CrossMark

Self-Adaptive Solution-Space Reduction Algorithm for Multi-Objective Evolutionary Design Optimization of Water Distribution Networks

Tiku T. Tanyimboh[1,2] · Anna Czajkowska[2,3]

Abstract An effective way to improve the computational efficiency of evolutionary algorithms is to make the solution space of the optimization problem under consideration smaller. A new reliability-based algorithm that does this was developed for water distribution networks. The objectives considered in the formulation of the optimization problem were minimization of the initial construction cost and maximization of the flow entropy as a resilience surrogate. After achieving feasible solutions, the active solution space of the optimization problem was re-set for each pipe in each generation until the end of the optimization. The algorithm re-sets the active solution space by reducing the number of pipe diameter options for each pipe, based on the most likely flow distribution. The main components of the methodology include an optimizer, a hydraulic simulator and an algorithm that calculates the flow entropy for any given network configuration. The methodology developed is generic and self-adaptive, and prior setting of the reduced solution space is not required. A benchmark network in the literature was investigated, and the results showed that the algorithm improved the computational efficiency and quality of the solutions achieved by a considerable margin.

Keywords Dynamic solution-space reduction · Maximum entropy formalism · Reliability-based design · Water distribution network · Self-adaptive boundary search · Failure tolerance and resilience

✉ Tiku T. Tanyimboh
   tiku.tanyimboh@wits.ac.za

   Anna Czajkowska
   anna.czajkowska.civeng@gmail.com

1   School of Civil and Environmental Engineering, University of the Witwatersrand,
    Johannesburg Private Bag 3, WITS 2050, South Africa

2   Department of Civil and Environmental Engineering, University of Strathclyde, 75 Montrose Street,
    Glasgow G1 1XJ, UK

3   RPS Group, Mercantile Building, 53 Bothwell Street, Glasgow G2 6TS, UK

1 Introduction

Genetic algorithms are a class of population-based approaches that can search different regions in the solution space of an optimization problem simultaneously. They are therefore well suited to complex multi-objective optimization problems. Various multi-objective evolutionary optimization algorithms have been developed in the last three decades. Gen et al. (2008) grouped multi-objective evolutionary algorithms (MOEAs) based on fitness assignment while Deb (2001) classified them as non-elitist and elitist. Elitism is important as it helps to achieve better convergence and ensure the fittest candidates are not lost even if they are achieved at an early stage in the optimization (Zitzler et al. 2000).

Vector Evaluated Genetic Algorithm (VEGA) (Schaffer 1985), Vector Optimized Evolution Strategy (Kursawe 1990), Multi-Objective Genetic Algorithm (MOGA) (Fonseca and Fleming 1993), Weight Based Genetic Algorithm (Hajela and Lin 1992), Nondominated Sorting Genetic Algorithm (NSGA) (Srinivas and Deb 1994) are examples of non-elitist evolutionary algorithms. Examples of elitist MOEAs include Nondominated Sorting Genetic Algorithm (NSGA) II (Deb et al. 2002), Distance Based Pareto Genetic Algorithm (Osyczka and Kundu 1995), Strength Pareto Evolutionary Algorithm (SPEA) (Zitzler and Thiele 1998) and Pareto-Archive Evolution Strategy (PAES) (Knowles and Corne 2000).

Deb et al. (2002) proposed Nondominated Sorting Genetic Algorithm (NSGA) II. NSGA II includes elitism and preserves diversity among the nondominated solutions based on the crowding distance. Solutions that have relatively large crowding distance values occupy less crowded regions of the solution space and preserving such solutions results in better diversity. Comparisons of NSGA II and other evolutionary algorithms are available in the literature. For example, Deb et al. (2002) assessed NSGA II, PAES and SPEA on nine difficult problems. NSGA II achieved the best convergence and most diverse solutions. Many researchers in various disciplines have employed NSGA II, and previous investigations on water distribution networks in the literature have demonstrated that it is practical and efficient. It was thus adopted here.

The population size, genetic operators, encoding and other factors influence the performance of genetic algorithms (McClymont et al. 2015) and there is no doubt that the size of the solution space has the greatest influence. Thus, it is advantageous to reduce the size of the solution space to improve the computational efficiency and solution quality (Maier et al. 2014). However, from the perspective of the development of rigorous, practical and computationally efficient algorithms, it has received very little attention, perhaps because it is complex and challenging. Furthermore, the approaches proposed so far for water distribution networks have involved cost minimization only.

Vairavamoorthy and Ali (2005) demonstrated the effectiveness of reducing the size of the solution space by limiting the number of candidate diameters for each pipe. The pipe index vector that was used to quantify the relative importance of each pipe was calculated repeatedly during the optimization. Calculation of the pipe index vector imposed a very high computational burden, as it required hydraulic simulations of the water distribution network and the solution of a complex system of linear equations.

Kadu et al. (2008) reduced the number of candidate diameters for each pipe based on the concept of the critical path in a single-source network. The limitations are such that it has not yet been applied to practical problems such as rehabilitation and networks with pumps, tanks, multiple operating conditions, and time varying demands.

Haghighi et al. (2011) reduced the solution space indirectly by reducing the number of decision variables through problem transformation. The transformation was achieved by

hybridizing the genetic algorithm (GA) with integer linear programming. Thus, a looped network was transformed into a branched network, in order to define a path from a source to each demand node. The GA optimized the loop-completing links, following optimization of the rest of the pipes using integer linear programming. The integer linear programming and GA sequence was applied iteratively.

Zheng et al. (2011) applied extensive problem simplification and pre-processing using graph theory and non-linear programming to find near-optimal solutions that were then used to define the reduced solution space. On the other hand, rather than reduce the solution space explicitly, Kang and Lansey (2012) used heuristic procedures to generate initial solutions for the evolutionary algorithm.

A new methodology that reduces the solution space during the design optimization of water distribution systems was developed. The reason for reducing the solution space is to increase the computational efficiency. In the algorithm proposed, the reduced solution space is not set a priori, and all the decision variables are optimized at once. The methodology developed is essentially an application of the most likely flow distribution that derives from the maximum entropy formalism (Jaynes 1957). Resilience considerations (Dunn and Wilkinson 2017) were addressed by maximizing the flow entropy at the same time. In other words, the algorithm carries out solution space reduction and entropy maximization simultaneously.

2 Network Flow Entropy

The primary design objectives for a water distribution network are cost minimization and adequate water supply and pressure at the demand nodes. In practice, it is likely that some network elements will be unavailable due to pipe breakage, pump failure, repairs and maintenance, etc. Therefore, some spare capacity needs to be included to enable the network to perform reasonably satisfactorily under both normal and abnormal operating conditions. Accordingly, other criteria besides the costs are often considered also (Tanyimboh and Templeman 1994).

The importance of considering failure tolerance and redundancy in addition to the hydraulic reliability has been emphasized previously in the literature (Kalungi and Tanyimboh 2003; Gheisi and Naser 2015). Redundancy comprises any surplus capacity in the components (i.e. pipes, pumps, service reservoirs, etc.) and/or multiple independent supply paths from the supply nodes to the demand nodes (Awumah et al. 1991; Herrera et al. 2016; Praks et al. 2015; Saleh and Tanyimboh 2014; Yazdani et al. 2011). Failure tolerance is a measure of the extent to which a water distribution network can satisfy the nodal demands at adequate pressure with one or more components unavailable (Tanyimboh and Templeman 1998).

However, due to the high computational burden associated with the evaluation of accurate measures of reliability, researchers frequently employ alternative measures (Wu et al. 2013). Probably the best definition of the hydraulic capacity reliability of a water distribution system, as the expectation of the ratio of the flow delivered at adequate pressure to the flow required, under normal and abnormal operating conditions, was proposed by Tanyimboh and Templeman (1998).

Various performance surrogates have been proposed in the last two decades. As in the wider context of surrogate modelling, where surrogates represent the response surface of simulation models (Wang et al. 2014), they have the advantages that they are easy to calculate and incorporate in optimization procedures. Awumah et al. (1990) used Shannon's informational entropy (Shannon 1948) to design and assess the performance of water distribution networks.

Tanyimboh and Templeman (1993a) provided the definitive formulation of the flow entropy based on the conditional entropy concept (Khinchin 1953, 1957). A discussion of the relative merits of the flow entropy functions proposed by Awumah et al. (1990) and Tanyimboh and Templeman (1993a) is available in Tanyimboh (1993). Variants of the flow entropy include Tsallis entropy (Singh and Oh 2015) and diameter-sensitive flow entropy (Liu et al. 2014).

The relationship between flow entropy and resilience (Dunn and Wilkinson 2017) as characterised by reliability and failure tolerance (Praks et al. 2015) was shown previously to be strong (Czajkowska and Tanyimboh 2013; Saleh and Tanyimboh 2016; Tanyimboh et al. 2016). Other reliability and redundancy surrogates in the literature include various resilience indices (Jayaram and Srinivasan 2008; Liu et al. 2016; Prasad and Park 2004; Todini 2000). Recent comparative assessments of the surrogates include Atkinson et al. (2014), Gheisi and Naser (2015), Liu et al. (2014, 2016) and Tanyimboh et al. (2016). Gheisi and Naser (2015) and Tanyimboh et al. (2016) observed that flow entropy had a significantly stronger positive correlation with reliability and failure tolerance than the available alternatives.

3 Water Distribution Network Design Optimization Problem

In general, optimization problems involve constraints while genetic algorithms often do not address constraints directly. The most common constraint-handling strategy (Michalewicz 1995) is to penalize infeasible solutions to reduce their fitness. However, in general the formulation of suitably responsive penalty functions is very challenging (Chang et al. 2010; Deb and Datta 2013). Moreover, searching through the feasible and infeasible regions of the solution space improves an algorithm's efficiency and identifies better solutions than searching through the feasible regions only (Glover and Greenberg 1989).

Deb (2000) introduced the tournament selection operator whereby feasible solutions are favoured over infeasible solutions. When comparing infeasible solutions, the solutions with the smallest constraint violations are preferred. If the solutions are feasible, the fittest solutions have preference. A major drawback in this approach is that it impedes the propagation of efficient infeasible solutions, which play an important role in the evolutionary processes that drive the optimization (Saleh and Tanyimboh 2014; Siew et al. 2014, 2016). A penalty-free approach was adopted in this investigation due to these considerations. However, other constraint handling strategies are available (Chootinan and Chen 2006; Oyama et al. 2007; Woldesenbet et al. 2009).

The objectives considered in the formulation of the optimization problem were: (a) minimization of the network's initial construction cost; and (b) maximization of the flow entropy, based on the maximum entropy formalism (Jaynes 1957). The constraints were the system of equations that govern the flow, and the nodal minimum residual pressure constraints. EPANET 2, a hydraulic solver (Rossman 2000), was used to analyse the candidate solutions and thus the conservation of mass and energy constraints were satisfied.

On the other hand, the nodal minimum residual pressure constraints were addressed by introducing an infeasibility objective that was minimized. Thus, the function f_2 in Eq. 2 identifies the largest shortfall in the required nodal residual pressure in the network. As the value of f_2 is zero for feasible solutions, minimizing the infeasibility aims to achieve feasible solutions by reducing the infeasibility to zero.

The decision variables of the optimization problem are the pipe diameters; i.e. the optimization seeks the optimal vector for the pipe diameters. The optimization problem may be summarised briefly as follows.

$$\text{Minimize the initial construction cost}: \quad f_1 = \sum_{i=1}^{np} C_i(d_i, l_i) \tag{1}$$

where $C_i(d_i, l_i)$ is the cost of pipe i with diameter d_i and length l_i, and np represents the number of pipes in the network.

$$\text{Minimize the infeasibility}: \quad f_2 = Max\{Max[0, (H_i^{req} - H_i)]; \forall i\} \tag{2}$$

where the head at node i, H_i, was obtained from the hydraulic simulation model (EPANET 2) and H_i^{req} is the head at node i above which the demand is satisfied in full.

$$\text{Maximize the flow entropy}: \quad f_3 = S \tag{3}$$

where S is the flow entropy (Tanyimboh and Templeman 1993a).

$$\text{Subject to}: \quad d_i \in D; \quad \forall i \tag{4}$$

where the set D comprises the available discrete pipe diameter options.

Shannon's statistical entropy function quantifies the amount of uncertainty that a probability distribution represents (Shannon 1948). The flow entropy function in Eq. 5 (Tanyimboh and Templeman 1993a, b, c, d) is an extension that measures the relative uniformity of the pipe flow rates in a water distribution network.

$$\frac{S}{K} = -\sum_{i=1}^{I} \frac{Q_i}{T} \ln\left(\frac{Q_i}{T}\right) - \frac{1}{T} \sum_{j=1}^{J} T_j \left[\frac{Q_j}{T_j} \ln\left(\frac{Q_j}{T_j}\right) + \sum_{ij \in N_j} \frac{q_{ij}}{T_j} \ln\left(\frac{q_{ij}}{T_j}\right) \right] \tag{5}$$

where S is the entropy; K is an arbitrary positive constant that is often taken as 1.0; T is the total supply; T_j is the total flow reaching node j; Q_i is the inflow at a supply node; Q_j is the demand at a demand node; q_{ij} is the flow rate in pipe ij; I is the number of supply nodes; J is the number of demand nodes including junctions; and N_j is the set of pipe flows from node j.

4 Procedure for Reducing The Solution Space

Among the nondominated solutions in each generation, a *feasible solution* with a specified value of entropy may be selected as the reference solution for the purposes of defining the reduced solution space. In order to determine an appropriate subset of candidate pipe diameters to allocate to a particular pipe, the diameter of the pipe in question in the reference solution is selected along with four additional diameters to provide five candidate diameters in total. The four additional diameters included are the nearest two diameters that are smaller than the reference diameter and the nearest two diameters that are larger than the reference diameter.

For the purposes of the proposed algorithm, the *reference diameter* of a pipe is the current diameter in the *reference solution*. The *reference solution* is the solution whose diameters inform the limits of the *active solution space*. The *active solution space* is dynamic, and is the *reduced solution space in the current generation*. The *reduced solution space* is the portion of the solution space that remains after excluding the pipe diameter options or segments of any

continuous decision variables that should not be considered or explored by the optimization algorithm.

Furthermore, if the diameter in the reference solution is the largest or smallest available diameter, it is allocated three times. Similarly, if the reference diameter is the next largest or next smallest diameter, the largest or smallest diameter is allocated two times. This is necessary to avoid bias; all the pipes should have an equal number of sizing options. This process of re-defining the active solution space is repeated in each subsequent generation until the end of the optimization process.

The smallest number of pipe diameter options that permits the diameters to increase or decrease is three. However, it may yield suboptimal solutions by being too restrictive as each pipe diameter may change by at most one pipe size only. Five pipe diameter options were adopted as a compromise that yields satisfactory results (Kadu et al. 2008; Siew et al. 2014; Zheng et al. 2011).

The source code of the genetic algorithm (NSGA II) (Deb et al. 2002) was modified and combined with the hydraulic solver (EPANET 2) (Rossman 2000) and a subroutine the authors developed to calculate the flow entropy value for any given network configuration. The hydraulic solver used was demand-driven (see e.g. Siew and Tanyimboh 2011) and so the entropy values of the infeasible solutions could be unrealistic and/or misleading. The reduction in the solution space was initiated only after achieving feasible solutions. When feasible solutions became available, the solution space was reduced to promote exploitation by intensifying the search around the feasibility boundaries. On the other hand, exploration and diversity were sustained by continuing the search for new solutions with higher entropy values.

4.1 Details of the Methodology

Previous studies have shown that the slope of the graph of entropy versus cost often approaches zero after around 99% of the maximum entropy value (Czajkowska 2016). Hence, 99% of the maximum entropy value may be considered the approximate point at which the entropy stabilizes. Moreover, previous studies in the literature have shown that solutions having different configurations but identical entropy values generally have similar performance characteristics including resilience (Czajkowska and Tanyimboh 2013; Tanyimboh and Setiadi 2008; Tanyimboh et al. 2011; Tanyimboh and Sheahan 2002). Consequently, taking the solution with the highest entropy value achieved in each generation as the reference solution may yield solutions with diameters that are larger and more expensive than necessary, if less expensive solutions with similar resilience properties are available.

Accordingly, the entropy value for the reference solution may be set as follows.

$$S = (1-\varepsilon)S^{*}; \quad 0 \leq \varepsilon \leq \delta; \quad 0 \leq \delta < 1.0 \tag{6}$$

where S^{*} is the global maximum entropy value; ε a tolerance parameter that is used to adjust the properties of the Pareto-optimal front; and δ is an upper limit for ε.

The value of S^{*} is unknown at the start of the optimization. However, it is evolved by assuming that it is the largest value achieved in the current generation. By definition, the combination of entropy maximization and cost minimization provides a diverse population of solutions with enhanced resilience properties. Consequently, a small value of δ, say about 0.02, should suffice. The new algorithm developed (SSRA) (self-adaptive solution-space reduction algorithm) has three main components as summarised briefly below. It is worth noting that each component is independent and may be applied on its own. The effectiveness of the

algorithm was assessed by comparing the results with the baseline conventional algorithm with no reduction in the solution space.

4.1.1 Part 1: Solution Space Reduction with ε of zero

The algorithm starts with the full solution space. In the first generation following the achievement of feasible solutions, the number of pipe diameter options for each pipe is reduced to five and the solution with the maximum entropy value (i.e. $\varepsilon = \varepsilon_0 \equiv 0$ in Eq. 6) is used to define the active solution space for the next generation. Then, the active solution space is re-defined for each subsequent generation until the end of the optimization. This is similar to the conventional approach, except for the solution space reduction. With the maximum entropy solution as the reference solution, this favours the high-entropy and high-cost end of the Pareto-optimal front.

4.1.2 Part 2: Solution Space Reduction with ε of 0.02

With $\varepsilon = \varepsilon_2 \equiv 0.02$ in Eq. 6, the solution closest to 98% of the highest entropy value achieved so far in the optimization is selected as the reference solution that is used to define the active solution space, starting from the first generation following the achievement of feasible solutions. This favours the low-entropy and low-cost end of the Pareto-optimal front.

4.1.3 Part 3: Solution Space Reduction with ε of 0.01

With $\varepsilon = \varepsilon_1 \equiv 0.01$ in Eq. 6, the solution closest to 99% of the highest entropy value achieved so far in the optimization is selected as the reference solution that is used to define the active solution space, starting from the first generation following the achievement of feasible solutions. This favours the inner portion of the Pareto-optimal front.

4.2 Conventional Solution Approach

The conventional solution approach employs the full solution space with no restriction in the entropy value. In other words, ε is zero and the entire solution space is active.

5 Results and Discussion

To illustrate the methodology and potential of the proposed approach the network shown in Fig. 1 (Kadu et al. 2008) was investigated. Its solution space is substantial and Kadu et al. (2008) used it previously to study the benefits of reducing the solution space. Other researchers, for example, Barlow and Tanyimboh (2014), Haghighi et al. (2011) and Siew et al. (2014) have provided solutions for this network also. However, previous studies addressed a single-objective cost minimization problem that did not include reliability or resilience considerations. Thus, the results are not directly comparable.

5.1 Network Description and Problem Specifications

The skeletonised network shown in Fig. 1 comprised 26 nodes, 34 pipes and 9 loops. There were two reservoirs at nodes 1 and 2 with constant heads of 100 and 95 m, respectively. The
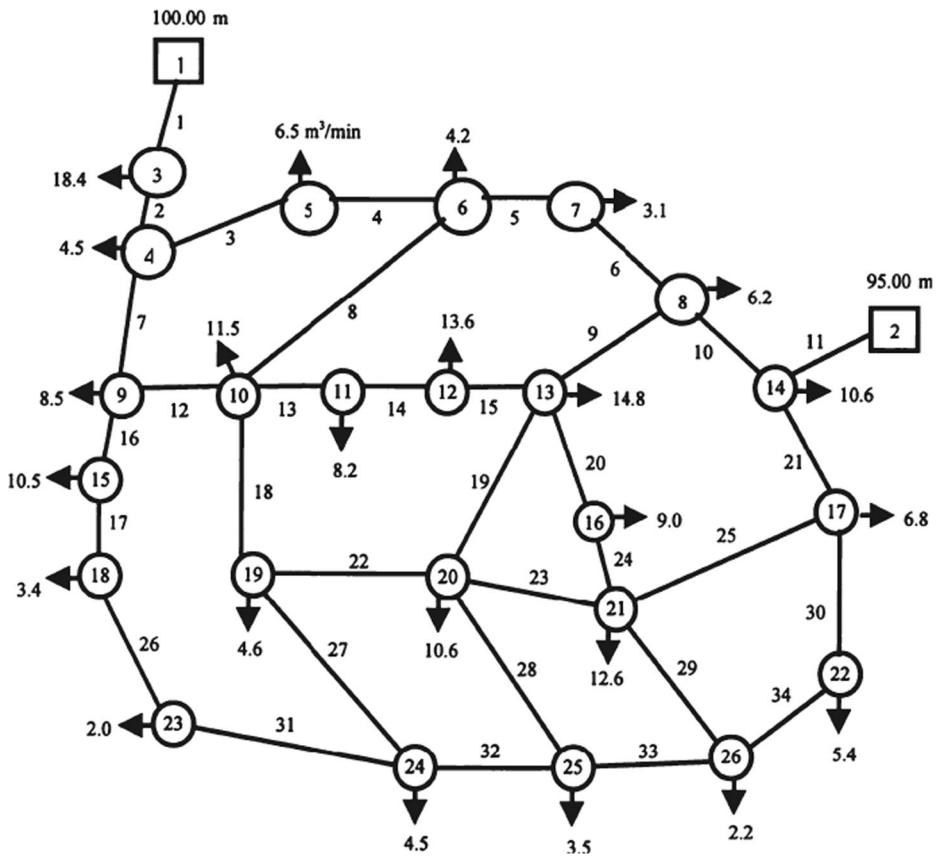
Fig. 1 Network topology. The rectangles represent supply nodes (reservoirs) while circles represent demand nodes. Nodal demands (m³/min), reservoir water levels (m) and pipe identifiers are as indicated

Hazen-Williams roughness coefficient was 130 for all the pipes. The nodal demands in m³/min are shown in Fig. 1. The pipe lengths and residual heads above which the nodal demands would be satisfied in full are in Table 1. There were 14 available pipe diameter options as shown in Table 2.

The objectives were to minimize the initial construction cost and maximize the flow entropy. The minimum node pressure constraints were addressed by introducing an infeasibility objective that was minimized. The decision variables were the pipe diameters. The solution space comprised 14^{34} or 9.30×10^{38} feasible and infeasible solutions, based on 14 pipe diameter options and 34 pipes.

5.2 Details of the Implementation of the Genetic Algorithm

For the full solution space with 14 pipe diameter options, a four-bit binary substring provided $2^4 = 16$ codes for the pipe diameters. There were thus two redundant substrings (Herrera et al. 1998). Redundant substrings that do not correspond to any of the available pipe diameter options arise if the number of decision variables is not a power of 2. To match the number of codes and achieve a balanced allocation of the redundant codes, both the 5th (350 mm) and 10th (700 mm) diameter options were doubled. In an ordinal sense, a balanced allocation of the

Table 1 Pipe lengths and minimum nodal heads for satisfactory water supply and pressure

Pipe lengths				Minimum nodal heads for full supply			
Pipes	Lengths (m)	Pipes	Lengths (m)	Nodes	Heads (m)	Nodes	Heads (m)
1	300	18	650	1	100	14	82
2	820	19	760	2	95	15	85
3	940	20	1100	3	85	16	82
4	730	21	660	4	85	17	82
5	1620	22	1170	5	85	18	85
6	600	23	980	6	85	19	82
7	800	24	670	7	82	20	82
8	1400	25	1080	8	82	21	82
9	1175	26	750	9	85	22	80
10	750	27	900	10	85	23	82
11	210	28	650	11	85	24	80
12	700	29	1540	12	85	25	80
13	310	30	730	13	82	26	80
14	500	31	1170				
15	1960	32	1650				
16	900	33	1320				
17	850	34	3250				

redundant codes, in terms of the sequence of pipe diameter options, aims to minimize the effects of representation bias (Herrera et al. 1998) that arises due to the over-representation of two codes out of 14. Czajkowska (2016) has shown that a balanced allocation of the redundant codes yields better results in terms of a more even distribution of the solutions on the Pareto-optimal front achieved.

On the other hand, when the number of pipe diameter options reduced to five, a three-bit binary substring provided $2^3 = 8$ codes for the pipe diameters. Similarly, to match the number of codes and achieve a balanced allocation of the three redundant codes, each of the 1st, 3rd and 5th pipe diameter options were doubled. Alternative techniques for handling redundant binary codes were discussed by Saleh and Tanyimboh (2014) who proposed an innovative approach based on natural selection.

A bitwise mutation operator was used to change the bits from zero to one or vice versa with a mutation probability of $1/n_g = 1/136 = 0.007$, where $n_g = 136$ was the chromosome length. A single-point crossover operator was used to produce two offspring from two parents with a crossover probability of 1.0.

There were four solution scenarios for the optimization, i.e. the reduced solution spaces with the three different ε values of zero, 0.01 and 0.02, and the full solution space with ε of zero. The termination criterion for the GA was 100,000 function evaluations, i.e 1000 generations based

Table 2 Available pipe diameter options and unit costs

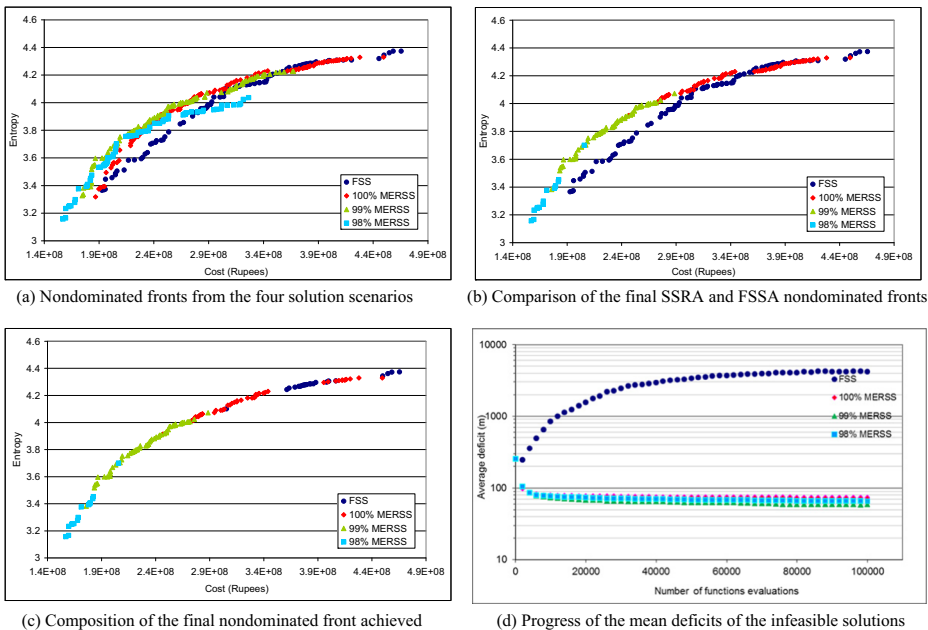| Diameter options (mm)	Unit costs (rupees)	Diameter options (mm)	Unit costs (rupees)	Diameter options (mm)	Unit costs (rupees)
150	1115	400	4255	750	11,874
200	1600	450	5172	800	13,261
250	2154	500	6092	900	16,151
300	2780	600	8189	1000	19,395
350	3475	700	10,670		

on a population of 100. The members of the initial populations were created randomly, and each optimization scenario had 50 independent optimization runs. The average CPU time on a personal computer (Intel Core 2 Duo @ 3.5 GHz and 3 GB RAM) for an optimization run with 100,000 function evaluations was almost the same for all the scenarios. It is worth noting that only 50 independent initial populations were created, and each initial population was used in the four solution scenarios. The reason was to simplify subsequent comparisons by minimizing any discrepancies due to differences in the initial populations.

After excluding infeasible solutions at the end of the optimization, the nondominated feasible solutions achieved were combined and sorted based on Pareto-dominance to provide the best Pareto-optimal front (Fig. 2c). Similarly, Pareto-optimal fronts were obtained from the respective sets of 50 runs for the full and reduced solutions spaces (Fig. 2a). In addition, the nondominated solutions from the three reduced solution spaces were combined and sorted based on Pareto-dominance (Fig. 2b). The results are summarised in Fig. 2 and Table 3, with additional details in Figs. 3 and 4, and the supplementary data. In Figs. 2, 3 and 4, the abbreviations FSS and MERSS represent full solution space and maximum entropy reduced solution space, respectively.

5.3 Results and Discussion

Recalling that the independent initial populations used were the same in the four solution scenarios, the individual Pareto-optimal fronts in the supplementary data reveal that the method with solution space reduction achieved superior results, compared to the full solution space.

Figure 2a shows that the Pareto-optimal fronts from the reduced solution spaces tended to dominate the full solution space. What is more, SSRA (self-adaptive solution-space reduction

(a) Nondominated fronts from the four solution scenarios

(b) Comparison of the final SSRA and FSSA nondominated fronts

(c) Composition of the final nondominated front achieved

(d) Progress of the mean deficits of the infeasible solutions

Fig. 2 Solution quality and computational efficiencies of SSRA and FSSA methods based of 50 independent optimization runs

Table 3 Computational efficiency based on 50 independent optimization runs

| Solution scenario	Minimum	Mean	Median	Maximum
CPU time to complete 100,000 function evaluations (minutes)				
FSS ($\varepsilon = 0$)	3.50	3.62	3.60	3.78
RSS ($\varepsilon = 0$)	3.50	3.65	3.67	3.78
RSS ($\varepsilon = 0.01$)	5.57	3.67	3.68	3.82
RSS ($\varepsilon = 0.02$)	3.53	3.65	3.63	3.78
Number of feasible solutions per optimization run ($\times 10^3$)				
FSS ($\varepsilon = 0$)	2.336	3.985	3.984	5.750
RSS ($\varepsilon = 0$)	6.990	11.831	11.362	18.389
RSS ($\varepsilon = 0.01$)	4.737	10.363	10.172	16.831
RSS ($\varepsilon = 0.02$)	4.939	10.183	9.571	16.597
[a]Number of function evaluations to achieve convergence of entropy ($\times 10^3$)				
FSS ($\varepsilon = 0$)	4.0	37.0	30.0	100.0[b]
RSS ($\varepsilon = 0$)	4.0	10.4	4.0	100.0[b]
RSS ($\varepsilon = 0.01$)	4.0	8.6	4.0	76.0
RSS ($\varepsilon = 0.02$)	4.0	5.8	4.0	28.0

[a] This refers to the point after which improvements in entropy were deemed insignificant

[b] This corresponds to the maximum number of function evaluations allowed, i.e. 100,000

algorithm) achieved additional solutions with lower entropy values between 3.1 and 3.3 that were not found by FSSA (full solution space algorithm). FSSA on the other hand produced the solutions with the highest entropy values. The highest entropy values achieved by the full and reduced solution spaces were 4.370 and 4.329, respectively. The difference is 0.041 or 0.94%. The shortfall could be because SSRA had only five pipe diameter options compared to 14 for
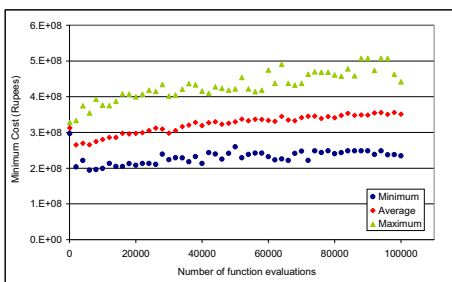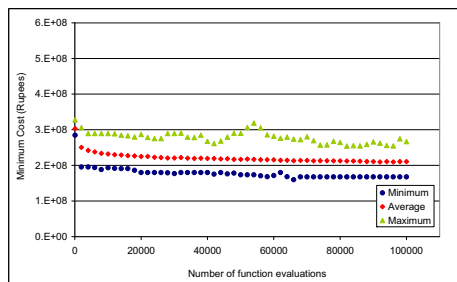

(a) Full solution space (FSS)
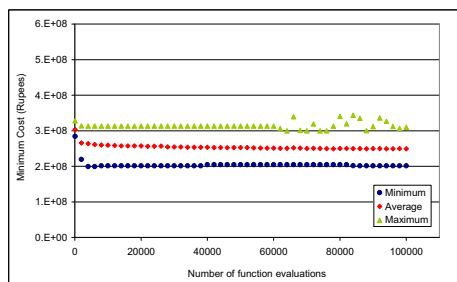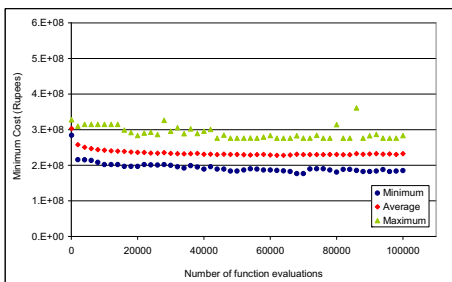
(b) Reduced solution space with ε of zero (100% MERSS)

(c) Reduced solution space with ε of 0.01 (99% MERSS)

(d) Reduced solution space with ε of 0.02 (98% MERSS)

Fig. 3 Progress of the least-cost feasible solution. The averages, minima and maxima are for 50 independent optimization runs

(a) Full solution space (FSS)

(b) Reduced solution space with ε of zero (100% MERSS)

(c) Reduced solution space with ε of 0.01 (99% MERSS)

(d) Reduced solution space with ε of 0.02 (98% MERSS)
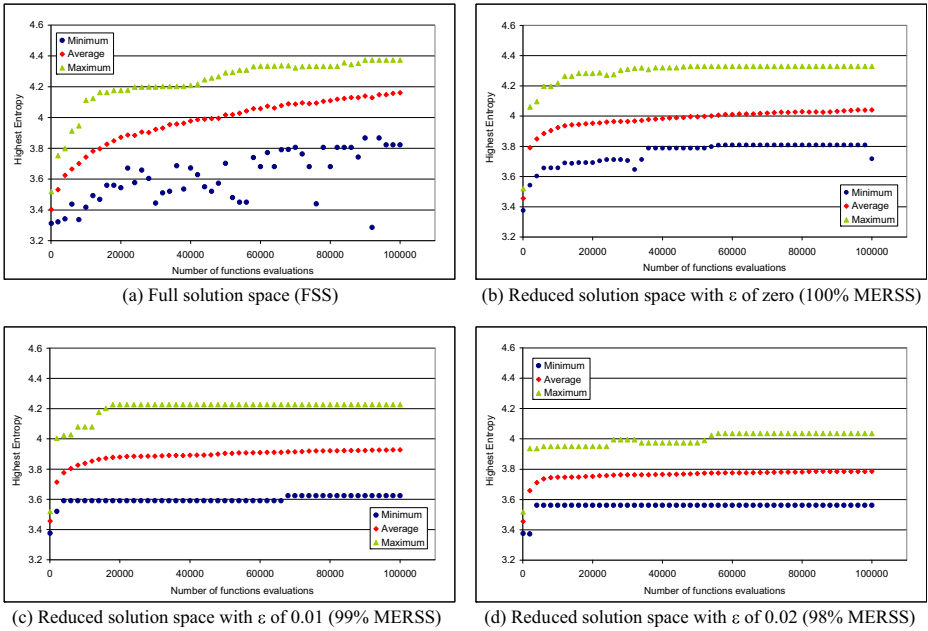
Fig. 4 Progress of the maximum entropy of the feasible solutions. The averages, minima and maxima are for 50 independent optimization runs

FSSA. However, the decision maker would likely not choose the solutions with the highest entropies because their costs are relatively uncompetitive.

It can be seen also that FSSA underperformed at the lower end of the Pareto-optimal front while SSRA with ε of 0.02 under-performed at the upper end. On the other hand, SSRA with ε of 0.01 was consistently optimal or near optimal and the most competitive. Overall, there were 147 nondominated solutions in total. Figure 2c shows the final Pareto-optimal front based on all the solutions achieved from all the optimization runs. FSSA contributed 23 solutions (16%); SSRA with ε of zero contributed 47 solutions (32%); SSRA with ε of 0.01 contributed 62 solutions (42%); and SSRA with ε of 0.02 contributed 15 solutions (10%).

Figure 3 shows the progress of the optimization in terms of cost. The cost decreased and stabilised rapidly in the reduced solution spaces. On the other hand, in the full solution space, the cost decreased rapidly initially, and then it increased gradually until the end. This is due to the entropy maximization with more pipe diameter options. Figure 3 shows that, compared to the full solution space, significantly lower costs were achieved in the reduced solution spaces.

Figure 4 shows the progress of the flow entropy based on the *feasible solutions only*. In the full solution space, after a rapid increase at the start, the entropy continued to increase slowly until the last generation, as the solution space is substantial. Convergence was considerably faster in the reduced solution spaces, where the entropy increased rapidly in the early generations. The maximum entropy values achieved in the full and reduced solution spaces were 4.370 and 4.329, respectively. The difference is only 0.041 or 0.94%.

Figure 2d demonstrates the progress in terms of the deficit in the residual head at the critical node. The critical node is the node that has the largest deficit in the residual head. It varies from one solution to the next and depends on the operating and loading conditions. In the full solution space, as the cost minimization progressed, the number of feasible solutions decreased

initially after which a gradual increase followed. In the reduced solution spaces, the average deficit decreased and then remained stable until the end. The average initial deficit was 251.13 m; recall that each initial population was used in the four solution scenarios.

The lowest average deficits achieved by SSRA were 58.59 m for ε of 0.01, 65.64 m for ε of 0.02 and 73.38 m for ε of zero. It is worth observing that the average deficit could not be reduced to zero; a fundamental *beneficial* property of the methodology is that the optimization algorithm retains nondominated infeasible solutions until the end. Compared to FSSA with an average deficit of 4161.93 m at the end of the optimization, SSRA deficits were relatively small (less than 2%). These results contribute to the evidence that SSRA intensifies the search around the feasibility boundaries.

Table 3 shows that that SSRA produced significantly more feasible solutions than FSSA. On average, SSRA produced approximately three times more feasible solutions per GA run than FSSA. This demonstrates clearly the efficiency of SSRA. Table 3 also presents the convergence and consistency statistics based on 50 independent optimization runs. There was almost no difference between the CPU times based on the entire optimization runs with 100,000 function evaluations. It is therefore logical to infer that the SSRA modifications did not impose a significant computational burden. In fact, the SSRA convergence was sufficiently fast to complete the three SSRA solution scenarios with ε values of zero, 0.01 and 0.02 in less time than one FSSA solution.

It was observed also that, in general, an entropy increase of less than 3% between successive generations implied the entropy had stabilised in the network considered. Table 3 presents the number of function evaluations needed to achieve convergence. The results show that, compared to FSSA, the SSRA convergence was extremely fast. For example, the median and minimum values for SSRA were identical. The rapid convergence observed was the direct result of reducing the number of candidate pipe diameters once the algorithm achieved feasible solutions. In the present example, after the reduction, the active solution space as a fraction of the full solution space was $(5/14)^{34}$ or 6.30×10^{-16}. This is the same as a ratio of $1:1.60 \times 10^{15}$.

6 Conclusions

A new reliability-based method that reduces the solution space in the design optimization of water distribution networks was developed. The formulation recognises the relative importance of every path through network, based on a fundamental property of flow entropy (Tanyimboh and Templeman 1993a).

The algorithm is generic, self-adaptive and prior initialisation or setting of the reduced solution space is not required. The algorithm has two phases. After achieving feasible solutions in the first phase that prioritises exploration, the second phase that prioritises exploitation intensifies the search around the active constraint boundaries. At the same time, in the second phase, the resilience of the individual solutions and diversity among the population of solutions are preserved due to the continuation of the entropy maximization.

The solutions that the proposed algorithm obtained were clearly superior to those from the full solution space in terms of cost and flow entropy. Furthermore, convergence in the reduced solution space was significantly faster. The new method also provided more solutions than the full solution space, and additional low-cost and low-entropy solutions were achieved in the reduced solution space that were not found in the full solution space. In other words, the new procedure extended the Pareto-optimal front at the lower end. The results demonstrated that,

depending on the circumstances of the investigation at hand, the algorithm enabled the top, middle or bottom of the Pareto-optimal front to be prioritised as required by varying the value of ε, an entropy-based tolerance or relaxation parameter.

More verification would be worth considering, in addition to other comparative analyses. The algorithm could be improved potentially by replacing the demand-driven hydraulic simulation model with pressure-driven simulation. Networks that are more complex may be worth considering also.

Compliance with Ethical Standards

Conflict of Interest   The authors declare that they have no conflict of interest.

References

Atkinson S, Farmani R, Memon FA, Butler D (2014) Reliability indicators for water distribution system design: comparison. J Water Res Pl-ASCE 140(2):160–168

Awumah K, Goulter I, Bhatt SK (1990) Assessment of reliability in water distribution networks using entropy based measures. Stoch Hydrol Hydraul 4(4):309–320

Awumah K, Goulter I, Bhatt SK (1991) Entropy-based redundancy measures in water distribution network design. J Hydraul Eng 117(5):595–614

Barlow E, Tanyimboh TT (2014) Multi-objective memetic algorithm applied to the optimization of water distribution systems. Water Resour Manag 28(8):2229–2242

Chang LC, Chang FJ, Wang KW, Dai SY (2010) Constrained genetic algorithms for optimizing multi-use reservoir operation. J Hydrol 390:66–74. https://doi.org/10.1016/j.jhydrol.2010.06.031

Chootinan P, Chen A (2006) Constraint handling in genetic algorithms using a gradient-based repair method. Comput Oper Res 33:2263–2281

Czajkowska AM (2016) Maximum entropy based evolutionary optimization of water distribution networks under multiple operating conditions and self-adaptive search space reduction method. PhD thesis, University of Strathclyde Glasgow, UK

Czajkowska AM, Tanyimboh TT (2013) Water distribution network optimization using maximum entropy under multiple loading patterns. Water Sci Technol Water Supply 13(5):1265–1271

Deb K (2000) An efficient constraint handling method for genetic algorithms. Comput Methods Appl Mech Eng 186(2):311–338

Deb K (2001) Multi-objective optimization using evolutionary algorithms. Wiley, Chichester

Deb K, Datta R (2013) A bi-objective constrained optimization algorithm using a hybrid evolutionary and penalty function approach. Eng Optim 45:503–527

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multi-objective genetic algorithm: NSGA II. IEEE T Evolut Comput 6(2):182–197

Dunn S, Wilkinson S (2017) Hazard tolerance of spatially distributed complex networks. Reliab Eng Syst Saf 157:1–12

Fonseca CM, Fleming PJ (1993) Genetic algorithms for multiobjective optimization: formulation, discussion and generalization. International conference on genetic algorithms, vol. 5:416–423. Los Altos CA, USA

Gen M, Cheng R, Lin L (2008) Network models and optimization: multiobjective genetic algorithm approach. Springer Science and Business Media, New York

Gheisi A, Naser G (2015) Multistate reliability of water-distribution systems: comparison of surrogate measures. J Water Res Pl-ASCE. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000529

Glover F, Greenberg H (1989) New approaches for heuristic search: a bilateral linkage with artificial intelligence. Eur J Oper Res 39:19–130

Haghighi A, Samani HM, Samani ZM (2011) GA-ILP method for optimization of water distribution networks. Water Resour Manag 25(7):1791–1808

Hajela P, Lin CY (1992) Genetic search strategies in multi-criterion optimal design. Struct Optim 4(2):99–107

Herrera F, Lozano M, Verdegay JL (1998) Tackling real-coded genetic algorithms: operators and tools for behavioural analysis. Artif Intell Rev 12(4):265–319

Herrera M, Abraham E, Stoianov I (2016) A graph-theoretic framework for assessing the resilience of sectorised water distribution networks. Water Resour Manag 30(5):1685–1699. https://doi.org/10.1007/s11269-016-1245-6

Jayaram N, Srinivasan K (2008) Performance-based optimal design and rehabilitation of water distribution networks using life cycle costing. Water Resour Res 44(1):W01417

Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106:620–630 and 108:171–190

Kadu MS, Gupta R, Bhave PR (2008) Optimal design of water networks using a modified genetic algorithm with reduction in search space. J Water Res Pl-ASCE 134(2):147–160

Kalungi P, Tanyimboh TT (2003) Redundancy model for water distribution systems. Reliab Eng Syst Saf 82(3):275–286

Kang D, Lansey K (2012) Revisiting optimal water-distribution system design: issues and a heuristic hierarchical approach. J Water Res Pl-ASCE 138(3):208–217

Khinchin AI (1953) The entropy concept in probability theory. Uspekhi Matematicheskikh Nauk 8(3):3–20

Khinchin AI (1957) Mathematical foundations of information theory. Dover, New York, pp 1–28

Knowles JD, Corne DW (2000) Approximating the non-dominated front using the Pareto archived evolution strategy. Evol Comput 8(2):149–172

Kursawe F (1990) A variant of evolution strategies for vector optimization. Parallel problem solving from. Nature 1:183–197

Liu H, Savić DA, Kapelan Z, Zhao M, Yuan Y, Zhao H (2014) A diameter-sensitive flow entropy method for reliability consideration in water distribution system design. Water Resour Res 50(7):5597–5610

Liu H, Savić DA, Kapelan Z, Creaco E, Yuan Y (2016) Reliability surrogate measures for water distribution system design: a comparative analysis. J Water Res Pl-ASCE. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000728

Maier HR, Kapelan Z et al (2014) Evolutionary algorithms and other metaheuristics in water resources: current status, research challenges and future directions. Environ Model Softw 62:271–299

McClymont K, Keedwell E, Savi D (2015) An analysis of the interface between evolutionary algorithm operators and problem features for water resources problems. A case study in water distribution network design. Environ Model Softw 69:414–424

Michalewicz Z (1995) A survey of constraint handling techniques in evolutionary computation methods. In: JR MD et al (eds) Evolutionary programming IV. MIT Press, Cambridge, pp 135–158

Osyczka A, Kundu S (1995) A new method to solve generalized multicriteria optimization problems using the simple genetic algorithms. Struct Optim 10(2):94–99

Oyama A, Shimoyama K, Fujii K (2007) New constraint-handling method for multi-objective and multi-constraint evolutionary optimization. Trans Jpn Soc Aero Space Sci 50(167):56–62

Praks P, Kopustinskas V, Masera M (2015) Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure. Reliab Eng Syst Saf 144:254–264

Prasad TD, Park NS (2004) Multiobjective genetic algorithms for design of water distribution networks. J Water Res Pl-ASCE 130(1):73–82

Rossman LA (2000) EPANET 2. User's manual. Water Supply and Water Resources Division, National Risk Management Research Laboratory, US EPA, Cincinnati

Saleh SH, Tanyimboh TT (2014) Optimal design of water distribution systems based on entropy and topology. Water Resour Manag 28(11):3555–3575

Saleh SH, Tanyimboh TT (2016) Multi-directional maximum-entropy approach to the evolutionary design optimization of water distribution systems. Water Resour Manag 30(6):1885–1901. https://doi.org/10.1007/s11269-016-1253-6

Schaffer JD (1985) Multiple objective optimization with vector evaluated genetic algorithms. 1st international conference on genetic algorithms, Pittsburgh PA, USA, 93–100

Shannon C (1948) A mathematical theory of communication. AT&T Tech J 27(3):379–428

Siew C, Tanyimboh T (2011) Practical application of the head dependent gradient method for water distribution networks. Water Sci Technol Water Supply 11(4):444–450

Siew C, Tanyimboh TT, Seyoum AG (2014) Assessment of penalty-free multi-objective evolutionary optimization approach for the design and rehabilitation of water distribution systems. Water Resour Manag 28(2):373–389

Siew C, Tanyimboh TT, Seyoum AG (2016) Penalty-free multi-objective evolutionary approach to optimization of Anytown water distribution network. Water Resour Manag 30(11):3671–3688. https://doi.org/10.1007/s11269-016-1371-1

Singh VP, Oh J (2015) A Tsallis entropy-based redundancy measure for water distribution networks. Phys A 421: 360–376

Srinivas N, Deb K (1994) Multiobjective optimization using non-dominated sorting in genetic algorithms. Evol Comput 2(3):221–248

Tanyimboh TT (1993) An entropy based approach to the optimum design of reliable water distribution networks. PhD thesis, University of Liverpool, UK

Tanyimboh TT, Setiadi Y (2008) Sensitivity analysis of entropy-constrained designs of water distribution systems. Eng Optim 40(5):439–457

Tanyimboh TT, Sheahan C (2002) A maximum entropy based approach to the layout optimization of water distribution systems. Civ Eng Environ Syst 19(3):223–253

Tanyimboh TT, Templeman AB (1993a) Calculating maximum entropy flows in networks. J Oper Res Soc 44(4): 383–396

Tanyimboh TT, Templeman AB (1993b) Maximum entropy flows for single-source networks. Eng Optim 22(1): 49–63. https://doi.org/10.1080/03052159308941325

Tanyimboh TT, Templeman AB (1993c) Optimum design of flexible water distribution networks. Civ Eng Syst 10(3):243–258. https://doi.org/10.1080/02630259308970126

Tanyimboh TT, Templeman AB (1993d) In: Coulbeck B (ed) Using entropy in water distribution networks. International Conference on Integrated Computer Applications for Water Supply and Distribution. De Montfort University, Leicester

Tanyimboh TT, Templeman AB (1994) Discussion of "redundancy-constrained minimum cost design of water distribution nets". J Water Res Pl-ASCE 120(4):568–571

Tanyimboh TT, Templeman AB (1998) Calculating the reliability of single-source networks by the source head method. Adv Eng Softw 29(7–9):499–505

Tanyimboh TT, Tietavainen MT, Saleh S (2011) Reliability assessment of water distribution systems with statistical entropy and other surrogate measures. Water Sci Technol Water Supply 11(4):437–443

Tanyimboh TT, Siew C, Saleh SH, Czajkowska AM (2016) Comparison of surrogate measures for the reliability and redundancy of water distribution systems. Water Resour Manag 30(10):3535–3552

Todini E (2000) Looped water distribution networks using resilience index based on heuristic approach. Urban Water 2(3):115–122

Vairavamoorthy K, Ali M (2005) Pipe index vector: a method to improve genetic algorithm based pipe optimization. J Hydraul Eng 131(12):1117–1125

Wang C, Duan Q, Gong W, Ye A et al (2014) An evaluation of adaptive surrogate modeling based optimization with two benchmark problems. Environ Model Softw 60:167–179

Woldesenbet YG, Yen GG, Tessema BG (2009) Constraint handling in multiobjective evolutionary optimization. Trans Evol Comput 13(3):514–525

Wu W, Maier HR, Simpson AR (2013) Multiobjective optimization of water distribution systems accounting for economic cost, hydraulic reliability, and greenhouse gas emissions. Water Resour Res 49:1211–1225. https://doi.org/10.1002/Wrcr.20120

Yazdani A, Otoo RA, Jeffrey P (2011) Resilience enhancing expansion strategies for water distribution systems: a network theory approach. Environ Model Softw 26(12):1574–1582

Zheng F, Simpson AR, Zecchin A (2011) A combined NLP-differential evolution algorithm approach for the optimization of looped water distribution systems. Water Resour Res 47:W08531

Zitzler E, Thiele L (1998) An evolutionary algorithm for multiobjective optimization. The strength Pareto approach. Technical Report 43, Zurich, Switzerland: Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH)

Zitzler E, Deb K, Thiele L (2000) Comparison of multiobjective evolutionary algorithms: empirical results. Evol Comput 8(2):173–195