http://doi.wiley.com/10.1111/j.1540-6261.1952.tb01525.x.

A. Mohammadi, M. N. Omidvar, and X. Li. Reference point based multi-objective optimization through decomposition. In *2012 IEEE Congress on Evolutionary Computation, CEC 2012*, 2012. ISBN 9781467315098. doi: 10.1109/CEC.2012.6256486.

J. M. Mulvey and B. Shetty. Financial planning via multi-stage stochastic op-

timization. *Computers and Operations Research*, 31(1):1–20, 1 2004. ISSN 03050548. doi: 10.1016/S0305-0548(02)00141-7.

J. Nicklow, P. Reed, D. Savic, T. Dessalegne, L. Harrell, A. Chan-Hilton, M. Karamouz, B. Minsker, A. Ostfeld, A. Singh, and E. Zechman. State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management*, 136(4):412–432, 2010. ISSN 07339496. doi: 10.1061/(ASCE)WR.1943-5452.0000053.

M. O'Connell, N. Voisin, J. Macknick, and Fu. Sensitivity of Western U.S. power system dynamics to droughts compounded with fuel price variability. *Applied Energy*, 2019. ISSN 03062619. doi: 10.1016/j.apenergy.2019.01.156.

V. Palakonda and R. Mallipeddi. Pareto Dominance-Based Algorithms with Ranking Methods for Many-Objective Optimization. *IEEE Access*, 5:11043–11053, 6 2017. ISSN 21693536. doi: 10.1109/ACCESS.2017.2716779.

A. Ponsich, A. L. Jaimes, and C. A. Coello. A survey on multiobjective evolutionary algorithms for the solution of the portfolio optimization problem and other finance and economics applications. *IEEE Transactions on Evolutionary Computation*, 17(3):321–344, 2013. ISSN 1089778X. doi: 10.1109/TEVC.2012.2196800.

W. B. Powell. A unified framework for stochastic optimization. *European Journal of Operational Research*, 275:795–821, 2019. doi: 10.1016/j.ejor.2018.07.014. URL `https://doi.org/10.1016/j.ejor.2018.07.014`.

Y. Qi, X. Li, J. Yu, and Q. Miao. User-preference based decomposition in MOEA/D without using an ideal point. *Swarm and Evolutionary Computation*, 44:597–611, 2 2019. ISSN 22106502. doi: 10.1016/j.swevo.2018.08.002.

J. D. Quinn, P. M. Reed, M. Giuliani, and A. Castelletti. What Is Controlling Our Control Rules? Opening the Black Box of Multireservoir Operating Policies Using Time-Varying Sensitivity Analysis. *Water Resources Research*, 55(7): 5962–5984, 7 2019. ISSN 0043-1397. doi: 10.1029/2018wr024177.

P. Reed, B. S. Minsker, and D. E. Goldberg. Simplifying multiobjective optimization: An automated design methodology for the nondominated sorted genetic algorithm-II. *Water Resources Research*, 39 (7), 7 2003. ISSN 00431397. doi: 10.1029/2002WR001483. URL http://doi.wiley.com/10.1029/2002WR001483.

P. M. Reed, D. Hadka, J. D. Herman, J. R. Kasprzyk, and J. B. Kollat. 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*, 2013. ISBN 9788890357428. doi: 10.1016/j.advwatres.2012.01.005.

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

J. B. Smalley, B. S. Minsker, and D. E. Goldberg. Risk-based in situ bioremediation design using a noisy genetic algorithm. *Water Resources Research*, 36(10): 3043–3052, 2000. ISSN 00431397. doi: 10.1029/2000WR900191.

M. C. Steinbach. Markowitz revisited: Mean-variance models in financial portfolio analysis. *SIAM Review*, 43(1):31–85, 3 2001. ISSN 00361445. doi: 10.1137/S0036144500376650.

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

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

N. Voisin, M. Kintner-Meyer, D. Wu, R. Skaggs, T. Fu, T. Zhou, T. Nguyen, and I. Kraucunas. Opportunities for joint water-energy management: Sensitivity of the 2010 western U.S. electricity grid operations to climate oscillations. *Bulletin of the American Meteorological Society*, 99(2):299–312, 2 2018. ISSN 00030007. doi: 10.1175/BAMS-D-16-0253.1.

V. L. Ward, R. Singh, P. M. Reed, and K. Keller. Confronting tipping points: Can multi-objective evolutionary algorithms discover pollution control tradeoffs given environmental thresholds? *Environmental Modelling and Software*, 73: 27–43, 11 2015. ISSN 13648152. doi: 10.1016/j.envsoft.2015.07.020.

J. Zatarain Salazar, P. M. Reed, J. D. Herman, M. Giuliani, and A. Castelletti. A diagnostic assessment of evolutionary algorithms for multi-objective surface water reservoir control. *Advances in Water Resources*, 2016. ISSN 03091708. doi: 10.1016/j.advwatres.2016.04.006.

Q. Zhang and H. Li. MOEA/D: A multiobjective evolutionary algorithm based on decomposition. *IEEE Transactions on Evolutionary Computation*, 11(6):712–731, 12 2007. ISSN 1089778X. doi: 10.1109/TEVC.2007.892759.

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

Q. Zhang, W. Liu, and H. Li. The performance of a new version of MOEA/D on CEC09 unconstrained MOP test instances. In *2009 IEEE Congress on Evolutionary Computation, CEC 2009*, pages 203–208, 2009. ISBN 9781424429592. doi: 10.1109/CEC.2009.4982949.

E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. Da Fonseca. Performance assessment of multiobjective optimizers: An analysis and review, 4 2003. ISSN 1089778X.