Tanyimboh, T. T., & Czajkowska, A. M. (2021).
Entropy maximizing evolutionary design optimization of water distribution networks under multiple operating conditions.
Environment Systems and Decisions, 41(2), 267-285.
https://doi.org/10.1007/s10669-021-09807-1

Entropy Maximizing Evolutionary Design Optimization of Water Distribution Networks under Multiple Operating Conditions

Tiku T. Tanyimboh[a,b*] and Anna M. Czajkowska[b,c]

[a]School of Civil and Environmental Engineering, University of the Witwatersrand, Johannesburg, Private Bag 3, Wits 2050, South Africa

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

[c]RPS Group, Merchantile Building, 53 Bothwell Street, Glasgow G2 6TS, UK

Authors

1. Tiku T. Tanyimboh, PhD (Corresponding author)

   [a]School of Civil and Environmental Engineering, University of the Witwatersrand, Johannesburg, Private Bag 3, Wits 2050, South Africa

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

   Email: tiku.tanyimboh@wits.ac.za; Telephone: +27 (0) 11 717 7105
   ORCID ID: https://orcid.org/0000-0003-3741-7689

   [*]Corresponding author

2. Anna M. Czajkowska, PhD

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

   [c]RPS Group, Merchantile Building, 53 Bothwell Street, Glasgow G2 6TS, UK

   Email: anna.czajkowska.civeng@gmail.com

Entropy Maximizing Evolutionary Design Optimization of Water Distribution Networks under Multiple Operating Conditions

Abstract

The informational entropy model for flow networks was formulated over 30 years ago by Tanyimboh and Templeman (University of Liverpool, UK) for a single discrete operating condition that typically comprises the maximum daily demands and was undefined for water distribution networks (WDNs) under multiple operating conditions. Its extension to include multiple discrete operating conditions was investigated experimentally herein considering the relationships between flow entropy and hydraulic capacity reliability and redundancy. A novel penalty-free multi-objective genetic algorithm was developed to minimize the initial construction cost and maximize the flow entropy subject to the design constraints. Furthermore, optimized designs derived from the maximum daily demands as a single discrete operating condition were compared to those derived from a combination of discrete operating conditions. Optimized designs from a combination of discrete operating conditions outperformed those from a single operating condition in terms of performance and initial construction cost. The best results overall were achieved by maximizing the sum of the flow entropies of the discrete operating conditions. The logical inference from the results is that the flow entropy of multiple discrete operating conditions is the sum of their respective entropies. Also, a crucial property of the resulting flow entropy model is that it is bias free with respect to the individual operating conditions; hitherto a fundamental weakness concerning the practical application of the flow entropy model to WDNs is thus addressed.

Keywords: Maximum entropy; redundancy and reliability; penalty-free multi-objective genetic algorithm; infrastructure resilience; water resources planning; uncertainty-based design optimization

1. INTRODUCTION

There is a general expectation of a safe and continuous supply of water even though there is enormous ***uncertainty*** associated with the design, operation and maintenance of water supply and distribution networks due to such ***stochastic*** processes as random (as distinct from diurnal and seasonal) nodal demand variations (Almeida et al. 2013; Wang et al. 2016; Zhao and Chen 2008), fire-fighting flows and pipe and other component failures. Uncertainty arises, also, from ***imprecise or incomplete*** information and knowledge on long-term changes in pipe roughness (Sharp and Walski 1988), socio-economic and demographic changes and their effects on demands (Chang et al. 2012; Qi and Chang 2011; Sebri 2016; Tanyimboh and Kalungi 2008; Wang et al. 2009a, b).

There is, also, increasing awareness of the need to consider ***uncertainty and resilience*** along with ***cost effectiveness*** in water resources planning and design (Amarasinghe et al. 2016; Chang et al. 2012; Constantine et al. 2017; Harrison and Williams 2016; Herrera et al. 2016; Mitchell and McDonald 2015; Watts et al. 2012; Wright et al. 2015; Yazdani et al. 2011). In the context of water distribution networks (WDNs), ***resilience*** relates to the ability to avoid service interruptions and the extent to which reasonably acceptable levels of service can be expected under both normal and unexpected operating conditions. Resilience is characterised by redundancy, failure tolerance and reliability. It is strongly related to components failure tolerance that, in turn, depends on the system's redundancy in the form of spare flow capacity in the individual components and multiple independent supply paths from the sources to the demand nodes. Therefore, it stems from the topology of the network, pipe diameters, reliability and availability of individual links and ultimately the network's flow re-routing properties (Gheisi and Naser 2015; Saleh and Tanyimboh 2016; Tanyimboh and Templeman 1994, 2000; Tanyimboh 2017).

However, it is computationally challenging to evaluate the ***reliability*** of water distribution networks, as it requires the simulation of such stochastic events as random pipe failures and demand fluctuations (Wagner et al. 1988; Xu and Goulter 1998). It is even more challenging to incorporate accurate measures of reliability in ***design optimization*** processes for water distribution networks, as the computational requirements are prohibitive (Eiger et al. 1994; Xu and Goulter 1999; Yates et al. 1984). For example, Marchi et al. (2014) explained the difficulties that arise from high dimensionality and computational complexity. They observed that the difficulties were addressed variously through engineering experience; parallel computing; reducing the number of decision variables; reducing the range of possible values for each decision variable; skeletonizing the network; sequential optimization; zone-by-zone design; and stage-wise optimization. High performance computing has been deployed also (Barlow and Tanyimboh 2014; Tanyimboh and Seyoum 2016).

Reliability-based design optimization of water distribution networks is particularly challenging because it addresses a complex, highly constrained, nonconvex, nonlinear, computationally intensive optimization problem with numerous decision variables, multiple conflicting objectives and probabilistic constraints. Evolutionary optimization algorithms are frequently used as a result, as they deploy populations of candidate solutions in the search for the globally optimal solutions. Consequently, the computational complexity increases enormously due to the need to evaluate each candidate solution accurately.

This is the reason that ***surrogate performance indicators*** (Forrester et al. 2008) such as ***flow entropy*** and resilience indices have been used widely (Atkinson et al. 2014; Gheisi and Naser 2015; Greco et al. 2012; Saleh and Tanyimboh 2014, 2016; Prasad and Park 2004; Baños et al. 2011; Jayaram and Srinivasan 2008; Liu et al. 2014, 2016; Recca et al. 2008; Singh and Oh 2015; Todini 2000; Lehký et al. 2017), wherein the system response that requires complex simulations is replaced by an approximation that is much faster to evaluate (Díaz et al. 2016).

Results in the literature show that ***flow entropy*** is suitable for consideration as a surrogate measure of resilience in water distribution networks (Awumah et al. 1990, 1991; Awumah and Goulter 1992; Tanyimboh Templeman 1993a). Flow entropy stems from informational entropy (Shannon 1948), a measure of the amount of uncertainty that a probability distribution represents (Zhao and Zhang 2011). It reflects the respective flow capacities and number of alternative supply paths between the source nodes and demand nodes (Awumah et al. 1990, 1991; Awumah and Goulter 1992; Ang and Jowitt 2005a, b; Tanyimboh and Templeman 1993b; Yassin-Kassab et al. 1999). Consequently, the flow entropy value increases if there are more supply paths and/or the uniformity of the flow capacities of the supply paths increases (Saleh and Tanyimboh 2014, 2016; Tanyimboh and Sheahan 2002).

Thus, ***flow entropy*** has been tested on different water distribution networks for over three decades (Atkinson et al. 2014; Gheisi and Naser 2015; Greco et al. 2012; Raad et al. 2010; Saleh and Tanyimboh 2014, 2016). Furthermore, recent studies have shown that flow entropy yields the most consistent results among all the surrogate performance measures for water distribution networks (Gheisi and Naser 2015; Saleh and Tanyimboh 2016; Liu et al. 2016; Tanyimboh et al. 2011, 2016).

However, previous research on flow entropy has focussed on networks with a ***single operating condition*** (SOC), essentially a steady-state condition in which it is assumed that the nodal demands are constant. While it is common in the literature to design water distribution networks considering the peak demands only (Kadu et al. 2008; Prasad and Park 2004), the demands generally follow a diurnal pattern and ***multiple loading conditions*** must be satisfied equally (Alperovits and Shamir 1977).

A novel entropy maximizing approach to the design optimization of water distribution networks (WDNs) is proposed here. Maximum entropy designs of WDNs counter

uncertainty and improve ***resilience*** as they transcend the design constraints by being maximally non-committal to the totality of the feasible operating conditions (Tanyimboh and Templeman 1993a, 2000; Zhao and Zhang 2011). The methodology is broadly comparable to the applications of the maximum entropy formalism in environmental management and decision making under conditions of uncertainty or incomplete information (Zhao and Zhang 2011; Ainslie et al. 2009; Gurupur et al. 2014; Hao et al. 2010; Kaplan et al. 2003; Lind 1997). Specifically, this article describes the development and assessment of a novel multi-objective evolutionary design optimization model for water distribution networks that maximizes the flow entropy under multiple discrete operating conditions. The algorithm developed was applied to networks in the literature and its effectiveness demonstrated.

2. FLOW ENTROPY FUNCTION FOR SINGLE OPERATING CONDITION

The flow entropy function (Tanyimboh 1993; Tanyimboh and Templeman 1993a-d) has been investigated extensively in the literature (Atkinson et al. 2014; Gheisi and Naser 2015; Saleh and Tanyimboh 2016; Liu et al. 2014; Tanyimboh et al. 2016). For the *k*th discrete operating condition, the flow entropy function may be expressed as in Eqs. 1 to 3.

$$S_k = S_{0,k} + \sum_{n=1}^{nn} p_{n,k} S_{n,k} \;;\forall k \qquad (1)$$

For the *k*th operating condition, S_k is the entropy; $S_{0,k}$ is the entropy due to the relative flow contributions from the supply nodes; $S_{n,k}$ is the entropy at node *n*; $p_{n,k} = T_{n,k}/T_k$ is the fraction of the total flow through the network that reaches node *n*; $T_{n,k}$ is the total flow that reaches node *n*; T_k is the sum of the nodal demands; and *nn* is the number of nodes in the network. Eq. 1 is predicated on flow continuity at the demand nodes and pipe junctions. Consequently, the total flow through the network, T_k, is also equal to the total flow from the supply nodes.

The entropy due to the flows from the supply nodes, i.e. their relative contributions, for the *k*th operating condition is

$$S_{0,k} = -\sum_{n \in I_k} \frac{Qn_{0n,k}}{T_k} \ln\left(\frac{Qn_{0n,k}}{T_k}\right); \quad \forall k \tag{2}$$

where $Qn_{0n,k}$ is the flow from supply node *n;* and I_k represents the set of supply nodes.

Similarly, the entropy at demand node *n* for the *k*th operating condition is

$$S_{n,k} = -\frac{Qn_{n0,k}}{T_{n,k}} \ln\left(\frac{Qn_{n0,k}}{T_{n,k}}\right) - \sum_{ij \in ND_{n,k}} \frac{Qp_{ij,k}}{T_{n,k}} \ln\left(\frac{Qp_{ij,k}}{T_{n,k}}\right); \quad n = 1, ..., nn; \ \forall k \tag{3}$$

$Qn_{n0,k}$ is the demand at node *n* and $Qp_{ij,k}$ the volume flow rate in pipe *ij* with nodes *i* and *j* as the upstream and downstream nodes, respectively. Set $ND_{n,k}$ represents the pipe flows from node *n.* $T_{n,k}$ is the total flow that reaches node *n.* It is also equal to the sum of the outflows from node *n* including any nodal demand, due to the continuity of the flows at the pipe junctions and demand nodes. Eqs. 1 to 3 describe a multi-probability space model the details of which are available in Tanyimboh (1993) and Tanyimboh and Templeman (1993a-d).

3. FORMULATION OF THE EVOLUTIONARY DESIGN OPTIMIZATION MODEL

A challenge that is associated with evolutionary algorithms in general is their poor ability to handle constraints directly. Most practical problems generally involve constraints that are usually addressed by introducing penalties that degrade infeasible solutions (Phan et al. 2013). There is a risk, however, that if poorly defined the penalties could hinder the search capabilities and so lead to suboptimal solutions (Woldesenbet et al. 2009). More fundamentally, the formulation of constraint violation penalty functions is extremely challenging (Coello Coello 2002; Saleh and Tanyimboh 2014; Siew et al. 2014). According to Siew et al. (2014), *"Dridi et al. (2008) observed that the results obtained are highly dependent on the penalty coefficients used, and user-specified constraint-violation penalties are not practical enough."* Also, constraint dominance tournaments (Deb et al. 2002) are used frequently as an alternative to penalty functions. They are relatively easy to apply but typically they favour the feasible solutions at the expense of infeasible solutions, which constitutes a weakness (Siew et al. 2014, Sheikholeslami and Talatahari 2016).

Furthermore, it has been established that evolutionary algorithms that exploit the information content of the infeasible solutions generated during the search generally outperform those that do not (Eskandar et al. 2012; Siew et al. 2014, 2016; Saleh and Tanyimboh 2013, 2014; Woldesenbet et al. 2009; Yang and Soh 1997). Tanyimboh and Seyoum (2016) observed that a constraint handling approach that promotes the development and crossbreeding of subpopulations of nondominated feasible and infeasible solutions provides a practical, seamless and effective boundary search mechanism. This is a crucial property whose significance is that the optimal solutions are generally located near the active feasibility constraint boundaries (Wu and Walski 2005). Also, the approach promotes diversity in the gene pool (Sheikholeslami and Talatahari 2016) by maintaining and exploiting the full spectrum of non-dominated infeasible solutions in every generation, a feature that helps to avoid premature convergence.

Therefore, to improve the computational efficiency and accuracy of the optimization results achieved, constraint violation penalties or constraint dominance tournaments were not used herein. The approach adopted is practical and has the added advantage that, besides the underlying GA parameters (e.g. mutation rate, crossover probability, etc.), additional parameters that require case-by-case calibration and time consuming trial runs are not introduced (Tanyimboh and Seyoum 2016; Moosavian and Lence 2017).

The objectives considered in the design optimization model developed were: (a) minimization of the initial construction cost; (b) maximization of flow entropy; and (c) minimization of any deficits in the residual pressures at the critical demand nodes. The critical demand node is the demand node with the largest shortfall in pressure. As the nodal demands vary spatially and temporally, the location of the critical demand node is dynamic. In other words, its location may vary from one hour to the next and from one design to the next, depending on the respective diurnal demand patterns of the nodes and the configurations of the various network components. Consequently, in general, it cannot be predicted accurately before completing the design, followed by detailed and extensive simulations.

The design constraints were the minimum residual pressure constraints at the demand nodes in addition to the conservation of mass and energy. The minimum nodal residual pressure constraints were incorporated as a ***single objective function*** and satisfied by reducing the residual node pressure deficits to zero (Eq. 5). In this way the minimum nodal residual pressure constraints were addressed seamlessly and efficiently. Siew and Tanyimboh (2012) similarly addressed the minimum node pressure constraints by maximizing the demand satisfaction ratio (DSR) as a single objective function, using pressure-driven analysis (He et al. 2016). The DSR takes values between zero and unity and is the fraction of the demand that is satisfied at adequate pressure (Tanyimboh et al. 2003). Barlow and Tanyimboh (2014) minimized the sum of the node pressure deficits, this being a single objective function.

Saleh and Tanyimboh (2013) included additional topological constraints that defined the reachability and redundancy of the demand nodes. The reachability constraint, in the topological optimization problem, was to avoid and account for possible node isolation. The redundancy constraint was to avoid and account for the absence of loops. The topological and minimum node pressure constraints were formulated as a single combined deficit objective function that was minimized. By contrast, Khu and Keedwell (2005) represented the minimum pressure constraint of each demand node as a separate objective function that minimized the nodal pressure deficit. This is not practical given the generally large number of demand nodes.

Siew and Tanyimboh (2012) demonstrated that faster convergence was achieved by minimizing the shortfall in flow and pressure or DSR at the critical node as opposed to the entire network, which is equivalent to the sum or average of the node pressure deficits. Eq. (5) minimizes the largest nodal head deficit accordingly.

The nodal mass balance and energy conservation equations were satisfied herein in the hydraulic simulation model, this being the EPANET 2 hydraulic solver (Rossman 2000). The decision variables of the optimization problem were the discrete pipe diameters, which were selected from a set of commercially available pipe diameters. Accordingly, the optimization problem may be summarized as follows:

Minimize the initial construction cost $f_1 = \sum_{i}^{np} C_i(d_i, L_i)$ (4)

Minimize the largest nodal head deficit $f_2 = Max\left\langle \max\left[0, (H_n^{req} - H_n)\right];\ \forall n \right\rangle$ (5)

Maximize the flow entropy $f_3 = S(k; \forall k)$ (6)

Subject to: $d_i \in D,\ \forall i;$ and $\mathbf{g}(\mathbf{Qp}, \mathbf{H}) = \mathbf{0}$ (7)

$C_i\ (d_i, L_i)$ is the cost of pipe i with diameter d_i and length L_i while np is the number of pipes in the network. The set D comprises the available discrete pipe diameter options. $S(k; \forall k)$ is the flow entropy considering the various operating conditions k. The entropy is essentially a function of the pipe flow rates, which were obtained from the hydraulic simulation model used, EPANET 2, which ensured energy and flow conservation, and provided the nodal head values. H_n and H_n^{req} are the available and minimum required heads, respectively, at node n. The minimum required head corresponds to the residual pressure above which the demand is satisfied in full. $\mathbf{g}(\mathbf{Qp}, \mathbf{H}) = \mathbf{0}$ represents the conservation of mass and energy system of equations, where **Qp** and **H** are the vectors of the pipe flow rates and nodal heads, respectively.

3.1 Flow Entropy under Multiple Operating Conditions

The fundamental issue addressed in the present investigation is that the flow entropy function in the literature was formulated considering only one operating condition, i.e. the maximum daily demands typically. Flow entropy was, thus, undefined for water distribution networks under multiple operating conditions. However, Alperovits and Shamir (1977) stated that when designing a water distribution network, the minimum and maximum daily demands and fire-fighting flows should be considered. Indeed, Prasad (2010) demonstrated that even if a network satisfies the maximum daily demands, it does not follow that other operating conditions will be feasible. Prasad (2010) also demonstrated that designs obtained using multiple operating conditions (MOCs) (Morgan and Goulter 1985; Walski et al. 1987) are more resilient than those based on a single operating condition.

Accordingly, the most appropriate formulation of the flow entropy function under multiple operating conditions was investigated experimentally herein. The three competing hypotheses with different interpretations of the maximum entropy formalism (Jaynes 1957) as characterised below were tested, viz:

(a) Maximize the maximum entropy value from all the operating conditions.

(b) Maximize the minimum entropy value from all the operating conditions.

(c) Maximize the total entropy value from all the operating conditions.

3.1.1 Maximizing the Maximum Entropy

The criterion used for maximizing the maximum entropy is the highest entropy value among all the entropy values achieved by a single candidate solution, considering all the operating conditions. Accordingly, the genetic algorithm seeks the highest entropy value possible, irrespective of the entropy values of the other operating conditions. In other words, the algorithm tries to find solutions that are very good, based on the entropy measure, for at least one operating condition. A solution with a very high entropy value for one operating condition and poor values for other operating conditions would be deemed superior to another solution with a slightly lower maximum entropy value, even if the entropy values are quite high for all the operating conditions. This criterion has the potential to overestimate the ***overall performance*** of a solution and, consequently, yield suboptimal results. The approach may be represented as follows.

$$\text{Maximize } f = Max(S_k, \forall k) \tag{8}$$

where S_k is the entropy of the kth operating condition (Eq. 1).

3.1.2 Maximizing the Minimum Entropy

Maximizing the minimum entropy, on the other hand, ensures that the entropy value for any operating condition would not be excessively low, by alleviating the worst case. However, it has the disadvantage that it ignores the operating conditions with higher entropy values than the worst case. Therefore, it would likely underestimate the overall performance of a solution and, consequently, yield suboptimal and/or inconsistent results. The approach may be summarized as follows.

$$\text{Maximize } f = Min(S_k, \forall k) \tag{9}$$

where S_k is the entropy of the kth operating condition.

3.1.3 Maximizing the Total Entropy

Maximizing the total entropy considers all the operating conditions simultaneously. Hence the resulting solutions can be expected to perform reasonably well in all the operating conditions. The total entropy value is obtained by adding together the entropy values of all the operating conditions considered and is maximized as follows.

$$\text{Maximize } f = \sum_k S_k \tag{10}$$

where S_k is the entropy of the kth operating condition.

4. COMPUTATIONAL SOLUTION METHODOLOGY

4.1 Penalty-free Multi-objective Genetic Algorithm

The genetic algorithm developed is a modified version of the elitist non-dominated sorting genetic algorithm (NSGA) II (Deb et al. (2002). NSGA II is efficient and used widely by many researchers in various disciplines (Alirezaei et al. 2019; Avci and Selim 2017; Lahsasna and Seng 2017; Sanodiya et al. 2019; Soui et al. 2019; Wang et al. 2018). It maintains diversity by seeking an even distribution of the non-dominated solutions in the objective space using the crowding distance (Deb et al. 2002), a measure of the spatial density of the solutions in the objective space, based on the average distance between a solution and its nearest neighbours.

The methodology developed herein is generic and can handle ***any number of discrete operating conditions for any network configuration.*** It uses Pareto-dominance sorting (Deb et al. 2002) but unlike NSGA II, the constraint dominance concept is not invoked. Instead, Pareto-dominance is applied strictly with respect to the objective functions $(f_1, f_2, f_3)^T$ only. The genetic algorithm is thus characterised as penalty free, which means the propagation of infeasible solutions is not impeded by any additional constraint-based penalties or criteria (Siew and Tanyimboh 2012; Saleh and Tanyimboh 2014; Tanyimboh and Seyoum 2016).

The evidence in the literature shows that algorithms that exploit the information content of any infeasible solutions generated generally outperform those that do not (Eskandar et al. 2012; Woldesenbet et al. 2009; Yang and Soh 1997). For example, Tanyimboh and colleagues (Siew et al. 2014, 2016; Saleh and Tanyimboh 2013, 2014; Barlow and Tanyimboh 2014) achieved numerous new best-known solutions for ***nine*** different benchmark optimization problems in the literature. Some of the optimization problems are as follows. The Balerma network (Reca and Matinez 2006) has 443 demand nodes, four supply nodes and 454 pipes. The Anytown network (Walski et al. 1987) involves capacity expansion, rehabilitation, upgrading, tanks siting and design, pumps design and scheduling, five operating conditions and extended period simulation. Also, new best solutions were achieved for the Kadu et al. (2008) network, and the least cost was improved by 4.72%.

Network design problems that optimize both the topology and components are extremely complex due to extra layers of difficulty. Saleh and Tanyimboh (2013) achieved two new best ***topologies*** and six new best ***designs*** in total, for three benchmark problems. Moreover, the novel penalty-free approach in Saleh and Tanyimboh (2014) yielded a ***complete spectrum*** of optimal and near-optimal ***designs and topologies*** for each benchmark problem considered, an unprecedented achievement. All the new best solutions achieved were feasible.

Thus, the source code of NSGA II in C++ was modified and used with the hydraulic simulation software for water distribution systems EPANET 2. A procedure that calculates the flow entropy for any given network topology was developed, tested and incorporated in the optimization algorithm. The penalty-free multi-objective genetic algorithm thus developed is shown diagrammatically in Figure 1.

It was observed that the number of infeasible solutions exceeded the feasible solutions in the Pareto-optimal sets achieved. The nodal pressure deficit function f_2 in Eq. 5 does not distinguish between solutions with different levels of surplus pressure, as all feasible solutions have a pressure deficit of zero. Hence, for *feasible* solutions, the Pareto-dominance is based on the initial construction cost and flow entropy only (f_1 and f_3). By contrast, for *infeasible* solutions, the Pareto-dominance is based on all three objectives (f_1, f_2 and f_3). In general, more solutions become non-dominated as the number of objectives increases (Ishibuchi et al. 2015). This is likely the reason for the greater number of infeasible solutions in the Pareto-optimal sets achieved. On discarding the non-dominated infeasible solutions at the end of the optimization, all the *non-dominated feasible* solutions achieved from the start to the end of the optimization were selected, based on Pareto dominance, using software developed herein for this purpose in the Perl language (Figure 2). All the solutions presented herein were derived in this way.

Furthermore, for each instance of the various design optimization problems considered, the non-dominated solutions from all the optimization runs were combined and sorted, based on Pareto dominance, to obtain the final cost vs. entropy non-dominated set of solutions.

Also, following the optimization, the final non-dominated solutions obtained from the above-mentioned union of the non-dominated sets were analysed further to determine their reliability and failure tolerance values as described in the next subsection. An outline of the overall procedure for the optimization and performance assessment is provided in Figure 2.

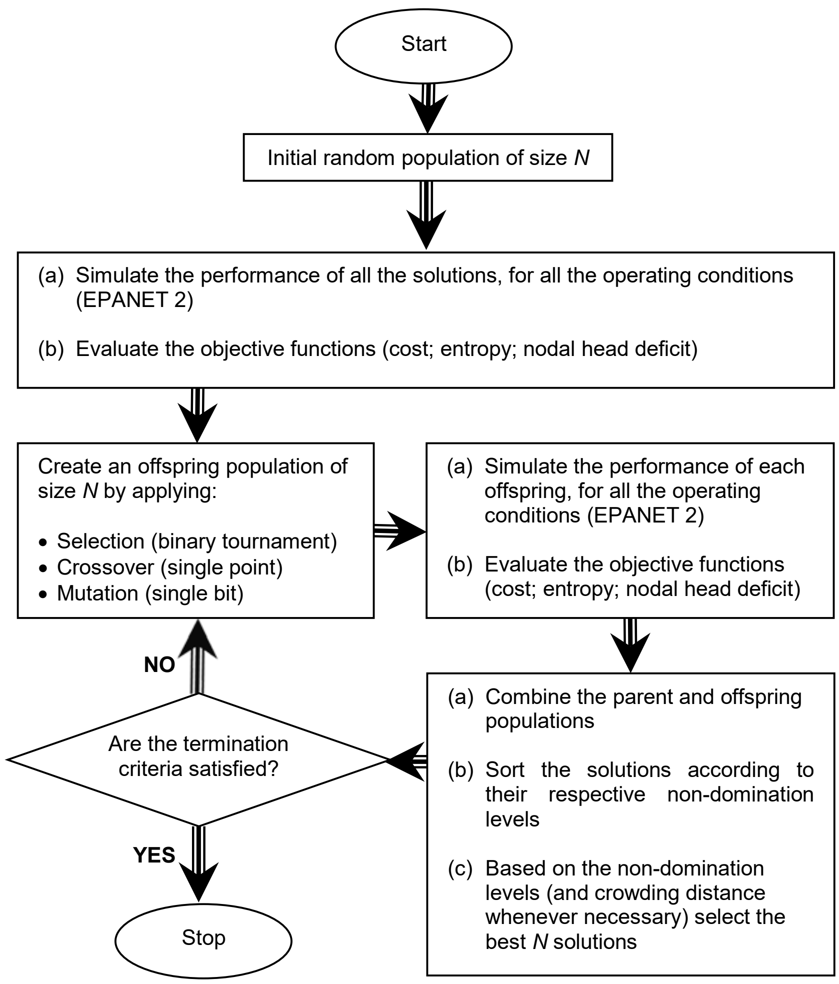


Fig. 1 Penalty-free multi-objective genetic algorithm procedure

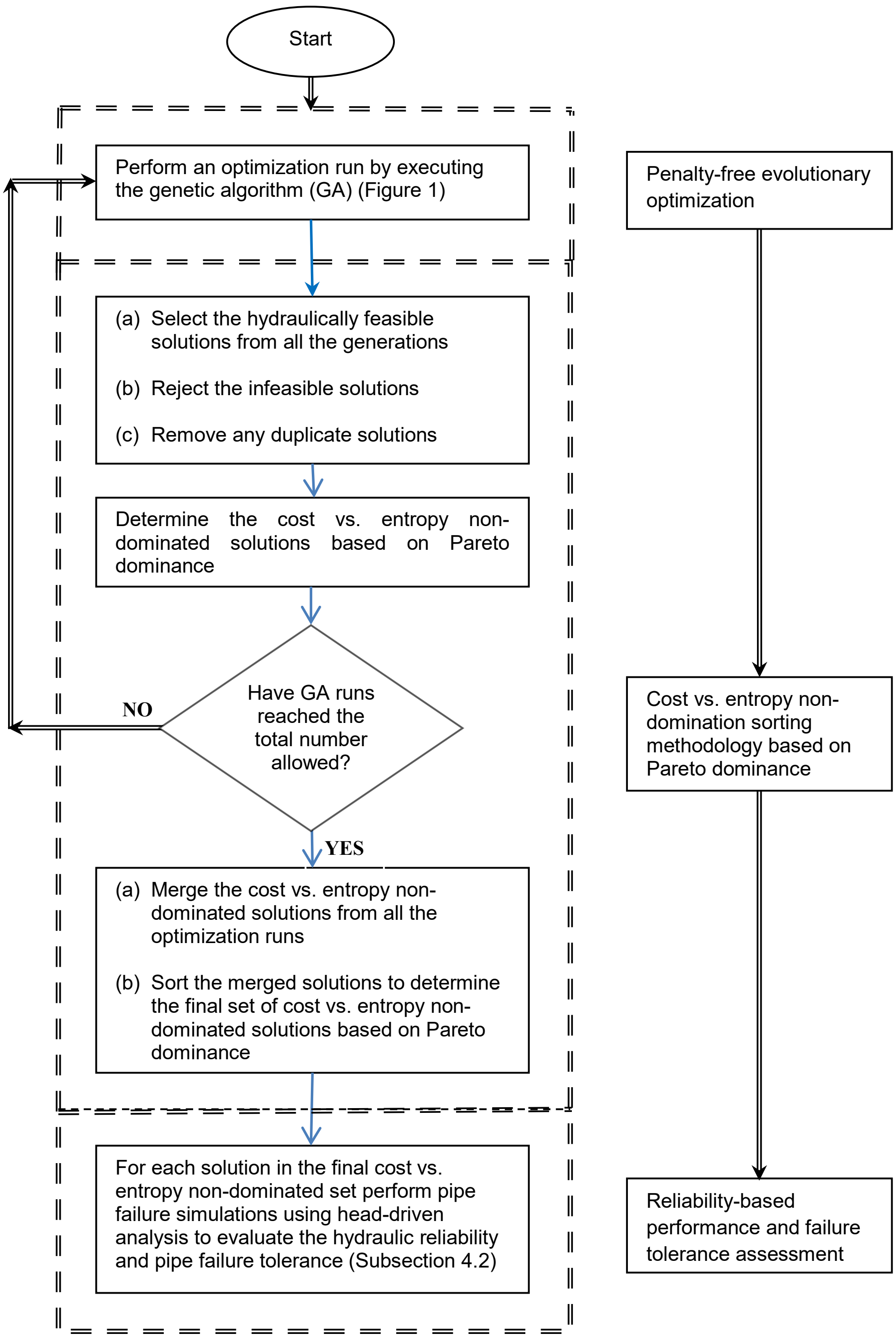


Fig. 2 Optimization and performance evaluation procedure

4.2 Post-Optimization Performance Evaluation

The solutions in the final cost vs. entropy non-dominated sets achieved were analysed to determine their reliability and failure tolerance values (Figure 2). Pressure-driven analysis (He et al. 2016; Rasekh and Brumbelow 2014) was carried out to simulate the effects of pipe failures or closures using PRAAWDS (program for the realistic analysis of the availability of water in distribution systems) (Tanyimboh et al. 2003; Tanyimboh and Templeman 2010). The program has been tested and used extensively over many years (e.g. Saleh and Tanyimboh 2016; Tanyimboh et al. 2016). It has the four pressure-dependent nodal discharge functions proposed by (a) Wagner et al. 1988; (b) Germanopoulos (1985) as amended by Gupta and Bhave (1996); (c) Fujiwara and Ganesharajah (1993); and (d) Tanyimboh and Templeman (2010). The logistic pressure dependent nodal discharge function in Tanyimboh and Templeman (2010) was adopted herein. In addition to its superior computational properties (Kovalenko et al. 2014), Vairagrade et al. (2015) and Ciaponi et al. (2015) have demonstrated that it is the most accurate approximation of the nodal pressure-discharge relationship.

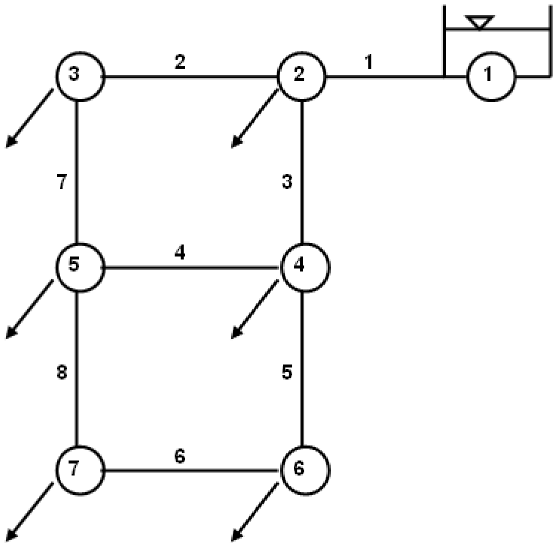
The reliability was calculated as in Tanyimboh and Templeman (2000) and Tanyimboh and Sheahan (2002). The probabilistic pipe failure model in Cullinane et al. (1992) was used to estimate the pipe failure rates. It was recommended in Tanyimboh et al. (2011) following an appraisal of the available alternatives in the literature, where it was also observed that the model in Khomsi et al. (1996) gave similar results. It may be noted, for example, that the pipe failure model put forward in Xu et al. (2018) is not valid for pipe diameters that are greater than 300 mm.

The reliability of WDNs is frequently defined as the mean value of the fraction of the required flow that is satisfied at adequate pressure, considering both normal and abnormal operating conditions (Tanyimboh and Templeman 2000). Failure tolerance is a complementary property that provides an estimate of the fraction of the total demand that the network can satisfy on average, with one or more components out of service (Tanyimboh and Templeman 2000). To calculate the reliability, the *minimum* number of pressure-driven analysis simulations required for each candidate solution (or network) under consideration is $np+1$, where np is the number of pipes (Tanyimboh and Templeman 2000). Generally, the procedure is highly computationally expensive, especially when dealing with many Pareto sets, where each set may have hundreds of solutions or more, as in the present investigation.

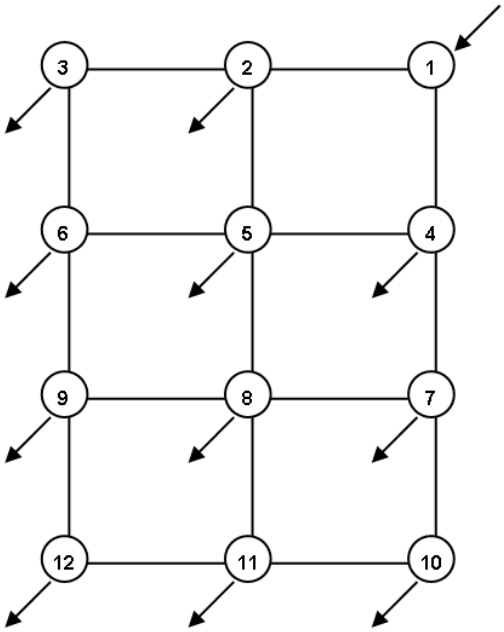
5. RESULTS

The methodology developed was assessed using several networks in the literature (Figure 3). Experiments were performed to determine suitable values for the genetic algorithm (GA) parameters including the population size, mutation rate and crossover probability. In addition, solutions based on (a) the peak demands only and (b) the peak demands plus additional operating conditions were compared.

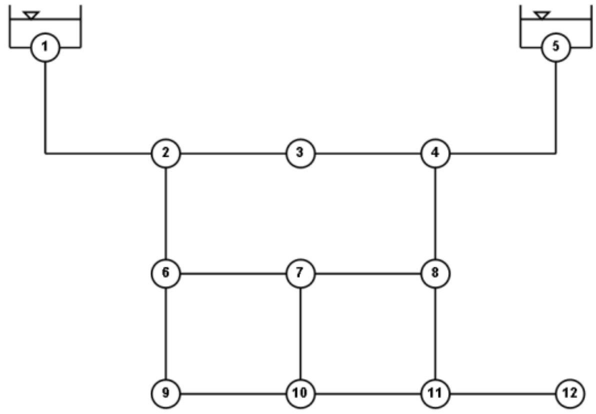
In the examples that follow, unless otherwise stated, the values of the parameters of the genetic algorithm were as follows. N_R = 30; N_E = 200,000; N_G = 1,000; N_S = 200; $p_m = 1/n_g$; and p_c = 1.0. The symbols represent the number of executions of the optimization algorithm (N_R); number of function evaluation allowed (N_E); number of generations (N_G); population size (N_S); mutation (p_m) and crossover (p_c) probabilities; and chromosome length (n_g). The results achieved suggest the values chosen were satisfactory with respect to the quality and consistency of the solutions. Any redundant binary codes (Saleh and Tanyimboh 2014) were allocated evenly (Czajkowska 2016) among the candidate pipe diameters to minimize the representational bias. A single-point crossover was used to produce two offspring from two parents.



(a)



(b)



(c)

Pipe Diameter (mm)	Unit Cost ($/m)
152	49.5
203	63.3
254	94.8
305	132.9
356	170.9
406	194.9
457	231.3
508	262.5

(d)

Fig. 3 Sample networks investigated. (a) Network 1 (b) Network 2 (c) Network 3 (d) Pipe diameter options for Network 3

A PC (personal computer) (Intel Core i3, 2.4 GHz, 3.0 GB RAM and Windows XP operating system) was used. Demand-driven simulation (EPANET 2) was used in the optimization. The performance assessment of the solutions in the final Pareto-optimal sets achieved employed pressure-driven simulation (PRAAWDS) (Tanyimboh and Templeman 2010). In total, 453 evaluations of reliability plus failure tolerance were performed for the networks in Figure 3. The total number of pressure-driven simulations was 6,690; i.e. 954, 3,186 and 2,550 for Network 1, 2 and 3, respectively.

5.1 Resilience Improvement due to Multiple Loading Conditions

This example is based on Network 1 in Figure 3a and Table 1. The available data in the literature had only one loading condition, and so two additional loading conditions were generated using demand multipliers from the literature (Walski et al. 1987; Surendran et al. 2005). Thus, the nodal demands in the literature were taken as the peak demands. The demand multiplier for the average demand was taken as 0.8. The average demand was then multiplied by 0.6 to obtain the minimum demand. The demand multipliers were applied uniformly to the demand nodes. The GA parameters were as follows: N_E = 200,000; N_S = 200; p_c = 1.0; and p_m = 0.03125. Four instances of the design optimization problem were considered, with N_R = 10, i.e. 40 GA runs in total, or 8 million function evaluations.

The entropy values for the total entropy maximization model were approximately three times the values of the other three cases, because they were the sums of the entropy values for three operating conditions. Thus, to facilitate the comparisons, and without loss of generality (Shannon 1948; Tanyimboh and Templeman 1993a-d), the total entropy values were divided by 3 before any subsequent analyses. The average CPU (central processing unit) time for one optimization run was approximately 11 minutes for a design with one operating condition and approximately 23 minutes for a design with three operating conditions.

Figure 4a illustrates the Pareto-optimal fronts of cost versus entropy for the single operating condition (SOC) and three different maximum entropy models for multiple operating conditions (MOCs). Figure 4b presents the same results, but only up to the point beyond which the entropy improvements are insignificant. There are many solutions with negligible increases in entropy and high increases in cost. For this network, the point at which the improvements in entropy become insignificant is, approximately, around 99% of the maximum entropy value achieved.

Table 1 Node data and pipe diameter options for Network 1

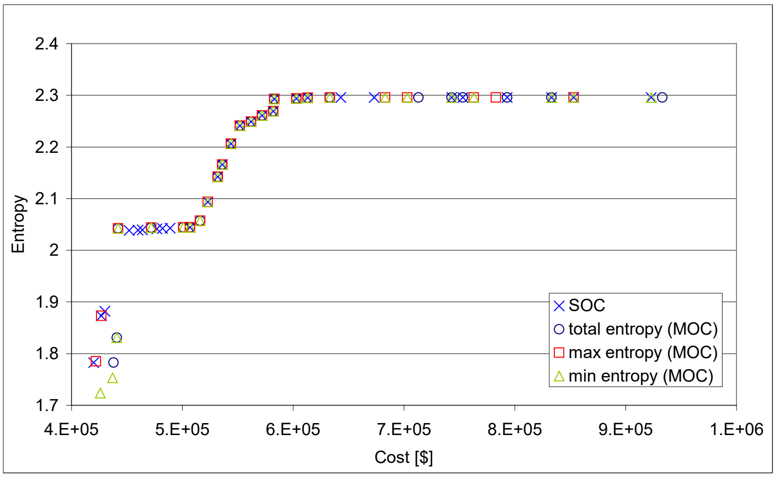
Nodes	Elevations (m)	Demands (l/s)	Required heads (m)	Diameters (mm)	Unit costs ($/m)	Diameters (mm)	Unit costs ($/m)
1	[a]210	-	-	25.4	2	304.8	50
2	150	27.78	30	50.8	5	355.6	60
3	160	27.78	30	76.2	8	406.4	90
4	155	33.33	30	101.6	11	457.2	130
5	150	75.00	30	152.4	16	508.0	170
6	165	91.66	30	203.2	23	558.8	300
7	160	55.56	30	254.0	32	609.6	350

[a]210 m is the total head at the supply node (reservoir).

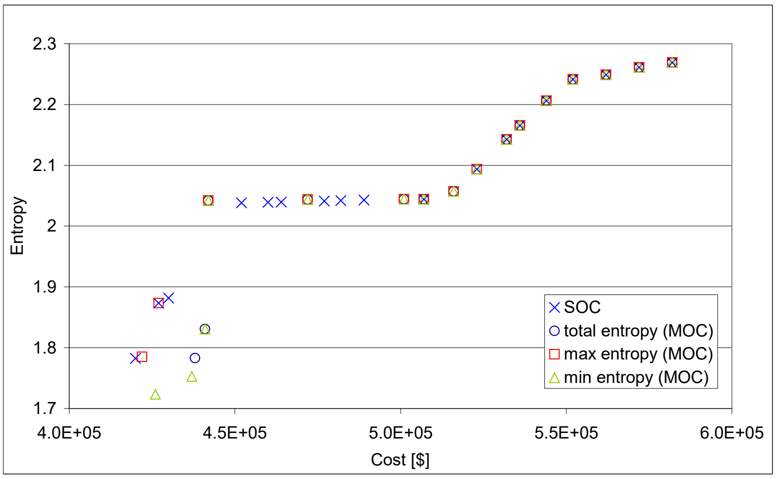
Figure 4a shows clearly that the top 1% of the entropy values achieved account for more than 60% of the total increase in cost relative to the least expensive feasible solution. Furthermore, the evidence in the literature shows that optimized solutions with similar maximum entropy values have similar hydraulic and pipe failure properties (Tanyimboh and Sheahan 2002; Tanyimboh and Setiadi 2008; Tanyimboh et al. 2011, 2016). In other words, it is reasonable to assume that the top 1% of the solutions achieved are virtually functionally the same, except for the differences in cost due to the larger pipe diameters that are not accompanied by any fundamental improvement in the flow distribution. Thus, the excessive increases in cost are not accompanied by any substantial performance improvements.

It can be observed also that the MOC and SOC solutions have comparable entropy values. Only five solutions in total (i.e. less than 5% approximately) (from the minimum and total entropy maximization approaches) are dominated by being located slightly below the common Pareto front.

Table 2 shows the strengths of the relationships between entropy, the distribution of the pipe diameters and reliability. The correlation coefficients are higher for the solutions up to 99% of the maximum entropy values achieved, except for the mean diameter in the case of the minimum entropy maximization model. The solutions with entropy values above 99% of the maximum entropy values achieved are essentially superfluous for practical purposes. Therefore, hereafter the presentation focuses on the solutions up to the assumed entropy cut-off point of 99%.



(a) Complete Pareto-optimal fronts



(b) Truncated Pareto-optimal fronts

Fig. 4 Pareto-optimal solutions for Network 1 based on ten optimization runs

Table 2 Correlation coefficients for selected performance indicators for Network 1

Relationships	Entropy range	Correlation Coefficients (R^2) SOC	Total entropy	Maximum entropy	Minimum entropy
Mean diameter vs. entropy	0-100%	0.842	0.794	0.835	0.966
	0-99%	0.893	0.899	0.904	0.918
Pipe size distribution vs. entropy	0-100%	0.663	0.714	0.674	0.773
	0-99%	0.874	0.892	0.875	0.922
Hydraulic reliability vs. entropy	0-100%	0.813	0.821	0.819	0.826
	0-99%	0.885	0.890	0.892	0.870

The range from 0 to 99% includes solutions up to 99% of the maximum entropy value achieved. The pipe size distribution parameter used was the coefficient of variation. The shading and boxes highlight the largest and smallest value in each row, respectively.

The methodology developed also solves the SOC design optimization problems as this example shows. As the pipe lengths are equal, the mean and coefficient of variation of the diameters are useful indicators. They show that, on average, the diameters become more uniform and larger as the flow entropy increases. Therefore, the correlation between flow entropy and resilience is expected to be positive. Strong positive correlation between entropy and reliability was observed for the four maximum entropy design approaches. The results demonstrate that the MOC solutions outperform the SOC solutions. Moreover, in general the SOC solutions achieved may be infeasible if additional loading conditions besides the peak demands are not included in the design specifications. Thus, SOC solutions are not considered hereafter.

The network considered in this example is peculiar in the sense that it is severely limited in terms of redundancy from the perspective of alternative flow paths. Therefore, as this network has very few alternative flow paths and very limited capacity for flow re-routing, the option of maximizing the minimum entropy seems to provide the best results.

5.2 Loading Conditions with Uniform Demand Multipliers

While Network 2 (Figure 3b) (Czajkowska and Tanyimboh 2013) does not reflect the scale and complexity of water distribution networks in general, it has been used extensively in previous flow entropy studies including Awumah *et al.* (1990, 1991) and has a considerable number of loops that offer many possibilities for alternative flow paths.

It has a single supply node with a head of 100 m. The elevation and minimum residual pressure required at the demand nodes are 0 m and 30 m, respectively. The pipes are 1000 m long with a Hazen-Williams roughness coefficient of 130. The pipe diameter options in mm are 100, 125, 150, 200, 250, 300, 350, 400, 450, 500, 550 and 600. The pipe cost per metre is $f(d) = \gamma d^{1.5}$, where d is the diameter in metres and γ = £800. The peak demands are as follows (node number and demand in litres/s): {(2, 27.8), (3, 41.7), (4, 41.7), (5, 41.7), (6, 27.8), (7, 55.5), (8, 55.5), (9, 55.5), (10, 27.8), (11, 41.7), (12, 27.8)}.

With 17 pipes and 12 pipe diameter options, the solution space has 12^{17} or 2.218×10^{18} feasible and infeasible solutions. A 4-bit binary substring was used. With $2^4 = 16$ substrings, four substrings were redundant. The four redundant substrings were allocated evenly to minimize the representational bias by doubling the 125, 250, 400 and 550 mm diameters (Czajkowska 2016). The GA parameters were $p_m = 1/n_g = 1/68 = 0.0147$; $p_c = 1.0$; $Ns = 200$; $N_E = 200{,}000$; $N_G = 1{,}000$; and $N_R = 30$, in each instance of the design optimization problem. There were thus 90 GA runs in total for the three alternative flow entropy maximization approaches with 18 million function evaluations or hydraulic simulations. The mean CPU time for a single optimization run was approximately 28 minutes.

The original nodal demands in the literature were taken as the peak demands and used to calculate the average and minimum demands. Due to the relatively small size of the network, identical demand multipliers were used for all the nodes. The entropy values achieved by adding together the entropy values of the minimum, average and peak loading conditions were divided by 3, as explained previously in Subsection 5.1.

Figure 5 shows the non-dominated solutions and relationship between entropy and mean pipe diameter. Notice the improvements in the entropy become negligible around a cost of £1.78×10^6 approximately. This point corresponds to 99% of the maximum entropy value achieved. Solutions up to 99% of the maximum entropy value are presented in Figure 5b. Compared to Czajkowska and Tanyimboh (2013) the present maximum entropy value is higher, and numerous new solutions have been added at the upper end of the Pareto front close to the maximum entropy value. Thus, the results herein are superior.

A few solutions based on maximizing the maximum entropy have higher entropy values for the same cost than maximizing the minimum or total entropy. However, there is a gap with no solutions between costs of £1.4×10^6 and £1.5×10^6 in the Pareto front for maximizing the maximum entropy. Moreover, above a cost of £1.5×10^6, a few solutions based on maximizing the maximum entropy have lower entropy values for the same cost than the other two Pareto fronts. Thus, the Pareto front achieved by maximizing the maximum entropy is the least consistent.

Figure 5c shows the relationship between the average pipe diameter and entropy, and Table 3 shows that the correlation coefficients are very similar. However, the total entropy maximization model achieved a slightly higher value than the other two. Previously, the importance of failure tolerance as a resilience indicator has been demonstrated very clearly in the literature (Gheisi and Naser 2015; Kalungi and Tanyimboh 2003). The peak demands were used in the pipe failure simulations as explained previously (Subsection 5.1). Although the correlation coefficients are generally comparable, maximizing the total entropy achieved the best result for each of the measures shown in Table 2. Conversely, maximizing the minimum entropy had the worst performance for each of the measures, which suggests that it may be suboptimal to attach undue importance to the worst-case scenario.

Table 3 Correlation coefficients for selected performance indicators for Network 2

	Correlation Coefficients (R^2)		
Relationships	Total entropy	Maximum entropy	Minimum entropy
Mean diameter vs. entropy	0.957	0.949	0.942
Hydraulic reliability vs. entropy	0.605	0.601	0.554
Pipe failure tolerance vs. reliability	0.699	0.672	0.622

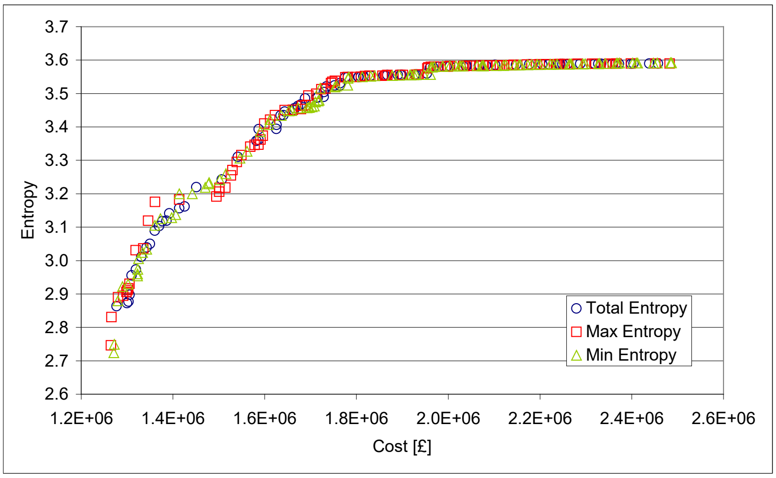
The shading highlights the largest value in each row.

Table 4 Nodal elevations and loading conditions for Network 3

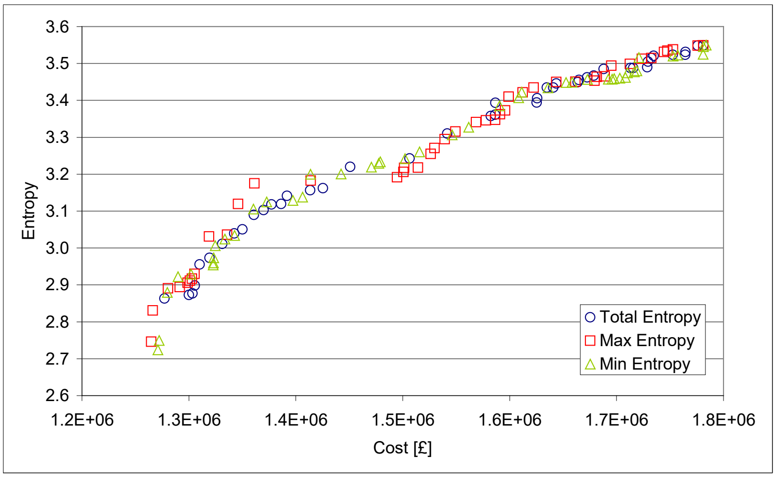
		Demand Pattern 1		Demand Pattern 2		Demand Pattern 3	
Nodes	Elevations (m)	Demands (l/s)	Heads (m)	Demands (l/s)	Heads (m)	Demands (l/s)	Heads (m)
2	320.04	12.62	28.18	12.62	14.09	12.62	14.09
3	326.14	12.62	17.61	12.62	14.09	12.62	14.09
4	332.23	0.00	17.61	0.00	14.09	0.00	14.09
6	298.70	18.93	35.22	18.93	14.09	18.93	14.09
7	295.66	18.93	35.22	82.03	10.57	18.93	14.09
8	292.61	18.93	35.22	18.93	14.09	18.93	14.09
9	289.56	12.62	35.22	12.62	14.09	12.62	14.09
10	289.56	18.93	35.22	18.93	14.09	18.93	14.09
11	292.61	18.93	35.22	18.93	14.09	18.93	14.09
12	289.56	12.62	35.22	12.62	14.09	50.48	10.57

5.3 Loading Conditions with Fire-Fighting Flows

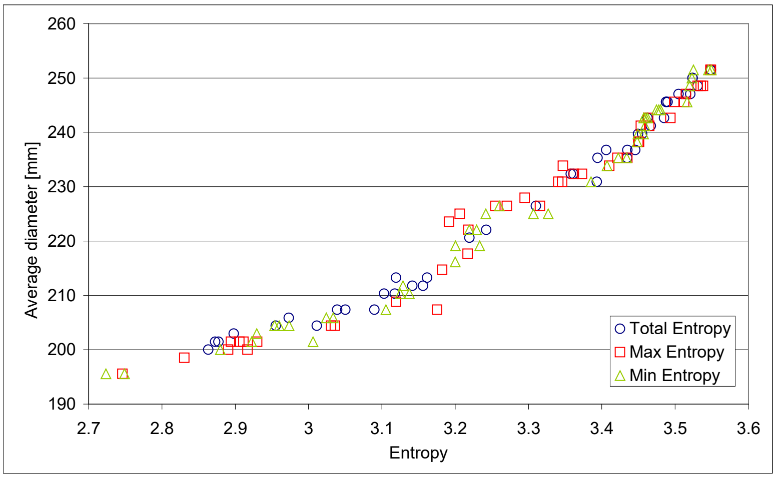
This example is based on Network 3 (Figure 3c) (Simpson et al. 1994). It has the peak demands and two fire-fighting flows. Each fire-flow condition has a fire flow added at one specified node, while the demands at the other nodes are the same as the peak demands. Moreover, the required residual pressures depend on the operating condition for the network as a whole and the location of the fire flow (Table 4). The residual pressure required is lower for the fire flow conditions, which is consistent with fire flow requirements in practice.



(a) Pareto-optimal fronts with all solutions included



(b) Truncated Pareto-optimal fronts



(c) Relationship between entropy and mean pipe diameter

Fig. 5 Pareto-optimal solutions for Network 2 based on 30 optimization runs

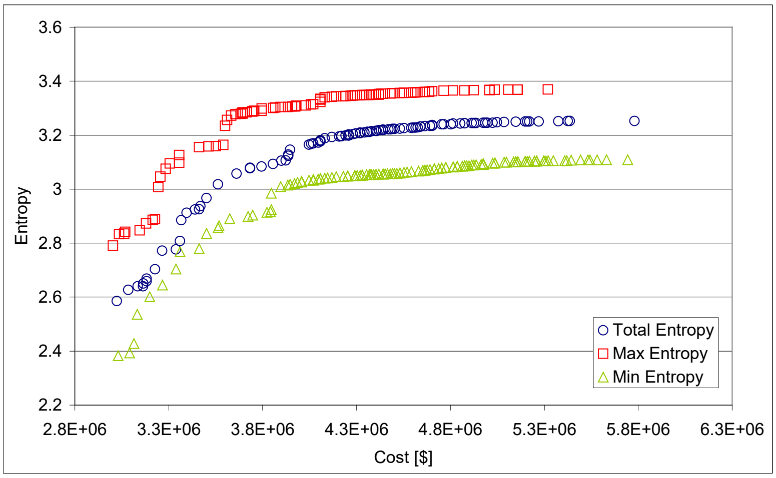
To avoid unnecessary complexity in the interpretation of the results, the design optimization problem was approached herein as a new network design problem with all pipes to be sized, as distinct from a rehabilitation and/or upgrading problem. The network consists of 10 demand nodes, 14 pipes and two supply nodes (reservoirs). The assumed Hazen-Williams roughness coefficient of all pipes was 120. All the pipes are 1,609 m long, except for pipe 1-2 (4,828 m) and 4-5 (6,437 m). The constant heads at the supply nodes (reservoirs) 1 and 5 were 365.76 m and 371.86 m, respectively.

There are eight candidate pipe diameters and 14 pipes, hence 8^{14} or 4.398×10^{12} feasible and infeasible solutions. A 3-bit binary code was used. There were no redundant codes as the number of substrings ($2^3 = 8$) matched the eight pipe diameter options. The values of the GA parameters were $N_E = 200{,}000$, $N_G = 1{,}000$, $N_S = 200$, $N_R = 30$, $p_c = 1.0$ and $p_m = 1/n_g = 1/42 = 0.0238$. With $N_R = 30$ for each of the three maximum entropy approaches, there were 90 optimization runs in total that comprised 18 million hydraulic simulations. The average CPU time for a single optimization run was approximately 17 minutes.

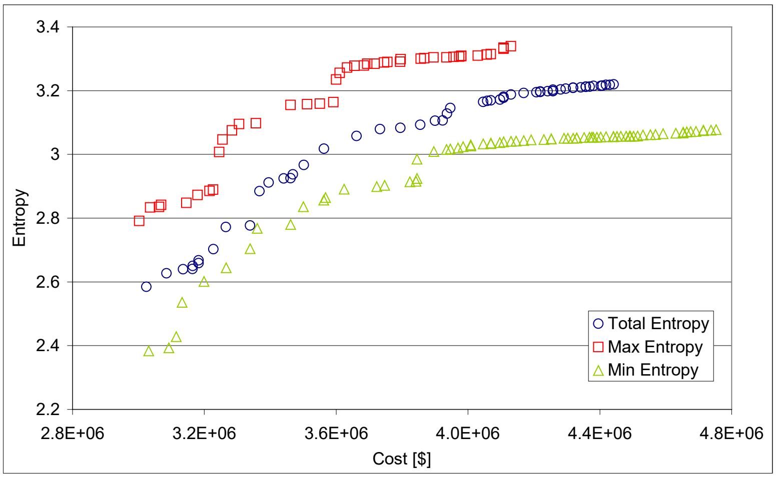
Differences between the Pareto fronts in Figure 6 can be seen in the ranges and shapes of the fronts, due to the deep differences between the loading conditions. With three loading conditions, the entropy values achieved by maximizing the total entropy were divided by 3. It is interesting that the resulting Pareto front for the total entropy is located between the other two fronts, thus reinforcing the notion that it represents a bias-free compromise.

It is easy to deduce that there is no need to analyse all the solutions, as numerous solutions have excessive cost with insignificant improvements in entropy. Hence, the entropy cut-off points were taken as 99% of the respective maximum entropy values achieved. The correlation coefficients for the average pipe diameter and its relationship to entropy were high, and the total entropy approach provided the highest value (Table 5).

Figure 7 shows the plot of the hydraulic capacity reliability against entropy for the solutions that are non-dominated based on cost and entropy (CEND) up to 99% of the maximum entropy value achieved. Also, the solutions that are non-dominated based on cost and reliability (CRND) among those that are non-dominated based on cost and entropy are highlighted. Similarly, the solutions that, additionally, are non-dominated based on cost and failure tolerance (CFTND) are highlighted. The correlation coefficients are shown in Table 5. The peak demands were used to evaluate the hydraulic reliability and failure tolerance. Unlike the fire flows that represent extreme and rare conditions, the peak loading occurs daily. Moreover, all the feasible solutions achieved satisfy all three operating conditions.



(a) Illustration of relatively low utility values of solutions with excessive marginal costs



(b) Solutions up to 99% of the maximum entropy values achieved

Fig. 6 Pareto-optimal solutions for Network 3 based on 30 optimization runs

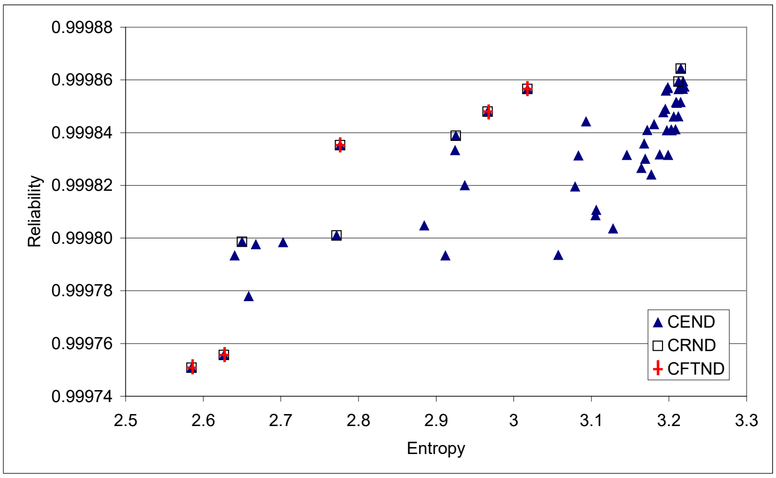
It is noticeable in Figure 7 that the three alternative maximum entropy approaches have clusters of solutions located near the highest entropy values. With the solutions above 99% of the achieved maximum entropy values excluded, at least half of the solutions retained were close to the highest entropy values (Figure 7). As a result, fewer solutions were non-dominated based on all the criteria considered (f_1, f_2 and f_3).

Table 5 Correlation coefficients for selected performance indicators for Network 3

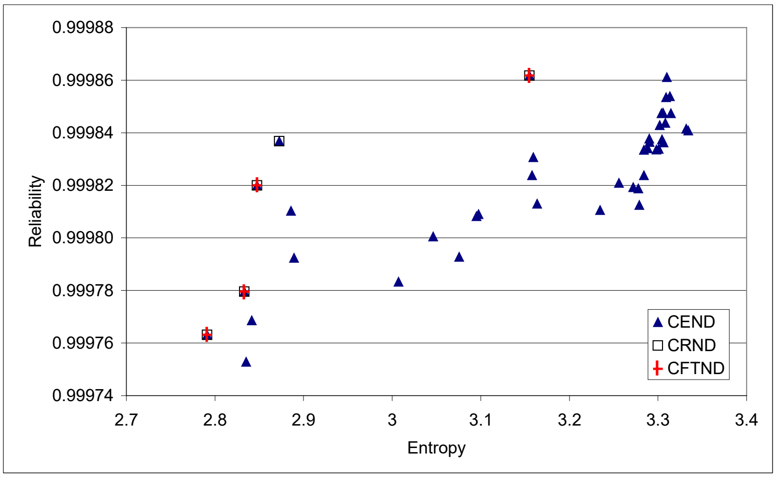
	Correlation Coefficients (R^2)		
Relationships and Pareto Categories	Total entropy	Maximum entropy	Minimum entropy
Mean diameter vs. entropy	0.952	0.857	0.872
Hydraulic reliability vs. entropy			
Level 1: Cost vs entropy (CEND)	0.651	0.599	0.007
Level 2: Cost vs. hydraulic reliability (CRND)	0.825	0.660	0.832
Level 3: Cost vs. pipe failure tolerance (CFTND)	0.898	0.814	0.832

The shading highlights the largest value in each row.

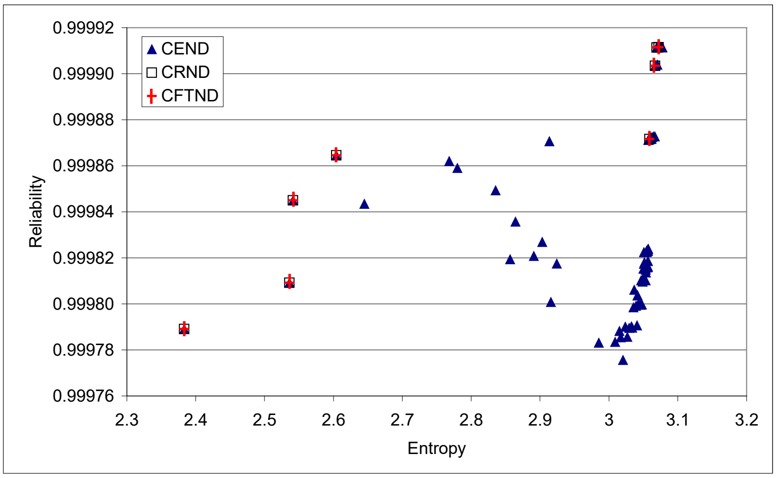
Furthermore, for the total entropy maximization, the correlation between the reliability and entropy increases from 0.651 for CEND and 0.825 for CRND to 0.898 for CFTND (Table 5). These results reinforce the hypothesis that, for cost-effective solutions, there is strong positive correlation between flow entropy and resilience. For the maximum entropy maximization, the correlation coefficients increase similarly, but the values are not as high. For the minimum entropy maximization, the correlation between hydraulic capacity reliability and entropy is only 0.007 for CEND, i.e. practically no correlation. Between the entropy values of 2.7 and 3.0, the hydraulic capacity reliability seems to decrease as the entropy increases. These results could be case specific or anomalous possibly, but 30 optimization runs with independent randomly generated initial populations were performed. Therefore, it is reasonable to infer that maximizing the minimum entropy value did not yield the best performing solutions.



(a) Maximizing the total entropy



(b) Maximizing the maximum entropy



(c) Maximizing the minimum entropy

Fig. 7 Relative effectiveness of alternative maximum entropy approaches on Network 3. CEND, CRND and CRTND, respectively, denote the sequence of non-dominated solutions obtained with respect to cost vs. entropy; cost vs. reliability; and cost vs. failure tolerance.

6. DISCUSSION

Overall, maximizing the total entropy seems the most effective and reliable formulation. This reinforces the hypothesis that the flow entropy function should include all the operating conditions and provides evidence that supports the central tenet of the maximum entropy formalism, i.e. the entropy should be maximized subject to all the available information (Jaynes 1957). While maximizing the total entropy adheres to this maxim, maximizing only the maximum or minimum entropy does not, as the maximum and minimum entropy values represent only the operating conditions from which they are derived.

As the best results were obtained by maximizing the sum of the entropies, it is posited here that the research has provided hitherto unavailable empirical evidence that the joint entropy of independent discrete operating conditions in a water distribution network is the sum of the separate entropies. Moreover, this is consistent with the formal definition of the joint entropy of two or more independent probabilistic schemes (Shannon 1948; Tanyimboh 1993). Additional research to verify this proposition further is thus required in the future.

The results also showed that, on the average, maximizing the maximum entropy was generally uncompetitive. This may provide a partial explanation for some of the inconsistent results that have been reported in Raad *et al.* (2010) and Atkinson *et al.* (2014), and relatively slow convergence properties reported in Liu *et al.* (2014). The reason is that maximizing the maximum entropy does not consider all the operating conditions simultaneously when selecting the entropy value that is used subsequently in the fitness assessment.

Moreover, based on a real-world network in the literature with 75 pipes, 19 loops and two source nodes (Creaco et al. 2010, 2012), Tanyimboh and Czajkowska (2018) demonstrated that the operating condition that yielded the maximum entropy value varied from one solution to the next, depending on the specific configurations of the various solutions. Consequently, the operating conditions having the maximum entropy values could change from one generation to the next during the optimization. More fundamentally perhaps, the respective critical operating conditions of the Pareto-optimal solutions belonged to all the operating conditions collectively rather than any single one.

Fundamentally, the critical operating conditions are design specific. Whether the peak, average or minimum demand, or any other loading condition is the most critical with respect to the residual pressures or any other design criteria depends on the configuration of the individual design. This includes, for example, the spatial distributions of the various components and their capacities, e.g. the pipe diameters and any service reservoirs, etc. Consequently, the critical operating conditions cannot be determined with certainty beforehand. Put differently, maximizing the total entropy ensures that all the active constraints are considered in each generation of the evolutionary optimization procedure. Thus, convergence towards any local optima would be less likely, as the active constraints at the forefront would drive the search, based on consistently reliable performance data. Additionally, slow convergence would be obviated, for the same reason.

The primary aim of the present research was to extend the flow entropy model to multiple operation conditions and an experimental investigation was carried out using a novel penalty-free multi-objective GA. The effectiveness of the optimization method can be gauged further using larger networks from the literature. For the network in Creaco et al. (2010, 2012) with 75 pipes, 19 loops, 2 source nodes, 3 operating conditions, 8 pipe diameter options and 500,000 function evaluations per optimization run, the average CPU time for the full optimization run with 500,000 function evaluations was about 90 minutes on a PC (Intel Core 2 Duo, 3.5 GHz, 3 GB RAM). Convergence was achieved at 200,000 function evaluations (Czajkowska 2016). For the network in Kadu et al. (2008) with the peak daily demands only, 34 pipes, nine loops, two source nodes and 14 pipe diameter options, the mean CPU time to complete 100,000 function evaluations was 3.63 minutes. The minimum number of function evaluations to achieve convergence was 4,000 (Czajkowska 2016) while NSGA II required 108,200 function evaluations in Barlow and Tanyimboh (2014).

Finally, the interested reader may refer to Czajkowska (2016: 101-110) for additional details on the properties of the algorithm including computational efficiency from the perspective of the objective function that minimizes the largest nodal head deficit (Eq. 5), for a real-world network from the literature (Creaco et al. 2010, 2012). The properties include: (a) the progress of the number of feasible and infeasible solutions from the start to the end of the optimization; (b) the surplus head at the critical node vs. cost for the Pareto-optimal solutions; (c) the surplus head at the critical node vs entropy for the Pareto-optimal solutions; (d) the progress of the cost considering both feasible and infeasible solutions from the start to the end of the optimization; and (e) the progress of the nodal head deficit function value (Eq. 5) from the start to the end of the optimization.

7. CONCLUSIONS

A flow entropy-based optimization approach for the design of water distribution networks under multiple operating conditions has been developed and demonstrated. The algorithm developed can handle both single and multiple operating conditions. Solutions based on multiple operating conditions generally outperformed those based on a single operating condition in terms of initial construction cost, hydraulic performance, entropy, reliability and redundancy. Under multiple operating conditions, there is a flow entropy value for each operating condition. In other words, each solution has a vector of entropy values. The design objectives investigated involved maximizing the largest, smallest or sum of the elements of the entropy vector. Maximizing the sum produced the best results overall. A crucial property of the sum unlike the maximum and minimum is that the sum responds to all the operating conditions simultaneously. It is thus bias free with respect to the individual operation conditions. Maximizing the sum, therefore, ensures that all the active constraints are considered in each generation of the evolutionary optimization process, thus reducing the chances of slow, local or premature convergence. Also, the results provide hitherto unavailable empirical evidence that the joint flow entropy of independent discrete operating conditions in a water distribution network is the sum of their separate entropies. Consequently, more clarity and consistency in future applications of the flow entropy model under multiple operating conditions can be expected. This research is in progress, and additional verification on larger networks is necessary. Some improvements in the methodology are desirable also, for example, to reduce the observed over-representation of solutions with the highest entropy values and the relatively high proportion of infeasible solutions in the Pareto-optimal fronts achieved.

Compared to traditional penalty-based approaches, the penalty-free multi-objective genetic algorithm developed in this research has many advantages. Additional parameters that require case-by-case calibration and numerous time consuming trial runs are not introduced. It is bias free with respect to constraint violations, as efficient and promising i.e. ***nondominated infeasible solutions*** are not discarded arbitrarily and prematurely. Consequently, diversity in the gene pool increases, premature convergence decreases, and a more even distribution of the frontier optimal solutions is achieved. Moreover, by maintaining and exploiting ***both feasible and infeasible*** nondominated solutions from the start to the end of the optimization, there is extensive and sustained boundary search the significance of which is that the best solutions generally occur near the active feasibility constraint boundaries. Finally, only one additional objective function was introduced, i.e. a two-objective constrained optimization problem was transformed and solved efficiently and seamlessly as a three-objective penalty-free unconstrained optimization problem.

DECLARATIONS

Funding: The UK Engineering and Physical Sciences Research Council (EPSRC Grant Number EP/G055564/1) funded this project.

Conflict of Interest: None

Availability of Data and Material: The essential data and results are provided in the manuscript.

Code Availability: The methodological details are described in full in the manuscript. EPANET 2 and NSGA II are available in the public domain in the Internet. Selected routines are available from the authors upon request.

REFERENCES

Ainslie B, Reuten C, Steyn DG, Le ND, Zidek JV (2009) Application of an entropy-based Bayesian optimization technique to the redesign of an existing monitoring network for single air pollutants. J. Environ. Manage. 90: 2715-2729

Alirezaei M, Niaki STA, Niaki SAA (2019) A bi-objective hybrid optimization algorithm to reduce noise and data dimension in diabetes diagnosis using support vector machines. Expert Syst. Appl. 127: 47-57

Almeida G, Vieira J, Sá Marques A, Kiperstok A, Cardoso A (2013) Estimating the potential water reuse based on fuzzy reasoning. J. Environ. Manage. 128: 883-892. https://doi.org/10.1016/j.jenvman.2013.06.048

Alperovits E, Shamir U (1977) Design of optimal water distribution systems. Water Resour. Res. 13: 885-900

Amarasinghe P, Liu A, Egodawatta P, Barnes P, McGree J, Goonetilleke A (2016) Quantitative assessment of resilience of a water supply system under rainfall reduction due to climate change. J. Hydrology 540: 1043-1052

Ang WK, Jowitt PW (2005a) Some new insights on informational entropy for water distribution networks. Eng. Optim. 37: 277-289

Ang WK, Jowitt PW (2005b) Path entropy method for multiple-source water distribution networks. Eng. Optim. 37: 705-715

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

Avci MG, Selim H (2017) A Multi-objective, simulation-based optimization framework for supply chains with premium freights. Expert Syst. Appl. 67: 95-106

Awumah K, Goulter I (1992) Maximizing entropy-defined reliability of water distribution networks. Eng. Optim. 20, 57-80

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

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

Baños R, Reca J, Martínez J, Gil C, Márquez AL (2011) Resilience indexes for water distribution network design: a performance analysis under demand uncertainty. Water Resour. Manage. 25: 2351-2366

Barlow E, Tanyimboh T (2014) Multiobjective memetic algorithm applied to the optimisation of water distribution systems. Water Resour. Manage. 28:2229-2242

Chang N-B, Qi C, Yang YJ (2012) Optimal expansion of a drinking water infrastructure system with respect to carbon footprint, cost-effectiveness and water demand. J. Environ. Manage. 110: 94-206

Ciaponi C, Franchioli L, Murari E, Papiri S (2015) Procedure for defining a pressure-outflow relationship regarding indoor demands in pressure-driven analysis of water distribution networks. Water Resour. Manage. 29: 817-832

Coello Coello CA (2002) Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art. Comput. Methods Appl. Mech. Eng. 191:1245-1287

Constantine K, Massoud M, Alameddine I, El-Fadel M (2017) The role of the water tankers market in water stressed semi-arid urban areas: Implications on water quality and economic burden. J. Environ. Manage. 188: 85-94

Creaco E, Franchini E, Alvisi S (2010) Optimal placement of isolation valves in water distribution systems based on valve cost and weighted average demand shortfall. Water Resour. Manage. 24: 4317-4338. https://doi.org/10.1007/s11269-010-9661-5

Creaco E, Franchini E, Alvisi S (2012) Evaluating water demand shortfalls in segment analysis. Water Resour. Manage. 26: 2301-2321. https://doi.org/10.1007/s11269-012-0018-0

Cullinane MJ, Lansey KE, Mays LW (1992) Optimization-availability-based design of water distribution networks. J. Hydraulic Engineering 118: 420-441

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. Wa. Sc. Technol. 13: 1265-1271

Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: NSGA II. Transactions on Evolutionary Computation. 6: 182-197

Díaz J, Montoya MC, Hernández S (2016) Efficient methodologies for reliability-based design optimization of composite panels. Advances in Engineering Software 93: 9-21

Dridi L, Parizeau M, Maihot A, Villeneuve J-P (2008) Using evolutionary optimization techniques for scheduling water pipe renewal considering a short planning horizon. J. Comput. Aided Civ. Infrastruct. Eng. 23: 625-635

Eiger G, Shamir U, Ben-tal A (1994) Optimal design of water distribution networks. Water Resour. Res. 30: 2637-2646

Eskandar H, Sadollah A, Bahreininejad A, Hamdi M (2012) Water cycle algorithm – A novel metaheuristic optimization method for solving constrained engineering optimization problems. Comput. Struct. 110-111: 151-166

Forrester AIJ, Sobester A, Keane AJ (2008) Engineering Design via Surrogate Modelling. Wiley

Fujiwara O, Ganesharajah T (1993) Reliability assessment of water supply systems with storage and distribution networks. Water Resources Research 29: 2917-2924

Germanopoulos G (1985) A technical note on the inclusion of pressure dependent demand and leakage terms in water supply network models. Civil Eng. Syst. 2: 171-179

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

Greco R, Di Nardo A, Santonastaso G (2012) Resilience and entropy as indices of robustness of water distribution networks. J. Hydroinformatics 14: 761-771

Gupta R, Bhave PR (1996) Comparison of methods for predicting deficient network performance. J. Water Res. Pl.-ASCE 122: 214-217

Gurupur VP, Sakoglu U, Jain GP, Tanik UJ (2014) Semantic requirements sharing approach to develop software systems using concept maps and information entropy: A personal health information system example. Adv. Eng. Softw. 70: 25-35

Hao JH, Kwok RC, Lau RY, Yu AY (2010) Predicting problem-solving performance with concept maps: an information-theoretic approach. Decision Support Systems 48: 613-621

Harrison CG, Williams PR (2016) A systems approach to natural disaster resilience. Simulation Modelling Practice and Theory 65: 11-31

He P, Tao T, Xin K et al. (2016) Modelling water distribution systems with deficient pressure: An improved iterative methodology. Water Resour. Manage. 30: 593-606

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

Ishibuchi H, Akedo N, Nojima Y (2015) Behaviour of multiobjective evolutionary algorithms on many-objective knapsack problems. Trans. Evol. Comput. 19: 264-283

Jayaram N, Srinivasan K (2008) Performance-based optimal design and rehabilitation of water distribution networks using life cycle costing. Water Resour. Res. 44: 1-15

Jaynes ET (1957) Information theory and statistical mechanics. Physical Review. 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: 147-160

Kalungi P, Tanyimboh TT (2003) Redundancy model for water distribution systems. Reliability Engineering and System Safety 82: 275-286

Kaplan JD, Howitt RE, Farzin YH (2003) An information-theoretical analysis of budget-constrained nonpoint source pollution control. J. Environ. Econ. Manag. 46: 106-130

Khomsi D, Walters GA, Thorley ARD, Ouazar D (1996) Reliability tester for water distribution networks. J. Computing in Civil Engineering 10: 10-19

Khu ST, Keedwell E (2005) Introducing choices (flexibility) in upgrading of water distribution network: the New York City tunnel network example. Eng Optim 37: 291-305

Kovalenko Y, Gorev NB, Kodzhespirova IF, Prokhorov E, Trapaga G (2014) Convergence of a hydraulic solver with pressure-dependent demands. Water Resour. Manage. 28: 1013-1031

Lahsasna A, Seng WC (2017) An improved genetic-fuzzy system for classification and data analysis. Expert Syst. Appl. 83: 49-62

Lehký D, Slowik O, Novák D (2017) Reliability-based design: Artificial neural networks and double-loop reliability-based optimization approaches. Advances in Engineering Software, https://doi.org/10.1016/j.advengsoft.2017.06.013

Lind NC (1997) Three information–theoretical methods to estimate a random variable. J. Environ. Manage. 49: 43-51

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

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: 5597-5610

Marchi A, Salomons E, Ostfeld A, et al. (2014) Battle of the water networks II. J. Water Res. Pl.-ASCE. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000378

Mitchell G, McDonald A (2015) Developing resilience to England's future droughts: Time for cap and trade? J. Environ. Manage. 149: 97-107

Moosavian N, Lence BJ (2017) Nondominated sorting differential evolution algorithms for multiobjective optimization of water distribution systems. J. Water Resour. Plann. Manage. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000741

Morgan DR, Goulter IC (1985) Optimal urban water distribution design. Water Resour. Res. 21: 642-652

Phan DT, Lim JBP, et al., (2013) Effect of serviceability limits on optimal design of steel portal frames. Journal of Constructional Steel Research 86: 74-84

Prasad TD (2010) Design of pumped water distribution networks with storage. J. Water Res. Pl.-ASCE 136: 129-132

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

Qi C, Chang N-B (2011) System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. J. Environ. Manage. 92: 1628-1641

Raad DN, Sinske AN, van Vuuren JH (2010) Comparison of four reliability surrogate measures for water distribution systems design. Water Resour. Res. 46: 1-11

Rasekh A, Brumbelow K (2014) Drinking water distribution systems contamination management to reduce public health impacts and system service interruptions. Environmental Modelling and Software 51: 12-25

Reca J, Martinez J (2006) Genetic algorithms for the design of looped irrigation water distribution networks. Water Resour. Res. 42, W05416

Recca J, Martinez J, Banos R, Gil C (2008) Optimal design of gravity-fed looped water distribution networks considering the resilience index. J. Water Res. Pl.-ASCE 134: 234-238

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

Sanodiya RK, Saha S, Mathew J (2019) A kernel semi-supervised distance metric learning with relative distance: Integration with a MOO approach. Expert Syst. Appl. 125: 233-248

Saleh SHA, Tanyimboh TT (2013) Coupled topology and pipe size optimization of water distribution systems. Water Resour. Manage. 30:3671-3688

Saleh S, Tanyimboh T (2014) Optimal design of water distribution systems based on entropy and topology. Water Resour. Manage. 28: 3555-3575. https://doi.org/10.1007/s11269-014-0687-y

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

Sebri M (2016) Forecasting urban water demand: A meta-regression analysis. J. Environ. Manage. 183: 777-785

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

Sharp WW, Walski TM (1988) Predicting internal roughness in water mains. J. AWWA. 80: 34-40

Sheikholeslami R, Talatahari S (2016) Developed swarm optimizer: A new method for sizing optimization of water distribution systems. J. Comput. Civ. Eng. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000552

Siew C, Tanyimboh TT (2012) Penalty-free feasibility boundary-convergent multi-objective evolutionary algorithm for the optimization of water distribution systems. Water Resour. Manage. 26: 4485-4507. https://doi.org/10.1007/s11269-012-0158-2

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. Manage. 28: 373-389. https://doi.org/10.1007/s11269-013-0488-8

Siew C, Tanyimboh TT, Seyoum AG (2016) Penalty-free multi-objective evolutionary approach to optimization of Anytown water distribution network. Water Resour. Manage. 30:3671-3688

Simpson AR, Dandy GC, Murphy LJ (1994) Genetic algorithms compared to other techniques for pipe optimization. J. Water Res. Pl.-ASCE 120: 423-443

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

Soui M, Gasmi I, Smiti S, Ghédira K (2019) Rule-based credit risk assessment model using multi-objective evolutionary algorithms. Expert Syst. Appl. 126: 144-157

Surendran S, Tanyimboh TT, Tabesh M (2005) Peaking demand factor-based reliability analysis of water distribution systems. Advances in Engineering Software 36: 789-796

Tanyimboh TT (1993) An entropy-based approach to the optimum design of reliable water distribution networks. PhD thesis, University of Liverpool, UK, pp 74-77

Tanyimboh TT (2017) Informational entropy: a failure tolerance and reliability surrogate for water distribution networks. Water Resources Management 31: 3189-3204

Tanyimboh TT, Czajkowska AM (2018) Joint entropy based multi-objective evolutionary optimization of water distribution networks. Water Resour. Manage. 32: 2569-2584. https://doi.org/10.1007/s11269-017-1888-y

Tanyimboh TT, Kalungi P (2008) Optimal long-term design, rehabilitation and upgrading of water distribution networks. Engineering Optimization 40: 637-654

Tanyimboh TT, Setiadi Y (2008) Sensitivity analysis of entropy constrained designs of water distribution systems. Engineering Optimization 40: 439-457

Tanyimboh TT, Seyoum AG (2016) Multiobjective evolutionary optimization of water distribution systems: Exploiting diversity with infeasible solutions. Journal of Environmental Management 183: 133-141

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

Tanyimboh TT, Siew C, Saleh S, Czajkowska A (2016) Comparison of surrogate measures for the reliability and redundancy of water distribution systems. Water Resources Management 30: 3535-3552. https://doi.org/10.1007/s11269-016-1369-8

Tanyimboh TT, Tahar B, Templeman AB (2003) Pressure-driven modelling of water distribution systems. Water Science and Technology – Water Supply 3: 255-261

Tanyimboh TT, Templeman AB (1993a) Optimum design of flexible water distribution networks. Civil Engineering Systems. 10: 243-258

Tanyimboh TT, Templeman AB (1993b) Calculating maximum entropy flows in networks. J. the Operational Research Society 44: 383-396

Tanyimboh TT, Templeman AB (1993c) Maximum entropy flows for single-source networks. Engineering Optimization 22: 49-63

Tanyimboh TT, Templeman AB (1993d) Using entropy in water distribution networks. International Conference on Integrated Computer Applications for Water Supply and Distribution, Coulbeck B (ed.), Leicester, UK, pp 14, 7-9 September 1993

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

Tanyimboh T, Templeman A (2000) A quantified assessment of the relationship between the reliability and entropy of water distribution systems. Eng. Optim. 33: 179-199

Tanyimboh TT, Templeman AB (2010) Seamless pressure-deficient water distribution system model. J. Water Management 163: 389-396

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

Todini E (2000) Looped water distribution networks design using a resilience index based heuristic approach. Urban Water 2: 115-122

Vairagade SA, Abdy Sayyed MAH, Gupta R (2015) Node head flow relationships in skeletonized water distribution networks for predicting performance under deficient conditions. World Environmental and Water Resources Congress, Austin, Texas

Wagner JM, Shamir U, Marks DH (1988) Water distribution reliability: analytical methods. J. Water Res. Pl.-ASCE 113: 253-275

Walski TM, Brill ED, Gessler J, Goulter IC, Jeppson RM, Lansey K, Lee HL, Liebman JC, Mays L, Morgan DR, Ormsbee L (1987) Battle of the network models: Epilogue. J. Water Res. Pl.-ASCE 113: 191-203

Wang Y, Assogba K, Liu Y, Ma X, Xu M, Wang Y (2018) Two-echelon location-routing optimization with time windows based on customer clustering. Expert Syst. Appl. 104: 244-260

Wang S, Huang GH, Zhou Y (2016) A fractional-factorial probabilistic-possibilistic optimization framework for planning water resources management systems with multi-level parametric interactions. J. Environ. Manage. 172: 97-106

Wang X, Sun Y, Song L, Mei C (2009a) An eco-environmental water demand based model for optimising water resources using hybrid genetic simulated annealing algorithms. Part I. Model development. J. Environ. Manage. 90: 2628-2635

Wang X, Sun Y, Song L, Mei C (2009b) An eco-environmental water demand based model for optimising water resources using hybrid genetic simulated annealing algorithms. Part II. Model application and results. J. Environ. Manage. 90: 2612-2619

Watts G, von Christierson B, Hannaford J, Lonsdale K (2012) Testing the resilience of water supply systems to long droughts. J. Hydrology 414-415: 255-267

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

Wright R, Abraham E, Parpas P, Stoianov I (2015) Control of water distribution networks with dynamic DMA topology using strictly feasible sequential convex programming. Water Resour. Res. 51, 9925-9941. https://doi.org/10.1002/2015WR017466

Wu ZY, Walski T (2005) Self-adaptive penalty approach compared with other constraint-handling techniques for pipeline optimization. J Water Res. Pl.-ASCE 131: 181-192

Xu C, Goulter IC (1998) Probabilistic model for distribution reliability. J. Water Res. Pl.-ASCE 124: 218-228

Xu C, Goulter IC (1999) Reliability based optimal design of water distribution networks. J. Water Res. Pl.-ASCE. 125: 352-362

Xu Q, Qiang Z, Chen Q, Liu K, Cao N (2018) A superposed model for the pipe failure assessment of water distribution networks and uncertainty analysis: a case study. Water Resources Management 32: 1713-1723

Yang J, Soh CK (1997) Structural optimization by genetic algorithms with tournament selection. J. Comput. Civ. Eng. 11: 195-200

Yassin-Kassab A, Templeman AB, Tanyimboh TT (1999) Calculating maximum entropy flows in multi-source, multi-demand networks. Eng. Optim. 31: 695-729

Yates DF, Templeman AB, Boffey TB (1984) The computational complexity of determining least capital cost designs for water supply networks. Eng. Optim. 7: 143-155

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

Zhao R, Chen S (2008) Fuzzy pricing for urban water resources: Model construction and application. J. Environ. Manage. 88: 458-466

Zhao Z, Zhang Y (2011) Design of ensemble neural network using entropy theory. Advances in Engineering Software 42: 838-845. https://doi.org/10.1016/j.advengsoft.2011.05.027

Figure Captions

Table Captions