Battling Arrow's Paradox to Discover Robust Water Management Alternatives

Joseph R. Kasprzyk[1]; Patrick M. Reed, A.M.ASCE[2]; and David M. Hadka[3]

Abstract: This study demonstrates how Arrow's Impossibility Theorem, a theory of social choice, is of direct concern when formulating water-resources systems planning problems. Traditional strategies for solving multiobjective water resources problems typically aggregate multiple performance measures into single composite objectives (e.g., a priori preference weighting or grouping-like measures by category). Arrow's Impossibility Theorem, commonly referred to as Arrow's Paradox, implies that a subset of performance concerns will inadvertently dictate the properties of the optimized design alternative in unpredictable ways when using aggregated objectives. This study shows how many-objective planning can aid in battling Arrow's Paradox. Many-objective planning explicitly disaggregates measures of performance while supporting the discovery of planning tradeoffs, using tools such as multiobjective evolutionary algorithms (MOEAs). An urban water portfolio planning case study for the Lower Rio Grande Valley, Texas is used to demonstrate how aggregate, lower objective-count formulations can adversely bias risk-based decision support. Additionally, this study employs a comprehensive diagnostic assessment of the Borg MOEA's ability to address Arrow's Paradox by enabling users to explore problem formulations with increasing numbers of objectives and decisions. Counter to conventional assumptions, the diagnostic analysis carefully documents that for modern self-adaptive MOEA searches, increasing objective counts can lead to more effective, efficient, reliable, and controllable searches. The increased objective counts are also shown to directly reduce decision biases that can emerge from problem formulation aggregation and simplification, related to Arrow's Paradox. **DOI:** 10.1061/(ASCE)WR.1943-5452.0000572.

Author keywords: Water management; Decision support; Multiobjective evolutionary algorithms.

Introduction

The U.S. Council on Environmental Quality recently discussed the need for communicating "tradeoffs among potential solutions" given the "complex and often conflicting water resources needs of today and the future" (U.S. Council on Environmental Quality 2013). Multiple planning objectives arise when allocating water to multiple competing uses, such as irrigation, hydropower, and municipal water supply. Also, multiple objectives may arise within one type of water allocation; for example, minimizing the frequency of small failures may conflict with minimizing the magnitude of failures (Hashimoto et al. 1982). Objective conflicts could also emerge when attempting to minimize modeling uncertainty, performance sensitivities, or adaptability (Haimes 1977). In all cases, the existence of multiple objectives requires the identification of the conflicts and relationships between the competing goals and objectives in the problem.

Starting with the Flood Control Act of 1936 (Reuss 2005), cost-benefit analysis has been historically used to evaluate multiple objective planning problems. Planners decide to fund projects with benefits "to whomsoever they may accrue" outweighing estimated project costs (Arnold 1988). Ultimately, performing cost-benefit analysis requires the commensuration or aggregation of multiple objectives together into a single benefit function. This aggregation represents a form of a priori decision-making that typifies classical multicriteria decision-making (Cohon and Marks 1975; Chankong and Haimes 1983; Borsuk et al. 2001). There have been a broad array of strategies proposed for aggregating a set of objectives f_1 through f_N through the assignment of weights, w_1 through w_N yielding a single objective function f_{agg}. These weights are assumed to fully represent all preferences between the competing objectives, as shown in Eq. (1)

$$f_{\text{agg}} = w_1 f_1 + \ldots + w_N f_N \tag{1}$$

Fundamentally, for a weighting procedure such as Eq. (1) to be valid, the equation must be an internally consistent measure of stakeholders' decision preferences. In his Nobel Prize-winning work, Arrow (1950) explores this issue for social choice problems such as a set of voters choosing among three or more alternatives (e.g., politicians running for office, or policy choices). The goal is to develop a single aggregate measure of value that encompasses the preferences of a group of voters (or decision-makers). Internal consistency in this context means that the aggregate preference function does not violate the intent of maximizing the group members' individual values (i.e., everyone's preferences are well considered). In the paper, Arrow proposes the Impossibility Theorem, which states that it is impossible to construct an internally consistent weighting function for the social choice problem. The Theorem is also called Arrow's Paradox, because although the intent is to promote each voter's preferences in the overall group, a single voter's preference will dictatorally dominate the final choice in a manner that would be difficult to predict a priori.

[1]Assistant Professor, Dept. of Civil, Environmental, and Architectural Engineering, Univ. of Colorado Boulder, ECOT 441, UCB 428, Boulder, CO 80309 (corresponding author). E-mail: joseph.kasprzyk@colorado.edu

[2]Professor, School of Civil and Environmental Engineering, Faculty Fellow, Atkinson Center for a Sustainable Future, Cornell Univ., 220 Hollister Hall, Ithaca, NY 14853. E-mail: patrick.reed@cornell.edu

[3]Associate Research Engineer, Applied Research Laboratory, Pennsylvania State Univ., P.O. Box 30, 4100D, State College, PA 16804. E-mail: dmh309@psu.edu

Note. This manuscript was submitted on July 24, 2014; approved on June 3, 2015; published online on September 1, 2015. Discussion period open until February 1, 2016; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, © ASCE, ISSN 0733-9496/04015053(12)/$25.00.

To connect the implications of Arrow's Paradox to multiobjective engineering design, Franssen (2005) provides a re-evaluation of Arrow's Paradox. In Franssen's framing of Arrow's Paradox, voters are equivalent to planning objectives and states of the world are then equivalent to the alternatives under consideration. The traditional process of creating an aggregated objective function for optimization problems (Heidari et al. 1971; Andreu et al. 1996; Huang and Loucks 2000) is therefore equivalent to the same aggregation in the social choice problem. Franssen's result means that when combining design objectives together, some of the objectives are rewarded or penalized in ways that are difficult to predict a priori.

This study seeks to show how many-objective analysis can help to address biases of aggregated and simplified formulations and avoid some of the pitfalls of Arrow's Paradox. Many-objective analysis exploits multiobjective evolutionary algorithms (MOEAs), quantitative simulation models, and interactive visualizations. MOEAs are a population-based search technique that use a process based on natural selection to evolve solutions that have good performance with respect to quantitative objectives (Coello et al. 2007; Nicklow et al. 2010; Reed et al. 2013; Maier et al. 2014). The MOEA search process embeds trusted simulation models to capture system performance, avoiding the simplifying assumptions that are often required for classical optimization (Labadie 2004). The goal of many-objective analysis is to discover, evaluate, and communicate policy designs that can balance conflicting objectives for a system. The output of the many-objective analysis process is a tradeoff set of solutions, allowing stakeholders to determine the full range of performance of the objectives (Cohon and Marks 1975). The tradeoffs are defined using Pareto optimality; informally a solution is Pareto optimal if it is better than all other feasible solutions in at least one objective.

In the context of Arrow's Impossibility Theorem, an aggregate value function that ranks alternatives should obey the following property: as new alternatives are added or removed from the analysis, the preferential ordering of the remaining alternatives should not change in unpredictable ways (Arrow 1950). The implication for engineering design is that a problem's definition (its objectives, constraints, and decisions) must be static and well-defined, or else some motivating assumptions will be violated. However, problem formulations are rarely fixed; from the very earliest systems planning efforts, it has been observed that objectives may actually be discovered during the planning process (Hitch 1960). In fact, planners are now shifting from single-objective, fixed problem formulations to a constructive approach where elements of the analysis are discovered in collaboration with stakeholders (Tsoukias 2008), and problems are reformulated to incorporate new information (Kasprzyk et al. 2012), including the structuring of variables that define decision alternatives.

Given the need to continually reformulate and revise many-objective problem formulations during the decision-making support process, MOEA searches on these problems must be efficient, reliable, and controllable. This is especially relevant given the computational constraints of real-world decision problems, where search performance can vary significantly from changes in problems' mathematical properties or internal algorithmic sensitivities (i.e., the population size, run duration, and probability of crossover or mutation) (Reed et al. 2013). This study uses a diagnostic evaluation of MOEA parameters to show whether or not aggregations of the problem formulation affect the MOEA's ability to search effectively. Specifically, the diagnostic framework tests the Borg MOEA, a recently introduced MOEA framework that adaptively uses multiple search operators to effectively solve difficult problems (Hadka and Reed 2013).

In summary, this paper will explore two fundamental questions for multiobjective water resources analysis. The first question deals with formulating a planning problem (i.e., choosing appropriate objectives and decisions). Specifically, are aggregated problem formulations negatively biasing our understanding and exploration of the tradeoffs in complex water management applications? The second question explores the difficulty of actually finding appropriate solutions to a given problem formulation. In other words, how does MOEA performance differ over four problem formulations that modify objectives (Many-Objective versus Only Cost and Reliability) and decisions (full complexity and simplified variables)? A water marketing test case, developing portfolios of market and nonmarket based instruments for the Lower Rio Grande Valley of Texas, is used to demonstrate the approach.

Water Marketing Case Study

Lower Rio Grande Valley Case Study

Water marketing (Anderson and Hill 1997) is an emerging nonstructural adaptation strategy for water scarcity. In a water market, users transfer water between different regions or user sectors within the same system. Although a market may be present, it is not straightforward how a city could best use this market to increase the reliability of their supplies. The city must decide how much water to take from traditional, nonmarket supplies as well as to what extent the market should be utilized. Moreover, each planning alternative is evaluated with respect to multiple conflicting objectives, including the cost of providing water, the engineering performance (i.e., can portfolios meet demands), the efficiency of the city's transfer schemes, and providing water for other uses.

This paper uses a case study of a single city's use of a water market in the Lower Rio Grande Valley (LRGV) in Texas, using multiple supply instruments in a portfolio approach for water supply planning (Characklis et al. 2006; Brown and Carriquiry 2007; Palmer and Characklis 2009; Zeff and Characklis 2013). These complex portfolio problems are appropriate for testing Arrow's Paradox because they pose challenging search properties, including severe constraints, stochasticisity, and mixed real and discrete decisions, suited to MOEA optimization (Reed et al. 2013).

The LRGV simulation model is written in C++, presented by Characklis et al. (2006), Kirsch et al. (2009), and Kasprzyk et al. (2009). The following is a brief overview of the LRGV test case. The city has a portfolio of three planning instruments: permanent rights, spot leases, and an adaptive options contract. In the case study, portfolios are controlled by up to eight decision variables that control market-based and nonmarket-based supplies. Permanent rights are assigned based on a percentage of reservoir inflow in every month, using the volume of the city's rights relative to the entire volume of regional rights. Additionally, leases and options control the city's market acquisition of water from the irrigation sector to municipal use. Leases are immediate transfers of water, in any month, at a variable price. The options contract guarantees a fixed price for acquisitions later in the year, and a fixed date of acquisition, similar to an European stock-option contract. In the model, a given supply portfolio is analyzed under an ensemble of uncertain hydrology, demands, and lease pricing, generated using a Monte Carlo simulation of monthly historical data. Therefore, objective function calculations from the simulation model are noisy, and provide distributions of performance for the supply portfolio.

Review of Prior Water Marketing Problem Formulations

Prior studies in water marketing mainly focus on maximizing net social welfare or payoff (Vaux and Howitt 1984; Smith and Marin 1993)

J. Water Resour. Plann. Manage., 2016, 142(2): 04015053

or minimizing cost (Lund and Israel 1995; Watkins and McKinney 1997; Characklis et al. 2006; Kirsch et al. 2009). The studies that maximize net social payoff seek to balance the producers' excess revenue of unused supplies that are sold on the market, and buyers' savings gained by participating in a market (Vaux and Howitt 1984). These approaches use economic equilibrium models that assume that each actor has perfect knowledge of the system, and that all actors exhibit perfect competition. Lund and Israel (1995) designed a water market not as a regional equilibrium model but rather as a multiple-stage decision-making problem for a single utility. The objective of the optimization was to minimize cost, including the cost of the initial contract, the cost of physically transferring the water, drought conservation cost, and the cost of emergency water for a drought. Constraints in the model addressed the reliability of the system; they ensured the amount of water used cannot exceed the requested contract and that the supply must meet the demand. Watkins and McKinney (1997) modified the objective function to explicitly consider variance in the cost distribution. The LRGV case study was also initially introduced using a single-objective cost minimization (Characklis et al. 2006; Kirsch et al. 2009), subject to reliability and cost-variability constraints.

This brief review is provided to help motivate the set of objective functions in the simplified formulations considered in this study. Broadly, prior water marketing optimization studies broadly focused on cost and reliability as two types of objectives. The cost objective uses monetary terms to express the fiscal requirements of a plan and the benefits incurred by users. Some studies also use a measure of reliability; this was explicitly included in the initial LRGV problem (Characklis et al. 2006; Kirsch et al. 2009) and implicitly included in the demand constraints of Lund and Israel (1995) and Watkins and McKinney (1997). Therefore, in this study, the set of objective functions in the simplified problem formulations will similarly seek to minimize the cost and maximize the reliability of LRGV supply portfolios.

Methods

Problem Formulations

Four problem formulations are used in this study in order to explore two types of simplification for the LRGV: lowering the number of decision variables considered, and simplifying and aggregating the objectives. Lowering the number of decision variables seeks to create a simpler supply policy that is easier to optimize. Furthermore, the simplified objective function set (with Aggregated Cost and Reliability objectives) seeks to clarify the implications of Franssen (2005) showing the linkage between Arrow's Paradox and multiobjective aggregation schemes. In other words, can a formulation based only on cost and reliability (based on prior literature) provide the same diversity of solutions as a full many-objective approach to the planning problem?

Table 1 shows the four formulations. Decision variable simplification is denoted by the following terms: Full refers to problems with all decision variables, and Simplified refers to the smaller set of three decision variables. Objective simplification is denoted by the following terms: Many Objective refers to problems where all six objectives are solved independently, and Cost and Reliability refers to problems that have two objectives, each of which combine several measures of cost, and reliability, respectively. The following sections describe these simplifications in detail.

Decision Variable Simplification

The main motivation of LRGV decision variable simplification comes from a prior study (Kasprzyk et al. 2012), in which the decision variables were reduced in order to make a water supply policy that is less complex and that could be easier to implement in practice. In the Full formulation, variables that controlled market acquisitions are adaptive and seasonal in nature, whereas in the Simplified formulation, the decisions are less complex and less adaptive.

The Full complexity formulations contain eight decision variables in total. Permanent rights volume is defined by N_R. Three decision variables control the adaptive options contract: the first specifies a low volume alternative $N_{o\,\mathrm{low}}$; the second variable gives a high volume alternative $N_{o\,\mathrm{high}}$; and the third variable is a threshold on the city's current supply used to select either alternative (ξ). The city's market acquisitions are controlled by a series of risk-based thresholds using sets of two variables. In a given month, α controls timing of water acquisitions (i.e., when water is bought), and β controls the volume of the acquisition (i.e., how much is bought). The Full complexity formulations specify four variables that are designed to take advantage of seasonal patterns: one value of α and β for January to April, and another set for May through December.

The Simplified decision variable formulations use only three decision variables, permanent rights (N_R), a simple nonadaptive options contract that has a single volume (N_O) and a single variable that controls both timing and volume of acquisitions in all months (α). In other words, in the Simplified formulations, α and β are effectively constrained to be the same value; if the value is 1.3, the city is triggered to acquire water when the value falls below 1.3, as well as using 1.3 as its volumetric acquisition.

Although there exist promising approaches for automatic selection and diagnosis of input variables (May et al. 2008a, b; Galelli and Castelletti 2013), there are several challenges for decision variable simplification in many objective analysis. Simplified decision variables are potentially easier for policymakers to understand and implement, and the specific set of Simplified variables were shown

Table 1. Problem Variants

Description	Objectives	Decisions
Cost reliability full	Aggregated cost Aggregated reliability	N_R, $N_{o\,\mathrm{low}}$, $N_{o\,\mathrm{high}}$, ξ, $\alpha_{\mathrm{Jan-Apr}}$, $\alpha_{\mathrm{May-Dec}}$, $\beta_{\mathrm{Jan-Apr}}$, $\alpha_{\mathrm{May-Dec}}$
Cost reliability simplified	Aggregated cost Aggregated reliability	N_R, N_O, α —
Many-objective full	(10-year) cost, critical reliability Surplus water, dropped transfers (Drought) transfer costs	N_R, $N_{o\,\mathrm{low}}$, $N_{o\,\mathrm{high}}$, ξ, $\alpha_{\mathrm{Jan-Apr}}$, $\alpha_{\mathrm{May-Dec}}$, $\beta_{\mathrm{Jan-Apr}}$, $\alpha_{\mathrm{May-Dec}}$
Many-objective simplified	(10-year) cost, critical reliability Surplus water, dropped transfers (Drought) transfer costs	N_R, N_O, α — —

J. Water Resour. Plann. Manage., 2016, 142(2): 04015053

to yield more robust performance in Kasprzyk et al. (2013) compared to the Full problem variants. However, the solutions' objective function performance was lower in some objectives than the full-complexity decision variable set. These results underscore that decision variables have a nuanced effect on many-objective decision support; lower numbers of decision variables can dramatically change the structure of the Pareto-approximate set (van Werkhoven et al. 2009).

Objective Function Simplification

Optimization problem formulations require objectives and constraints, both of which broadly are constructed using quantitative performance metrics, denoted here with f and a subscript. A given performance metric can either be optimized (treated as an objective) or set as an equality or inequality constraint. The following text provides a brief description of all performance metrics used in the formulations, with the full definitions provided in Kasprzyk et al. (2012). As discussed in introducing the objective names, two time horizons are used in the study: a 10-year planning horizon and a single year of severe drought. Each formulation in this study is constructed from performance metrics in each of the three informal categories: cost, reliability, and efficiency.

$f_{10\,\text{yr. cost}}$ refers to the expected annual cost for the full 10-year planning horizon. Cost variability ($f_{10\text{ yr. costvar}}$) calculates the mean of the costs falling above the 90th percentile, dividing by the expected annual cost. An additional metric quantifies the cost of transfers during the drought simulation ($f_{\text{dr. trans. cost}}$). When any of these metrics are treated as objectives, they are to be minimized.

Generally, reliability quantifies the likelihood of supply failure, with each individual metric using a different definition of what constitutes a failure. First, the 10-year reliability ($f_{10\,\text{yr. rel}}$) uses a strict comparison of whether expected supply is higher than expected demand, using the 10-year planning horizon. The drought reliability ($f_{\text{dr. rel}}$) performs a similar calculation within the drought simulation. The 10-year critical reliability ($f_{10\,\text{yr. crit. rel}}$) extends the definition of failure to constitute severe cases in which supply falls lower than 60% of demand. This measure was introduced to help portfolios avoid situations that could not be easily mitigated by demand conservation (Kasprzyk et al. 2009). Metrics used here as objectives are maximized.

Third, a set of measures quantifies the efficiency of each portfolio. $f_{10\,\text{yr. surplus}}$ determines the average amount of water left over from year to year, $f_{10\,\text{yr. dropped}}$ determines the volume of transfers that are dropped due to non-use, and $f_{10\,\text{yr. num. leases}}$ determines the number of leases that are specified by each portfolio, as a proxy for transactions cost. When measures from this group are chosen as objectives, they are to be minimized.

Based on these performance metrics, a Many-Objective formulation for the LRGV has been adapted from prior work (Kasprzyk et al. 2012, 2013) and is shown in the following set of equations:

$$\mathbf{F_{MO}}(\mathbf{x}) = (f_{10\,\text{yr. cost}}, f_{10\,\text{yr. surplus}}, f_{10\,\text{yr. crit. rel}}, f_{10\,\text{yr dropped}}, f_{10\,\text{yr num. leases}}, f_{\text{dr. trans. cost}}) \quad \forall\ \mathbf{x} \in \Omega \tag{2}$$

$$\text{Subject to: } c_{10\,\text{yr. rel}}{:}f_{10\,\text{yr. rel}} \geq 0.98 \tag{3}$$

$$c_{10\,\text{yr. costvar}}{:}f_{10\,\text{yr. costvar}} \leq 1.1 \tag{4}$$

$$c_{10\,\text{yr. crit. rel}}{:}f_{10\,\text{yr. crit. rel}} \geq 0.99 \tag{5}$$

$$c_{\text{dr. rel.}}{:}f_{\text{dr. rel.}} = 1.00 \tag{6}$$

In the formulation, $\mathbf{x}$ = vector of decision variables, which must be in the feasible space of upper and lower bounds, denoted by Ω. The Many-Objective formulation is a vector-valued objective function, F_{MO}, that optimizes multiple objectives simultaneously (minimizing or maximizing them depending on the variable).

Each solution must meet a set of constraints indicated in Eqs. (3) through (6). Unlike classical optimization techniques that require constraints to meet mass balance restrictions or physical realities within a system, constraints are present here to set acceptable limits on performance. Because these constraints are assumed to be strict requirements on portfolio performance, the constraint set is equivalent in both the Many-Objective and Cost Reliability formulations [Eqs. (8)–(11)].

The objective function simplification used to create the Cost Reliability formulations is motivated by the existing body of work in water marketing optimization, where formulations focused on cost and reliability objectives. In the objective function simplification, two weighted objectives are created: one for cost, and one for reliability. In each objective, multiple constituent metrics are added together [Eq. (1)]; a body of literature in water planning provides examples of adding metrics together to make aggregate objectives (Cohon and Marks 1975; Gershon and Duckstein 1983; Palmer and Lund 1985; Watkins and McKinney 1997). The Cost Reliability formulation is presented in the following equations:

$$\mathbf{F_{cost-rel}}(\mathbf{x}) = (f_{\text{agg. cost}}, f_{\text{agg. rel}}) \tag{7}$$

$$\forall \mathbf{x} \in \Omega \qquad \text{Subject to: } c_{10\,\text{yr. rel}}{:}f_{10\,\text{yr. rel}} \geq 0.98 \tag{8}$$

$$c_{10\,\text{yr. costvar}}{:}f_{10\,\text{yr. costvar}} \leq 1.1 \tag{9}$$

$$c_{10\,\text{yr. crit. rel}}{:}f_{10\,\text{yr. crit. rel}} \geq 0.99 \tag{10}$$

$$c_{\text{dr. rel.}}{:}f_{\text{dr. rel.}} = 1.00 \tag{11}$$

The Aggregated Cost objective combines multiple cost metrics into one equation: the 10-year cost and cost variability, with the drought transactions cost. Due to issues of scale between the three constituent performance metrics, each term is scaled to have commensurate magnitudes in the objective function calculation

$$f_{\text{agg. cost}}(\mathbf{x}) = \frac{f_{10\,\text{yr. cost}}}{1.0 \times 10^7} + \frac{f_{10\,\text{yr. costvar}}}{10} + \frac{f_{\text{dr. trans. cost}}}{1.0 \times 10^7} \tag{12}$$

Similarly, Aggregated Reliability ($f_{\text{agg. rel}}$) combines the three reliabilty measures. The metrics' similar magnitudes means that they do not need to be scaled as in the cost aggregation

$$f_{\text{agg. rel}}(\mathbf{x}) = \frac{1}{3}(f_{10\,\text{yr. rel}} + f_{10\,\text{yr. crit. rel}} + f_{\text{dr. rel}}) \tag{13}$$

The Cost Reliability formulation represents a reduction in the number of objectives (as well as some objectives adding constituent metrics together). Various techniques have been presented for reducing the number of objectives (Brockhoff and Zitzler 2006, 2007, 2009; Deb and Saxena 2006; López Jaimes et al. 2008; Jaimes et al. 2009). Notably, a recent water systems application (Giuliani et al. 2014a) demonstrated that even when using dimensionality reduction, many-objective optimization is often still needed. These approaches are complimentary, in the sense that the current generation of MOEAs is limited beyond 10 objectives (Reed et al. 2013; Teytaud 2007).

Borg Multiobjective Evolutionary Algorithm

Although MOEAs have gained prominence in the water resources community (Nicklow et al. 2010; Reed et al. 2013), they are not guaranteed to always find high-quality solutions in a reasonable number of function evaluations. Algorithm failure mechanisms can be classified into two dominant modes: (1) the mathematical structure of a problem formulation is simply too difficult given an MOEA's search operators, and (2) the MOEA algorithmic user choices (i.e., population size, search duration, mating/mutation probabilities, etc.) diminish searches. Recently introduced autoadaptive algorithm technologies (Vrugt and Robinson 2007; Hadka and Reed 2013) seek to alleviate these failure modes by combining multiple algorithms or multiple search operators; the Borg MOEA used in this study is autoadaptive.

The Borg MOEA framework (Hadka and Reed 2013) was chosen because it has exhibited superior search performance relative to other state-of-the-art MOEAs (Hadka and Reed 2012; Woodruff et al. 2012) and its performance has been extensively explored in the context of the LRGV problem (Reed et al. 2013; Reed and Hadka 2014; Hadka and Reed 2015). The Borg MOEA tailors its use of alternative search operators subject to their effectiveness for a given problem (Hadka and Reed 2012, 2013) and helps the user control the precision of output by using epsilon-dominance archiving to scale objective precision (Laumanns 2002). The algorithm's adaptive population sizing (Kollat and Reed 2006) represents a form of time continuation (Goldberg 2002), in that the search is allowed to explore new regions of the decision space with each generation.

Objective function calculations in the LRGV are subjected to noise due to the Monte Carlo simulation in each objective function calculation; in other words, the objective function calculation for a given solution is different each time the calculation is performed. The Borg MOEA's performance has been shown to be resilient to these types of mathematical challenges (Reed et al. 2013), and it has been used successfully in several recent applications (Woodruff et al. 2013; Zeff et al. 2014; Giuliani et al. 2014b). The objective functions average across a large Monte Carlo simulation of uncertainties (Kollat and Reed 2006; Kasprzyk et al. 2009); however the Borg MOEA framework supports other types of treatment of uncertainty that use different sampling of the uncertainty space (Singh and Minsker 2008; Bayer et al. 2010).

Diagnostic Framework and Performance Criteria

A key contribution of this paper is to show whether or not the simplifications proposed in Table 1 affects a MOEA's ability to find high-quality alternatives to the problem formulations. To truly characterize a MOEA's performance, an analyst must track both the algorithm's convergence (how close is the MOEA's approximation set to the true set) and diversity (can the algorithm find points well-spread across the full range of objective function values). For a difficult optimization problem, it is computationally infeasible to find the true Pareto-optimal set. Therefore, algorithm performance criteria on a given problem are calculated with respect to a best-known approximation to the true Pareto set, hereafter termed the problem's reference set.

This study uses a diagnostic framework based on Hadka and Reed (2012) on each of the four problem formulations considered. The framework proceeds as follows. First, feasible ranges for the algorithm parameters are selected (Table 2). Second, a uniform random sample (URS) of size n_{samp} is created of the parameter space. In other words, a MOEA run will yield different results for each of the n_{samp} samples due to algorithm parameterization. Additionally, consideration must be given to random effects within MOEA search (both the initial random population and the random operators), so typically MOEA runs are repeated for n_{seed} random seeds. The total number of algorithm runs is therefore $n_{\text{samp}} \times n_{\text{seed}}$. Let PF_k denote a Pareto-approximate set for a single algorithm run on problem k and PF'_k denote the reference set (the best possible solutions for problem k). To create the reference set, a nondominated sorting procedure is performed across all $n_{\text{samp}} n_{\text{seed}}$ samples to find PF'_k; each problem is considered separately. Last, quantitative algorithm performance metrics show how well each PF_k compares to PF'_k. The remainder of this section defines the three quantiative algorithm performance metrics used in this analysis. In each equation, a given PF is compared to the appropriate problem's reference set PF', following notation of Reed et al. (2013) (i.e., the k subscripts are not used in the subsequent discussion).

Table 2. Sampled Parameter Ranges and Default Settings for the Borg MOEA

Parameter	Minimum	Maximum	Default
Initial population size	10	1,000	100
Maximum evaluations	10,000	100,000	N/A
Injection rate	0.1	1.0	0.25
SBX rate	0.0	1.0	1.0
SBX distribution index	0.0	500.0	15.0
PM rate	0.0	1.0	$1 n_{\text{decvar}}$
PM distribution index	0.0	500.0	20.0
DE crossover rate	0.0	1.0	0.1
DE step size	0.0	1.0	0.5
UM rate	0.0	1.0	$1 n_{\text{decvar}}$
PCX number of parents	2	10	3
PCX number of offspring	1	10	2
PCX eta	0.0	1.0	0.1
PCX zeta	0.0	1.0	0.1
UNDX number of parents	2	10	3
UNDX number of offspring	1	10	2
UNDX eta	0.0	1.0	0.5
UNDX zeta	0.0	1.0	0.35
SPX number of parents	2	10	3
SPX number of offspring	1	10	2
SPX epsilon	0.0	1.0	0.5

Generational Distance (GD) (Van Veldhuizen and Lamont 1998) uses the minimum Euclidean distance between points in PF and the closest corresponding points in PF'

$$GD(PF, PF') = \frac{\sqrt{\sum_{\mathbf{x}\in PF} d_{\mathbf{x}}^2}}{P} \tag{14}$$

and

$$d_{\mathbf{x}} = \min_{\mathbf{y}\in PF'} \sqrt{\sum_{i=1}^{M} [f_i(\mathbf{x}) - f_i(\mathbf{y})^2} \tag{15}$$

where M = number of objectives and P = number of solutions in the approximation set, PF. If there is only one point in the approximation set, and that point is close to one of the points in the reference set, then the value of GD will be near-optimal although the search is not usually considered a success; in other words, GD is often the easiest metric to satisfy.

The additive ε-indicator (EI) (Zitzler et al. 2003) is the worst case distance that would be required to translate PF such that it dominates PF'

$$EI(PF, PF') = \max_{\mathbf{y}\in PF'} \min_{\mathbf{x}\in PF} \max_{1\le i\le M} [f_i(\mathbf{x}) - f_i(\mathbf{y})] \tag{16}$$

Prior work has shown that the additive ε-indicator is sensitive to gaps in the approximation set (Hadka and Reed 2012). In order to

J. Water Resour. Plann. Manage., 2016, 142(2): 04015053

dominate a reference set with gaps, solutions have to be translated further than if gaps were not present. The EI metric can be considered a measure of diversity, since it focuses on the worst-case distance, and solutions have to be well-represented across the full range of objective function values.

The third measure considered is hypervolume (HV) (Zitzler et al. 2003). The hypervolume is a volume filled by a set of points in multidimensional space. The HV metric compares the hypervolume of the approximation set to the hypervolume of the reference set. Because the metric comprehensively determines how much space is filled from every point in the set, it is the most difficult metric to meet (Knowles and Corne 2002)

$$HV(PF, PF') = \frac{\int_V a_{PF}(\mathbf{x})dx}{\int_V a_{PF'}(\mathbf{y})dy} \tag{17}$$

where the volume integral is calculated using normalized objective function values, and a = an attainment function to represent the part of objective space dominated by objectives in a set, S

$$a_S(\mathbf{x}) = \begin{cases} 1 & \text{if } \exists \mathbf{z} \in S \text{ such that } \mathbf{z} \preceq \mathbf{x} \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

Many-Objective Visual Analytics

Many-objective visual analytics (MOVA) refers to a complete framework for MOEA-based decision support where multiple problem formulations are solved with a MOEA and visualized using multiple linked views of multidimensional data (Keim et al. 2010). Two types of visualizations are used in this study. First, objective function tradeoffs are visualized using three-dimensional glyph plots, in which coordinates, in addition to the glyphs' size, orientation, and transparency, directly show objective function values (Fig. 1). These glyphs have been used effectively in a number of MOEA studies (Kollat and Reed 2007; Woodruff et al. 2013; Kasprzyk et al. 2009; Fu et al. 2013). A second type of visualization (Figs. 2–3) summarizes the results of the diagnostic experiments (Hadka and Reed 2012; Reed et al. 2013). These colormaps and distributional results demonstrate a large amount of quantitative information from thousands of MOEA runs, seeking to characterize the most-important components of MOEA performance during search process. Each type of visualization is highly interactive, allowing users to brush solutions that do not meet their criteria (Inselberg 1997) or focus on certain subsets of the plotted data, facilitating stakeholder learning about system properties.

Computational Experiment

Computational experiments in this study seek to determine whether the Borg MOEA framework can reliably solve the four LRGV problem variants of Table 1; each problem here is considered separately. Therefore, the diagnostic framework does not consider optimal or tuned parameterization for the MOEA but rather samples across ranges of the parameter values. Table 2 shows these ranges. To sample the parameter space, a Latin Hypercube Sample (LHS) of size $n_{\text{samp}} = 256$ was used according to the ranges in Table 2. Therefore, the algorithm was run 256 times with a different search parameterization each time, for each problem. Then, all runs were repeated for $n_{\text{seed}} = 50$ random seed replicates. In total, there are four problems, 256 Borg runs per problem, and 50 replicates of each experiment for a total of 51,200 algorithm runs and more than 150,000 hours of computation in this study. For each problem k, the reference set PF'_k represents the best-known solutions for the problem across all these runs.

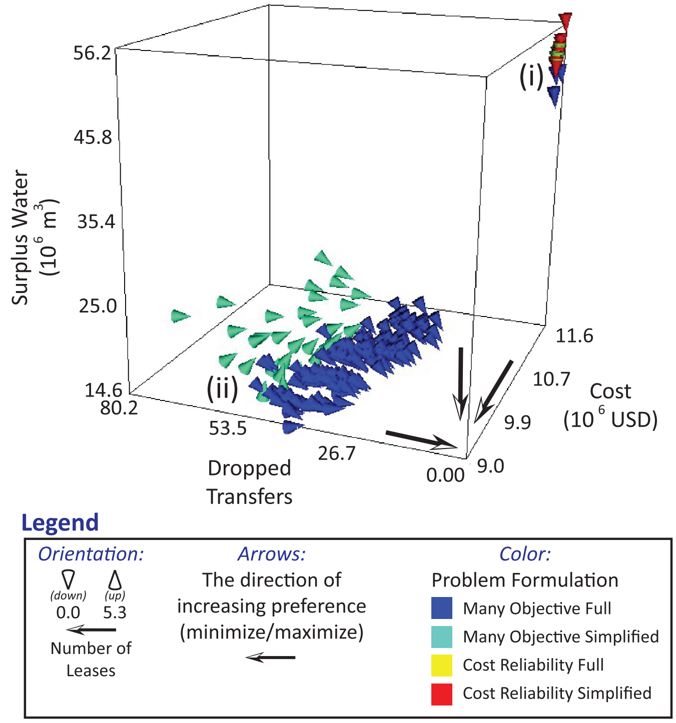

Fig. 1. (Color) Glyph plot comparing the formulations, shown with the color of each cone; spatial axes plot cost, dropped transfers, and surplus water with the arrows indicating the direction of increasing preference; orientation plots the number of leases; solutions in Group 1 lack market usage, whereas solutions in Group 2 use the market to a higher degree; only the Many-Objective formulations were able to find points in Group 2

A key aspect of this study is focused on the biases that aggregation can introduce in decision support. Consequently, the rigor of our diagnostic analysis ensures that the structural effects reported for the alternative formulations' Pareto-approximate sets are not artifacts of the Borg MOEA but emerge as unintended consequences of aggregation (i.e., Arrow's Paradox). The study was also enabled by the use of a massively parallel version of the Borg MOEA that has been shown to scale well even when used on thousands of processors (Hadka et al. 2013, 2015).

A final consideration in setting up the computational experiment is Monte Carlo sampling of hydrology, demands, and lease pricing within the LRGV simulation model itself. Monte Carlo sample sizes were also chosen to preserve computational efficiency. Following Reed et al. (2013), 1,000 Monte Carlo replicates were used within the LRGV simulation model. When calculating the final reference sets, though, a larger Monte Carlo sample size of 10,000 was used, to ensure accurate calculation of the LRGV's performance objectives. Using a larger Monte Carlo size in a reevaluation procedure balances the desire to both have a robust sampling of the LRGV's uncertainties and also preserve the computational requirements for carrying out the diagnostic experiment.

Results and Discussion

Fig. 1 shows the reference sets for each of the four formulations provided in Table 1, obtained by a nondominated sort of all $n_{\text{samp}}n_{\text{seed}}$ algorithm runs for each problem. Each cone is an individual solution from the reference set for each problem variant. The spatial coordinates of each cone indicate the solution's cost

J. Water Resour. Plann. Manage., 2016, 142(2): 04015053

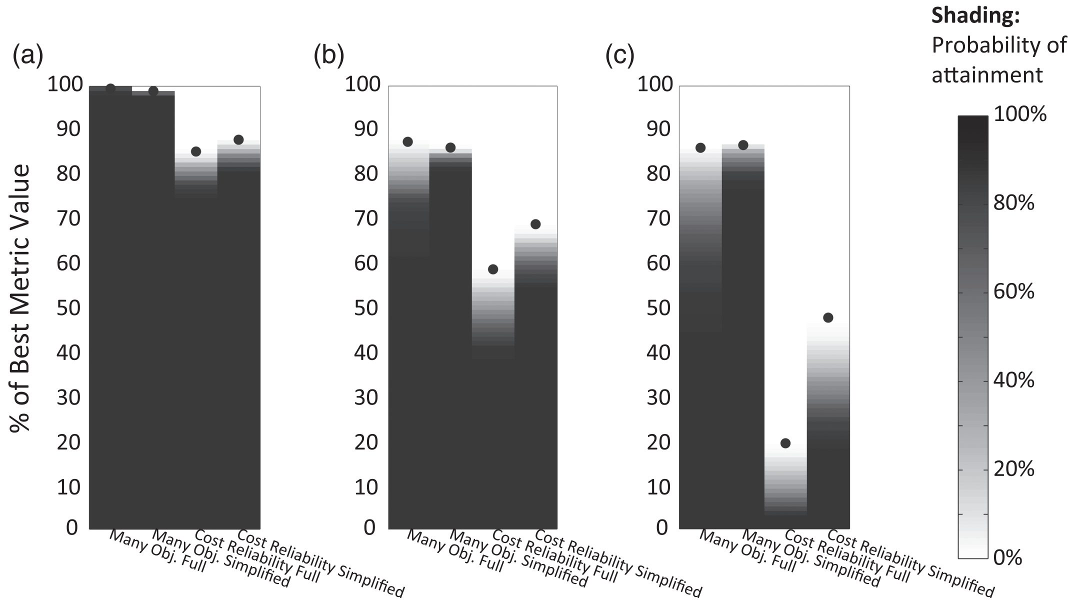

Fig. 2. Plot of the best metric value (shown with a dot), and the probability of attaining the best metric value (shown with shading) across all algorithm parameterizations, for three performance metrics (panels a–c); columns of each panel show the four problem formulations considered in this study: (a) generational distance; (b) epsilon indicator; (c) hypervolume

($f_{10\,\text{yr. cost}}$), dropped transfers ($f_{10\,\text{yr. dropped}}$), and surplus water ($f_{10\,\text{yr. surplus}}$) objective function values. Arrows show the direction of increasing preference for each objective. The ideal solution, with optimal values for all objectives, would be in the bottom right corner of the cube. The orientation of the cones shows their objective function value on the number of leases ($f_{10\,\text{yr. num. leases}}$) objective, with cones pointing toward the top of the figure having more leases than cones pointing downward. The color of the cones designates the solutions attained for each formulation. Each problem variant was considered separately within the Borg MOEA optimization. The solutions attained for the Cost Reliability problems were re-evaluated using the full set of objectives used in the Many-Objective formulation to enable Fig. 1 to show the broader decision implications of aggregation (i.e., Arrow's Paradox).

Similar to results in prior work (Kasprzyk et al. 2009), two groups of solutions emerge. Solutions in Group 1 are similar to what is currently done in practice; the city purchases a large amount of water rights in excess of actual usage (Characklis et al. 1999), which causes poor performance with respect to minimizing their cost or minimizing surplus water. Solutions in Group 2 use the market to a larger degree, lowering their cost and surplus water while trading off slightly higher volumes of dropped or wasted transfers; however, their market use is manageable with relatively low numbers of lease acquisitions on average over the 10 years of simulation. The existence of solutions from Group 2 are important because they show that the MOEA can find innovative, diverse solutions that represent new management alternatives for the system. Furthermore Kasprzyk et al. (2013) showed that solutions from this group are robust to uncertainties, such as sampling from different hydrological inflow distributions, modified assumptions about demand growth, and modified initial conditions.

Fig. 1 demonstrates that there are no solutions from the Cost Reliability formulations that are represented in Group 2. In other words, the aggregation used to create the cost and reliability objective functions has inadvertently caused the search to be biased toward solutions that have high permanent rights usage. From a policy standpoint, these solutions represent traditional management that cannot be adaptive to conditions of water shortage. This provides support for Franssen (2005), who argued that objective aggregation schemes penalize certain objectives in a manner that is difficult or impossible to predict a priori. In this instance, aggregation of cost and reliability measures would inadvertently bias decision-makers from even knowing that market use could dramatically improve the system across all of its measures (i.e., Group 2 in Fig. 1). In absence of the higher dimensional information from the Many-Objective formulation, it would be difficult for stakeholders to understand the value of water transfers, a major mechanism for improving the adaptability of the LRGV system to severe droughts.

The second part of this study presents the results of the diagnostic exploration of the Borg MOEA. By comprehensively testing how difficult it is to solve the Cost Reliability and Many-Objective problems, the analysis challenges the widely-held belief that reducing the number of objectives considered makes a problem easier to solve. The hypothesis in this study design is that the proven difficulty of finding feasible solutions within the LRGV test case (Reed et al. 2013) could actually cause the lower-dimensional aggregate formulations to be more difficult because the Pareto-approximate sets are dramatically smaller and create a needle-in-a-haystack challenge. The higher-dimensional formulation provides more pathways (i.e., decision variable combinations) to attain Pareto-approximate solutions. Critical to this hypothesis is the Borg MOEA's underlying theoretical exploitation of epsilon-dominance archiving to ensure diverse, convergent search for Many-Objective problems (Hadka and Reed 2012, 2013). The first exposition of the results focuses on how likely it is that any parameterization of the algorithm performs well, relative to the three quantitative performance metrics introduced previously. Fig. 2 is termed a best-attainment plot (Reed et al. 2013). The black dot plots the best performance metric value attained across all runs of the Borg MOEA for each problem variant. The shaded area shows the probability of attaining a given value of performance, for the three performance metrics and four problem formulations. The performance metrics are quantitative representations of how close a run of the algorithm on problem k, PF_k, is to the reference set, PF_k'. The plot is similar to a cumulative distribution function of algorithm performance.

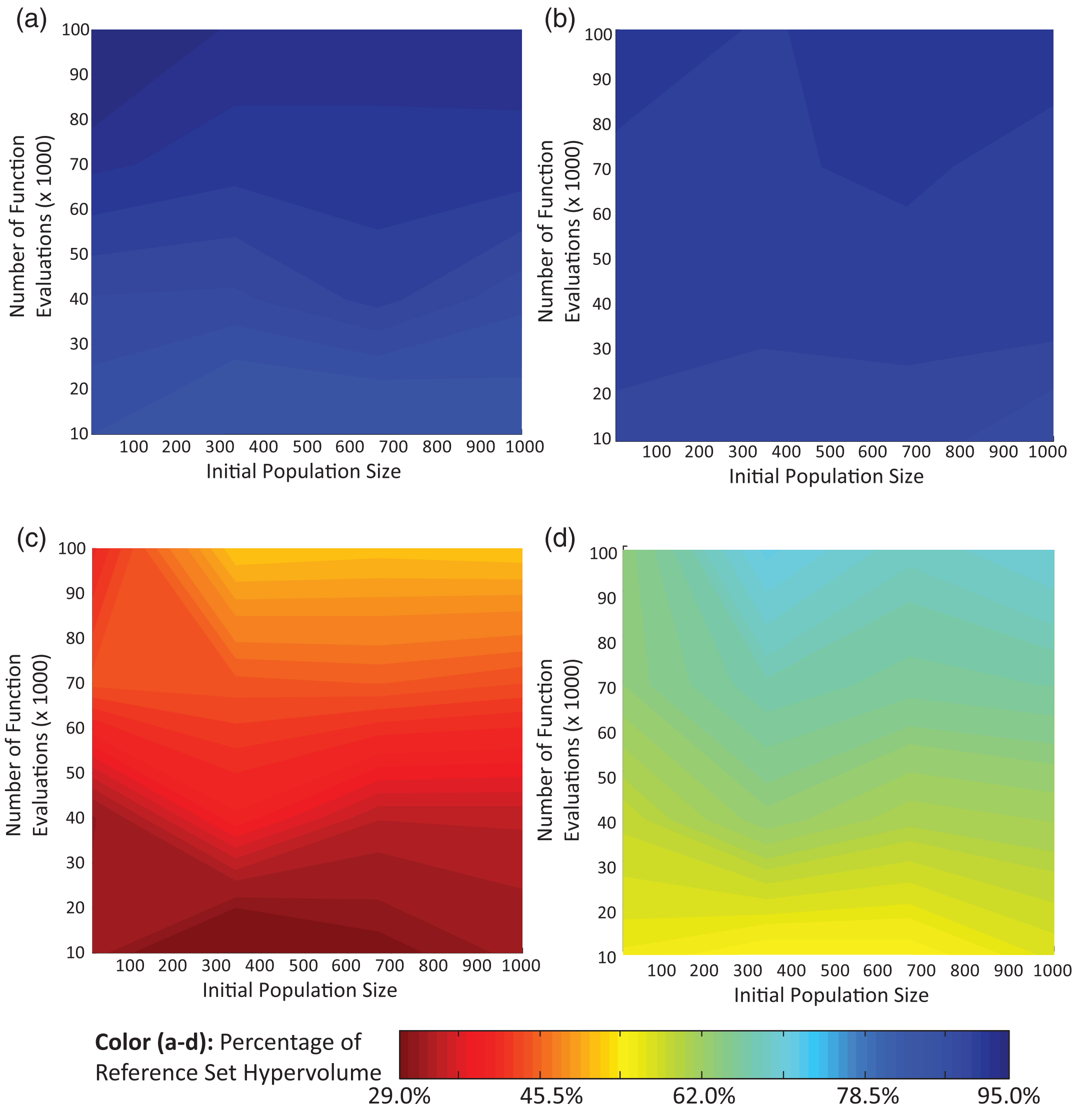

Fig. 3. (Color) Hypervolume control maps for the four problems; the color represents how the hypervolume performance varies for increasing population size and search duration; ideal performance is dark blue, while poor performance is shown with dark red; panels (a)–(d) show the problem formulations considered in the study: (a) Many-Objective Full; (b) Many-Objective Simplified; (c) Cost Reliability Full; (d) Cost Reliability Simplified

Because each of the n_{samp} algorithm runs are replicated n_{seed} times, each result in the MOEA diagnostics must address the replicate runs. In Fig. 2, a performance metric value was calculated for each approximation set, and then the expected value performance was calculated across all random seeds. In essence, this plot shows the probability of a single random seed achieving a given level of search performance. Ideal performance would be that the Borg MOEA would have 100% probability of attaining its best performance (i.e., the dark black shading that meets the best performance black dots). If a problem formulation's black dot falls below 100%, it indicates that no run of the algorithm was able to capture the ideal metric performance on a given problem. To summarize, the plot shows both the effectiveness of the algorithm as well as the reliability of attaining good performance.

Generational Distance (GD) is the first metric considered, in Fig. 2(a). GD is often considered the easiest metric to meet, requiring only one solution that is close to the reference set to obtain good metric performance. Results show that 100% of the algorithm parameterizations were able to obtain 100% of the best possible metric value, on the Many-Objective Full problem. The Many-Objective Simplified problem exhibited similar performance, with a best value of slightly less than 100%. This indicates that the Borg MOEA shows reliable performance on these problem formulations; in almost every algorithm parameterization, the algorithm is able to find a solution close to the problem's reference set. Transitioning to the aggregated problems, the plot shows that the algorithm exhibited a lower level of performance on both the Cost Reliability Full and Cost Reliability Simplified problems. On the Cost Reliability Full problem, the best-performing parameterization of the Borg MOEA obtained less than 90% of the best generational distance, and approximately 60% of the parameterizations obtained 80% of the best metric value. The Cost Reliability Simplified problem showed slightly more reliable performance, but the best value is still significantly lower than on the Many-Objective problems. The implication of poor GD values on the simple problems is that many parameterizations are fully failing, by not getting a single solution close to the problem's reference set. The Cost Reliability problems have created a reference set that is difficult to obtain for many parameterizations of the algorithm, which means that in practice it would be difficult to reliably solve this problem given limitations in computational resources.

Fig. 2(b) shows the algorithm's performance on the four problems, using the Epsilon Indicator (EI) metric. EI is the worst-case distance that one would have to translate an approximation set PF_k to dominate the reference set, PF'_k on a given problem k. The metric is sensitive to gaps, or missing solutions, in the approximation set. Consequently, it is a more difficult metric to satisfy and this is demonstrated in the plot. For the Many-Objective Full problem, the best metric value is slightly less than 90%, which means that the expected value of performance for the best parameterization attained was less than 90% of the best possible EI value. Similar to Fig. 2(a), though, the cost-reliability problems exhibit reduced performance compared to the Many-Objective problems, with the algorithm only achieving a high level of attainment up to about 40% of the best value on the Cost Reliability Full problem, and 60% of the Cost Reliability Simplified problem. Interpreting the EI results indicates that even the best approximation set generated by the algorithm would need to be translated significantly to match the values in the reference set, showing that the aggregation scheme has created solutions that are difficult to replicate.

The performance differences between problems are most apparent on the Hypervolume plot, shown in Fig. 2(c). Because Hypervolume is a metric that captures both convergence and diversity, it is both the most difficult metric to meet and also one of the most useful to diagnose algorithm performance. Here, it is apparent that the Borg MOEA was able to perform well on Many-Objective Full and Many-Objective Simplified problems, but it was unable to find a representative distribution of points on the Cost Reliability problems. The high values of hypervolume on the Many-Objective problems indicates that many runs of Borg MOEA are able to acheive effective performance on these problems; a single run of the algorithm is not expected to succeed in finding good solutions to the Cost Reliability problems, however.

Recall that the problem formulations in Table 1 used two types of simplification: Many-Objective and Cost Reliability (based on objective function) and Full and Simplified (based on decision variable). Looking across all three parts of Fig. 2 can be used to ascertain which of the two types of simplification affected the search performance to a higher degree. On each metric, it appears that the simplified decision variable formulation causes slightly more reliable performance, likely due to the fact that there are fewer degrees of freedom for search. However, the algorithm was always able to solve the Many-Objective problems more reliably and effectively than the Cost Reliability problems. Good performance for the Many-Objective problems, as quantified by these metrics, means that the algorithm can reliably find the full suite of diverse solutions for the LRGV, exhibited by the two regions in Fig. 2. However, for the Cost Reliability problems, the reference set is a limited number of solutions in Region 1 of Fig. 1, and the analysis presented in this study suggests that these solutions are difficult to find and the aggregation and simplification is indeed biasing MOEA search. The implication of these results is that the Cost Reliability formulations are collectively more difficult to address and fully ignore the market use inadvertently in a manner that corresponds well with Franssen (2005)'s concerns related to Arrow's Paradox.

Fig. 2 is useful because it explains how likely it is for any algorithm parameterization to attain good performance. However, such figures do not lend insight as to the value of algorithm parameters that yield good performance. In other words, a visualization is needed that explains the ease-of-use of the algorithm, where users can attain high-quality results in a straightforward manner. Goldberg (2002) characterized this concept as controllability. If an algorithm is termed controllable for a problem, many values of the algorithm parameters will still yield good performance on the problem (Reed et al. 2013). Fig. 3 provides a control map, which is a way to ascertain algorithm controllability. The plots use a surface of color to illustrate Hypervolume values across dimensions of algorithm parameters. The color indicates the percentage of each reference set's hypervolume that was found by each parameterization. Dark blue color indicates a high level of hypervolume, whereas dark red indicates that the algorithm's parameterization achieved lower hypervolume values. Specifically, dark red values indicate that a parameterization was only able to find 29% of the reference set's hypervolume. Ideal performance would be a dark blue color over the whole map, indicating all parameterizations found all of the problem's hypervolume. The two axes of the plots illustrate the algorithm's initial population size versus the number of function evaluations, or search duration. Although the LHS ensemble sampled all of Borg's search parameters, the two visualized parameters were chosen because they are often the most-important determinants of an algorithm's success (Reed et al. 2013).

As discussed previously, the best-attainment plot in Fig. 2 focuses on the probability that a single random seed will achieve good performance on a problem, so the results represent an expected value across all n_{samp} replicates. In the control maps of Fig. 3, the approximation sets PF that are being evaluated using the Hypervolume metric represent an individual nondominated sorting procedure across all of the random seed replicates of a given parameterization of a problem. In other words, n_{samp} different approximation sets are generated, and the best known results from those runs are combined together into an approximation set for each problem that is then tested with the metrics. The reference set of each problem is the same as previously described. This is done to provide a better demonstration of the effects of the actual parameterization values on performance. If an analyst were to run a given parameterization of the Borg MOEA, he or she would likely replicate these runs over multiple random seeds, so this study's nondominated sort of the results mimics this process (i.e., it is standard practice to perform random seed analysis).

The four panels of Fig. 3 show the results for the four problem variants. Observing the results, it can be seen that the Borg MOEA performed very well on both of the Many-Objective problems, with all parameterizations finding more than 90% of each problem's Hypervolume. In both Figs. 3(a and b), one of the only discernible trend in the results is that the algorithm's performance improves slightly with a longer search duration. Figs. 3(c and d) reiterate that the aggregated problem has created a situation where Borg cannot find a large percentage of the Cost Reliability problem's reference set. The algorithm found between 29 and 50% of the reference set hypervolume of the Cost Reliability Full problem, and the performance was slightly better on the Cost Reliability Simplified problem. Looking at the response surface of both of the plots, it can also be seen that there were regions of the parameter space that performed slightly better than others. In other words, there is a more complicated relationship than simply increasing search duration, as seen in Figs. 3(a and b).

Looking across all four problems in Fig. 3, the Borg MOEA can be considered controllable for the two Many-Objective problems. Although its performance is consistent on the Cost Reliability problems (i.e., no drastic changes across population size or run duration), the lower Hypervolume metric values suggests that is difficult to find a high percentage of the reference set on any given run or parameterization of the algorithm. Therefore, the results tend to suggest that the additional objectives in the Many-Objective problems (surplus water, dropped transfers, etc.) have actually made the LRGV problem easier to solve. Furthermore, even if solutions in the Cost Reliability problem are desirable, our control maps suggest that it is very difficult to actually search for and find the best-known approximation set for these problems. In other words, the simplification and aggregation of objectives has biased the search and made it more difficult.

Conclusion

Although Franssen (2005) advanced the argument that aggregated formulations can severely bias decisions in multiobjective systems, this concept was not yet explored in water resources planning and management. This study sought to answer two fundamental questions regarding problem formulation simplification and aggregation. The first question wanted to determine whether or not simplified formulations with aggregated objectives and simplified decisions were biasing our ability to find diverse tradeoffs and innovative solutions for complex water management problems. The second question explored a practical consideration: how does this simplification improve the ability for a search algorithm to find solutions to a problem? The paper used MOEA-based decision support of a water marketing problem in the Lower Rio Grande Valley of Texas to explore the issues.

The results showed that the Cost Reliability formulation was severely limited in its ability to yield diverse solutions. In fact, the solutions that were found using this formulation strongly mapped to current management in the LRGV system. Without expanding the problem definition to include many planning objectives, market-based solutions would never have been found.

The diagnostic framework used to evaluate the Borg MOEA yielded some surprising results. In addition to having low diversity, the Cost Reliability formulations also presented a challenge to the search algorithm. There were consistent failures on the aggregated problems, exhibited by consistently lower values in the quantitative performance metrics from the diagnostic study. Therefore, the results of the study show that for the LRGV, increased objective counts lead to more effective, efficient, reliable, and controllable searches. This result, which is counter to conventional assumptions in the field, shows promise for similar problem formulations to be pursued. This finding was made possible by the self-adaptive properties of the Borg MOEA, an algorithm that was shown to do better than many other state-of-the-art MOEAs on the LRGV problem (Reed et al. 2013). The Borg MOEA's efficient performance means that it can be used in rapid fashion to solve multiple versions of a problem, allowing analysts to discover new objectives as part of the optimization process itself (Reed and Kasprzyk 2009; Kasprzyk et al. 2012).

Several recent advances in water resources planning and management are enabling many new advances in the field, including more-reliable and easy-to-use search tools, a larger amount of access to high-performance and cloud computing, and immersive, interactive visualizations. These advances are both an opportunity and a challenge in this field. Results of this study show that some classic systems assumptions such as aggregated, low-dimensional problem formulations may need to be revisited in light of emerging technologies. Many-Objective analysis shows promise to overcome Arrow's Paradox and contribute to better, more sustainable water management solutions in the future.

Acknowledgments

This research was supported in part by the Extreme Science and Engineering Discovery Environment (XSEDE) supported by the National Science Foundation, and by the University of Texas Advanced Computing Center under Grant No. TG-EAR090013. The contributions of anonymous reviewers are also graciously acknowledged.

References

Anderson, T. L., and Hill, P. J., eds. (1997). *Water marketing: The next generation*, Rowman and Littlefield, Lanham, MD.

Andreu, J., Capilla, J., and Sanchis, E. (1996). "AQUATOOL, a generalized decision-support system for water-resources planning and operational management." *J. Hydrol.*, 177(3–4), 269–291.

Arnold, J. L. (1988). "The evolution of the 1936 flood control act." U.S. Army Corps of Engineers, Fort Belvoir, VA.

Arrow, K. J. (1950). "A difficulty in the concept of social welfare." *J. Polit. Econ.*, 58(4), 328–346.

Bayer, P., de Paly, M., and Burger, C. M. (2010). "Optimization of high-reliability-based hydrological design problems by robust automatic sampling of critical model realizations." *Water Resour. Res.*, 46(5), in press.

Borsuk, M., Clemen, R., Maguire, L., and Reckhow, K. (2001). "Stakeholder values and scientific modeling in the neuse river watershed." *Group Decis. Neg.*, 10(4), 355–373.

Brockhoff, D., and Zitzler, E. (2006). "Are all objectives necessary? On dimensionality reduction in evolutionary multiobjective optimization." *Parallel problem solving from nature IX*, T. Olivier,, et al., eds., Springer, Berlin, 533–542.

Brockhoff, D., and Zitzler, E. (2007). "Improving hypervolume-based multiobjective evolutionary algorithms by using objective reduction methods." *IEEE Congress on Evolutionary Computation, CEC 2007*, IEEE, 2086–2093.

Brockhoff, D., and Zitzler, E. (2009). "Objective reduction in evolutionary multiobjective optimization: Theory and applications." *Evol. Comput.*, 17(2), 135–166.

Brown, C., and Carriquiry, M. (2007). "Managing hydroclimatological risk to water supply with option contracts and reservoir index insurance." *Water Resour. Res.*, 43, W11423.

Chankong, V., and Haimes, Y. (1983). *Multiobjective decision making: Theory and methodology*, North-Holland, New York.

Characklis, G., Kirsch, B. R., Ramsey, J., Dillard, K., and Kelley, C. T. (2006). "Developing portfolios of water supply transfers." *Water Resour. Res.*, 42(5), W0540.

Characklis, G. W., Griffin, R. C., and Bedient, P. B. (1999). "Improving the ability of a water market to efficiently manage drought." *Water Resour. Res.*, 35(3), 823–831.

Coello Coello, C. A., Lamont, G. B., and Van Veldhuizen, D. A., eds. (2007). "Evolutionary algorithms for solving multi-objective problems." *Genetic and evolutionary computation*, 2nd Ed., Springer, New York.

Cohon, J., and Marks, D. (1975). "A review and evaluation of multiobjective programming techniques." *Water Resour. Res.*, 11(2), 208–220.

Deb, K., and Saxena, D. K. (2006). "Searching for Pareto-optimal solutions through dimensionality reduction for certain large-dimensional multiobjective optimization problems." *2006 IEEE Congress on Evolutionary Computation (CEC'2006)*, IEEE, 3353–3360.

Franssen, M. (2005). "Arrow's theorem, multi-criteria decision problems and multi-attribute preferences in engineering design." *Res. Eng. Des.*, 16(1–2), 42–56.

Fu, G., Kapelan, Z., Kasprzyk, J., and Reed, P. (2013). "Optimal design of water distribution systems using many-objective visual analytics." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000311, 624–633.

Galelli, S., and Castelletti, A. (2013). "Tree-based iterative input variable selection for hydrological modeling." *Water Resour. Res.*, 49(7), 4295–4310.

Gershon, M., and Duckstein, L. (1983). "Multiobjective approaches to river basin planning." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1983)109:1(13), 13–28.

Giuliani, M., Galelli, S., and Soncini-Sessa, R. (2014a). "A dimensionality reduction approach for many-objective Markov decision processes: Application to a water reservoir operation problem." *Environ. Modell. Software*, 57, 101–114.

Giuliani, M., Herman, J. D., Castelletti, A., and Reed, P. (2014b). "Many-objective reservoir policy identification and refinement to reduce policy inertia and myopia in water management." *Water Resour. Res.*, 50, 3355–3377.

Goldberg, D. E. (2002). *The design of innovation: Lessons from and for competant genetic algorithms*, Kluwer Academic, Boston.

Hadka, D., and Reed, P. (2012). "Diagnostic assessment of search controls and failure modes in many-objective evolutionary optimization." *Evol. Comput.*, 20(3), 423–452.

Hadka, D., and Reed, P. (2013). "Borg: An auto-adaptive many-objective evolutionary computing framework." *Evol. Comput.*, 21(2), 231–259.

Hadka, D., and Reed, P. (2015). "Large-scale parallelization of the borg multiobjective evolutionary algorithm to enhance the management of complex environmental systems." *Environ. Modell. Softw.*, 69, 353–369.

Haimes, Y. Y. (1977). *Hierarchical analyses of water resources systems, modeling and optimization of large scale systems*, McGraw-Hill, New York.

Hashimoto, T., Stedinger, J. R., and Loucks, D. P. (1982). "Reliability, resiliency and vulnerability criteria for water resource system performance evaluation." *Water Resour. Res.*, 18(1), 14–20.

Heidari, M., Chow, V. T., Kokotovic, P. V., and Meredith, D. D. (1971). "Discrete differential dynamic programming approach to water resources systems optimization." *Water Resour. Res.*, 7(2), 273–282.

Hitch, C. J. (1960). "On the choice of objectives in systems studies." *Rep. No. P-1955*, RAND, Santa Monica, CA.

Huang, G. H., and Loucks, D. P. (2000). "An inexact two-stage stochastic programming model for water resources management under uncertainty." *Civ. Eng. Environ. Syst.*, 17(2), 95–118.

Inselberg, A. (1997). "Multidimensional detective." *Proc., IEEE Symp. on Information Visualization, 1997*, IEEE, 100–107.

Jaimes, A. L., Coello, C. A. C., and Barrientos, J. E. U. (2009). "Online objective reduction to deal with many-objective problems." *Evolutionary multi-criterion optimization*, M. Ehrgott, C. M. Fonseca, X. Gandibleux, J.-K. Hao, and M. Sevaux, eds., Springer, Berlin, 423–437.

Kasprzyk, J. R., Nataraj, S., Reed, P. M., and Lempert, R. J. (2013). "Many objective robust decision making for complex environmental systems undergoing change." *Environ. Modell. Software*, 42, 55–71.

Kasprzyk, J. R., Reed, P. M., Characklis, G. W., and Kirsch, B. R. (2012). "Many-objective de novo water supply portfolio planning under deep uncertainty." *Environ. Modell. Software*, 34, 87–104.

Kasprzyk, J. R., Reed, P. M., Kirsch, B. R., and Characklis, G. W. (2009). "Managing population and drought risks using many-objective water portfolio planning under uncertainty." *Water Resour. Res.*, 45(12), in press.

Keim, D. A., Kohlhammer, J., Ellis, G., and Mansmann, F., eds. (2010). *Mastering the information age—Solving problems with visual analytics*, Eurographics Association, Goslar, Germany.

Kirsch, B. R., Characklis, G. W., Dillard, K. E. M., and Kelley, C. T. (2009). "More efficient optimization of long-term water supply portfolios." *Water Resour. Res.*, 45(3), W03414.

Knowles, J., and Corne, D. (2002). "On metrics for comparing non-dominated sets." *Proc., 2002 World Congress on Computational Intelligence (WCCI)*, IEEE, 711–716.

Kollat, J. B., and Reed, P. M. (2006). "Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design." *Adv. Water Resour.*, 29(6), 792–807.

Kollat, J. B., and Reed, P. M. (2007). "A framework for visually interactive decision-making and design using evolutionary multiobjective optimization (VIDEO)." *Environ. Modell. Software*, 22(12), 1691–1704.

Labadie, J. W. (2004). "Optimal operation of multireservoir systems: State-of-the-art review." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(2004)130:2(93), 93–111.

Laumanns, M. (2002). "Combining convergence and diversity in evolutionary multiobjective optimization." *Evol. Comput.*, 10(3), 263–282.

López Jaimes, A., Coello Coello, C. A., and Chakraborty, D. (2008). "Objective reduction using a feature selection technique." *Proc., 10th Annual Conf. on Genetic and Evolutionary Computation*, Association for Computing Machinery (ACM), New York, 673–680.

Lund, J. R., and Israel, M. (1995). "Optimization of transfers in urban water supply planning." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1995)121:1(41), 41–48.

Maier, H., et al. (2014). "Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions." *Environ. Modell. Software*, 62, 271–299.

May, R. J., Dandy, G. C., Maier, H. R., and Nixon, J. B. (2008a). "Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems." *Environ. Modell. Software*, 23(10–11), 1289–1299.

May, R. J., Maier, H. R., Dandy, G. C., and Fernando, T. M. K. G. (2008b). "Non-linear variable selection for artificial neural networks using partial mutual information." *Environ. Modell. Software*, 23(10–11), 1312–1326.

Nicklow, J. (2010). "State of the art for genetic algorithms and beyond in water resources planning and management." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000053, 412–432.

Palmer, R. N., and Characklis, G. W. (2009). "Reducing the costs of meeting regional water demand through risk-based transfer agreements." *J. Environ. Manage.*, 90(5), 1703–1714.

Palmer, R. N., and Lund, J. R. (1985). "Multi-objective analysis with subjective information." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1985)111:4(399), 399–416.

Reed, P., Hadka, D., Herman, J., Kasprzyk, J., and Kollat, J. (2013). "Evolutionary multiobjective optimization in water resources: The past, present and future." *Adv. Water Resour.*, 51, 438–456.

Reed, P. M., and Hadka, D. (2014). "Evolving many-objective water management to exploit exascale computing." *Water Resour. Res.*, 50(10), 8367–8373.

Reed, P. M., and Kasprzyk, J. R. (2009). "Water resources management: The myth, the wicked, and the future." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)WR.1943-5452.0000047, 411–413.

Reuss, M. (2005). "Ecology, planning, and river management in the United States: Some historical reflections." *Ecol. Soc.*, 10(1), 34.

Singh, A., and Minsker, B. S. (2008). "Uncertainty-based multiobjective optimization of groundwater remediation design." *Water Resour. Res.*, 44(2), W02404.

Smith, M. G., and Marin, C. M. (1993). "Analysis of short-run domestic water supply transfers under uncertainty." *Water Resour. Res.*, 29(8), 2909–2916.

Teytaud, O. (2007). "On the hardness of offline multi-objective optimization." *Evol. Comput.*, 15(4), 475–491.

Tsoukias, A. (2008). "From decision theory to decision aiding methodology." *Eur. J. Oper. Res.*, 187(1), 138–161.

U.S. Council on Environmental Quality. (2013). "Principles and requirements for federal investments in water resources." *Rep. No.*

van Veldhuizen, D. A., and Lamont, G. B. (1998). "Evolutionary computation and convergence to a Pareto front." *Proc., 3rd Annual Conf.Genetic Programming 1998*, J. R. Koza, W. Banzhaf, K. Chellapilla, K. Deb, M. Dorigo, and D. B. Fogel, eds., Morgan Kaufmann, San Francisco, 22–25.

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y. (2009). "Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models." *Adv. Water Resour.*, 32(8), 1154–1169.

Vaux, H. J., Jr., and Howitt, R. E. (1984). "Managing water scarcity: An evaluation of interregional transfers." *Water Resour. Res.*, 20(7), 785–792.

Vrugt, J. A., and Robinson, B. A. (2007). "Improved evolutionary optimization from genetically adaptive multimethod search." *Proc. Natl. Acad. Sci.*, 104(3), 708–711.

Watkins, D. W., Jr., McKinney, D. C. (1997). "Finding robust solutions to water resources problems." *J. Water Resour. Plann. Manage.*, 10.1061/(ASCE)0733-9496(1997)123:1(49), 49–58.

Woodruff, M., Hadka, D., Reed, P. M., and Simpson, T. W. (2012). "Auto-adaptive search capabilities of the new Borg MOEA: A detailed comparison on alternative product family design problem formulations." *14th AIAA/ISSMO Multidisicplinary Analysis and Optimization Conf.*, American Institute of Aeronautics and Astronautics (AIAA), Reston, VA.

Woodruff, M. J., Reed, P. M., and Simpson, T. (2013). "Many objective visual analytics: Rethinking the design of complex engineered systems." *Struct. Multidiscip. Optim.*, 48(1), 201–219.

Zeff, H. B., and Characklis, G. W. (2013). "Managing water utility financial risks through third-party index insurance contracts." *Water Resour. Res.*, 49(8), 4939–4951.

Zeff, H. B., Kasprzyk, J. R., Herman, J. D., Reed, P. M., and Characklis, G. W. (2014). "Navigating financial and supply reliability tradeoffs in regional drought management portfolios." *Water Resour. Res.*, 50, 4906–4923.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and Grunert da Fonseca, V. (2003). "Performance assessment of multiobjective optimizers: An analysis and review." *IEEE Trans. Evol. Comput.*, 7(2), 117–132.