Least-cost design of water distribution systems under demand uncertainty: the robust counterpart approach

Lina Perelman, Mashor Housh and Avi Ostfeld

ABSTRACT

In this study, a non-probabilistic robust counterpart (RC) approach is demonstrated and applied to the least-cost design/rehabilitation problem of water distribution systems (WDSs). The uncertainty of the information is described by a deterministic user-defined ellipsoidal uncertainty set that implies the level of risk. The advantages of the RC approach on previous modelling attempts to include uncertainty are in making no assumptions about the probability density functions of the uncertain parameters and their interdependencies, having no requirements on the construction of a representative sample of scenarios, and the deterministic equivalent problem preserves the same size (i.e. computational complexity) as the original problem. The RC is coupled with the cross-entropy heuristic optimization technique for seeking robust solutions. The methodology is demonstrated on an illustrative example and on the Hanoi network. The results show considerable promise of the proposed approach to incorporate uncertainty in the least-cost design problem of WDSs. Further research is warranted to extend the model for more complex WDSs, incorporate extended period simulations, and develop RC schemes for other WDSs related management problems.

Key words | least-cost design, robust optimization, uncertainty, water distribution systems

Lina Perelman
Avi Ostfeld (corresponding author)
Faculty of Civil and Environmental Engineering,
Technion – Israel Institute of Technology,
Haifa 32000,
Israel
E-mail: *ostfeld@tx.technion.ac.il*

Mashor Housh
Faculty of Civil and Environmental Engineering,
University of Illinois at Urbana-Champaign,
2524 Hydrosystems Laboratory 301 N. Mathews Ave.,
Urbana,
IL 61801,
USA

NOMENCLATURE

AC	ant colony
CE	cross entropy
cms	cubic metres per second
GA	genetic algorithm
GRG2	generalized reduced gradient 2
NSGAII	non-dominated sorting genetic algorithm II
PDF	probability density function
RC	robust counterpart
WDS	water distribution system
a	uncertain parameter
$\hat{a}$	nominal value of the uncertain parameter
A_{11}	diagonal matrix of pipe's resistance
A_{12}	connectivity matrix
b	right-hand-side parameter
B	box
c	coefficient of a linear objective function
D	diameter of a pipe
E	ellipsoid
f_0	objective function
f_i	constraint function
f_P	nonlinear energy conservation equations
f_Q	linear mass conservation equations
H	nodal head
$H^{\min}$	minimum nodal heads
I	indicator function
L	length of a pipe
N	number of samples
$\hat{p}$	probability
$\hat{p}_0$	initial probability
P	mapping matrix
q	nodal demand
$\hat{q}$	mean nodal demand
q_{D}	deterministic demand
q_{E}	ellipsoidal uncertainty demand
q_{I}	interval uncertainty demand
Q	flow in pipe

doi: 10.2166/hydro.2013.138

R	reliability
S	performance measure function
t	iteration counter
u	perturbation vector
U	uncertainty set
x	decision variable vector
X	random sample
z	vector of ones
α	smoothing parameter
β	elite sample percentage
δ	interval uncertainty
γ	safety parameter of the uncertainty set
λ	Lagrangian multiplier
μ	Lagrangian multiplier
σ	standard deviation
Σ	covariance matrix
ρ	correlation coefficient
ω	elite samples performance
$	unit cost US$

INTRODUCTION

The least-cost design problem of a water distribution system (WDS) is to find the WDS's component characteristics: pipe diameters, pump heads and maximum power, and tank storage that minimize the total system cost, such that constraints at the consumer nodes are fulfilled and hydraulic laws are maintained. Over the last four decades, numerous models for least-cost design of WDSs have been published in the research literature ranging from early linear, nonlinear, and dynamic programming (Schaake & Lai 1969; Alperovits & Shamir 1977; Quindry et al. 1979, 1981; Kessler & Shamir 1989; Eiger et al. 1994; Ostfeld & Shamir 1996) to the more recent evolutionary/meta-heuristic schemes such as genetic algorithms (GAs) (Holland 1975) or ant colony (AC) (Dorigo 1992) optimization (Simpson et al. 1994; Dandy et al. 1996; Savic & Walters 1997; Salomons 2001; Vairavamoorthy & Ali 2005; Wu & Walski 2005; Ostfeld & Tubaltzev 2008). The former methods typically suffer from limited computational strength and problem representation (e.g. continuous diameters). The latter, on the other hand, encounter difficulties when dealing with constraints and must use weighted objective penalty functions, and a large number of simulations have to be performed since only zero-order information is used.

The approach in most previous studies was to treat the problem as deterministic, assuming perfectly known parameters. Consequently, such models are likely to perform poorly when implemented in reality when the actual problem parameters are revealed, hence the need to find more 'robust' solutions. Walski (2001) stressed that optimization based on cost minimization has serious limitations, and that there is a need for developing new network design strategies accounting for other benefits such as capacity, response to uncertainty, and reliability. It is generally accepted that there is a strong correlation between robustness, reliability, and capacity, and that increased reliability comes at a price. Typically, WDS reliability concerns the system's capability to supply consumer demands with adequate heads or pressures. In this work, WDS reliability is evaluated as the probability that the minimum allowable nodal pressures are met under demand variability with the assumption that the required nodal demand flows are satisfied (Xu & Goutler 1999).

More recently, new methodologies for optimal design/rehabilitation of WDSs under uncertainty have been developed. Various uncertainty handling techniques have been integrated with different optimization models for both single- and multi-objective formulations. Typically, uncertainty quantification was classified as: (1) the surrogate approach – reliability indexes; (2) the stochastic approach – analytical and sampling-based techniques; (3) fuzzy logic – stochastic simulation-based techniques; and (4) the deterministic equivalent – worst-case oriented techniques and uncertainty sets. Previous work has mainly concentrated on the surrogate approach and stochastic formulation of the uncertain problem, and all have evaluated reliability as the system's capability to supply consumer demands with adequate pressures.

Some researchers have developed and included surrogate reliability measures in their optimization formulation. Todini (2000) introduced a resilience index, which, for a given topology is expressed as an increase in energy redundancy (i.e. head surplus). Then, intuitively, increasing resilience increases reliability, regardless of the mechanism of failure and not requiring statistical analysis of the different types of uncertainty, the optimization problem is then formulated and solved for minimizing cost and maximizing resilience using the heuristic approach providing Pareto fronts. Prasad & Park (2004) further extended the resilience index to 'network resilience', which incorporates the effect of both the surplus head and the topology of the systems (reliable loops). The objective problem was again formulated as a two-objective optimizing cost and network resilience. A similar problem was formulated by Farmani *et al.* (2005) and solved using a non-dominating sorting genetic algorithm II (NSGAII). The optimization problem was solved using multi-objective GA. Raad *et al.* (2009) provided a performance comparison of different multi-objective heuristic methods for minimum cost versus reliability indexes (resilience index and network resilience). Other surrogate measures were proposed such as flow entropy (Tanyimboh & Templeman 1993) and mixed reliability index (Raad *et al.* 2010). Prasad & Tanyimboh (2008) demonstrated the use of a flow entropy index, which is a measure of the uniformity of pipe flows, as a surrogate measure for network reliability combined with the NSGAII two-objective optimization scheme. Raad *et al.* (2010) introduced a mixed reliability measure involving the flow entropy and the resilience indexes and compared the four surrogate reliability indexes for the multi-objective design problem using evolutionary algorithms. The results demonstrated a mixed performance of the three indexes depending on the system and the type of failure, whereas the flow entropy measure performed poorly in all circumstances. Lately, Jung *et al.* (2012) introduced reliability and robustness indexes and solved a two-objective optimization problem using NSGAII and the uncertainty was quantified using the first-order second moment method. All of these works were governed by the notion that energy (head) redundancy in the system will provide a more robust design of the system in the case of hydraulic uncertainty.

Stochastic formulation of the problem, explicitly expressing uncertainty, was first carried out by Lansey *et al.* (1989). Lansey *et al.* (1989) presented a chance-constraint formulation of the least-cost design problem considering uncertainty in nodal demands, minimum required heads, and pipe friction coefficients. A deterministic form of the chance-constraints method assuming the uncertain variables are independent and following normal distribution was formulated, resulting in a non-linear programming problem with continuous pipe diameters. The optimization problem was solved using the generalized reduced gradient (GRG2) method. Xu & Goutler (1999) formulated a single-optimization problem considering least-cost design and system reliability in a form of chance constraint. System reliability was calculated using the first-order reliability method, originally developed to evaluate structural reliability, and the optimization problem was solved using GRG2 considering uncertain demands and continuous diameters. Later, Tolson *et al.* (2004) improved the formulation of Xu & Goutler (1999) by formulating a single objective of the weighted cost and reliability index and using the GA to solve the optimization problem for discrete pipe diameters. The uncertainty was considered in the nodal demands, assumed to be normally distributed with given means and covariances, and pipe roughness coefficients were assumed to take their worst values at the end of the design horizon. Kapelan *et al.* (2005) formulated a two-objective optimization problem minimizing cost and maximizing system robustness, which was measured as its ability to simultaneously maintain minimum pressures expressed in a probabilistic form. Nodal demands and pipe roughnesses were considered uncertain with given probability density functions (PDFs). To handle the computational burden imposed by the stochastic nature of the uncertain problem formulation and the optimization technique: (1) Latin hypercube sampling was used to obtain a reliable representation of the underlying stochastic variables requiring fewer samples than a standard Monte Carlo sampling technique; and (2) a robust NSGAII was proposed requiring a much smaller number of candidate solutions in each iteration compared with NSGAII, thus further reducing the computational complexity. Babayan *et al.* (2007) expressed system robustness in terms of system probability to maintain adequate pressures, which was weighted in the single-objective formulation. The uncertainties were quantified using an integration method allowing for a limited number of hydraulic simulations and thus computationally attractive compared with straightforward stochastic techniques. Giustolisi *et al.* (2009) proposed a single- and multi-objective GA sampling-based methodology for the robust design of WDSs. The problem was formulated as a two-objective design optimizing cost and reliability, where the solution of the single-objective deterministic problem served as the starting population for the multi-objective GA to decrease running times. The reliability was evaluated based on stratified sampling (Latin hypercube), achieving significant computational savings compared with Monte Carlo sampling, and chance-constraint formulation. Nodal demands and pipe roughnesses were considered to be uncorrelated with a known PDF.

A fuzzy logic approach was proposed by Fu & Kapelan (2011). The optimal design problem was formulated as a two-objective optimization problem minimizing cost and maximizing system reliability, and solved using NSGAII. System reliability was represented as fuzzy random and was calculated as the expected value of the system acceptable performance level using the Latin hypercube sampling method. The nodal demands were assumed to be fuzzy random variables characterized by a normal distribution with a fuzzy mean, represented as a triangular fuzzy number with the original demand and its deviation as the centre and the base of the triangle, and percentage of the original demand as a standard deviation.

Babayan *et al.* (2005, 2007) proposed two deterministic equivalent formulations for the stochastic problem both solved using the GA: (1) a single-objective formulation replacing the stochastic chance constraint with a deterministic equivalent adding safety margins to the selected nodal head constraints. The approach requires identifying a small subset of critical nodes and assumes that head fluctuations are caused by demand fluctuations at immediate nodes only (Babayan *et al.* 2005); and (2) adding safety margins to uncertain parameters that resulted in a deterministic optimization problem. The safety margins were determined through iterative sensitivity analysis of the optimal solution of the deterministic problem using Monte Carlo sampling (Babayan *et al*. 2007).

Most previous work, not including the surrogate approach, is stochastic-oriented integrated with simulation-based techniques. The drawbacks of a stochastic-oriented approach are three-fold: PDFs and statistical dependencies between the uncertain variables need to be specified (which is obviously typically not available and the uncertain variables are assumed to be independent, uncorrelated, and to follow normal distribution); analytical solutions are generally not available and thus the uncertainty is evaluated through the large number of samples (Monte Carlo, Latin hypercube), which substantially increases the size of the original optimization problem; widely used simulation-based optimization techniques (e.g. GA), capable of handling large discrete problems, are known for their computational complexity, which is further enhanced due to the size of the stochastic problem.

This study proposes formulating a deterministic equivalent of the stochastic problem of optimal design/rehabilitation of WDSs, namely, a non-probabilistic robust counterpart (RC) (Ben-Tal & Nemirovski 1998, 1999), a more recent approach to optimization under uncertainty. The uncertainty of the information is quantified by a deterministic user-defined ellipsoidal uncertainty set, which can be probabilistically justified, with the decision maker seeking a solution that is optimal for all possible realizations in the uncertainty set. The RC makes no assumptions about the PDF of the uncertain variables and their dependencies, does not require the construction of a representative sample of scenarios, and has the same size as the original model. The RC approach was previously applied to an optimal multi-year management problem of a water supply system under recharge uncertainty (Housh *et al.* 2011) and is adapted herein to treat the least-cost design problem of WDSs. The resulting equivalent deterministic problem is solved using the cross entropy (CE) optimization method (Rubinstein 1999) previously successfully applied to single- and multi-objective optimal design of WDSs (Perelman & Ostfeld 2007; Perelman *et al.* 2008).

The proposed methodology is demonstrated on an illustrative example and on the Hanoi network (Fujiwara & Khang 1990). The results show considerable promise of the RC approach in terms of its tractability and size of model, as well as for being able to explore the trade-off between demand uncertainty and cost.

METHODOLOGY

Throughout this work, 'Robust optimization' (Bertisimas *et al.* 2011) is referred to as a *non-stochastic* approach for optimization under uncertainty and 'RC' (Ben-Tal & Nemirovski 1999) as the *deterministic equivalent* of the uncertain problem. The application of the robust approach to least-cost design of a WDS involves formulating the deterministic RC, solving the optimization problem, and evaluating the reliability of the attained design. The robust optimization, RC, optimization technique, and application to WDSs are described next.

ROBUST OPTIMIZATION

The goal of robust optimization is to find single solutions that guarantee feasibility independent of the data, as opposed to a sensitivity analysis approach aimed at quantifying the sensitivity of optimal solutions to small perturbations in the underlying problem.

The general robust optimization formulation is:

Minimize $f_0(x)$

$$\text{Subject to: } f_i(x, a_i) \geq 0 \;\forall a_i \in U_i, i = 1, \ldots, m \tag{1}$$

where x is a vector of decision variables, $f_0(\cdot)$ and $f_i(\cdot)$ are the objective and the ith constraint functions, respectively, a_i are uncertain parameters and U_i are uncertainty sets.

The goal is to find the optimal solutions among all feasible realizations of the disturbances within U_i. The formulation in Equation (1) also incorporates the cases in which: (1) the objective function is affected by the parameter uncertainty; (2) some constraints are without uncertainty; and (3) all parameters belong to a single uncertainty set $\forall a_i \in U, i = 1, \ldots, m$. To incorporate the above, an auxiliary variable can be introduced with additional constraint to represent uncertainty in the objective function, constraints without uncertainty can be represented by corresponding singletons, and an equivalent problem is with the u_i taken as the projection of u along its corresponding dimensions (Ben-Tal & Nemirovski 1998, 1999).

Typically, the addition of robustness to an optimization problem is intractable as it significantly increases computational complexity because of the large number of scenarios required to correctly estimate the stochastic process. However, for classes of functions f_i combined with types of uncertainty sets u_i, many robust problems can yield tractable RCs, in other words, the size of the robust deterministic equivalent remains similar to the original optimization problem. Bertisimas et al. (2011) provided a broad review of theory and application of robust optimization. Below, we will review the structure, tractability, and flexibility of robust optimization for WDS least-cost design.

The RC (Equation (1)) for a linear optimization problem can be written, without loss of generality, in a constraint-wise form:

Minimize $c^T x$

$$\text{Subject to: } a_i^T x \geq 0 \;\forall a_i \in U_i, \forall i; f^T x = 1 \tag{2}$$

where $f = (0, \ldots, 0, 1)^T$ is used to include uncertainty on the right-hand side.

Uncertainty set

The next question is how to define the uncertainty sets U_i. Consider that a_i lies in the interval $[\hat{a}_i - \delta_i, \hat{a}_i + \delta_i]$, where $\hat{a}_i$ is the nominal value, and δ_i is a deviation from $\hat{a}_i$. The uncertainty set is in the form of a box $|a_i - \hat{a}_i| \leq \delta_i$, and the straightforward robust optimal state is with the largest worst-case scenario where all parameters are simultaneously at their extremes.

Ben-Tal & Nemirovski (1999) considered ellipsoidal uncertainty sets such that a less conservative solution can be derived. It is plausible to assume that the uncertain data are not expected to be at their worst values simultaneously, consequently, the uncertainty can be quantified by an m-dimensional ellipsoid (m is the number of uncertain parameters) instead of a box.

The ellipsoidal uncertainty set considers all demands that are within some specified distance γ from the mean demands. The distance is measured using the Mahalanobis distance, which is an adapted version of the Euclidian distance measure, and incorporates the correlations between the variables. For uncertain parameters a, with nominal value $\hat{a}$ and covariance matrix Σ, the ellipsoidal uncertainty set is written in the form:

$$U(\gamma) = \left\{a | (a - \hat{a})^T \Sigma^{-1} (a - \hat{a}) \leq \gamma^2 \right\} \tag{3}$$

⇓

$\|P^{-1}(a-\hat{a})\|^2 \leq \gamma^2$ with a change of variable $\Sigma = P \cdot P^T$

$\Downarrow$

$\|u\| \leq \gamma$ where $u = P^{-1}(a-\hat{a}) = P^{-1}a - P^{-1}\hat{a}$. Rearranging the term, the affine mapping of the uncertain variable is:

$$a = a(u) = \hat{a} + Pu \tag{4}$$

Equation (3) can be rewritten as:

$$U(\gamma) = \{a|\hat{a} + Pu, \|u\| \leq \gamma\} \tag{5}$$

where P is the mapping matrix, u is the perturbation vector, $\|\cdot\|$ is the Euclidean norm, and γ is a value controlling the size of the ellipsoid.

Robust equivalent

To derive the robust equivalent, consider the uncertain ith constraint in Equation (2):

$$a^T x \geq 0 : \forall a \in U, \forall i \tag{6}$$

where a is uncertain with the mapping derived from Equation (4), i.e. $a(u) = \hat{a} + Pu$ and $\|u\| \leq \gamma \Rightarrow u^T u \leq \gamma^2$.

Because the constraint (Equation (6)) is uncertain, its minimum needs to be non-negative, hence:

$$\begin{aligned} &\min_{u:u^Tu\leq\gamma^2}\{a^T(u)x\} \geq 0 \\ &\Rightarrow \min_{u:u^Tu\leq\gamma^2}\{a^T(u)x\} = \hat{a}^T x + \min_{u:u^Tu\leq\gamma^2}\{u^T P^T x\} \end{aligned} \tag{7}$$

To get the robust constraint, we need to solve the optimization problem:

$$\min_{u:u^Tu\leq\gamma^2}\{u^T P^T x\} \tag{8}$$

The analytical solution to the problem can be attained through Lagrangian duality (see the Appendix, available online at http://www.iwaponline.com/jh/015/238.pdf) and is equal to:

$$-\gamma\|P^T x\| \tag{9}$$

Substituting Equation (9) into Equation (7), the uncertain constraint (Equation (6)) is replaced by its robust equivalent:

$$\hat{a}^T x - \gamma\|P^T x\| \geq 0 \tag{10}$$

where P is the affine mapping matrix, which can be computed by Cholesky decomposition of the covariance matrix $\Sigma = P \cdot P^T$ and γ is a safety factor representing the uncertainty set size. For larger values of γ, the design is more conservative and for a zero value, the nominal design is attained not considering the uncertainty.

LEAST-COST DESIGN OF WDSS

The least-cost design objective of a WDS is to find its minimum cost with discrete diameters as decision variables, linear mass conservation equations (Equation (11)), nonlinear energy conservation equations (Equation (12)), and head bounds constraints (Equation (13)). Todini & Pilati (1988) generalized the mass and energy constraints in a matrix form:

$$f_Q(Q, H) = A_{21}Q - q = 0 \tag{11}$$

$$f_P(Q, H) = A_{11}Q + A_{12}H = 0 \tag{12}$$

$$H \geq H^{\min} \tag{13}$$

where $A_{21} = A_{12}^T$ is a topological matrix with elements of the ith row equal to $\{1, -1, 0\}$, which depend on the network connectivity, A_{11} is a diagonal matrix with nonlinear elements representing the pipe's resistance, Q and H are flows and heads, q are the consumer's demands, and $H^{\min}$ are the minimum desired nodal heads.

Considering uncertainty in the demands, the mass balance constraint (Equation (10)) is uncertain corresponding to the constraint in Equation (6), with only the right-hand-side parameters being uncertain, i.e. $q \in U$. Then, the deterministic equivalent of the mass balance equation according to Equations (7)–(10) is:

$$A_{21}Q - \hat{q} - \gamma \|P^T\| = 0 \qquad (14)$$

and Equation (12) is replaced with Equation (14) in the RC formulation. So to compute the RC of the uncertain demand constraint: the mean nodal demands $\hat{q}$, the covariance matrix of the demands Σ, and the safety parameter γ need to be specified. The mapping matrix P, again, can be computed by Cholesky decomposition of the covariance matrix $\Sigma = P \cdot P^T$. It can be seen that the current formulation of the RC to WDSs results in a degenerate case, comparing Equation (14) to Equation (10), because, in each mass balance equation, only one demand is apparent. Consequently, correlations between demands are not explicitly accounted for.

Finally, to solve the RC, the existing optimization approach for the least-cost design of WDSs can be used (GA, AC) without any modifications keeping the same problem size. New robust demands are computed based on Equation (14) and the problem is solved using, for example, the EPANET-GA iterative scheme.

MOTIVATING EXAMPLE

The motivating example is presented to answer the questions of how to define the uncertainty set and why using ellipsoidal uncertainty. An ellipsoid E is incomparably less than a box B and the difference becomes more dramatic the larger the dimension m. It follows, that the ellipsoidal RC of the random constraint is much less conservative than the interval RC. To demonstrate this point, consider a single source supplying three demand zones, as shown in Figure 1. The demand of one zone is assumed to be known with certainty, and the demands of the remaining two zones are uncertain, their mean demands and deviations from the means are estimated and given in a form of $\hat{q}$ (mean) and Σ (covariance matrix). In deterministic design, the system is optimized taking the following total demand: $q_D = q_1 + q_2 + q_3$. In robust design, the system is optimized taking uncertain demands into consideration. Consider two cases: (1) interval uncertainty; and (2) ellipsoidal uncertainty.

Figure 1 | Motivating example.

Interval uncertainty

Suppose that each demand lies in the interval $[\hat{q}_i - \sigma_i, \hat{q}_i + \sigma_i]$, then the deterministic constraint is replaced with it RC, as a result all the demands take their worst-case values:

$$\min_{u:|u|\le\delta}\{a^T x - b(u)\} = a^T x + \min_{u:|u|\le\delta}\{-b(u)\} = a^T x - q_I \ge 0$$

where b is the uncertain parameter of the right-hand side of the constraint, $q_I = q_1 + \hat{q}_2 + \sigma_2 + \hat{q}_3 + \sigma_3$ is the interval uncertainty demand equivalent, $\hat{q}_2, \hat{q}_3$ are means, and σ_2, σ_3 are standard deviations of nodal demands 2 and 3, respectively.

Ellipsoidal uncertainty

$$\min_{u:u^T u\le\gamma^2}\{a^T x - b(u)\} = a^T x - \hat{b} + \min_{u:u^T u\le\gamma^2}\{-u^T P^T\}$$
$$= a^T x - \hat{b} - \gamma\|P^T z\| = a^T x - q_E \ge 0$$

where z is a vector of ones. Thus, the robust equivalent demand with ellipsoidal uncertainty is:

$$q_E = q_1 + \hat{q}_2 + \hat{q}_3 + \gamma \left\| P^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|$$

Given the mean and the covariance matrix of the uncertain demands:

$$\hat{q} = \begin{bmatrix} 2 \\ 2.5 \end{bmatrix}, \sigma^2 = \begin{bmatrix} 0.48 \\ 0.75 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.48 & \rho\sigma_2\sigma_3 \\ \rho\sigma_2\sigma_3 & 0.75 \end{bmatrix}$$

where $-1 \le \rho \le 1$ is a correlation coefficient.

The deterministic demand is:

$$q_D = q_1 + q_2 + q_3 = q_1 + 4.5$$

The interval uncertainty equivalent is:

$$q_I = q_1 + \hat{q}_2 + \hat{q}_3 + \sigma_2 + \sigma_3 = q_1 + 4.5 + 1.558$$

The ellipsoidal uncertainty equivalent, as a function of correlation coefficient ρ and for a maximum uncertainty set size $\gamma = 1$, is:

$$q(\rho)_E = q_1 + 4.5 + 1 \cdot \left\| P^T(\rho) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = q_1 + 4.5 + \left\| P^T(\rho) \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|$$

where

$$\rho = 1 \Rightarrow \Sigma = \begin{bmatrix} 0.48 & 0.6 \\ 0.6 & 0.75 \end{bmatrix}, P^T = \begin{bmatrix} 0.693 & 0 \\ 0.866 & 0 \end{bmatrix}$$

$$\Rightarrow \left\| P^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = 1.558$$

$$q(\rho = 1)_E = q_1 + 4.5 + \left\| P^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = q_1 + 4.5 + 1.558$$

$$\rho = 0.5 \Rightarrow \Sigma = \begin{bmatrix} 0.48 & 0.5 \times 0.6 \\ 0.5 \times 0.6 & 0.75 \end{bmatrix}, P^T = \begin{bmatrix} 0.693 & 0 \\ 0.433 & 0.75 \end{bmatrix}$$

$$\Rightarrow \left\| P^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = 1.353$$

$$q(\rho = 0.5)_E = q_1 + 4.5 + \left\| P^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = q_1 + 4.5 + 1.353$$

Table 1 lists robust demand equivalents for interval and ellipsoidal uncertainties for different values of correlation and maximum uncertainty size. Table 1 demonstrates that for $\rho = 1$, the interval and ellipsoidal uncertainties coincide, complying with the worst-case scenario. For lower values of ρ, the robust ellipsoidal design is incomparably less conservative, and the difference will be more evident if the number of dimensions is higher. The results also agree with the reasoning that if the demand zones, for example, represent different consumer types that do not reach their peak demand at the same time, the design should be based accordingly. From the example above, the ellipsoidal counterpart accounts for the correlations between the demands, whereas the interval counterpart does not.

Table 1 | Robust equivalent demand

	Correlation between demand zones				
Uncertainty	$\rho = 1$	$\rho = 0.5$	$\rho = 0$	$\rho = -0.5$	$\rho = -1$
Interval	1.558	–	–	–	–
Ellipsoidal	1.558	1.353	1.109	0.794	0.173

CE METHOD FOR OPTIMIZATION

Any typical simulation-based optimization scheme (GA, AC) can be utilized to solve the deterministic equivalent problem. In this work, the CE method for combinatorial optimization (Rubinstein 1999; Rubinstein & Kroese 2004), previously successfully applied to single- and multi-objective optimal design of WDS (Perelman & Ostfeld 2007; Perelman et al. 2008), is used. The CE scheme seeks to find an optimal probability, such that the Kullback–Leibler distance (Kullback & Leibler 1951) between the sampling probability and the theoretical optimal probability is minimal.

The CE algorithm is a two-stage iterative procedure involving the following stages:

1. Choose an initial probability $\hat{p}_0$. Set an iteration counter $t = 1$.
2. Generate a random sample $X_1, X_2, \ldots, X_N$ from the density $\hat{p}_{t-1}$ and evaluate each sample using some measure function (e.g. cost $S(X_i)$). Sort their performances in a descending order and select β percentage of samples with the best performance, i.e. the elite samples and set $\hat{\omega}_t = S_{[(1-\beta)N]}$.
3. Update $\hat{p}_t$ using the same elite samples by:

   $$\hat{p}_{t,j} = \frac{\sum_i^N I_{\hat{\omega}_t} X_{i,j}}{\sum_i^N I_{\hat{\omega}_t}}$$

   where $I_{\hat{\omega}_t}$ is an indicator function. Set $\hat{p}_t \leftarrow \alpha \cdot \hat{p}_t + (1 - \alpha) \cdot \hat{p}_{t-1}$, where α is a smoothing parameter between 0 and 1.
4. If stopping criteria are met, then STOP; otherwise set $t = t + 1$ and return to step 2.

The CE algorithm parameters are the sample size N, the elite sample percentage β, and the smoothing parameter α.

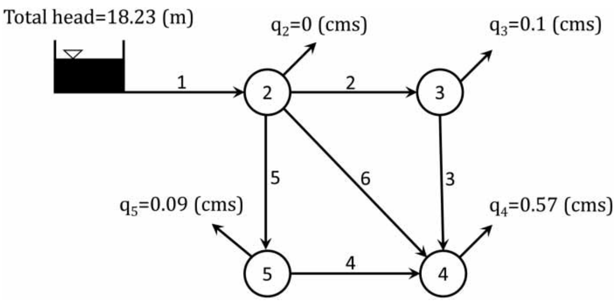

Figure 2 | Illustrative example layout.

Parameter values are generally set through model calibration. A full description of the CE method and its application to single- and multi-objective optimal design of WDSs can be found in Perelman & Ostfeld (2007) and Perelman *et al.* (2008).

APPLICATIONS

The proposed methodology is demonstrated on an illustrative example and tested on the Hanoi network (Fujiwara & Khang 1990).

Illustrative example

The layout of the network and its base demands are shown in Figure 2. It consists of a single constant head source, four demand nodes, and six pipes. The length and friction coefficient of all links, and elevation of all nodes are 304.8 m, 100, and 0 m, respectively. For ease of representation, the demands at nodes 2 and 5 are considered to be constant, whereas the demands at nodes 3 and 4 are assumed uncertain, with given means, standard deviations, and correlation $\hat{q} = (0.1, 0.057), \sigma = (0.0343, 0.0196), \rho = 0$, respectively. The available diameters considered in this example are 203.2, 254, 304.8, 355.6, 406.4 mm, with corresponding costs of 75.46, 105, 164, 196.85, 295.28 unit cost/m. Next, setting the safety parameter, for example, to $\gamma = 0.4$, the goal is to find the optimal corresponding design.

As explained previously, the entire process roughly involves the following: (1) formulating the RC based on the mean and covariance matrix of the uncertain data and user selected safety factor ($\hat{q}, \Sigma, \gamma$ supplied above); (2) solving the optimization problem – the solution space in this small example is 5^6, which can be easily enumerated to find the optimal solutions for different sizes of uncertainty sets $U(\gamma)$(without the need for optimization); and (3) evaluate reliability of the design through simulation under demand variability. Practically, the attained design is simulated under varying demands using EPANET (USEPA 2002), and the minimum head constraints are evaluated for each simulation (i.e. feasibility of the solution based on Equation (13)). The reliability is computed as the fraction of scenarios that the heads resulting from the

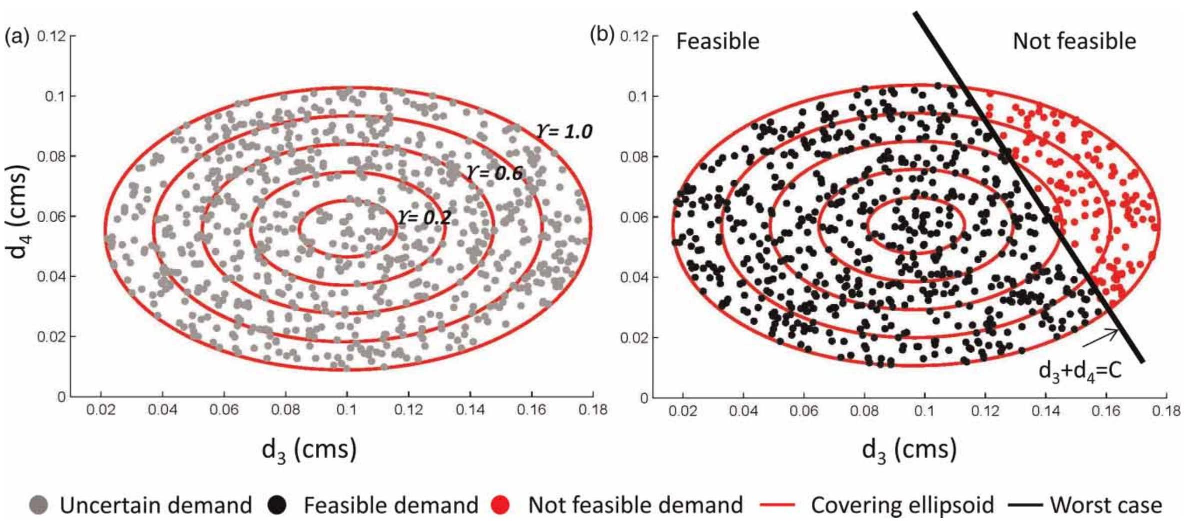

Figure 3 | (a) Uncertainty ellipsoids and (b) feasibility of optimal design for uncertainty ellipsoid $U(\gamma = 0.4)$.

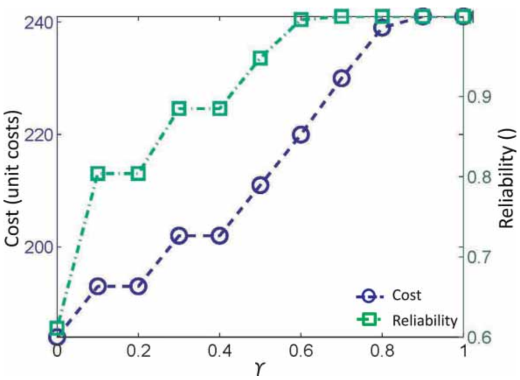

Figure 4 | Cost and reliability versus γ.

Table 2 | Illustrative example of pipe diameters

		Diameter (mm)					
γ	**Cost (unit cost)**	**Pipe 1**	**Pipe 2**	**Pipe 3**	**Pipe 4**	**Pipe 5**	**Pipe 6**
0	184	355.6	203.2	203.2	203.2	203.2	254
0.1,0.2	193	355.6	254	203.2	203.2	203.2	254
0.3,0.4	202	355.6	254	203.2	203.2	254	254
0.5	211	355.6	304.8	203.2	203.2	254	203.2
0.6	220	355.6	304.8	203.2	203.2	254	254
0.7	230	355.6	355.6	203.2	203.2	254	254
0.8	239	355.6	355.6	254	203.2	254	254
0.9,1	241	406.4	304.8	203.2	203.2	203.2	254

simulation were above the minimum required, according to:

$$R = \frac{\sum_{i=1}^{N} I_{i\{H>H^{\min}\}}}{N} \tag{15}$$

where R is the reliability, $I_{\{H>H^{\min}\}}$ is an indicator function indicating design feasibility under demand scenario, and N is the number of scenarios.

Figure 3(a) demonstrates plotted uncertainty ellipsoids $U(\gamma)$ of different sizes corresponding to different values of γ and Figure 3(b) depicts the feasibility of the optimal robust design for $U(\gamma = 0.4)$. It can be seen that the optimal design is feasible within a covering ellipsoid of $U(0.4)$ and infeasible outside this region where the ellipsoid intersects with the worst-case combination of demands (with some redundancy because of the discrete nature of the pipe diameters). Next, optimal designs were computed for different values of γ, and the reliability of each design was computed based on 1,000 different demand scenarios assuming uniform uncorrelated demands. Figure 4 shows the cost and reliability of the robust optimal solutions versus γ. From the figure, it can be seen that both functions are non-decreasing and the reliability of a design increases with its cost,

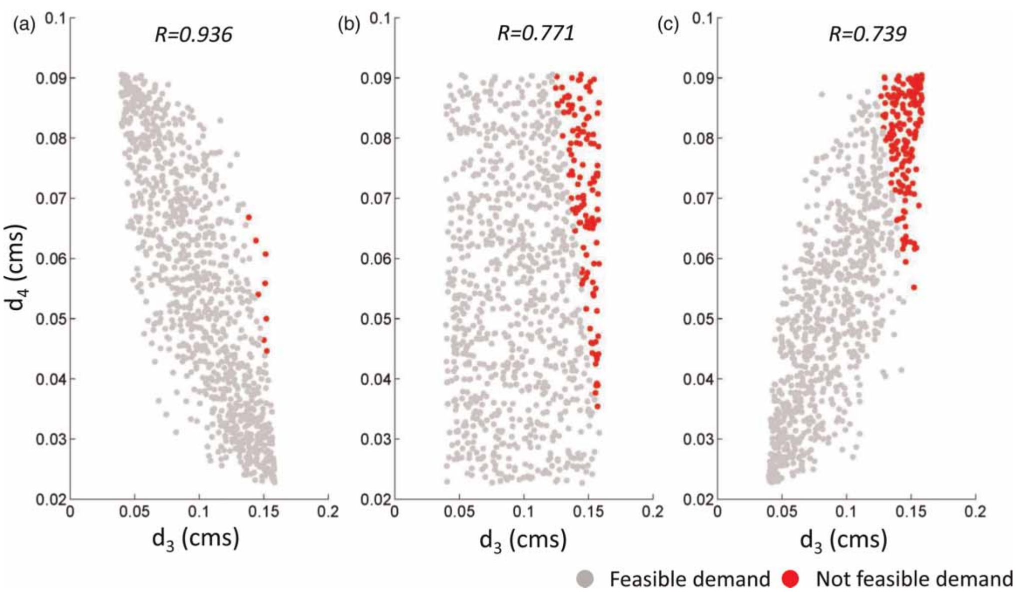

Figure 5 | Design reliability (R) for $\gamma = 0.4$: (a) negatively correlated demands, (b) uncorrelated demands and (c) positively correlated demands.

which implicates the worth of the approach. Table 2 lists the selected diameter for each design.

Next, the reliability of the optimal design for $\gamma = 0.4$ was evaluated for uncorrelated, positively, and negatively correlated demands, $\rho = 0, 0.8, -0.8$, respectively, with respect to the same mean and standard deviation of the demands. The first (uncorrelated) corresponds to random demands, the second (positively correlated) corresponds to similar types of consumers (e.g. domestic), and the third (negatively correlated) corresponds to possible different types of consumers (e.g. domestic and industry). Feasibility and reliability of the robust design are shown in Figure 5. It can be seen from Figure 5 that the hydraulic reliability of the robust design is significantly affected by the relationship between the demands, which is not explicitly expressed in this formulation of the optimization problem (Equation (13)). Figure 6

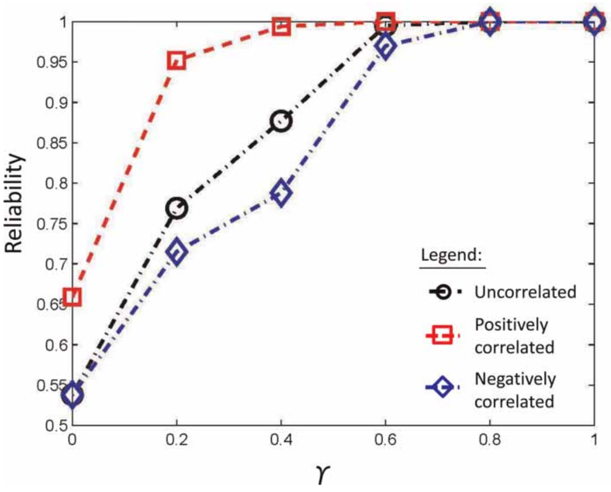

Figure 6 | Reliability versus ρ for positively, negatively, and uncorrelated demands.

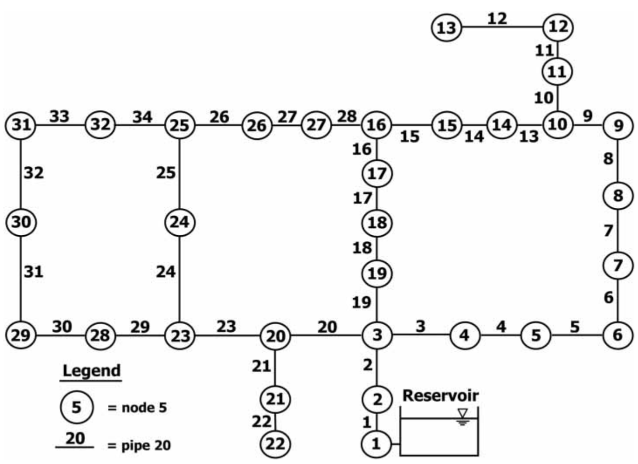

Figure 7 | The Hanoi network layout (Fujiwara & Khang 1990).

Table 3 | Robust solutions of the Hanoi network

Link	Diameter (mm) $\gamma = 0$	$\gamma = 0.05$	$\gamma = 0.10$	$\gamma = 0.15$
1	1,016	1,016	1,016	1,016
2	1,016	1,016	1,016	1,016
3	1,016	1,016	1,016	1,016
4	1,016	1,016	1,016	1,016
5	1,016	1,016	1,016	1,016
6	1,016	1,016	1,016	1,016
7	1,016	1,016	1,016	1,016
8	1,016	762	1,016	1,016
9	1,016	762	1,016	762
10	762	762	762	1,016
11	609.6	762	762	762
12	609.6	609.6	609.6	609.6
13	508	406.4	304.8	508
14	406.4	508	508	609.6
15	304.8	609.6	508	609.6
16	304.8	762	762	1,016
17	406.4	762	1,016	1,016
18	609.6	1,016	1,016	1,016
19	508	1,016	1,016	1,016
20	1,016	1,016	1,016	1,016
21	508	508	508	508
22	304.8	304.8	508	406.4
23	1,016	762	762	762
24	762	609.6	508	508
25	762	508	304.8	304.8
26	508	406.4	609.6	762
27	304.8	762	762	609.6
28	304.8	609.6	762	762
29	406.4	508	508	508
30	304.8	508	406.4	406.4
31	304.8	304.8	304.8	304.8
32	406.4	304.8	406.4	406.4
33	406.4	508	508	406.4
34	609.6	609.6	609.6	609.6
Cost ($\times 10^6$ \$)	6.08	6.59	6.84	7.18

further shows the design reliability for different values of γ for the three correlations. The results demonstrate that, as expected, system reliability is higher for negative correlation between demands and lower for positive correlation, since in the former case, the demands do not reach their peak simultaneously and in the latter, they do.

Hanoi network

The Hanoi network is a relatively large gravitational system introduced by Fujiwara & Khang (1990). The network (Figure 7) is subject to a one demand loading condition, and consists of 34 links and 32 demand nodes supplied by a single reservoir at a constant head of +100 m. The minimum pressure head requirement at all nodes is 30 m. All nodes are at zero elevation. Six candidate pipe diameters, 304.8, 406.4, 508, 609.6, 762 and 1,016 mm with a Hazen–Williams coefficient of 130, are considered for each of the links. The full data can be found in CWS (2001). The goal of the optimization is least-cost design where the design cost Cost ($) of installing a pipe of diameter D (mm) and length L (m) is:

$$\text{Cost} = 8.593 \times 10^{-3} D^{1.5} L \tag{16}$$

To formulate the RC, the mean demands are assumed to be the nominal demand of the original optimization problem (Fujiwara & Khang 1990) with a standard deviation of 10% of the mean demand. The CE optimization algorithm (Perelman & Ostfeld 2007, Perelman et al. 2008) was used to find robust optimal solutions for different sizes of covering ellipsoids, with 100,000 evaluations on average until convergence (around 20 min on Intel® 8 GB 2.80 GHz).

Table 3 lists attained robust solutions for different γ values, with $\gamma = 0$ being the deterministic solution. Figure 8(a) shows the design cost and reliability as a function of the size of the uncertainty set γ. Figure 8(b) describes the gain in reliability as a function of the additional invested cost. It can be seen from Figure 8 that an initial additional 8% invested cost will increase the reliability by more than 250% compared with the deterministic design. However, as the invested cost increases, the net benefit of the reliability substantially decreases.

CONCLUSIONS

In this study, an RC approach was suggested for least-cost design of WDSs. The results show considerable promise of the robust approach, which exhibits several advantages that need to be further investigated: (1) in contrast to stochastic optimization, the RC does not assume the parameter uncertainty is stochastic with underlying PDF; (2) it results in a deterministic equivalent, which is especially appealing due to the number of required simulations. This becomes even

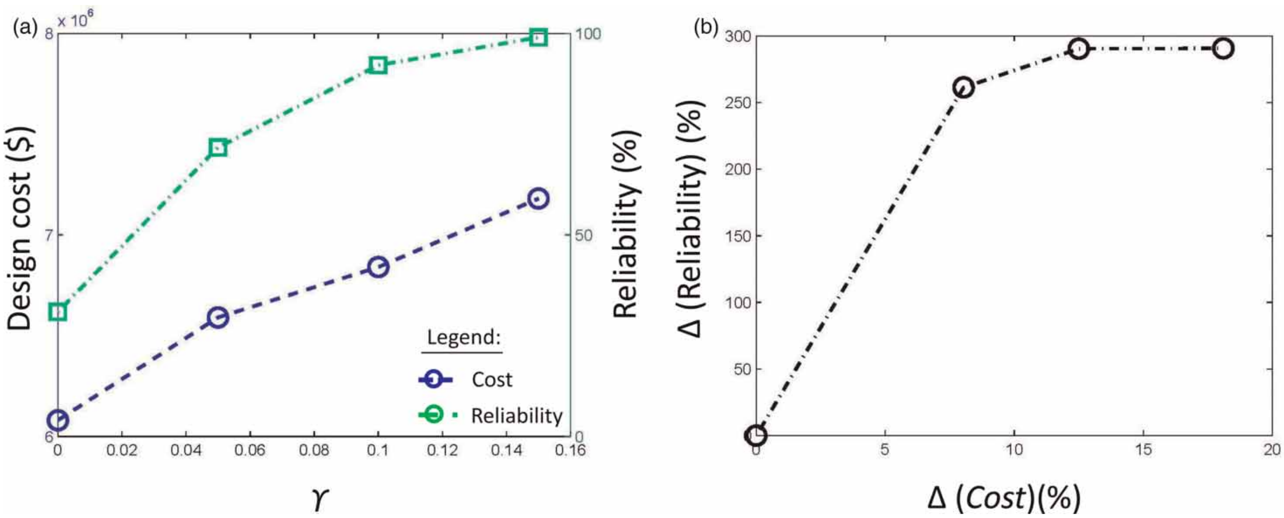

Figure 8 | Hanoi network: (a) design cost ($) and reliability (%) versus ρ, (b) gain in reliability (%) versus invested cost (%).

more evident when the problem is solved using simulation-based optimization techniques; (3) it can provide a trade-off between reliability and cost as a function of safety parameter γ set by the engineer; and (4) its implementation is rather straightforward and flexible.

It is important to note some of the drawbacks of the RC approach, particularly in the current formulation: (1) it is a min–max oriented approach, thus the solution may be conservative; (2) the current formulation of the RC to WDS results in a degenerate case, Equation (14) compared with Equation (10), because, in each mass balance equation, only one demand is apparent. Consequently, correlations between demands are not explicitly accounted for; and (3) the current formulation does not explicitly relate the uncertainty to the reliability, i.e. γ to R, and the reliability is computed after the problem is solved. The RC application to WDS design should be further investigated addressing all the above-mentioned issues including application to more complex networks.

REFERENCES

Alperovits, E. & Shamir, U. 1977 Design of optimal water distribution systems. *Water Resources Research* **13** (6), 885–900.

Babayan, A. V., Kapelan, Z., Savic, D. A. & Walters, G. A. 2005 Least-cost design of water distribution networks under demand uncertainty. *Journal of Water Resources Planning and Management* **131** (5), 375–382.

Babayan, A. V., Savic, D. A., Walters, G. A. & Kapelan, Z. S. 2007 Robust least-cost design of water distribution networks using redundancy and integration-based methodologies. *Journal of Water Resources Planning and Management* **133** (1), 67–77.

Ben-Tal, A. & Nemirovski, A. 1998 Robust convex optimization. *Mathematics of Operational Research* **23**, 769–805.

Ben-Tal, A. & Nemirovski, A. 1999 Robust solutions of uncertain linear programs. *Operational Research Letters* **25**, 1–13.

Bertisimas, D., Brown, D. B. & Caramanis, C. 2011 Theory and applications of robust optimization. *SIAM Review* **53** (3), 464–501.

CWS Benchmarks 2001 Available online at: www.projects.ex.ac.uk/cws/.

Dandy, G. C., Simpson, A. R. & Murphy, L. J. 1996 An improved genetic algorithm for pipe network optimization. *Water Resources Research* **32** (2), 449–458.

Dorigo, M. 1992 Optimization, Learning and Natural Algorithms [in Italian]. PhD Thesis, Dipartimento di Electtronica Politecnico di Milano, Milan.

Eiger, G., Shamir, U. & Ben-Tal, A. 1994 Optimal design of water distribution networks. *Water Resources Research* **30** (9), 2637–2646.

Farmani, R., Walters, G. A. & Savic, D. A. 2005 Trade-off between total cost and reliability for Anytown water distribution network. *Journal of Water Resources Planning and Management* **131** (3), 161–171.

Fu, G. & Kapelan, Z. 2011 Fuzzy probabilistic design of water distribution networks. *Water Resources Research* **47**, W05538.

Fujiwara, O. & Khang, D. B. 1990 A two-phase decomposition method for optimal design of looped water distribution networks. *Water Resources Research* **26** (4), 539–549.

Giustolisi, O., Laucelli, D. & Colombo, A. F. 2009 Deterministic versus aleatoric design of water distribution networks. *Journal of Water Resources Planning and Management* **135** (2), 117–127.

Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.

Housh, M., Ostfeld, A. & Shamir, U. 2011 Optimal multi-year management of water supply system under uncertainty: a robust counterpart approach. *Water Resources Research* **47**, W105.

Jung, D., Kang, D. & Lansey, K. 2012 Water distribution network design using a robustness index. In: *Proceedings of the World Environmental and Water Resources Congress*, EWRI, ASCE, Albuquerque, New Mexico, USA.

Kapelan, Z. S., Savic, D. A. & Walters, G. A. 2005 Multiobjective design of water distribution systems under uncertainty. *Water Resources Research* **41**, W11407.

Kessler, A. & Shamir, U. 1989 Analysis of the linear programming gradient method for optimal design of water supply networks. *Water Resources Research* **25** (7), 1469–1480.

Kullback, S. & Leibler, R. A. 1951 On information and sufficiency. *Annals of Mathematical Statistics* **22** (1), 79–86.

Lansey, K. E., Duan, N., Mays, L. W. & Yeou-Kung, T. 1989 Water distribution system design under uncertainties. *Journal of Water Resources Planning and Management* **115** (5), 630–645.

Ostfeld, A. & Shamir, U. 1996 Design of optimal reliable multiquality water supply systems. *Journal of Water Resources Planning and Management* **122** (5), 322–333.

Ostfeld, A. & Tubaltzev, A. 2008 Ant colony optimization for least cost design of water distribution systems. *Journal of Water Resources Planning and Management* **134** (2), 107–118.

Perelman, L. & Ostfeld, A. 2007 An adaptive heuristic cross entropy algorithm for optimal design of water distribution systems. *Engineering Optimization* **39** (4), 413–428.

Perelman, L., Ostfeld, A. & Salomons, E. 2008 Cross entropy multiobjective optimization for water distribution systems design. *Water Resources Research* **44**, W094.

Prasad, T. D. & Park, N. S. 2004 Multiobjective genetic algorithms for design of water distribution networks. *Journal of Water Resources Planning and Management* **130** (1), 73–82.

Prasad, T. D. & Tanyimboh, T. T. 2008 Entropy based design of 'Anytown' water distribution network. In: *Proceedings of the 10th Annual Water Distribution Systems Analysis Conference*, American Society of Civil Engineering, Kruger National Park, South Africa.

Quindry, G. E., Brill, E. D. & Liebman, J. C. 1981 Optimization of looped water distribution systems. *Journal of Environmental Engineering* **107** (EE4), 665–679.

Quindry, G. E., Brill, E. D., Liebman, J. C. & Robinson, A. R. 1979 Comment on 'Design of optimal water distribution systems' by Alperovits E. and Shamir U. *Water Resources Research* **15** (6), 1651–1654.

Raad, D. N., Sinske, A. N. & van Vuuren, J. H. 2009 Robust multiobjective optimization for water distribution system design using a meta-metaheuristic. *International Transactions in Operational Research* **16**, 595–626.

Raad, D. N., Sinske, A. N. & van Vuuren, J. H. 2010 Comparison of four reliability surrogate measures for water distribution systems design. *Water Resources Research* **46**, W05524.

Rubinstein, R. Y. 1999 The cross-entropy method for combinatorial and continuous optimization. *Methodology and Computing in Applied Probability* **1**, 127–190.

Rubinstein, R. Y. & Kroese, D. P. 2004 *The Cross-Entropy Method: a Unified Approach to Combinatorial Optimization, Monte Carlo Simulation, and Machine Learning*. Springer Science and Business Media Inc., USA.

Salomons, E. 2001 Optimal Design of Water Distribution Systems Facilities and Operation. MS Thesis, Technion, Haifa, Israel (in Hebrew).

Savic, D. & Walters, G. 1997 Genetic algorithms for least cost design of water distribution networks. *Journal of Water Resources Planning and Management* **123** (2), 67–77.

Schaake, J. C. & Lai, D. 1969 Linear Programming and Dynamic Programming Application to Water Distribution Network Design. Report No. 116, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA.

Simpson, A. R., Dandy, G. C. & Murphy, L. J. 1994 Genetic algorithms compared to other techniques for pipe optimization. *Journal of Water Resources Planning and Management* **120** (4), 423–443.

Tanyimboh, T. T. & Templeman, A. B. 1993 Calculating maximum entropy flows in networks. *Journal of the Operational Research Society* **44** (4), 383–396.

Todini, E. 2000 Looped water distribution networks design using a resilience index based heuristic approach. *Urban Water* **2**, 115–122.

Todini, E. & Pilati, S. 1988 A gradient algorithm for the analysis of pipe network. In: *Computer Applications in Water Supply* (B. Coulbeck & O. Chun-Hou, eds), Vol. 1, System Analysis and Simulation. Research Studies Press Ltd, Taunton, UK, pp. 1–20.

Tolson, B. A., Maier, H. R., Simpson, A. R. & Lence, B. J. 2004 Genetic algorithms for reliability-based optimization of water distribution systems. *Journal of Water Resources Planning and Management* **130** (1), 63–72.

USEPA 2002 EPANET 2.0. Available online at: www.epa.gov/ORD/NRMRL/wswrd/epanet.html.

Vairavamoorthy, K. & Ali, M. 2005 Pipe index vector: a method to improve genetic-algorithm-based pipe optimization. *Journal of Hydraulic Engineering* **131** (12), 1117–1125.

Walski, T. M. 2001 The wrong paradigm – why water distribution optimization doesn't work. *Journal of Water Resources Planning and Management* **127** (2), 203–205.

Wu, Z. Y. & Walski, T. 2005 Self-adaptive penalty approach compared with other constraint-handling techniques for pipeline optimization. *Journal of Water Resources Planning and Management* **131** (3), 181–192.

Xu, C. & Goulter, I. C. 1999 Reliability-based optimal design of water distribution networks. *Journal of Water Resources Planning and Management* **125** (6), 352–362.

First received 23 August 2012; accepted in revised form 13 December 2012. Available online 31 January 2013