

# Optimal Design of Water Distribution Systems Based on Entropy and Topology

Salah H. A. Saleh · Tiku T. Tanyimboh

Received: 22 December 2013 / Accepted: 22 May 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** A new multi-objective evolutionary optimization approach for joint topology and pipe size design of water distribution systems is presented. The algorithm proposed considers simultaneously the adequacy of flow and pressure at the demand nodes; the initial construction cost; the network topology; and a measure of hydraulic capacity reliability. The optimization procedure is based on a general measure of hydraulic performance that combines statistical entropy, network connectivity and hydraulic feasibility. The topological properties of the solutions are accounted for and arbitrary assumptions regarding the quality of infeasible solutions are not applied. In other words, both feasible and infeasible solutions participate in the evolutionary processes; solutions survive and reproduce or perish strictly according to their Pareto-optimality. Removing artificial barriers in this way frees the algorithm to evolve optimal solutions quickly. Furthermore, any redundant binary codes that result from crossover or mutation are eliminated gradually in a seamless and generic way that avoids the arbitrary loss of potentially useful genetic material and preserves the quality of the information that is transmitted from one generation to the next. The approach proposed is entirely generic: we have not introduced any additional parameters that require calibration on a case-by-case basis. Detailed and extensive results for two test problems are included that suggest the approach is highly effective. In general, the frontier-optimal solutions achieved include topologies that are fully branched, partially- and fully-looped and, for networks with multiple sources, completely separate sub-networks.

**Keywords** Statistical entropy · Reliability-based design · Evolutionary optimization · Redundant binary codes

## 1 Introduction

The planning for a water distribution system may include topology design and sizing of components to evaluate the hydraulic properties of the system. Since failures may occur due to

---

S. H. A. Saleh · T. T. Tanyimboh (✉)

Department of Civil and Environmental Engineering, University of Strathclyde Glasgow, 107 Rottenrow,  
Glasgow G4 0NG, UK

e-mail: Tiku.tanyimboh@strath.ac.uk

S. H. A. Saleh

e-mail: salah.saleh@strath.ac.uk

pipe material deterioration with time or sudden increase in pressure, for example, the system's reliability is worth considering also. With regard to the topology, branched systems are suitable for small and low-density rural areas, while fully or partially looped systems are proper for urban areas (Swamee and Sharma 2008). Branched systems have the disadvantage that a break in any pipe puts all consumers downstream out of service. In fully looped systems, each demand node can be supplied from the source(s) through at least two independent paths. Two supply paths are said to be independent if they do not have a pipe in common. In the literature, the joint effects of topology and pipe size optimization were dealt with typically as two separate stages in which topology design followed by pipe sizing was carried out (Rowel and Barnes 1982; Morgan and Goulter 1982, Kessler et al. 1990, Cembrowicz 1992). However, such methods neglect the strong coupling between topology and components design to varying degrees.

Also, the relationship between topology, pipe sizes and hydraulic reliability is strong. However, previous studies that included reliability generally did not optimize the topology. Various reliability measures that are easy to calculate have been suggested including statistical entropy (Tanyimboh and Templeman 1993), resilience index (Todini 2000), network resilience (Prasad and Park 2004), modified resilience index (Jayaram and Srinivasan 2008) and surplus power factor (Vaabel et al. 2006). Among these measures, statistical entropy has been shown to be the most consistent (Reca et al. 2008; Raad et al. 2010; Baños et al. 2011; Tanyimboh et al. 2011; Saleh et al. 2012). For water distribution systems, the statistical entropy may be considered a measure of the uniformity of the pipe flow rates (Tanyimboh and Templeman 1993).

Awumah et al. (1989) developed a two-stage model for optimizing the pipe sizes and topology. In the first stage, a topology model determines whether a link is to be included using integer programming. In the second stage, pipe diameters are adjusted. Awumah and Goulter (1992) also proposed an alternative approach using statistical entropy theory. Tanyimboh and Sheahan (2002) also used statistical entropy in an approach in which the topology, pipe sizing, reliability and redundancy were considered in successive stages.

Evolutionary optimization algorithms have been used also (Davidson and Goulter 1995; Walters and Smith 1995; Geem et al. 2000; Afshar and Jabbari 2007). Evolutionary algorithms often generate infeasible solutions when solving problems that involve constraints. Case-specific constraint-violation penalties (Kougias and Theodossiou 2013) that require calibration are frequently introduced to address this issue. Saleh and Tanyimboh (2013) introduced an approach that optimizes both the topology and pipe sizes. The algorithm provides a single optimal solution and reliability aspects beyond the topology were not addressed.

This paper describes a new multi-objective evolutionary approach for the simultaneous topology, pipe size and entropy-based optimization of water distribution systems. Unlike previous entropy-based approaches such as Tanyimboh and Sheahan (2002), the pipe *flow directions* and *candidate topologies* are not specified in advance. Also, the algorithm promotes full exploitation of all *feasible and infeasible* solutions generated to guide the search. Our algorithm includes a robust measure for the *infeasibility* of any solution and a seamless generic procedure for *redundant binary codes*. Results for two test problems in the literature are included.

## 2 Optimization Approach

The difficulties associated with constraint-violation penalties that are commonly used in evolutionary algorithms include time-consuming trial runs and parameter calibration (Dridi

et al. 2008). On the other hand, penalty-free methods eliminate the need to design penalty functions and are relatively straightforward to implement without sacrificing the computational efficiency (Siew and Tanyimboh 2012). Also, penalty-free methods can maintain infeasible solutions that may have useful properties that may not be common in feasible solutions in successive generations of the optimization. Other constraint handling methods have been proposed (Deb et al. 2002). For example, Ray et al. (2001) suggested three stages of nondomination ranking using different combinations of the objective and constraint functions. Constraint handling in Deb et al. (2002) involves a binary tournament in which feasible solutions automatically dominate infeasible solutions. We developed a penalty-free strategy that exploits all efficient solutions generated, without introducing additional measures aimed at reducing the propagation of infeasible solutions.

### 2.1 Details of the Optimization Model

We used the EPANET 2 hydraulic simulation model (Rossman 2000) to determine the hydraulic properties of all solutions generated in the optimization process and to ensure the solutions satisfy conservation of mass and energy. The optimization model minimizes the initial construction cost,  $f_1$ , the infeasibility measure,  $f_2$ , and the number of pipes,  $f_3$ , as explained below.

$$f_1 = \sum_{ij} f(L_{ij}, D_{ij}) \tag{1}$$

$$f_2 = l + h + (S^* - S) + (S_g^* - S) : l = \sum_{i=1}^N \max(0, R_i^{req} - R_i), \quad h = \sum_i \max(0, H_i^{req} - H_i) \tag{2}$$

$$f_3 = \sum_{ij} p_{ij} \tag{3}$$

in which  $N$ =number of nodes; for pipe  $ij$ ,  $L_{ij}$ =length;  $D_{ij}$ =diameter;  $p_{ij}$ =1 if pipe  $ij$  is included in the topology and  $p_{ij}$ =0 otherwise;  $H_i$  and  $H_i^{req}$ =available and required residual head at demand node  $i$ , respectively;  $R_i$  and  $R_i^{req}$ =actual and required number of independent supply paths to node  $i$ , respectively;  $S$ =entropy;  $S^*$ =maximum entropy; and  $S_g^*$  = global maximum entropy.

The function  $l$  in Eq. 2 represents the total topological infeasibility of a candidate solution. The topological infeasibility at node  $i$  was taken as the shortfall in the number of independent supply paths  $R_i$ . The required number of independent supply paths,  $R_i^{req}$ , is typically 1 and 2, respectively, for branched and fully looped configurations. The function  $h$  in Eq. 2 represents the residual head infeasibility. If  $H_i \geq H_i^{req}$  for all demand nodes, then the solution is hydraulically feasible. The required residual head  $H_i^{req}$  is the head at a node above which demands are satisfied in full.  $H_i^{req}$  is typically not less than a *minimum* of about 7 m (OFWAT 2008).

For any feasible topology that has loops, there are multiple feasible sets of flow directions each of which has a maximum entropy value.  $S^*$  is the *theoretical* maximum value of entropy for a particular feasible set of *flow directions* while  $S_g^*$  is the *global* maximum entropy value considering all permissible topologies. The *global* maximum entropy value  $S_g^*$  is not known a priori; our algorithm evolves the global maximum entropy solution by assuming it corresponds to the largest entropy value it has so far identified. The infeasibility measure  $f_2$  seeks feasible

solutions that have high values of entropy (a proxy for hydraulic reliability and redundancy). Minimizing the infeasibility measure  $f_2$  promotes the inclusion of a range of maximum entropy solutions for which, *by definition*,  $S=S^*$ , in the nondominated set in addition to  $S_g^*$ .

To complete the characterization of the infeasibility function  $f_2$ , the entropy functions are described here briefly (Tanyimboh and Templeman 1993).

$$S = S_0 + \sum_{i=1}^N P_i S_i; \quad (4)$$

$S$ =entropy;  $S_0$ =entropy of source supplies;  $S_i$ =entropy of node  $i$ ;  $P_i=T_i/T$ =fraction of the total flow through the network that reaches node  $i$ ;  $T_i$ =total flow that reaches node  $i$ ;  $T$ =total demand;

$$S_0 = -\sum_{i \in I} \frac{Q_{0i}}{T} \ln\left(\frac{Q_{0i}}{T}\right); \quad (5)$$

$Q_{0i}$ =inflow rate at source node  $i$ ;  $I$ =the set of source supply nodes;

$$S_i = -\frac{Q_{i0}}{T_i} \ln\left(\frac{Q_{i0}}{T_i}\right) - \sum_{ij \in \text{out}(N_i)} \frac{Q_{ij}}{T_i} \ln\left(\frac{Q_{ij}}{T_i}\right), \quad i = 1, \dots, N; \quad (6)$$

$Q_{i0}$ =demand at node  $i$ ;  $Q_{ij}$ =flow rate in pipe  $ij$ ; and  $\text{out}(N_i)$ =set of all pipe flows from node  $i$ .

For a typical node with, say, two incident pipes downstream, it can be shown that  $S_i \leq \ln(3) \approx 1.1$  (Shannon 1948). Given that  $P_i=T_i/T \leq 1.0$ , it is expected that the value of the network entropy  $S$  in Eq. 4 will be relatively small for the typical water distribution system. Therefore, it is expected that the contributions of the entropy terms  $(S^*-S)$  and  $(S_g^*-S^*)$  to the infeasibility measure  $f_2$  in Eq. 2 will be relatively small. The objective function  $f_2$  may be considered an *entropy-augmented infeasibility measure*. Minimizing  $f_2$  aims simultaneously to satisfy residual head and topology requirements and maximize entropy. Eqs. 4–6 are an extension of the statistical entropy function that is a measure of uncertainty (Shannon 1948). In a probabilistic system the uncertainty is a maximum if all possible system states or outcomes are equally likely. Conversely, the uncertainty decreases as the probabilities associated with the states or outcomes become more unequal. The term  $[(S^*-S)+(S_g^*-S^*)]=(S_g^*-S)$  in the infeasibility measure  $f_2$  may be considered an estimate of the *unrealized entropy potential*; by definition its value is zero for  $S=S^*=S_g^*$ .

## 2.2 Practical Topology Confirmation and Redundant Binary Codes

We developed a topology confirmation algorithm coded in C, to enable a consistent and bias-free fitness assessment of all feasible and infeasible solutions. The total number of paths  $NP_i$  supplying demand node  $i$  from all sources collectively was determined with regard to the pipe flow directions obtained from EPANET 2. We used an efficient path enumeration algorithm proposed by Yassin-Kassab *et al.* (1999). If  $NP_i=0$ , the node cannot be supplied. If  $NP_i=1$ , the node can be supplied. If  $NP_i \geq 2$ , for all nodes, a path inter-dependency investigation is carried out to check whether the network is fully looped. We adopted a practical procedure that does not involve an exhaustive enumeration of all the paths supplying each node. For a pair of independent supply paths, removing a pipe from one path does not affect the other path. Therefore, the procedure entails removing all pipes one at a time and in each case observing whether all nodes can be reached. If all nodes can be supplied from one or more sources after the removal of all pipes one by one with replacement, then all nodes have at least two independent supply paths. It is worth observing that EPANET 2 sets default values of node

pressures and pipe flows within parts of a network that are not connected to a source. We addressed this by assigning zero flows and pressures, respectively, to such pipes and nodes.

In order to represent the vector of decision variables in a genetic algorithm, an  $n$ -bit binary string gives rise to  $2^n$  different  $n$ -bit codes and, depending on the number of decision variables, some codes may be redundant. We assumed redundant codes represent closed pipes whose flow-carrying capacity is zero. The closed pipes are allocated pipe sizes taken from just above the upper end of the *real* set of available pipe diameters. The data required to implement the procedure are the unit costs for the *fictitious or assumed* diameters. As the fictitious diameters have no functional value, it is anticipated they will become extinct through evolution and natural selection. The benefits of this novel approach are that it is entirely generic and very practical; additional parameters that require special calibration are not introduced and pre-optimization trial runs are not required. The premature loss of potentially useful genes is thus avoided, and the genetic code that is transmitted in successive generations is not degraded (Herrera *et al.* 1998).

### 3 Computational Solution

We used the Nondominated Sorting Genetic Algorithm (NSGA) II that has been used extensively, and its merits have been reported elsewhere (Deb *et al.* 2002; Dridi *et al.* 2008). Selection for crossover was carried out with a binary tournament. Single-point crossover was used to produce two offspring from two parents. Once the offspring population was created, the mutation operator reversed the selected bits. The optimization problem was posed as:

$$\text{Minimize } \mathbf{f} = (f_1, f_2, f_3)^T \tag{7}$$

The decision variables are the pipe diameters  $D_{ij}$  and link selection variables  $p_{ij}$  for the entire network. To make all three objectives in Eq. 7 roughly similar in magnitude, each  $f_i^m$ , i.e. the value of objective  $m$  for solution  $i$ , was normalized as

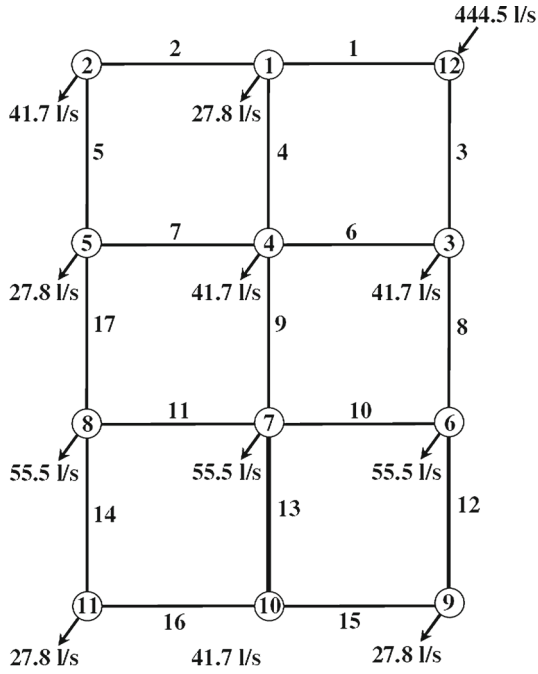
$$fn_i^m = (f_i^m - f_{\min}^m) / (f_{\max}^m - f_{\min}^m); \forall i, \forall m \tag{8}$$

In the generation in question,  $f_{\min}^m$  and  $f_{\max}^m$  = minimum and maximum value of objective  $m$ , respectively; and  $fn_i^m$  = normalized value of objective  $m$  for solution  $i$ .

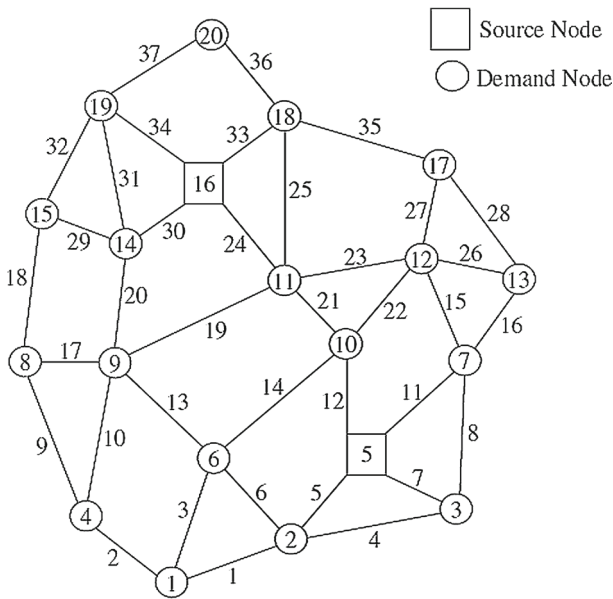
In each generation of the optimization algorithm, each solution in the population is analysed using EPANET 2. The resulting pipe flow rates are used to calculate the entropy (Eq. 4). In general, numerical nonlinear optimization is required to calculate the maximum value of the entropy  $S^*$ . However, computationally efficient path entropy methods that do not involve numerical optimization directly are available. We used the “simplified path entropy method” developed by Ang and Jowitt (2005) for the single-source network example (Section 4.1) and an algorithm known as the “ $\alpha$ -method” developed by Yassin-Kassab *et al.* (1999) for the multiple-source network example (Section 4.2). Application of the  $\alpha$ -method involves solving a non-linear system of equations and, for a two-source network, it reduces to the solution of a single nonlinear equation for which we used the bisection method (Press *et al.* 2003).

### 4 Results and Discussion

Two networks from the literature were considered. The Hazen-Williams roughness coefficient for *all* pipes is 130. For each network, the optimization algorithm was executed 30 times on a



(a) Network 1



(b) Network 2

**Fig. 1** Topologies of Networks 1 and 2 with all the candidate pipes

**Table 1** Results and convergence statistics for 30 optimization runs

(a) Network 1					
Properties	Minimum	Mean	Median	Maximum	Standard deviation
GME entropy	3.380570	3.560733	3.561684	3.592494	0.041524
SME entropy	2.401622	2.489328	2.476941	2.660135	0.059723
GME cost (£10 <sup>6</sup> )	2.177413	2.787246	2.730261	3.552885	0.349057
SME cost (£10 <sup>6</sup> )	1.181715	1.293399	1.292801	1.496615	0.074389
Number of fully looped feasible solutions (out of 100)	36	48.533	50	57	5.778
Number of partially looped and branched feasible solutions per 100	0	6.2	5	13	3.219
Smallest surplus residual head for feasible solutions (m)	0.007	0.629	0.464	2.691	0.654
Function evaluations (FEs) for convergence	314,700	733,413	806,050	979,200	208,529
Extinction of all fictitious pipes (FEs)	1,500	4,600	3,850	17,300	3,725
Extinction of 750 mm pipes (FEs)	500	2,050	1,650	8,500	1,594
Extinction of 700 mm pipes (FEs)	900	2,583	1,950	5,700	1,406
Extinction of 650 mm pipes (FEs)	900	4,003	2,600	17,300	3,885
Hypervolume	0.653	0.661	0.660	0.682	0.004
CPU time for convergence (minutes)	27.35	63.75	70.06	85.11	18.13
(b) Network 2					
Properties	Minimum	Mean	Median	Maximum	Standard deviation
GME entropy	4.476402	4.872415	4.895875	5.190007	0.185378
SME entropy	2.981072	3.140691	3.136344	3.306349	0.089291
GME cost (millions of CU)	5.626414	6.814409	6.843641	7.738914	0.561343
SME cost (millions of CU)	2.253554	2.549198	2.502468	2.925169	0.179179
Number of fully looped feasible solutions (out of 100)	37	46.207	46	52	2.631
Number of partially looped and branched feasible solutions per 100	2	8.586	9	14	3.275
Smallest surplus residual head for feasible solutions (m)	0.002	0.082	0.058	0.335	0.083
Function evaluations (FEs) for convergence	552,000	905,224	949,800	997,500	106,861
Extinction of all fictitious pipes (FEs)	13,400	39,314	41,100	80,600	14,452
Extinction of 800 mm pipes (FEs)	5,400	25,914	24,600	78,200	16,222
Extinction of 750 mm pipes (FEs)	11,800	39,028	41,100	80,600	14,800
Hypervolume	0.642	0.645	0.645	0.648	0.002
CPU time for convergence (minutes)	69.10	113.32	118.90	124.87	13.38

desktop personal computer (Processor: Intel Core 2 Duo, CPU: 2.99 GHz, RAM: 3.21 GB). The population size, cross-over probability and stopping criterion were: 100, 1.0 and 10<sup>6</sup> hydraulic simulations, respectively. The 100 solutions in each of the 30 nondominated sets achieved were then merged. Out of the 30×100 i.e. 3,000 solutions the final set of 100 nondominated solutions was obtained by a screening procedure that considers the Pareto-optimality and diversity (i.e. crowding distance) of the solutions in the objective space (as in NSGA II). The convergence point in the optimization was taken as the point after which there

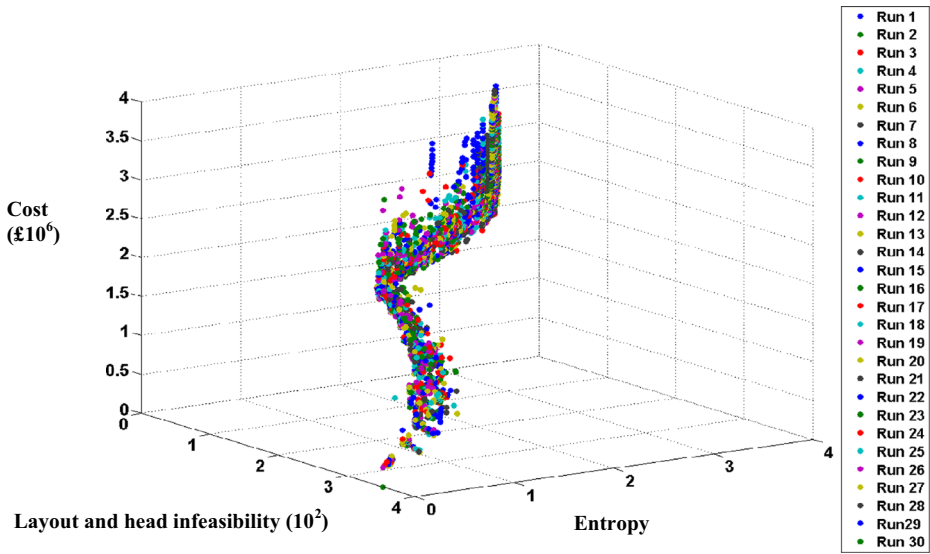


Fig. 2 Pareto-optimal fronts for Network 1 showing 30 optimization runs

was no further improvement in both the entropy and cost for the feasible solution with the highest entropy value.

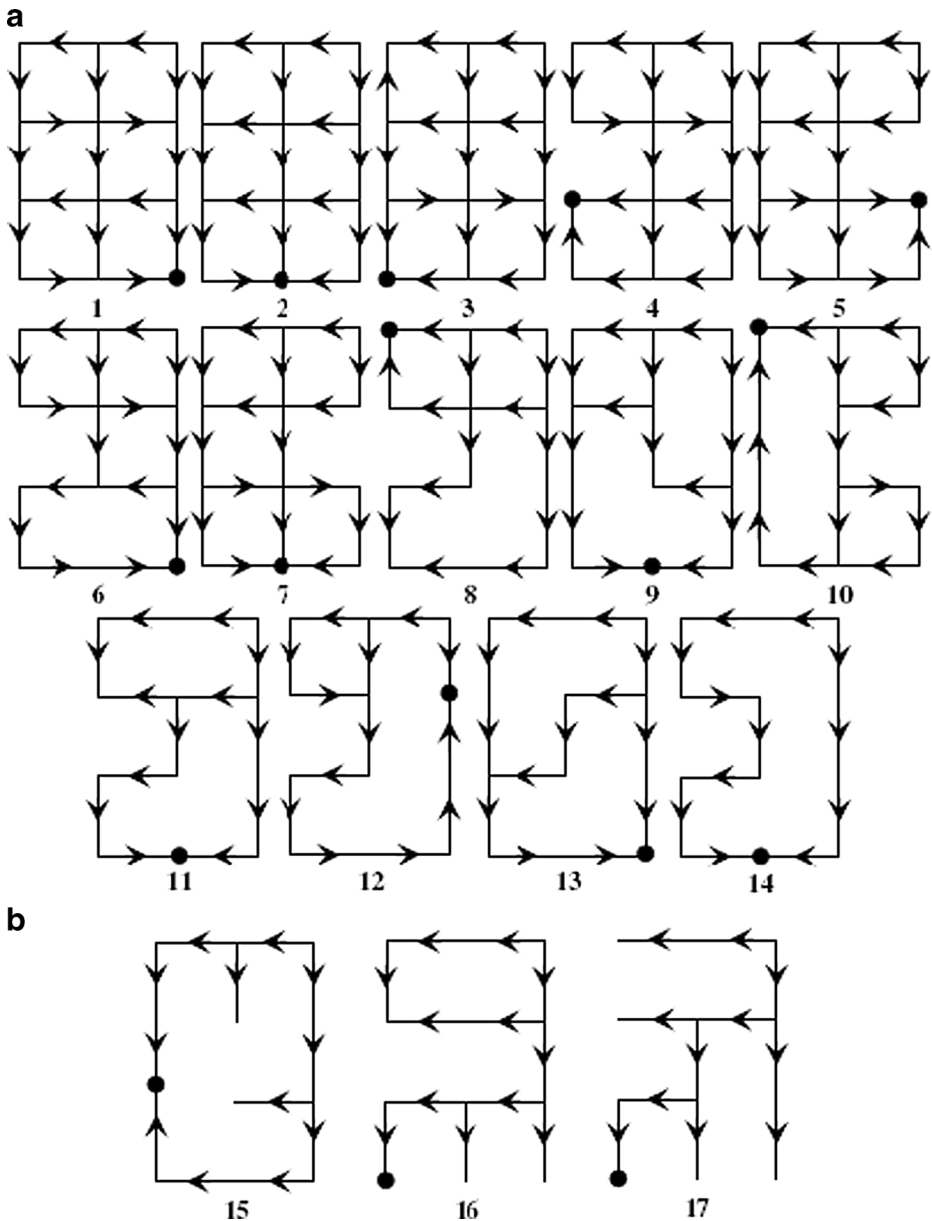
Given a set of nondominated solutions, the hypervolume is a measure of the fraction of the objective space dominated by the said solutions. Its value increases as the achieved solutions approach the real Pareto-optimal front. The value increases also as the range of solutions in the nondominated set increases or their distribution becomes more uniform. Larger hypervolume values are thus preferred (Knowles 2005). The hypervolume was calculated after normalizing the objectives according to Eq. 8, for each optimization run and the union of all the 30 runs.

#### 4.1 Sample Network 1

The network shown in Fig. 1a (Awumah *et al.* 1990) has one supply node, 17 pipes and 11 demand nodes. The elevation of the nodes is 0 m. The head at the supply node is 100 m.  $H_i^{req} = 30$  m for all demand nodes. All pipes have length of 1,000 m.  $R_i^{req} = 2$  specifies a fully looped topology. We used 12 pipe diameters (100, 125, 150, 200, 250, 300, 350, 400, 450, 500, 550 and 600 mm) i.e.  $13^{17} = 8.65 \times 10^{18}$  solutions including pipe omission. Given  $10^6$  hydraulic simulations the sampling ratio was  $10^6 / 8.65 \times 10^{18} = 8.65 \times 10^{-12}$ . Each solution was represented by a 68-bit chromosome based on a 4-bit pipe-size representation scheme. A 4-bit binary string produces  $2^4 = 16$  codes three of which are redundant as there are 13 pipe-size alternatives. We allocated three *assumed* pipe diameters of 650, 700 and 750 mm to the three redundant codes. The pipe costs were taken as  $800D^{1.5}$  (£/m) where  $D$  is the pipe diameter (in metres). The absolute probability of bit mutation was  $1/68 \approx 0.015$ .

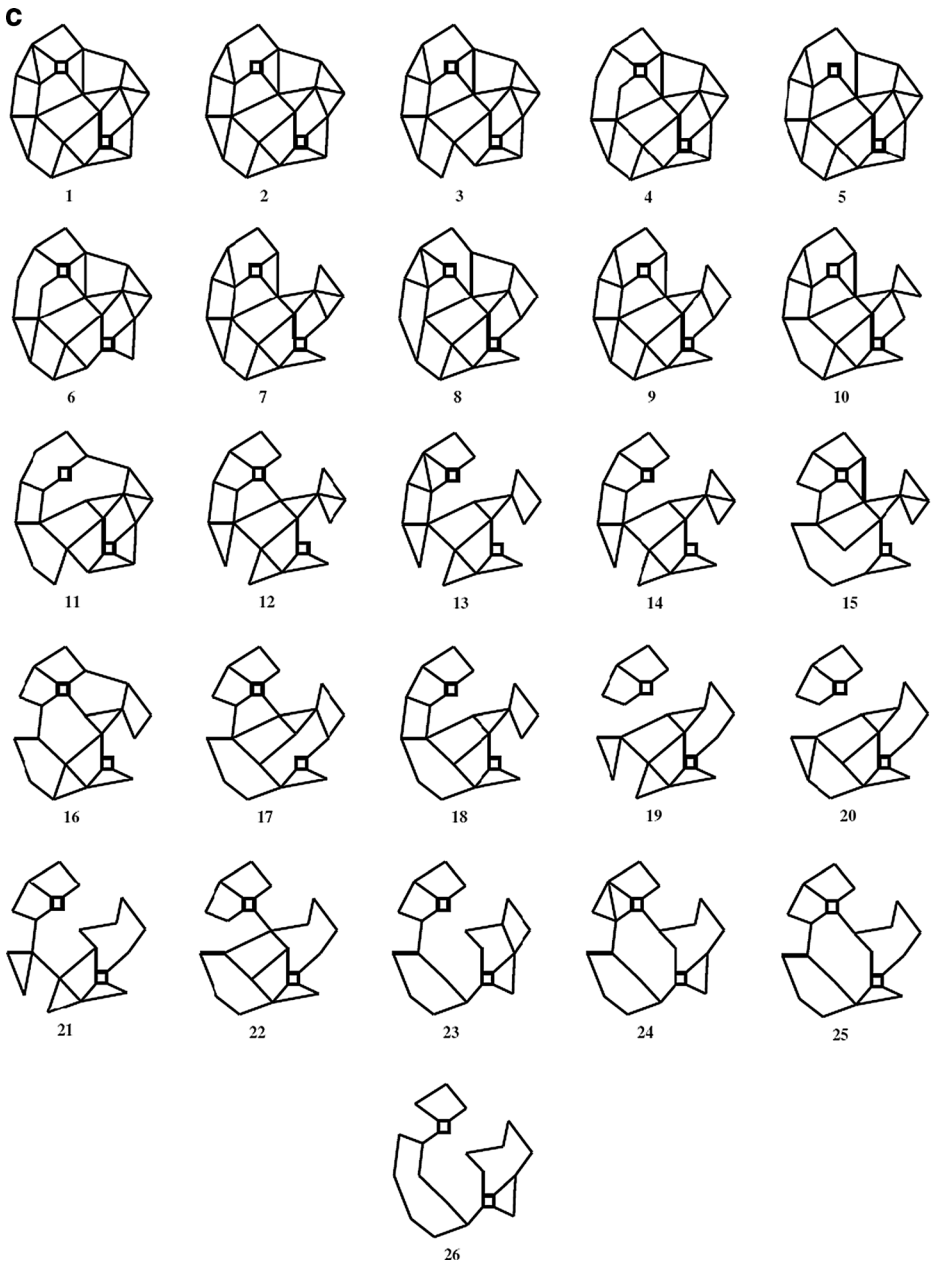
Table 1 shows the general characteristics of the optimization algorithm. The *minimum* cost achieved for the *global* maximum entropy (GME) solution was £2,177,413. The *maximum* value of entropy for the GME solution was 3.592494. The *mean* number of function evaluations and CPU time required to achieve convergence were 733,413 and about 64 min, respectively. There is a multiplicity of maximum entropy values and one of the aims of the optimization is to





**Fig. 3** **a** Topologies and flow directions of fully looped hydraulically feasible maximum entropy families for Network 1. The solid circles represent nodes with the smallest residual heads. **b** Topologies and flow directions of branched and partially looped hydraulically feasible maximum entropy families for Network 1. The solid circles represent nodes with the smallest residual heads. **c** Topologies of looped and partially looped feasible solutions for Network 2. The rectangles represent sources. **d** Topologies of branched and partially branched feasible solutions for Network 2. The rectangles represent sources

provide a wide range of maximum entropy solutions. The maximum entropy value that is the *smallest* gives rise to the *smallest* Maximum Entropy (SME) solution. The *minimum* cost of the



**Fig. 3** (continued)

SME solution was £1,181,715. The *maximum* value of entropy for the SME solution was 2.660135. The minimum surplus head at the critical node was 0.007 m. The optimization model includes multiple conflicting objectives. Therefore, it is not guaranteed that any minimum node pressure constraints will be active. Furthermore, the slack for a limiting minimum node pressure constraint need not be exactly zero, due to the discrete pipe diameters.

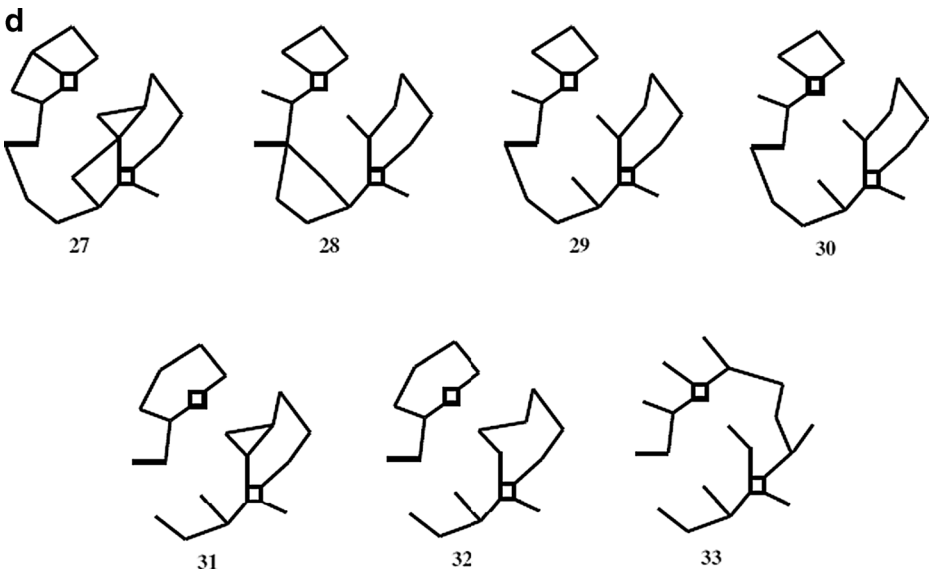


Fig. 3 (continued)

Figure 2 shows the frontier-optimal solutions achieved of which the most infeasible solution has cost = 0; entropy = 0; topological infeasibility = 24 (i.e. 2 independent paths per node  $\times$  12 nodes); and residual head infeasibility = 330 m (i.e. 11 demand nodes  $\times$  30 m of residual head for each demand node). This solution survives until the end of the optimization because the algorithm is bias-free with respect to constraint violations. Any crossover between this solution and another solution will likely create new layouts. Also, the hypervolume value for the final merged Pareto-optimal front was 0.676. This is similar to the values in Table 1 for the individual optimization runs.

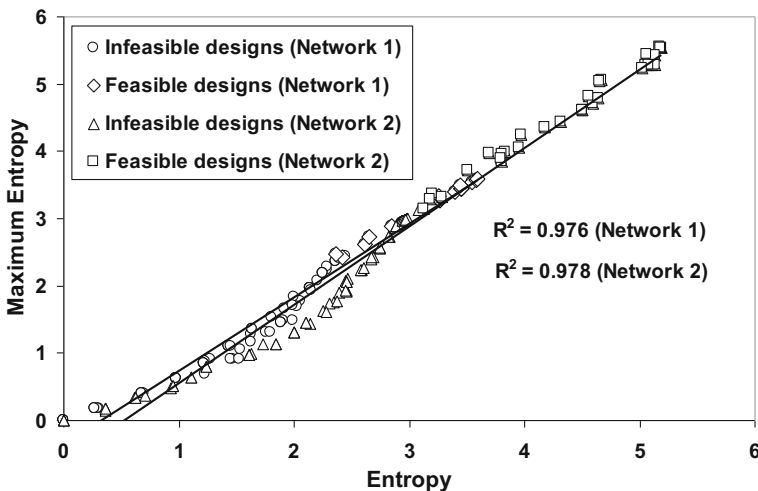


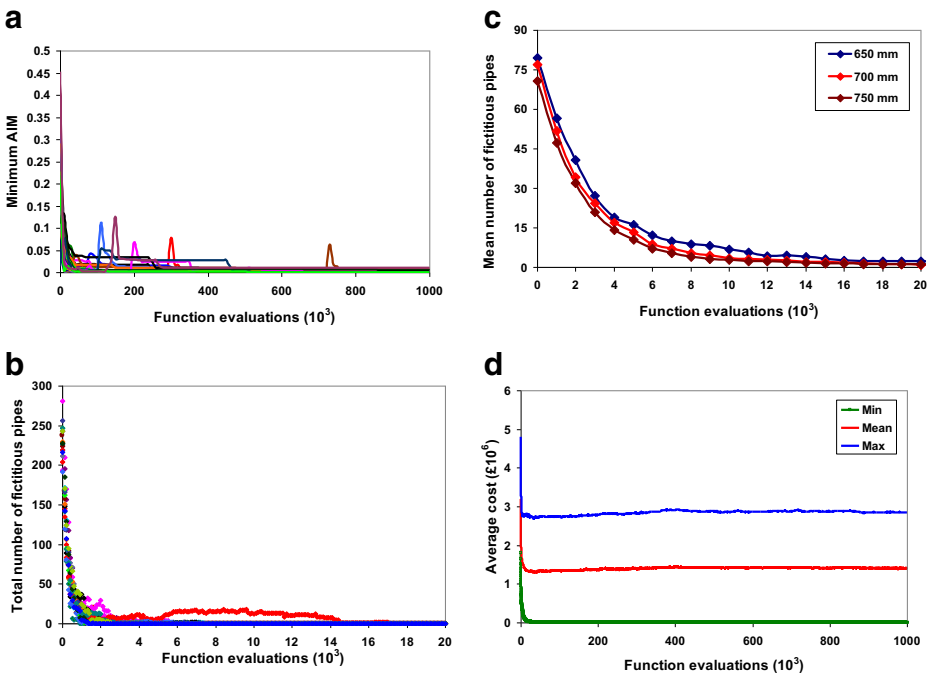
Fig. 4 Achieved vs theoretical maximum entropy values of Networks 1 and 2

Tables 2 and 3 (in the appendix) illustrate the range of feasible solutions achieved. The final Pareto-optimal set has 23 hydraulically feasible fully-looped solutions and 11 different fully looped topologies (see Fig. 3a). All infeasible solutions in the final Pareto-optimal set were found to be topologically infeasible (i.e.  $\exists i: R_i < R_i^{req} = 2$ ), of which only three were hydraulically feasible (i.e.  $H_i \geq H_i^{req} = 30$  m;  $\forall i$ ) (see Fig. 3b). Fig. 4 provides further confirmation that the solutions achieved are essentially maximum entropy solutions.

Figure 5 shows the progress of the optimization. The fictitious pipe diameters were eliminated in the early stages consistently (Fig. 5b-c). Prior to their complete elimination, fictitious pipe diameters were present in both hydraulically feasible and infeasible solutions. Also, the observed rates of elimination reflected the pipe sizes and costs (Table 1 and Fig. 5c). On average the larger more expensive assumed diameters were eliminated more quickly. These results suggest convergence of the algorithm is very quick and the proposed procedure for handling redundant binary strings is highly effective.

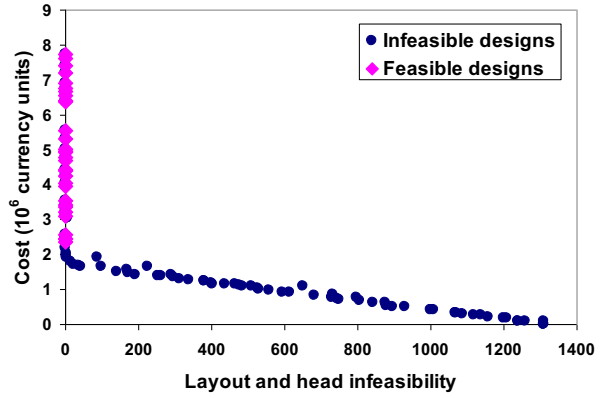
## 4.2 Sample Network 2

The network shown in Fig. 1b has two supply nodes, 18 demand nodes and 37 pipes. The node demands, required residual heads, pipe lengths and costs are available in Morgan and Goulter (1985). There are 13 pipe sizes, i.e.  $14^{37} = 2.55 \times 10^{42}$  solutions including pipe omission. Given  $10^6$  hydraulic simulations the sampling ratio was  $10^6 / 2.55 \times 10^{42} = 3.92 \times 10^{-37}$ . A 4-bit binary substring for each pipe size gave a chromosome with length of 148 bits. The absolute probability of bit mutation was  $1/148$ . With 14 options for each pipe, two

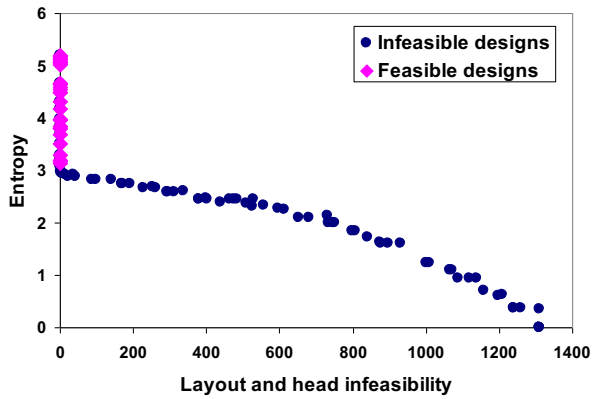


**Fig. 5** Illustration of convergence characteristics with Network 1. (a) and (b) show 30 individual optimization runs while (c) and (d) show averages based on the 30 optimization runs. AIM abbreviates the entropy-augmented infeasibility measure

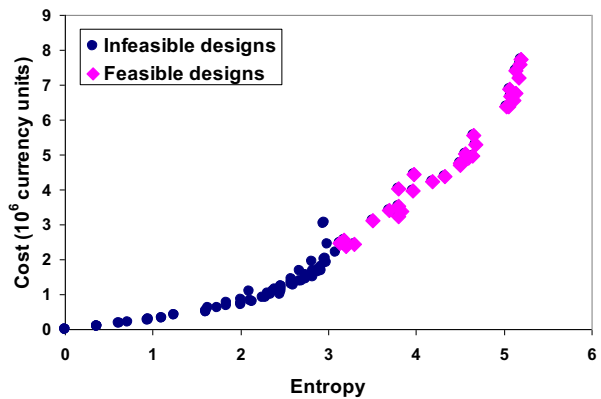
**Fig. 6** Pareto-optimal fronts for Network 2 based on the union of the results of 30 optimization runs



(a) Infeasibility vs. cost



(b) Infeasibility vs. entropy



(c) Entropy vs. cost

codes (out of  $2^4=16$ ) were redundant. Two *fictitious* pipe diameters of 750 mm and 800 mm, with costs of 520.9/m and 591.7/m respectively, were allocated to the two redundant codes by extending the cost function of the *real* pipe diameters. The costs are in generic currency units (CU).

Table 1 summarizes the results achieved. The final Pareto-optimal front had 31 feasible solutions based on 26 layouts (Fig. 3c) that are fully non-dendritic (i.e. layouts with no dead ends). Additionally, seven branched and partially-branched feasible solutions were achieved (Fig. 3d). The cheapest fully-looped feasible solution (Layout 24 in Fig. 3c) with a cost of 2,374,070 CU (Solution 29 in Table 4) had 12 pipes removed. The most expensive fully-looped feasible solution (Layout 2 in Fig. 3c) with a cost of 7,738,914 CU (Solution 4 in Table 4) had one pipe removed. Figure 6 shows the relationship between the cost, entropy and infeasibility.

## 5 Conclusions

A new approach to the simultaneous topology and reliability-based pipe-size optimization of water distribution systems has been developed. The method provides a multiplicity of cost-effective candidate solutions distributed among a diverse range of optimal topologies. We used statistical entropy as a computationally efficient surrogate measure of the hydraulic reliability/redundancy and reduced the computational complexity by introducing a new *entropy-augmented infeasibility measure*. Our optimization model includes the following essential features: (a) entropy maximization within individual feasible sets of flow directions; (b) entropy maximization across all feasible sets of flow directions within individual topologies; (c) entropy maximization across all topologies; (d) minimization of initial construction cost; (e) promotion of a wide variety of alternative solutions; (f) satisfaction of minimum topological adequacy (i.e. supply node and demand node reachability); (g) satisfaction of minimum topological redundancy (i.e. alternative independent supply paths); and (h) adequacy of nodal flows and pressures.

Clearly, many complex objectives and constraints are involved. The entropy-augmented infeasibility measure introduced here simplifies the optimization and reduces the computational complexity considerably as the objectives have been reduced to only three (Saxena *et al.* 2013; Deb *et al.* 2002). The optimization problem addressed has six objectives. Sinha *et al.* (2013) emphasize that the computational solution of a six-objective optimization problem is a '*formidable task*' for most evolutionary multi-objective optimization algorithms that aim to generate the entire Pareto-optimal front. Some of the challenges include: difficulties in achieving at once both *diversity of solutions* and convergence on the true Pareto-optimal front; and difficulties arising from the inability to visualize the Pareto-optimal front geometrically.

The genetic algorithm approach proposed allows full exploitation of all the *efficient feasible and infeasible solutions* generated in the optimization. Any *redundant binary codes* created are eliminated in a seamless and generic way through natural selection. This avoids arbitrary loss of potentially useful genetic material and preserves the quality of the information that is transmitted from one generation to the next. The results for the two test problems considered are sufficiently encouraging to suggest further research to improve and extend the algorithms proposed may be beneficial.

**Acknowledgments** The first author's PhD programme was funded by the Government of Libya. This research was funded in part by the UK Engineering and Physical Sciences Research Council (EPSRC Grant Reference EP/G055564/1).

Appendix

**Table 2** Nondominated feasible solutions achieved for Network 1

(a) Fully looped feasible solutions								
Solution number	Cost (£10 <sup>6</sup> )	<sup>a</sup> Surplus head (m)	<sup>b</sup> Critical node	Achieved entropy ( <i>S</i> )	Maximum entropy (ME)	<i>ME</i> - <i>S</i>	<i>GME</i> - <i>ME</i>	<sup>d</sup> ME family
1	2.52235	19.734	9	3.592494	3.592800 <sup>c</sup>	0.000306	0.000000	1
2	1.592572	9.817	10	3.257305	3.330590	0.073285	0.262210	2
3	1.823409	2.083	11	3.439667	3.489588	0.049921	0.103213	3
4	2.756591	21.594	8	3.581014	3.581115	0.000101	0.011685	4
5	1.977641	11.642	6	3.545526	3.546760	0.001234	0.046041	5
6	3.138684	38.903	9	3.449581	3.449665	0.000084	0.143135	6
7	3.067751	38.729	9	3.449580	3.449665	0.000085	0.143135	6
8	2.930185	36.096	9	3.449268	3.449665	0.000397	0.143135	6
9	2.235305	8.626	9	3.448149	3.449665	0.001517	0.143135	6
10	1.871266	5.864	10	3.395190	3.398060	0.002870	0.194740	7
11	1.292923	1.350	2	2.850790	2.891747	0.040958	0.701053	8
12	1.271386	3.443	10	2.647933	2.723657	0.075724	0.869144	9
13	2.969728	33.350	2	3.265855	3.269803	0.003947	0.322998	10
14	1.173127	1.910	10	2.367236	2.469176	0.101940	1.123624	11
15	2.849506	35.237	3	2.940251	2.940255	0.000004	0.652545	12
16	2.839981	34.462	3	2.940250	2.940255	0.000005	0.652545	12
17	2.769049	34.041	3	2.940247	2.940255	0.000008	0.652545	12
18	2.729939	32.628	3	2.940223	2.940255	0.000033	0.652545	12
19	2.683465	32.394	3	2.940159	2.940255	0.000097	0.652545	12
20	2.605798	30.773	3	2.940102	2.940255	0.000154	0.652545	12
21	2.535235	23.781	3	2.939597	2.940255	0.000659	0.652545	12
22	1.529674	9.394	9	2.614232	2.624432	0.010200	0.968369	13
23	1.721305	6.330	10	2.419348	2.425825	0.006477	1.166975	14
(b) Branched or partially looped feasible solutions								
Solution number	Cost (£10 <sup>6</sup> )	Surplus head (m)	Critical node	Achieved entropy ( <i>S</i> )	Maximum entropy (ME)	<i>ME</i> - <i>S</i>	<i>GME</i> - <i>ME</i>	ME family
24	1.508359	18.840	8	2.447343	2.447345	0.000002	1.145456	15
25	1.075510	0.842	11	2.387568	2.404150	0.016582	1.188651	16
26	1.050212	0.503	11	2.360799	2.360799	0.000000	1.232002	17

<sup>a</sup> Refers to the excess residual head at the node with the smallest residual head

<sup>b</sup> Refers to the node with the smallest residual head

<sup>c</sup> The global (i.e. greatest) maximum entropy (GME) value found

<sup>d</sup> The maximum entropy family here refers to any subset of solutions whose topology and pipe flow directions are identical

**Table 3** Details of selected nondominated feasible solutions achieved for Network 1

Pipes	Pipe diameters (mm) <sup>a</sup>						GME solution	
	Least-cost solutions							
	Fully branched	Partly branched	Fully looped	Three loops	Four loops	Five loops		Six loops
1-12	200	200	200	350	450	350	350	450
1-2	150	200	150	300	550	250	250	400
1-4	-	-	-	300	300	300	300	300
2-5	-	200	150	200	350	300	300	450
3-12	400	400	400	300	200	300	300	200
3-4	350	150	300	-	300	300	350	350
3-6	200	350	250	300	400	-	200	300
4-5	150	100	150	200	400	300	300	550
4-7	300	-	300	100	200	200	350	200
5-8	-	-	-	250	-	350	150	125
6-7	-	300	-	200	300	400	200	400
6-9	200	150	200	200	100	250	200	100
7-8	250	250	250	-	350	350	200	450
7-10	200	200	-	-	-	-	100	200
8-11	200	200	250	200	300	200	200	300
9-10	-	-	250	200	200	200	125	250
10-11	-	-	100	100	350	250	300	300
Cost (£10 <sup>6</sup> )	1.050	1.076	1.173	1.272	2.235	1.871	1.593	2.522
Critical node	11	11	10	10	9	10	10	9
Surplus head at critical node (m)	+0.50	+0.84	+1.91	+3.44	+8.63	+5.86	+9.82	+19.73
Number of pipes	11	12	13	14	15	16	17	17
Number of loops	0	1	2	3	4	5	6	6
Entropy	2.361	2.389	2.367	2.648	3.448	3.395	3.257	3.593



Table 3 (continued)

	Pipe diameters (mm) <sup>a</sup>										GME solution
	Least-cost solutions										
	Fully branched	Partly branched	Fully looped	Three loops	Four loops	Five loops	Six loops				
<i>ME – S</i>	0.000	0.017	0.102	0.076	0.002	0.003	0.073				0.000
<i>GME – ME</i>	1.232	1.189	1.124	0.869	0.143	0.195	0.262				0.000
ME family	17	16	11	9	6	7	2				1
Entropy-augmented infeasibility	12.232	7.205	1.226	0.945	0.145	0.198	0.335				0.000
Layout infeasibility	11	6	0	0	0	0	0				0
Residual head infeasibility (m)	0	0	0	0	0	0	0				0
Solution number	26	25	14	12	9	10	2				1

<sup>a</sup> A missing diameter shown as dash (–) means the link is not included

**Table 4** Nondominated feasible solutions achieved for Network 2

(a) Looped and partially looped feasible solutions									
Solution number	Cost (CU)	<sup>a</sup> Surplus head (m)	Critical node	Achieved entropy (S)	Maximum entropy (ME)	ME - S	GME - ME	Layout number	
1	7,209,000	0.306	15	5.171049	5.557610	0.386561	0.000000	1	
2	6,563,651	0.267	19	5.109483	5.277704	0.168221	0.279906	1	
3	6,392,398	0.643	19	5.051366	5.281130	0.229764	0.276480	1	
4	7,738,914	0.470	15	5.190007	5.541571	0.351563	0.016039	2	
5	6,756,341	0.175	19	5.137941	5.280588	0.142647	0.277022	2	
6	6,672,388	0.669	19	5.075552	5.284013	0.208461	0.273597	2	
7	7,601,494	0.148	15	5.181323	5.536762	0.355439	0.020848	3	
8	7,399,096	0.082	15	5.130746	5.422957	0.292211	0.134653	4	
9	6,376,680	0.411	15	5.022152	5.232052	0.209900	0.325558	5	
10	6,894,839	0.024	15	5.061369	5.437145	0.375776	0.120465	6	
11	4,916,521	0.293	15	4.594931	4.709163	0.114232	0.848447	7	
12	5,307,348	0.628	8	4.672370	5.062133	0.389763	0.495477	8	
13	4,957,599	0.853	15	4.641112	4.786899	0.145787	0.770711	9	
14	4,774,924	1.594	15	4.498214	4.604216	0.106002	0.953394	10	
15	4,698,694	0.701	15	4.498006	4.604195	0.106189	0.953415	10	
16	5,552,754	1.241	15	4.655285	5.039450	0.384165	0.518160	11	
17	5,025,859	3.256	13	4.556314	4.818961	0.262647	0.738649	12	
18	4,394,286	0.515	17	4.320600	4.434738	0.114138	1.122872	13	
19	4,235,979	3.219	20	4.177757	4.350449	0.172692	1.207161	14	
20	3,372,806	2.753	7	3.836534	3.997189	0.160655	1.560421	15	
21	3,230,154	1.063	1	3.802677	3.851815	0.049138	1.705795	16	
22	3,960,139	0.866	7	3.956164	4.051982	0.095818	1.505628	17	
23	3,538,884	0.162	7	3.800675	3.955395	0.154720	1.602215	18	
24	4,433,439	6.644	7	3.976741	4.238134	0.261393	1.319476	19	
25	3,426,289	3.712	7	3.693540	3.974741	0.281202	1.582869	20	

Table 4 (continued)

26	4,039,451	1.481	17	3.795224	3.891766	0.096542	1.665844	21
27	3,107,144	1.568	13	3.503619	3.723526	0.219906	1.834085	22
28	2,454,604	0.435	20	3.287140	3.317104	0.029964	2.240506	23
29	2,374,070	0.066	19	3.203094	3.359538	0.156445	2.198072	24
30	2,561,895	1.473	19	3.179116	3.291627	0.1112511	2.265983	25
31	2,460,335	1.141	8	3.130207	3.137692	0.007485	2.419918	26
(b) Branched and partially branched feasible solutions <sup>c</sup>								
Solution number	Cost (CU)	Surplus head (m)	Critical node	Achieved entropy (S)	Maximum entropy (ME)	ME - S	GME - ME	Layout number
32	2,197,697	1.550	4	3.079714	3.120364	0.040650	2.437246	27
33	1,922,166	1.045	20	2.980528	2.992726	0.012198	2.564884	28
34	2,027,732	0.040	17	2.968468	2.990397	0.021929	2.567213	29
35	1,991,802	0.045	17	2.957860	2.983705	0.025845	2.573905	30
36	3,052,542	1.358	1	2.954783	2.963028	0.008244	2.594582	31
37	3,042,543	0.371	4	2.942952	2.963028	0.020076	2.594582	32
38	1,808,731	0.371	4	2.915428	2.915428	0.000000	2.642182	33

<sup>a</sup> Refers to the excess residual head at the node with the smallest residual head

<sup>b</sup> The global maximum entropy (GME) value found

<sup>c</sup> A branched configuration in this context is one that has a dead end

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Afshar MH, Jabbari I (2007) Simultaneous layout and size optimization of pipe networks using genetic algorithms. *Arab J Sc Eng* 33(2B):391–409
- Ang WK, Jowitt PW (2005) Path entropy method for multiple-source water distribution networks. *Eng Optim* 37(7):705–715
- Awumah K, Goulter I (1992) Maximizing entropy-defined reliability of water distribution networks. *Eng Optimization* 20(1):57–80
- Awumah K, Bhatt SK, Goulter IC (1989) An integer programming model for layout design of water distribution networks. *Eng Optim* 15(1):57–70
- Awumah K, Goulter I, Bhatt SK (1991) Entropy-based redundancy measures in water distribution network design. *J Hydraul Eng* 117(5):595–614
- Baños R, Reca J, Martínez J, Gil C, Márquez A (2011) Resilience indexes for water distribution network design: a performance analysis under demand uncertainty. *Water Resour Manag* 25(10):2351–2366
- Cembrowicz RG (1992) Water supply systems optimization for developing countries. In: Coulbeck B, Evans E (eds) *Pipeline systems*. Kluwer, London, pp 59–76
- Davidson JW, Goulter IC (1995) Evolution program for design of rectilinear branched networks. *J Comput Civ Eng* 9(2):112–121
- Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans Evol Comput* 6(2):182–197
- Dridi L, Parizeau M, Maihot A, Villeneuve J-P (2008) Using evolutionary optimization techniques for scheduling water pipe renewal considering a short planning horizon. *Comput-Aided Civ Infrastruct Eng* 23(8):625–635
- Geem ZW, Kim JH, Yoon YN (2000) Optimal layout of pipe networks using harmony search. *Proceedings of 4th International Conf. on Hydro-Science and Engineering* Seoul, South Korea
- Herrera F, Lozano M, Verdegay JL (1998) Tackling real-coded genetic algorithms: operators and tools for behavioural analysis. *Artif Intell Rev* 12:265–319
- Jayaram N, Srinivasan K (2008) Performance-based optimal design and rehabilitation of water distribution networks using life cycle costing. *Water Resour Res* 44(1)
- Kessler A, Ormsbee L, Shamir U (1990) A methodology for least-cost design of invulnerable water distribution networks. *Civ Eng Syst* 7(1):20–28.
- Knowles J (2005) A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE Trans Evol Comput* 10(1):50–66
- Kougias IP, Theodossiou NP (2013) Multi-objective pump scheduling optimization using harmony search algorithm and polyphonic HSA. *Water Resour Manag* 27(5):1249–1261
- Morgan DR, Goulter IC (1982) Least cost layout and design of looped water distribution systems. *Proceedings of 9th International Symposium on Urban Hydrology, Hydraulics and Sediment control*, Lexington, KY, USA, 27–30
- Morgan DR, Goulter IC (1985) Optimal urban water distribution design. *Water Resour Res* 21(5):642–652
- OFWAT (2008) Guaranteed Standards Scheme [Website accessed 26 February 2013] <http://www.ofwat.gov.uk/consumerissues/rightsresponsibilities/waterpressure>.
- Prasad TD, Park NS (2004) Multi-objective genetic algorithms for design of water distribution networks. *J Water Resour Plan Manag* 130(1):73–82
- Press WH, Teukolski SA, Vetterling WT, Flannery BP (2003) *Numerical Recipes in FORTRAN 77*, Vol. 1, Cambridge University Press, p. 346–7
- Raad DN, Sinske AN, van Vuuren JH (2010) Comparison of four reliability surrogate measures for water distribution systems design. *Water Resour Res* 46(5):W05524
- Ray T, Tai K, Seow C (2001) An evolutionary algorithm for multiobjective optimization. *Eng Optim* 33(3):399–424
- Reca J, Martínez J, Banos R, Gil C (2008) Optimal design of gravity-fed looped water distribution networks considering the resilience index. *J Water Resour Plan Manag* 134(3):234–238
- Rossman LA (2000) *EPANET 2 users manual*. Water supply and water resources division, national risk management research laboratory. U.S. EPA, Cincinnati
- Rowel WF, Barnes JW (1982) Obtaining the layout of water distribution systems. *J Hydraul Div ASCE* 108(1): 137–148

- Saleh HAS, Tanyimboh TT (2013) Coupled topology and pipe size optimization of water distribution systems. *Water Resour Manag*. doi:10.1007/s11269-013-0439-4
- Saleh S, Barlow E, Tanyimboh TT (2012) Unbiased and accurate assessment of surrogate measures of hydraulic reliability of water distribution systems. 14th Water Distribution Systems Analysis Conference, Adelaide, Australia, ISBN 978-1-922197-58-9, 148–157
- Saxena KS, Duro JA, Tiwari A, Deb K (2013) Objective reduction in many-objective optimization: linear and non-linear algorithms. *Trans Evol Comput* 17(1):77–99
- Shannon C (1948) A math. theory of communication. *Bell Syst Tech J* 27(3):379–428
- Siew C, Tanyimboh TT (2012) Penalty-free feasibility boundary convergent multi-objective evolutionary algorithm for the optimization of water distribution systems. *Water Resour Manag* 26(15):4485–4507
- Sinha A, Saxena DK, Deb K, Tiwari A (2013) Using objective reduction and interactive procedure to handle many-objective optimization problems. *Appl Soft Comput* 13(1):415–427
- Swamee PK, Sharma AK (2008) *Design of water supply pipe networks*. Wiley, New Jersey
- Tanyimboh TT, Sheahan C (2002) A maximum entropy based approach to the layout optimization of water distribution systems. *Civ Eng Environ Syst* 19(3):223–254
- Tanyimboh TT, Templeman AB (1993) Calculating maximum entropy flows in networks. *J Oper Res Soc* 44(4):383–396
- 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
- Todini E (2000) Looped water distribution networks design using a resilience index based approach. *Urban Water* 2(2):115–122
- Vaabel J, Ainola L, Koppel T (2006) Hydraulic power analysis for determination of characteristics of a water distribution system. *Proceedings of 8th Annual Water Distribution Systems Analysis Symposium, Cincinnati, Ohio, USA*
- Walters GA, Smith DK (1995) Evolutionary design algorithm for optimal layout of tree networks. *Eng Optim* 24(4):261–281
- Yassin-Kassab A, Templeman AB, Tanyimboh TT (1999) Calculating maximum entropy flows in multi-source, multi-demand networks. *Eng Optim* 31(6):695–729