

## Optimal design of water distribution systems using many-objective visual analytics

Guangtao Fu<sup>1</sup>; Zoran Kapelan<sup>2</sup>; Joseph Kasprzyk<sup>3</sup>; and Patrick Reed, M. ASCE<sup>4</sup>

### Abstract

This study investigates the use of many-objective optimization for water distribution system (WDS) design or rehabilitation problems. The term many-objective optimization refers to optimization with four or more objectives. The increase in the number of objectives brings new challenges for both optimization and visualization. This study uses a multi-objective evolutionary algorithm termed the epsilon Nondominated Sorted Genetic Algorithm II ( $\epsilon$ -NSGAII) and interactive visual analytics to reveal and explore the tradeoffs for the Anytown network problem. The many-objective formulation focuses on a suite of six objectives: (1) capital cost, (2) operating cost, (3) hydraulic failure, (4) leakage, (5) water age and (6) fire fighting capacity. These six objectives are optimized based on decisions related to pipe sizing, tank siting, tank sizing, and pump scheduling under five different loading conditions. Solving the many-objective formulation reveals complex tradeoffs that would not be revealed in a lower-dimensional optimization problem. Visual analytics are used to explore these complex tradeoffs and identify solutions that simultaneously improve the overall WDS performance but with reduced capital and operating costs. This study demonstrates that the many-objective visual analytics approach has clear advantages and benefits in supporting more informed, transparent decision making in the WDS design process.

---

<sup>1</sup> Lecturer, Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, Univ. of Exeter, North Park Road, Harrison Building, Exeter EX4 4QF, UK (corresponding author). E-mail: g.fu@exeter.ac.uk.

<sup>2</sup> Professor, Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, Univ. of Exeter, North Park Road, Harrison Building, Exeter EX4 4QF, UK.

<sup>3</sup> Ph.D. candidate, Dept. of Civil and Environmental Engineering, Pennsylvania State Univ., University Park, PA 16802-1408.

<sup>4</sup> Associate Professor, Dept. of Civil and Environmental Engineering, Pennsylvania State Univ., University Park, PA 16802-1408.

**Keywords:** Genetic algorithms, leakage, many-objective optimization, hydraulic failure, water distribution system, water age

## Introduction

Multi-objective optimal design and rehabilitation of Water Distribution Systems (WDS) has attracted increasing attention over recent years (e.g., Walski and Gessler 1985; Walski et al. 1990; Halhal et al. 1997; Farmani et al. 2003; Perelman et al. 2008; Fu and Kapelan 2011). The shift from least cost design to multi-objective performance-based design advances decision makers' understanding of tradeoff relationships between conflicting design objectives (Walski 2001). For multi-objective problems, the Pareto optimal set of solutions are sought, where each solution is better than all the others in at least one objective. Plotting these solutions in their objective space yields an explicit representation of design tradeoffs termed the Pareto frontier. Discovery of the Pareto approximate frontier can aid decision makers in discovering the diminishing returns of alternative trade-off levels, which can be exploited in transparent, informed decision support in the design process (Giustolisi and Berardi 2009).

To date, most applications of multi-objective WDS design have considered two objective formulations focused on a cost related objective and one additional performance related objective. The WADISO program (Walski and Gessler 1985; Walski et al. 1990) was possibly the first tool to address the multi-objective nature of pipe sizing. Halhal et al. (1997) and Walters et al. (1999) considered an aggregated benefit measure including hydraulic benefit (pressure shortfalls or excesses). Kapelan et al. (2005) formulated a two-objective problem with cost and system reliability, which is defined as the probability of simultaneously satisfying minimum pressure head constraints at all nodes under the uncertainties in nodal demands and pipe roughness. Jayaram and Srinivasan (2008) considered minimization of life cycle cost and revised resilience index, which measures the surplus power of the WDS that could be potentially utilised in the case of increased water demand

and/or system failure. Generally these methods demonstrate the benefit of multi-objective optimization in revealing trade-off relationships between cost and system performance (Xu and Goulter 1999; Fu and Kapelan 2011).

It has been increasingly recognised that there is a need to include more than two objectives in the WDS design process. Some attempts have been made to reveal the potentially complex tradeoffs between conflicting objectives in various design conditions. For example, in addition to cost and reliability, Farmani et al. (2006) added a third objective – water age (residence time) – to address the water quality concern in the design and operation of a WDS with storage tanks. Redesigning a WDS to improve fire fighting capability, Kanta et al. (2012) considered three objectives – potential fire damage, water quality, and mitigation cost. In a pipe replacement prioritization problem, Giustolisi and Berardi (2009) demonstrated the need to consider four objectives: capital cost, future pipe break risk (represented by expected cost of pipe breaks), pipe material, and system reliability. Environmental impacts have been considered as an additional objective for WDS design through various representations such as greenhouse gas emissions and integrated impact index (e.g., Wu et al. 2010; Herstein et al. 2011; Kang and Lansey 2012). These prior multi-objective studies have provided an insight into the multi-objective nature of WDS design and rehabilitation problems and understanding of the implications to the decision making process. The difficulty in the above prior multi-objective studies often arises from choosing an appropriate metric to measure the WDS system performance as a few objectives can hardly capture the performance of the WDS under various design conditions (Fu et al. 2012).

This paper proposes the use of many-objective visual analytics to improve the WDS design, which blends improved high dimensional multi-objective optimization with highly interactive visual decision support (Kollat et al., 2011, Reed et al., in press). This has significant value for the WDS community as WDS problems usually involve many objectives and highly diverse stakeholders.

Formulations with one or a very small sample of performance objectives are myopic in their reflection of real-world engineering design requirements (Brill et al. 1990). This approach can help bridge the gaps between water distribution optimization and engineering design practice, and make optimization a useful tool for decision support in the WDS design process (Walski 2001).

Prior literature has defined problems as being “many-objective” if they have four or more objectives (Fleming et al. 2005). Recent studies have demonstrated the use of many-objective visual analytics in water supply risk management (Kasprzyk et al. 2009; Kasprzyk et al. 2012) and groundwater monitoring network design (Kollat et al. 2011), and have yielded new design insights and demonstrated the potentially highly negative consequences that could result from lower dimensional formulations. The increase in the number of objectives brings new challenges to multi-objective optimization: deterioration of search ability, exponential increase in non-dominated Pareto approximate solutions and difficulty in solution visualization. Popular multiobjective genetic algorithms, such as NSGAI (Deb et al. 2002), lose their effectiveness as the number of objectives increases, although they have been shown as effective for multi-objective WDS design (Farmani et al. 2003). The epsilon Nondominated Sorted Genetic Algorithm II ( $\epsilon$ -NSGAI) (Kollat and Reed 2006) has been demonstrated as effective and efficient for solving many objective problems (Kasprzyk et al. 2009; Kollat et al. 2011, Hadka and Reed in press, Reed et al. in press).

This study investigates the use of many-objective optimisation methods for optimal design and rehabilitation of the WDS. The  $\epsilon$ -NSGAI is used to solve a high-dimensional optimization problem – a six objective problem in this study. The case study used to demonstrate the many-objective method is the benchmark Anytown network rehabilitation problem (Walski et al. 1987), which addresses various design concerns from different stakeholders, including capital and operating costs, hydraulic failure, leakage, water age and firefighting capacity. The many-objective visual analytics approach is demonstrated as one way forward to address the challenges identified in the context of water

distribution system optimization by Walski (2001), particularly in revealing and balancing the conflicts from different stakeholders.

### Pressure-driven Model

The pressure-driven demand extension of EPANET (EPANETpdd) is used in this study to evaluate the performance of the WDS with an Extended Period Simulation (EPS) (Morley and Tricarico 2008). This has a clear advantage when evaluating reliability and leakage related objectives since the pressure driven simulations allow an accurate prediction of pressures and flows in a network. With EPANETpdd, the actual supply is calculated using an iterative procedure described in the above reference and the following head flow relationship proposed by Wagner et al. (1998):

$$Q_{i,t} = \begin{cases} 0, & \text{if } H_{i,t} < H_i^* \\ Q_{i,t}^d \left( \frac{H_{i,t} - H_i^*}{H_i^s - H_i^*} \right)^{1/2}, & \text{if } H_i^* < H_{i,t} < H_i^s \\ Q_{i,t}^d, & \text{if } H_{i,t} \geq H_i^s \end{cases} \quad (1)$$

where  $Q_{i,t}$ =the supply of node  $i$  at time step  $t$ ;  $Q_{i,t}^d$ =the full demand requirement of node  $i$  at time step  $t$ ;  $H_i^s$ =the desirable service pressure head at node  $i$ , above which the demand is completely supplied;  $H_i^*$ =the minimum required pressure head at node  $i$ , below which there is no outflow;  $H_{i,t}$ =the calculated pressure head of node  $i$  at time step  $t$ ;  $i = 1, 2, \dots, nn$  and  $nn$ =the total number of demand nodes.

### Many-objective Problem Formulation

The WDS design problem is formulated as a many-objective optimization problem, i.e., minimizing many objectives simultaneously. In the conventional least cost design, explicit constraints are often used to ensure that the optimal solutions can satisfy specific design requirements such as minimum pressure requirements at demand nodes. In the many-objective design formulation, the explicit constraints are transformed into objectives, to allow generation of a diverse set of alternatives that can

explore multiple values for the constrained variables, thus providing the decision maker with a better understanding of key trade-offs.

Many-objective analysis allows a suite of objectives to be considered that best capture concerns from different stakeholders. This set of objectives differs depending on the network of concern, but the following list was chosen to be representative of typical objectives in the WDS field.

### **Capital and Operating Costs**

There are many different cost metrics that can be considered in WDS. For example, a designer would consider a different suite of costs depending on whether an existing system is being considered or a completely new system is being designed. Costs considered in this study are capital cost for network expansion/rehabilitation and operating cost during a design period. The capital cost accounts for the cost of network components such as pipes, storage tanks, and pumps, while the operating cost measures the energy cost for pump operation. In the conventional least cost design, different types of cost are combined into one single cost as the optimization objective. Since many-objective optimization can consider multiple objectives that may or may not conflict, this study separates capital costs and operating costs in the Anytown network case study. However in some situations, the aggregated cost can also be useful especially when many other objectives are considered.

### **Hydraulic Failure**

Hydraulic failure at a demand node is defined as the occurrence of reduced water supply (i.e., below the required demand) due to low nodal pressure. It can be quantified for different components in the WDS network. In an EPS, the failure fraction can be defined as the fraction of time during which pressure drops below the required pressure. The consequence of failure of node  $i$  at time step  $t$  is defined as water shortage at this node relative to the total demand of the entire WDS at that time step:

$$NC_{i,t} = \frac{Q_{i,t}^d - Q_{i,t}}{\sum_{i=1}^m Q_{i,t}^d} \quad (2)$$

where  $NC_{i,t}$  = the consequence of failure of node  $i$  at time step  $t$ .

The nodal hydraulic failure index  $R_{\text{node}}$  is defined as the product of failure fraction and consequence as follows:

$$R_{\text{node}} = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^{nn} NF_{i,t} NC_{i,t} \quad (3)$$

where  $NF_{i,t}$  = the failure fraction of node  $i$  at time step  $t$  (assuming one hour time step for simulation), and it is calculated as the percentage of time when there is a demand shortage during time step  $t$ , that is, the estimated head is below the service head required ( $H_{i,t} < H_i^s$ ).  $T$  = the total hours in an EPS, and  $T=24$  in the Anytown network case study. Note that EPANETpdd automatically choose to use a shorter time step than a specified one (for example, one hour in this study) during an EPS due to overfilling or emptying tanks or active devices switching their states, particularly at the presence of deficient pressure. Thus, it is important to use the actual time step rather than the specified one when calculating the two terms  $NF_{i,t}$  and  $NC_{i,t}$  in Eq. (3).

In addition to nodal demand, the operation of tank is also considered for failure evaluation. Tank hydraulic failure occurs when the water level at the end of EPS is lower than at the beginning of simulation as this can cause potential problems for the following time period. Similar to the nodal failure index, the tank failure index  $R_{\text{tank}}$  is calculated as

$$R_{\text{tank}} = \frac{1}{nt} \sum_{i=1}^{nt} TF_{i,t} TC_{i,t} \quad (4)$$

where  $TF_{i,t}$  = the failure fraction of tank  $i$  at time step  $t$ , which has a value of 1 when hydraulic failure occurs and 0 otherwise;  $nt$  = the total number of tanks; and  $TC_{i,t}$  = the consequence calculated as the ratio of tank volume deficit between the beginning and end simulation to the total tank volume.

The total system failure index (SFI) can be defined by combining nodal and tank failure indices, and a definition is given for the Anytown network in the Case Study section.

### Fire Flow Deficit

In the U.S., the water flow to suppress an urban fire is required at a residual pressure of 137.9 kPa for a specific duration (American Water Works Association 1998). Similar requirements exist in other countries around the world. Failure to deliver the required fire flow poses a severe risk for assets and public life. The potential fire damage can be represented by fire flow deficit

$$f_{fire} = \sum_{t=1}^{TF} \sum_{i=1}^{nh} (Q_{i,t}^d - Q_{i,t}) \quad (5)$$

where  $Q_{i,t}$ =the supply of hydrant  $i$  at time step  $t$ ;  $Q_{i,t}^d$ =the full demand requirement of hydrant  $i$  at time step  $t$ ;  $TF$ =the duration required for fire fighting; and  $nh$ =the number of hydrants.

### Leakage

Water loss in a WDS is a severe and complex problem for many cities, particularly those with ageing networks. WDS leakage is often treated by water utilities as an operational cost but here it is evaluated separately to emphasise its environmental aspect. WDS leakage can result from mechanical failure of pipes or other components and from background leaks located in the pipe walls and around the pipe junctions.

The leakage objective in this study considers background leakage from pipes only and is calculated based on the pipe pressure. The leakage along a pipe is allocated to two end nodes for simulation. According to Tucciarelli et al. (1999), the background leakage in half-pipes linked to node  $i$  is calculated using

$$q_{i,t}^{leak} = (H_{i,t})^\lambda \sum_{j=1}^{N_i} \frac{\pi}{2} D_{ij} \theta_{ij} L_{ij} \quad (6)$$



where  $\lambda$ =loss exponent;  $D_{ij}$ =pipe diameter;  $L_{ij}$ =pipe length; and  $\theta_{ij}$ =leak per unit surface of the pipe linking nodes  $i$  and  $j$ ; and  $N_i$ =the total number of nodes linked to node  $i$ . The leakage pressure component  $\lambda$  is related to pipe characteristics, and normally has a value within [0.5, 2.5].  $\theta_{ij}$  depends on pipe characteristics and some external factors, such as environmental conditions and traffic loading. Ideally, the two parameters should be calibrated for each network. However, to demonstrate the many-objective approach,  $\lambda$  is assumed to 1.18, and  $\theta_{ij}$  is  $1 \times 10^{-9} \text{ m}^{1-\lambda}/\text{s}$  for all pipes in the Anytown network according to Tucciarelli et al. (1999) and Giustolisi et al. (2008). The total leakage is calculated as sum of respective leakages associated with each network node (see Eq. (6)) averaged over an EPS,

$$f_{leak} = \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^{nn} q_{i,t}^{leak} \quad (7)$$

### Water Age

Water quality problems in the WDS often arise from interactions between water within the pipe and the pipe wall, and within the bulk water of storage tanks. As the water stays in the system for a longer period of time, there is a greater chance for contaminant formulation and subsequent adverse health effects. The time required for the water to reach the customer from water sources through the network therefore influences the water quality. Water age at a demand node is used as an indicator of water quality, and defined as the average travel time from water sources to the demand node (Farmani et al. 2006). The dynamics of water quality is simulated in EPANETpdd (as in EPANET2) and water age over time at each demand node is provided after running the water quality module of EPANETpdd. According to Farmani et al. (2006), the water quality objective is to minimize the maximum water age across all demand nodes (and time steps) in the WDS

$$f_{age} = \text{Max}_{i,t}(WA_{i,t}) \quad (8)$$

where  $WA_{i,t}$  is water age of node  $i$  at time step  $t$  directly calculated in EPANETpdd.

## Many-objective Optimization Method

Evolutionary algorithms have emerged as a widely-used method for solving problems in complex engineering systems characterized by conflicting objectives such as WDS design (Nicklow et al. 2010). In the multi-objective context, many algorithms work very well on two-objective problems. For example, NSGAI is very effective and has received many applications (Farmani et al. 2003; Herstein et al. 2011). The algorithms' convergence can be, however, severely deteriorated by the increase in the number of objectives.

To improve the performance of the original NSGAI, the  $\epsilon$ -NSGAI was developed by introducing  $\epsilon$ -dominance archiving, adaptive population sizing, and automatic termination with the intent of improving computational efficiency and search reliability (Kollat and Reed 2006). Use of  $\epsilon$ -dominance enables a more even search of the objective space and significantly reduces the size of the archive. The  $\epsilon$ -NSGAI has been tested in a wide variety of applications and shown to be very efficient and effective for complex optimization problems (e.g., Kollat and Reed 2006; Tang et al. 2006; Tang et al. 2007; Kasprzyk et al. 2009; Kollat et al. 2011; Fu et al. 2012; Kasprzyk et al. 2012). Building on this body of work, this algorithm is chosen to solve the many-objective WDS design problems described above.

The  $\epsilon$ -NSGAI's parameter values used in this study are shown in Table 1. These settings reflect the recommendations by Kollat and Reed (2006). Because of the random nature of genetic algorithms five random seed replicate runs were used to solve the Anytown network problem. For each random seed run the  $\epsilon$ -NSGAI searched for one million model evaluations. Visual analysis showed that beyond one million evaluations there was little or no substantive improvement in the Pareto approximate sets attained. The Pareto approximate set analysed in this study was generated across all five random seed optimization runs.

## Case Study: Anytown Network Problem

The Anytown water distribution system has many typical features and challenges found in real-world networks, such as pump scheduling, tank storage provision and fire fighting capacity provision. The existing network has 35 pipes, two tanks and three identical pumps delivering water from the treatment plant into the system. To meet the city expansion and increasing demands, 77 decision variables are considered, including 35 variables related to existing pipes (with options of cleaning and lining or duplication with a parallel pipe), 6 new pipe diameters, 12 variables for two potential tanks, and 24 variables for the number of pumps in operation during 24 hours. Five loading conditions are considered: average day flow, instantaneous flow, and three fire flow conditions. The reader is referred to the study of Farmani et al. (2006) for detailed encoding of the decision variables.

The capital cost  $f_{capital}$  and operating cost  $f_{operating}$  (i.e., annual pumping energy cost) are calculated according to Walski et al. (1987) and Farmani et al. (2006). The Anytown network is required to supply water to all demand nodes at a minimum pressure of 275.8 kPa at the average day (24 hours) and instantaneous flow conditions. So in Eq. (1) the minimum required pressure ( $H_i^s$ ) is set to 0 for all nodes and the service pressure ( $H_i^s$ ) is 275.8 kPa in the two conditions. These two flow conditions are considered for calculating the SFI

$$f_{risk} = w_1 R_{node}^a + w_2 R_{tank}^a + w_3 R_{node}^b \quad (9)$$

where  $R_{node}^a$  and  $R_{tank}^a$  are the nodal and tank failure indices at the average day flow calculated using Eqs. (3) and (4), respectively;  $R_{node}^b$  is the nodal failure index at the instantaneous flow calculated using Eq. (3); and the weights  $w_1$ ,  $w_2$  and  $w_3$  are set to 0.4, 0.2 and 0.4, respectively.

According to Farmani et al. (2006), three fire flow conditions were defined as: (1) 2,500 gpm (0.158 m<sup>3</sup>/s) at one node (node 19); (2) 1,500 gpm (0.0946 m<sup>3</sup>/s) at three nodes (nodes 5, 6 and 7);

and (3) 1,000 gpm (0.0631 m<sup>3</sup>/s) at two nodes (nodes 11 and 17). Except the above specified nodes, all the other nodes are required to supply a flow of 500 gpm (0.0316 m<sup>3</sup>/s) under the three fire flow conditions. The fire flows must last for 2 hours and be met while supplying the peak day flow at a minimum pressure of 137.9 kPa (used as the service pressure  $H_i^s$  for all nodes in EPANETpdd). The pressure requirements must be met with one pump out of service and the tank water levels at their low level during a normal day. The fire flow deficit objective is calculated as the average deficit across the three fire flow conditions

$$f_{\text{fire-avg}} = \frac{1}{3}(f_{\text{fire}}^c + f_{\text{fire}}^d + f_{\text{fire}}^e) \quad (10)$$

where  $f_{\text{fire}}^c$ ,  $f_{\text{fire}}^d$ , and  $f_{\text{fire}}^e$  are the deficits under the three fire flow conditions using Eq. (5).

The leakage  $f_{\text{leak}}$  and water age  $f_{\text{age}}$  are calculated for the average day (24 hours) flow condition using Eqs. (7) and (8). The leakage is assumed to exist in old, existing pipes only and no leakage is assumed for new pipes.

The Anytown network problem has been solved using different multi-objective formulations, for example, total cost and aggregated benefit (Walters et al. 1999), total cost and pressure deficit (Fu et al. 2012), total cost and reliability as well as water age (Farmani et al. 2006), and separate capital and operating cost as well as environmental impact index (Herstein et al. 2011). This study is the first to define a many-objective formulation, in which a total of six objectives are minimized simultaneously:

$\{f_{\text{capital}}, f_{\text{operating}}, f_{\text{risk}}, f_{\text{leak}}, f_{\text{age}}, f_{\text{fire-avg}}\}$ . This problem definition is to move away from the paradigm of cost minimization only, and represents a step further towards addressing the real concerns of practicing engineers in reality (see Walski et al. 1987).

## Results and Discussion

### Six-dimensional Objective Tradeoffs

A global view of the six-objective tradeoffs are first analysed to illustrate the benefits of the many-objective approach. Fig. 1 shows the Pareto approximate set of 989 solutions generated by applying the  $\epsilon$ -nondominated sorting algorithm to all of the Pareto approximate solutions attained over five random seed optimization runs. This set represents the best known approximation to the true Pareto-optimal set from a total of 5 million model simulations, which took about 180 hours using a desktop with a 3.00 GHz processor in Windows XP. The leakage, SFI, and capital cost, are plotted on the  $x$ ,  $y$  and  $z$  axes, respectively. The water age objective is shown by the color of the cones with color ranging from blue to red, representing the increasing water age from 6.85 to 24 hours. The fire flow deficit objective is shown by the orientation of the cones ranging over  $180^\circ$  of rotation. Cones pointing up represent the highest deficit and cones pointing down represent the lowest deficit. The operating cost objective is represented by the size of the cones, and the large cones represent high costs and small cones represent low costs. Note that all six objectives are minimization objectives, thus an ideal solution would be located toward the rear lower corner (low capital cost, low SFI, low leakage) of the plot in Fig. 1 and represented by a small (low operating cost), blue (low water age) cone pointing down (low fire flow deficit). The arrows in Fig. 1 highlight directions of increasing preference.

Region (i) in Fig. 1 captures higher cost solutions (capital cost  $>10\text{M}\$$ ), with reduced system failure ( $<0.1$ ) and low fire flow deficit ( $<0.34\text{ m}^3/\text{s}$ ). However, these solutions can have a wide range of water ages as illustrated by their color from green (13 hours) to red (24 hours). Region (i) solutions have leakages from  $0.05\text{ m}^3/\text{s}$  to  $0.09\text{ m}^3/\text{s}$ , and their operating costs are relatively high as shown by the large sizes of cones. Region (ii) in Fig. 1 is composed of solutions that have low capital and operating costs, that tradeoff with high failure indices and fire flow deficits. Similar to the

solutions in region (i), these solutions also have a wide range of water ages and leakages. This highlights the importance of considering the water age and leakage objectives in the optimization process. Particular attention should be given to water age as it is not linked to the levels of capital and operating costs, that is, solutions with a wide range of water age can be found at any cost level.

Fig. 2 shows a parallel line plot in which each line represents the 6 objective values for a single solution, and its color shading from blue (\$2.4 M) to red (\$22.0 M) shows the variation of the capital costs. The red lines represent the high cost solutions, i.e., those shown in the region (i) of Fig. 1, and the blue lines represent the low cost solutions in the region (ii). The line crossing between two objectives represents the tradeoff of these two objectives. There is a significant tradeoff between two pairs of objectives: operating cost vs. SFI and fire flow deficit vs. water age, as shown by the significant number of line crossings between each pair of objectives. On the contrary, the majority number of lines between capital and operating costs do not cross, that is, low capital costs are linked to low operating costs, and high capital costs are linked to high operating costs. Previous research has shown that a negative relationship (tradeoff) exists between capital and operating costs in a lower-dimensional objective space (e.g., Herstein et al. 2011). However, the results in Fig. 2 reveal a more positive rather than negative relationship between these two cost objectives when required to balance more objectives in a higher dimensional space. A similar relationship is also observed for SFI and fire flow deficit.

The trends for high cost and low cost solutions in Fig. 1 can also be clearly seen in Fig. 2. Further, Fig. 2 reveals that the medium cost solutions (represented by light blue and yellow lines) have a wide range of failure indices, fire flow deficits, water ages and leakages. This indicates the complexity and difficulty in decision making when selecting a solution from the medium-cost region, which is generally of particular concern to the decision maker in practice (due to typical high increase in benefits for marginal increase in costs).

## Lower-dimensional Objective Tradeoffs

In addition to the six-dimensional view, the lower-dimensional sub-problem tradeoffs within the approximate set are further explored to illustrate the benefits of the many-objective approach. It should be noted that solving the six-objective optimization problem automatically solves 62 smaller subproblems at the same time: 6 single-objective problems, 15 two-objective problems, 20 three-objective problems, 15 four-objective problems, and 6 five-objective problems. That is, the Pareto approximate set obtained from the full six-objective problem contains all of the tradeoffs for the 62 sub-problems. This allows the comparison of the solution sets from different-dimensional problem definitions with the results from the full six-objective optimization.

Fig. 3 shows some selected two-dimensional (three-dimensional) tradeoffs of the approximate set. The full suite of Pareto approximate solutions are shown as transparent cones in each two-dimensional plot, and the relevant sub-problem tradeoff curve is highlighted where possible. The cones are shown in colors representing an additional objective.

In Fig. 3a, a clear tradeoff curve between capital cost and fire flow deficit can be observed, and it represents the approximate Pareto front had only these two objectives been used for optimization (highlighted with black squares). Note that the cones in the Pareto approximate front have a very different color varying from blue to red, representing a significant variation in the objective of water age. This implies that many of these seemingly high-performing solutions fail in the objective of water age, and a decision maker may choose a solution with a very high water age had only the capital cost and fire flow deficit objectives been considered for optimization.

Fig. 3b shows the trade off in the objective space of capital cost and failure with the Pareto approximate solutions highlighted with red squares. Similar to Fig. 3a, the optimal solutions in Fig. 3b are also very different in water age as illustrated by colors. The Pareto approximate solutions for the *Capital Cost-Fire Flow Deficit* sub-problem highlighted in Fig. 3a are also shown in Fig. 3b,

highlighted with black squares. Most of these solutions are not non-dominated in the space of capital cost and system failure. Consequently, the lower dimensional formulations would bias decision makers towards neglecting potential improvements in fire flow deficit performance.

The tradeoffs between capital cost and system failure/fire flow deficit can be explained below. A higher capital cost means larger pipe diameters and tank volumes, resulting in higher nodal pressures, and thus a low frequency and consequence of hydraulic failure (i.e., a lower failure index). The fire flow deficit objective measures the consequence of a low pressure. Thus the two objectives are correlated and they have a similar tradeoff relationship with capital cost. As revealed by Figs. 3a and 3b, however, there is a distinction at the higher cost tails: a wider range of failure indices and a narrower range of fire flow deficits. This implies that a high cost design can generally achieve a low fire flow deficit but does not necessarily result in a low system failure index.

Moving to Fig. 3c, the Pareto approximate set is shown in the space of operating cost and water age. The sub-problems' lower dimensional tradeoffs analyzed in Figs. 3a and 3b are also shown in Fig. 3c. Clearly these solutions fail in the operating cost and water age. Most of them have a very high operating cost. Figs. 3c and 3d show different relationships between the two different types of cost and water age. Although there is a small Pareto approximate front between capital cost and water age, the more important relationship is that water age increases with capital cost. This is because larger pipe diameters cause slower flow through the network, resulting in higher water ages (Kanta et al. 2012). In the space of operating cost and water age, a clear tradeoff relationship can be observed. Not only the specific designs such as better designed loops have an important influence on water age, adjusting pump operations can also effectively reduce residence time through increasing storage turnover rates, resulting in a lower water age. The different relationships between the two costs and water age could not be revealed had the two costs been aggregated into one cost in a lower-dimensional problem definition.



Further Figs. 3c and 3d also show different relationships between the two types of cost and leakage, represented by the cone colors. There is a close positive relationship between leakage flow with operating cost, as revealed by the color variation along the  $x$ -axis of Fig. 3c. This is because pumping has a direct effect on pressure that is closely linked to leakage flow (see Eq. (6)). However, the leakage is not closely linked with the capital cost as the cones in the same color are distributed over a wide range of costs along the  $x$ -axis of Fig. 3d. This is an effect of capital type works (e.g., new pipes and tanks) on system pressures that generally increases pressures in the system but may also lead to pressure decrease in part(s) of the system, e.g. where the old/leaky pipes are predominantly located. This illustrates that operating cost has a more significant impact on leakage than capital cost because it is more directly and closely related to nodal pressures that consequently affect leakage flows generated from the old, leaky pipes.

### **Exploration to Inform Decision Making**

A visual approach has been developed to explore the complex tradeoffs by successively adding more objectives into the tradeoffs to aid the decision maker in better understanding objective interactions (Khu and Madsen, 2005; Kollat and Reed, 2007a; Kollat et al., 2011). This approach is used here to demonstrate how to reveal the interactions that may not be fully captured in a lower-dimensional problem formulation, and how to discover high-performing solutions for decision support.

Figs. 4a-4e mimic the possible steps that a decision maker could take to identify satisfying solutions starting from a two-dimensional tradeoff. Fig. 4a shows the tradeoff curve in the space of capital cost and fire flow deficit. Recall that this curve represents the Pareto-optimal solutions had only these two objectives been used for optimization in a two-objective formulation. Considering the tradeoff between the two objectives, a decision maker might want to choose a solution at the point of diminishing return on the tradeoff curve because after this point there is little improvement in fire flow deficit with the increase in capital cost. In this way, Solution 1 is first identified for comparison.

Fig. 4b shows the *Capital Cost-SFI* tradeoff highlighted with red squares. Although the failure index and fire flow deficit are correlated in measuring the performance of the network, the *Capital Cost-SFI* tradeoff curve has a much lower diminishing point around the cost of \$10 M. It can be seen that Solution 1 fails in achieving a very low failure index. Thus Solution 2 is identified at the diminishing point of the *Capital Cost-SFI* tradeoff, and it achieves a much lower failure index with a lower capital cost compared with Solution 1.

Moving to the objective of leakage, Fig. 4c shows the tradeoff curve between system failure and leakage. The *Capital Cost-Fire Flow Deficit* and *Capital Cost-SFI* tradeoffs are also shown in Fig. 4c for reference purposes. Note that Solution 1 does not represent the best *SFI-Leakage* tradeoff given its distance from the ideal sub-problem tradeoff. Although Solution 2 lies on the *SFI-Leakage* tradeoff curve, it has a very high leakage flow. Solution 3 represents a compromise for the *SFI-Leakage* tradeoff, and thus is selected for comparison.

Having moved up to the four-objective space so far, Fig. 4d shows the tradeoff in the two objectives left: operating cost and water age, with an additional objective of capital cost represented by the colors. Solutions 1, 2 and 3 have a medium water age and medium capital cost. At this stage the decision maker might want to choose a lower capital cost solution for comparison. Solution 4 is chosen from the lower end of the *Operating Cost-Water Age* tradeoff. Solution 5 is located at the edge of the *Operating Cost-Water Age* space with a low water age and a relatively high operating cost, but it has a lower capital cost and operating cost than Solution 2. Solution 5 is chosen to compare its performance in other objectives.

Fig. 4e shows the projection of all the selected solutions in the six dimensional plot. The capital cost, water age and failure index are plotted on the  $x$ ,  $y$  and  $z$  axes, respectively. The directions of increasing preference are shown by arrows. The leakage objective is shown by the color of the cones with color ranging from blue to red, representing the decreasing preference. Similar to Fig. 1, the fire

flow deficit and operating cost objective are shown by the orientation and size of the cones, respectively. In addition to the five solutions (marked with cubes), the Pareto approximate front for the *Capital Cost-SFI-Water Age* tradeoff is represented by the unmarked cones. The remaining solutions within the full Pareto approximate set are shown as transparent cones.

Fig. 4e shows the relative location of the five solutions to the Pareto approximate set. Solutions 2, 4 and 5 are located on or very close to the *Capital Cost-SFI-Water Age* sub-problem's tradeoff. Solution 1 is distant from the *Capital Cost-SFI-Water Age* sub-problem's tradeoff, implying poor performance for water age and capital cost. Solution 3 provides better in several other objectives, represented by a smaller (low operating cost), lighter blue (low leakage) cone in comparison to Solution 1. Solution 3 offers the best compromises among all the six objectives, thus is recommended through the use of visual analytics described above.

The differences in the five solutions are further clarified in the parallel line plot in Fig. 5. Compared with Solution 1, Solution 3 achieves a substantial improvement in SFI, water age and leakage with a lower capital cost and lower operating cost, although it does have a slightly higher fire flow deficit. Solution 2 has a much higher operating cost that helps it achieve improved performance for the SFI and fire flow objectives, while reducing its performance for the leakage objective (i.e., it results in a substantial increase in leakage flow). Solution 4 has the lowest capital and operating costs, and has the best performance in water age and leakage but has the worst performance in SFI and fire flow deficit. Solutions 2 and 5 have the similar operating cost and leakage, but with a much lower capital cost, solution 5 has a much higher failure index and fire flow deficit. The parallel line plot shows that besides Solution 3 all the other solutions have bad performance in at least one objective. Solution 3 achieves a well-balanced overall good performance with reduced capital and operating costs.

## Understanding of the Design Space

It is vital to understand how the system performance is affected by different decision variables. Fig. 5 shows Solution 1 has a very similar objective trajectory to the recommended solution 3, thus it is selected for comparison to understand the important domain knowledge in improving system performance.

Fig. 6 shows the network layouts for the selected solutions 1 and 3. Although solution 1 is more expensive than solution 3, it has fewer pipes duplicated than Solution 3 but with larger pipe diameters. To reduce capital cost, solution 3 has more pipes selected for cleaning. Also, solution 3 has a smaller total tank volume, i.e., 7400m<sup>3</sup> compared with 13300m<sup>3</sup> in solution 1. While the two solutions choose similar areas for installing two new tanks, they have very different provisions for the two tank volumes. Solution 1 has a much larger tank located at node 18, providing pressures for the old central areas of the network where several old pipes (pipes 37, 38 and 41) are not duplicated. The high pressures due to the tank result in a high rate of leakage from these old pipes, contributing to a higher leakage generated from solution 1 than solution 3. In comparison, Solution 3 allocates a much smaller tank to node 12, as more pipes in this old area of the network are duplicated or cleaned with a low pressure loss. Further, Solution 3 allocates a large tank to node 4, providing high pressures to the proposed industrial development area of the city. As the pipes are new in this area, the high pressures have little impacts on leakage.

One notable difference between these two solutions is that Solution 3 has 3 pumps operating at the 24th hour to fill the tanks while solution 1 has a more uniform pump provision for extra pumping through the day, i.e., 2 pumps operating at 5 different hours (i.e., 8th, 14th, 16th, 18th, and 22nd) while one pump operates at all the other hours. Solution 3 has an advantage in reducing the operating cost while reducing the high pressure periods. Recall that operating cost (i.e., pumping) has a more

significant impact on leakage than capital cost, thus the reduced pump operating hours also explain the low leakage from Solution 3.

## Conclusions

This study proposes use of the many-objective visual analytics approach for water distribution system design or rehabilitation problems. The many-objective approach combines a multi-objective evolutionary algorithm ( $\epsilon$ -NSGAI) and interactive visual analytics to reveal and explore the tradeoffs underlying the WDS design problem. This approach was demonstrated using the well-known benchmark Anytown network, which was designed for a suite of six objectives: capital cost, operating cost, system hydraulic failure, leakage, water age and fire fighting capacity.

In this study, the optimization results demonstrate the benefits of considering many objectives in the design process. The Pareto optimal solutions identified in a lower dimensional problem usually have a worse performance in other objectives considered in a higher dimensional problem. The capital cost and operating cost have a very different relationship with water age and leakage and this would not be revealed had the costs been aggregated into one objective in a lower dimensional problem formulation.

In addition, this study illustrates the use of visual analytics to explore the complex tradeoffs between various conflicting design objectives and identify the satisfying solutions in the design process. The solutions identified achieve a significant improvement in overall WDS performance with reduced capital and operating costs when compared with the solutions that would have been chosen from lower dimensional (two-objective) problems.

The many-objective visual analytics approach provides a powerful tool for discovering high-performing solutions that achieve the best tradeoffs between all the objectives considered. This can be used to support more informed, transparent decision making in the WDS design process. Thus this

approach is suggested as one way forward to address the challenges in the context of water distribution system optimization, particularly in revealing and balancing the tradeoffs between various design objectives. However, it should be noted that many-objective optimization poses both benefits and challenges in terms of the amount of problem information generated for decision makers. The consideration of many-objectives can lead to important insights, but there needs to be a strong focus in future research on effective human-computer interaction frameworks (e.g., visual analytics) that enhance discovery and decision making. Though this approach can effectively handle as many as six objectives or even more, this may not always be required in practice. The objectives used when solving a specific real-life problem have to be carefully selected on a case by case basis, to address the key issues and concerns from various stakeholders involved. The WDS design problem formulation used in the paper cannot fully represent the complexity of real-world design problems, which might have a different set of objectives addressing different stakeholders' concerns and involve considerable uncertainties in demands and system characteristics. Thus future work will involve further testing and verification of the many-objective methodology developed and presented here on more complex, real-life water distribution systems with additional/different objectives or uncertainties considered.

## References

- American Water Works Association (1998). *Distribution System Requirements for Fire Protection*. AWWA, Denver, CO.
- Brill, E. D., Flach, J. M., Hopkins, L. D., and Ranjithan, S. (1990). "MGA: A Decision Support System for Complex, Incompletely Defined Problems." *IEEE Transactions on Systems, Man, Cybernetics*, 20(4), 745-757.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE Trans. Evol. Comput.*, 6(4), 182-197.

Farmani, R., Savic, D. A. & Walters, G. A. (2003). "Multi-objective optimization of water system: a comparative study." In *Pumps, Electromechanical Devices and Systems Applied to Urban Water Management* vol. 1 (ed. Cabrera E. & Cabrera E. Jr), pp. 247–256. Balkema, Lisse, The Netherlands.

Farmani, R., Walters, G. and Savic, D. (2006). "Evolutionary multi-objective optimization of the design and operation of water distribution network: total cost vs. Reliability vs. Water quality." *J. Hydroinformatics*, 8(3), 165-179.

Fleming, P. J., Purshouse, R. C., and Lygoe, R. J. (2005). "Many-Objective Optimization: An engineering design perspective." *Lecture Notes in Computer Science*, 3410, 14-32.

Fu, G., and Kapelan, Z. (2011). "Fuzzy probabilistic design of water distribution networks." *Water Resour. Res.*, 47, W05538, doi:10.1029/2010WR009739.

Fu, G, Kapelan, Z, and Reed, P. (2012). "Reducing the complexity of multi-objective water distribution system optimization through global sensitivity analysis." *J. Water Resour. Plann. Manage.*, 138(3), 196-207.

Giustolisi, O. and Berardi, L. (2009). "Prioritizing pipe replacement: From multiobjective genetic algorithms to operational decision support." *J. Water Resour. Plann. Manage.*, 135(6), 484-492.

Giustolisi, O., Savic, D., and Kapelan, Z. (2008). "Pressure-driven demand and leakage simulation for water distribution networks." *J. Hydraul. Eng.*, 134(5), 626–635.

Hadka, D., Reed, P.M. (in press) "Diagnostic Assessment of Search Controls and Failure Modes in Many-Objective Evolutionary Optimization." *Evolutionary Computation*, in press.

Halhal, D., Walters, G. A., Ouazar, D., and Savic, D. A. (1997). "Water network rehabilitation with structured messy genetic algorithms." *J. Water Resour. Plann. Manage.*, 123(3), 137–146.

Herstein, L. M., Filion, Y. R., and Hall, K. R. (2011). "Evaluating the Environmental Impacts of Water Distribution Systems by Using EIO-LCA-Based Multiobjective Optimization." *J. Water Resour. Plann. Manage.*, 137, 162-172.

Jayaram, N., and Srinivasan, K. (2008). "Performance-based optimal design and rehabilitation of water distribution networks using life cycle costing." *Water Resour. Res.*, 44, W01417, doi:10.1029/2006WR005316.

Kang, D., and Lansey, K. (2012). "Dual Water Distribution Network Design under Triple-Bottom-Line Objectives." *J. Water Resour. Plann. Manage.*, 138(2), 162-175.

Kanta, L., Zechman, E., and Brumbelow, K. (2012). "A Multi-Objective Evolutionary Computation Approach for Redesigning Water Distribution Systems to Provide Fire Flows." *J. Water Resour. Plann. Manage.*, 138(2), 144-152.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2005). "Multiobjective design of water distribution systems under uncertainty." *Water Resour. Res.*, 41, W11407, doi:10.1029/2004WR003787.

Kasprzyk, J. R., Reed P. M., Kirsch B., and Characklis G. (2009). "Managing population and drought risks using many-objective water portfolio planning under uncertainty." *Water Resour. Res.*, 45, W12401, doi:10.1029/2009WR008121.

Kasprzyk, J. R., Reed, P. M., Kirsch, B. R., and Characklis, G. W. (2012). "Many-Objective de Novo Water Supply Portfolio Planning Under Deep Uncertainty." *Environmental Modelling & Software*, 34, 87-104.

Khu, S. T., and Madsen, H. (2005). "Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration." *Water Resour. Res.*, 41, W03004, doi:10.1029/2004WR003041.



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., Reed, P. M., and Maxwell, R. M. (2011), "Many-objective groundwater monitoring network design using bias aware ensemble Kalman filtering, evolutionary optimization, and visual analytics." *Water Resour. Res.*, 47, W02529, doi:10.1029/2010WR009194.

Morley, M.S., and Tricarico C. (2008). "Pressure Driven Demand Extension for EPANET (EPANETpdd)." Technical Report 2008/2, Centre for Water Systems, University of Exeter, UK.

Nicklow, J., Reed, P. M., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., and Zechman, E. (2010). "State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management." *J. Water Resour. Plann. Manage.*, 136(4), 412-432

Perelman, L., Ostfeld, A., and Salomons, E. (2008). "Cross entropy multiobjective optimization for water distribution systems design." *Water Resour. Res.*, 44, W09413, doi:10.1029/2007WR006248.

Reed, P. M., Hadka, D., Herman, J., Kasprzyk, J., Kollat, J. "Evolutionary Multiobjective Optimization in Water Resources: The Past, Present, and Future", *Advances in Water Resources*, in press.

Tang, Y., Reed P. M., and Wagener T. (2006). "How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?" *Hydrol. Earth Syst. Sci.*, 10, 289-307.

Tang, Y., Reed P., and Kollat J. B. (2007). "Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications." *Adv. Water Resour.*, 30(3), 335-353.

Tucciarelli, T., Criminisi, A., and Termini, D. (1999). "Leak analysis in pipeline systems by means of optimal valve regulation." *J. Hydraul. Eng.*, 125(3), 277–285.

Wagner, J. M., Shamir, U., and Marks, D. H. (1998). "Water distribution reliability: Simulation methods." *J. Water Resour. Plann. Manage.*, 114(3), 276–294.

Walski, T.M. and Gessler, J., (1985). "Water Distribution System Optimization," *Tech. Rep. EL 85-11*, US Army Corps of Engineers Waterways Experiment Station, Vicksburg, Miss.

Walski, T.M, Gessler, J. and Sjostrom, J.W., (1990). *Water Distribution Systems: Simulation and Sizing*, Lewis Publishers, Ann Arbor.

Walski, T.M., (2001). "The wrong paradigm - why water distribution optimization doesn't work." *J. Water Resour. Plann. Manage.*, 127(4), 203-205.

Walski, T., Brill E., Gessler J., Goulter I., Jeppson R., Lansey K., Lee H., Liebman J., Mays L., Morgan D., and Ormsbee L. (1987). "Battle of the network models: epilogue." *J. Water Resour. Plann. Manage.*, 113 (2), 191-203.

Walters, G. A., Halhal D., Savic D., and Ouazar D. (1999). "Improved design of Anytown distribution network using structured messy genetic algorithms." *Urban Water*, 1(1), 23-38.

Wu, W., Simpson, A. R., and Maier, H. R. (2010). "Accounting for greenhouse gas emissions in multiobjective genetic algorithm optimization of water distribution systems." *J. Water Resour. Plann. Manage.*, 136(2), 146-155.

Xu, C., and Goulter I. C. (1999). "Reliability-based optimal design of water distribution networks." *J. Water Resour. Plann. Manage.*, 125(6), 352–362.

## Figure Captions

Fig. 1. The approximate Pareto set from six objective minimization – capital cost, operating cost, system failure index (SFI), fire flow deficit, leakage, and water age. The arrows show directions of increasing preference.

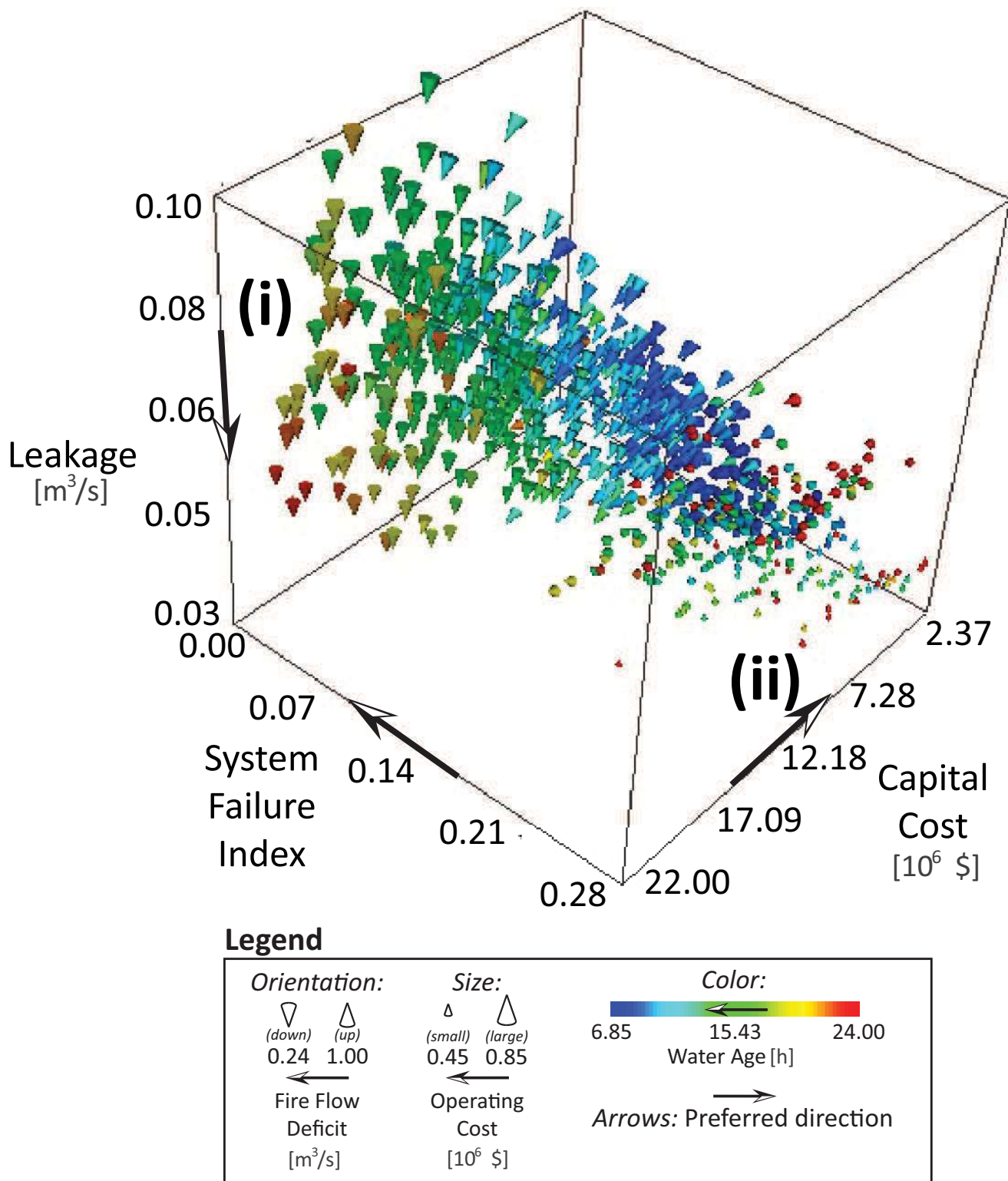
Fig. 2. The parallel line plot for the approximate Pareto set. Each solution is represented by a line across six objectives in the  $x$ -axis, with the line colors representing the variation in capital cost.

Fig. 3. Selected lower-dimensional tradeoffs from the full six-objective Pareto approximate set.

Fig. 4. A possible exploratory analysis to identify five interesting solutions from the full six-objective Pareto approximate set.

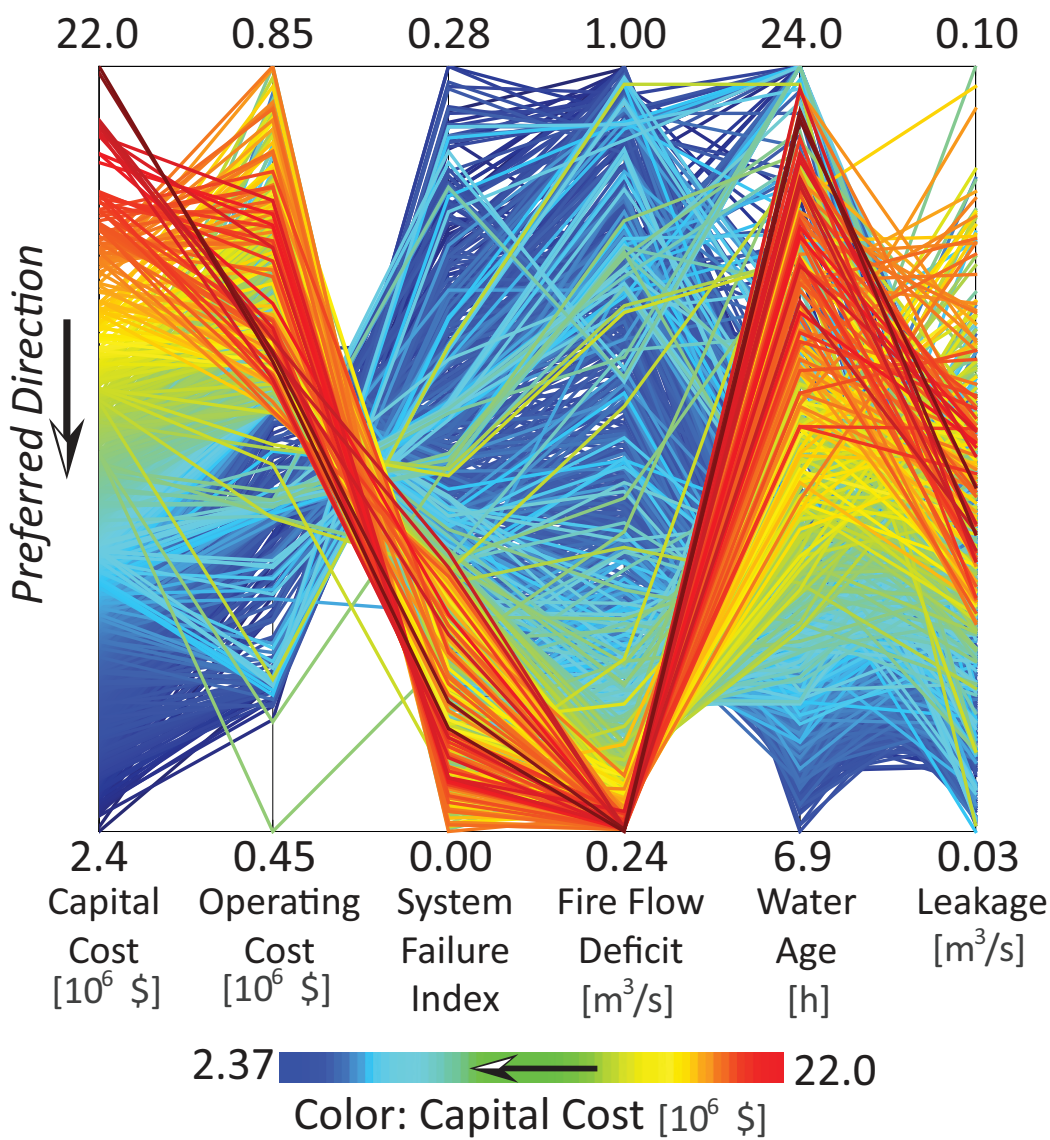
Fig. 5. The parallel line plot for the five solutions identified. Each solution is represented by a line across six objectives in the  $x$ -axis.

Fig. 6. Layout of two selected solutions. V=tank volume, B=tank bottom elevation, T=tank top elevation, and M=Minimum normal day elevation.



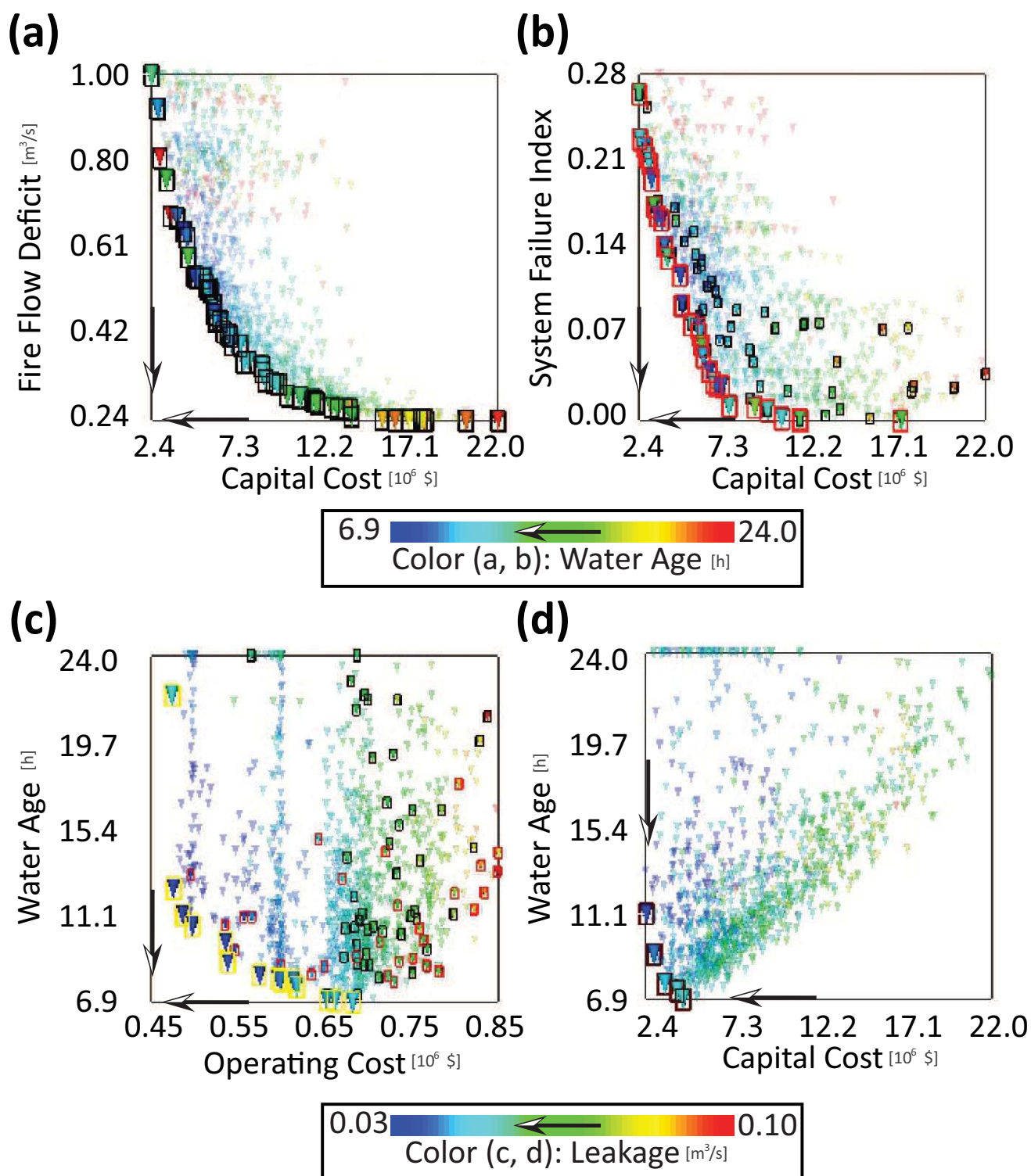
Downloaded from ascelibrary.org by University of Exeter on 04/09/13. Copyright ASCE. For personal use only; all rights reserved.

Accepted Manuscript  
Not Copyedited

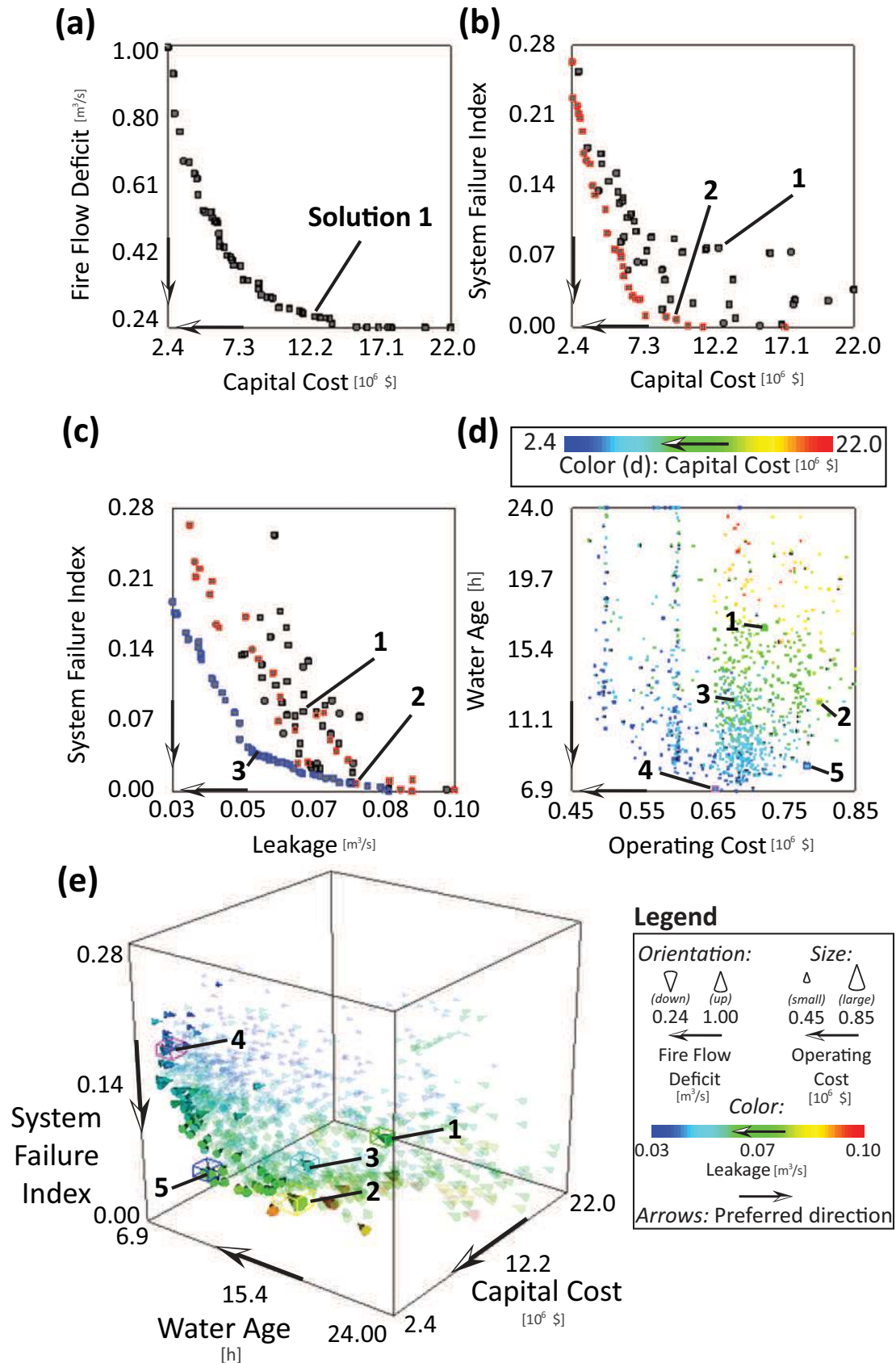


Accepted Manuscript  
Not Copyedited

Downloaded from ascelibrary.org by University of Exeter on 04/09/13. Copyright ASCE. For personal use only; all rights reserved.



Accepted Manuscript  
Not Copyedited



Accepted Manuscript  
Not Copied

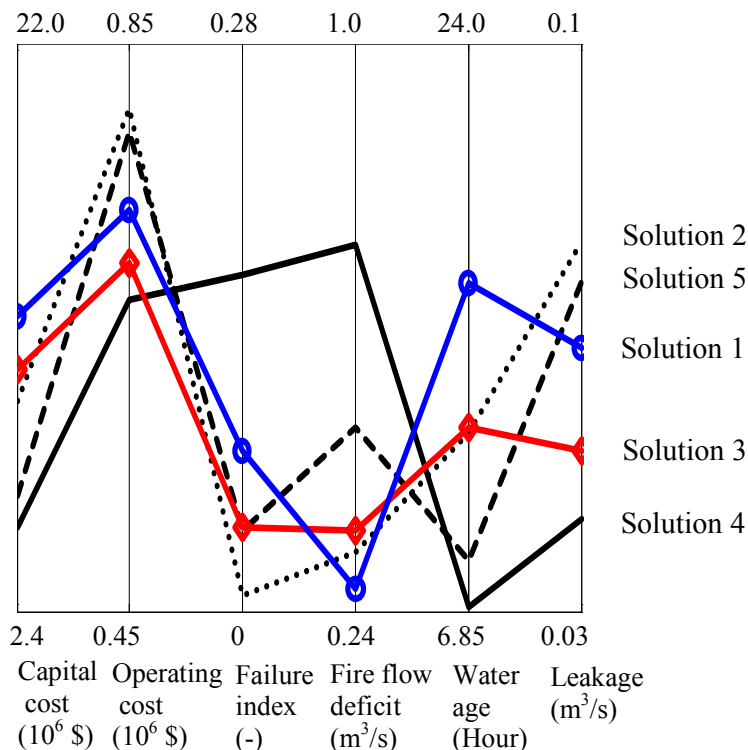


Fig. 5. The parallel line plot for the five solutions identified. Each solution is represented by a line across six objectives in the x-axis.

Accepted Manuscript  
Not Copyedited



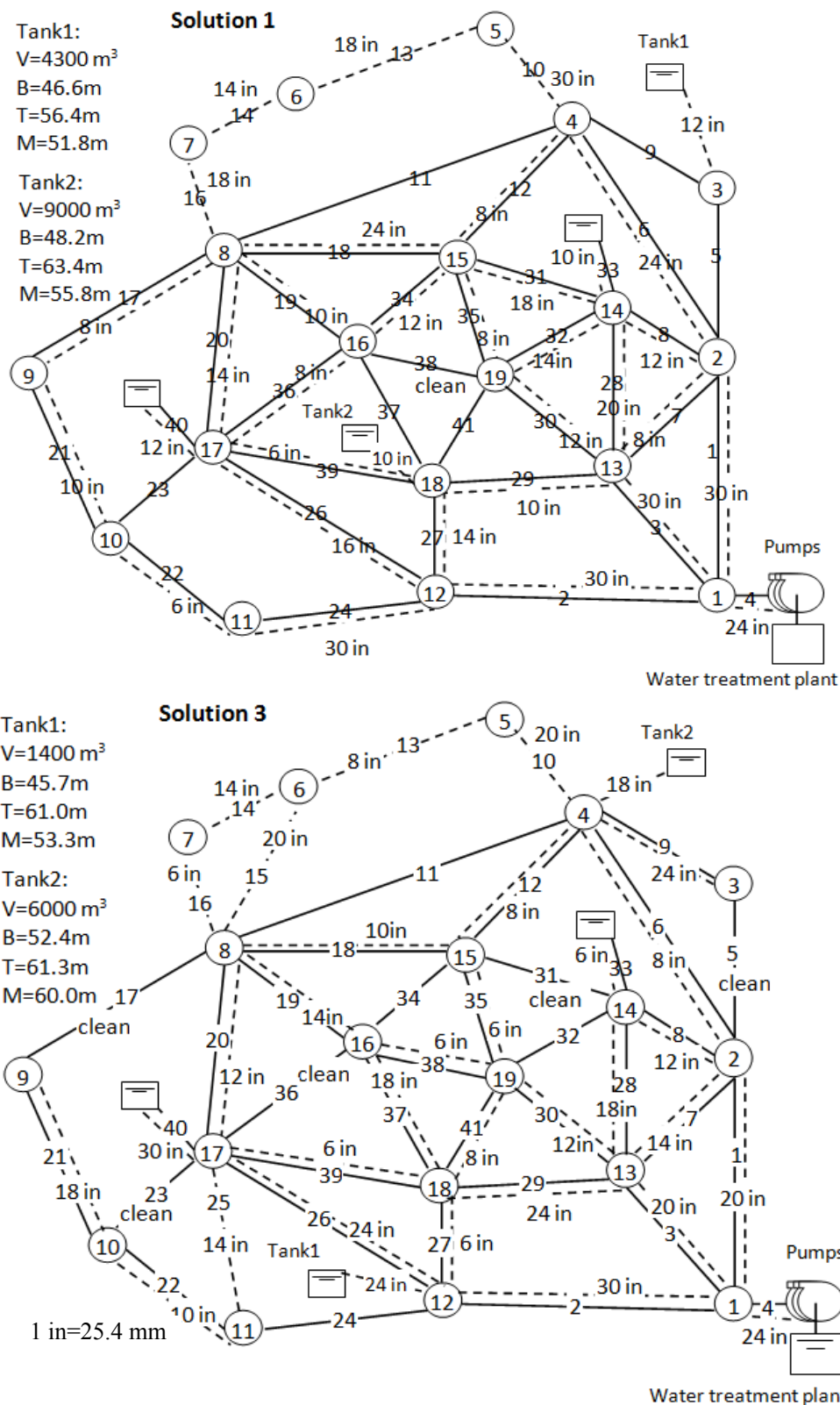


Fig. 6. Layout of two selected solutions. V=tank volume, B=tank bottom elevation, T=tank top elevation, and M=Minimum normal day elevation.

Accepted Manuscript  
Not Copied

**Table 1.** Parameter values of the  $\epsilon$ -NSGAI algorithm

Symbol	Value	Description
$n_{initial}$	12	Initial population size
$n_{generation}$	250	The maximum number of generations in each run
$n_{maximum}$	1 million	The maximum number of model simulations
$\eta_m$	20	Distribution index for mutation
$\eta_c$	15	Distribution index for crossover
E	\$10,000	Objective resolution: capital and operating costs
	0.01	Objective resolution: system failure index
	$2.83 \times 10^{-3} \text{ m}^3/\text{s}$	Objective resolution: fire flow deficit
	$2.83 \times 10^{-4} \text{ m}^3/\text{s}$	Objective resolution: leakage
	0.1 hour	Objective resolution: water age

Accepted Manuscript  
Not Copyedited