

Assessment of Penalty-Free Multi-Objective Evolutionary Optimization Approach for the Design and Rehabilitation of Water Distribution Systems

Calvin Siew · Tiku T. Tanyimboh ·
Alemtsehay G. Seyoum

Received: 29 August 2013 / Accepted: 3 December 2013 /
Published online: 20 December 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract This paper describes a penalty-free multi-objective evolutionary optimization approach for the phased whole-life design and rehabilitation of water distribution systems. The optimization model considers the initial construction, rehabilitation and upgrading costs. Repairs and pipe failure costs are included. The model also takes into consideration the deterioration over time of both the structural integrity and hydraulic capacity of every pipe. The fitness of each solution is determined from the trade-off between its lifetime costs and its actual hydraulic properties. The hydraulic analysis approach used, known as pressure-dependent modelling, considers explicitly the pressure dependency of the water supply consumers receive. Results for two sample networks in the literature are included that show the algorithm is stable and finds optimal and near-optimal solutions reliably and efficiently. The results also suggest that the *evolutionary sampling efficiency* is very high. In other words, the number of solutions evolved and analysed on average before finding a near-optimal solution is small in comparison to the total number of feasible and infeasible solutions. We found better solutions than those reported previously in the literature for the two networks considered. For the Kadu network, for example, the *new best solution costs Rs125,460,980*—a significant improvement. Additional statistics that are based on extensive testing are included.

Keywords Multi-objective evolutionary optimization · Pressure-dependent analysis · Water distribution system · Genetic algorithm · Hydraulic reliability · Penalty-free boundary search

Electronic supplementary material The online version of this article (doi:10.1007/s11269-013-0488-8) contains supplementary material, which is available to authorized users.

C. Siew · T. T. Tanyimboh (✉) · A. G. Seyoum
Department of Civil and Environmental Engineering, University of Strathclyde Glasgow,
John Anderson Building, 107 Rottenrow, Glasgow G4 0NG, UK
e-mail: tiku.tanyimboh@strath.ac.uk

C. Siew
e-mail: calvin.siew@strath.ac.uk

A. G. Seyoum
e-mail: alemtsehay.seyoum@strath.ac.uk

1 Introduction

A water distribution system is designed to meet the current regulatory standards for the quantity, quality and pressure of the water supplied to consumers. However, subjected to continuous environmental and operational stresses, an aging network will inevitably experience a declining ability to transport water due to its diminishing hydraulic capacity as encrustation builds up in pipes. The structural integrity of the network will also deteriorate, making it prone to bursts and leakage. The implementation of a planned rehabilitation and upgrading strategy is crucial to meet both current and future demands. Failure to do so would lead to adverse effects such as water quality problems, increase in operating costs due to high head losses, increase in water loss through leakage, low water supply pressure and unforeseen disruption of water supply to consumers. The rehabilitation and upgrading of a water distribution system is complex and involves a large amount of capital. Consequently, optimization models are being developed to address this problem. As an illustration of the scale and urgency of the problem, investment needs for buried drinking water infrastructure in the US are projected to increase from \$30 billion per year in 2010 to \$50 billion per year in 2040, with many utilities requiring sustained investment at this level *at least for several decades* (AWWA 2012).

Rehabilitation and upgrading models fall broadly into three main groups (Dandy and Engelhardt 2001). The first group focuses exclusively on the financial aspects. The second group consists of models that yield rehabilitation and upgrading decisions for individual components without considering the hydraulic performance of the network as a whole. The third group includes the system-wide models that consider the hydraulic performance of the entire network. Also, models for the design optimization of water distribution systems frequently do not consider reliability. Reliability relates to the probability that the system can supply the required amount of water at the required pressure. Two relatively robust probabilistic reliability measures are the hydraulic reliability and pipe (or component) failure tolerance. The hydraulic reliability has been defined as the fraction of the total demand that is satisfied on average (Tanyimboh and Templeman 2000). The failure tolerance is a complementary reliability measure defined by Tanyimboh and Templeman (1998) as the statistical mean of the fraction of the total demand that is satisfied when one or more links or other components are out of service.

Reliability evaluation procedures for water distribution systems are highly complex and have been categorised as NP-hard (Wagner et al. 1988). Generally, they require large numbers of time-prohibitive hydraulic simulations that render their direct inclusion in the solution of optimization problems impracticable. Consequently, many researchers have used surrogate reliability measures instead (Saleh et al. 2012). For example, Tanyimboh and Kalungi (2008, 2009) maximized the statistical entropy of the pipe flow rates. Compared to hydraulic reliability, calculation of the statistical entropy is a relatively simple procedure that requires only the pipe flow rates and nodal demands (Tanyimboh and Templeman 2000). Also, extensive research has shown strong positive correlation between statistical entropy and hydraulic reliability (Tanyimboh and Templeman 2000; Setiadi et al. 2005; Saleh et al. 2012). For water distribution systems, the statistical entropy is generally considered to be a measure of the uniformity of the pipe flow rates (Awumah et al. 1990; Tanyimboh and Templeman 1993) derived from a probabilistic measure of uncertainty known as informational entropy (Shannon 1948).

This paper *extends the scope* of the work of Tanyimboh and Kalungi (2008, 2009) in which the pipe flow rates were determined by maximizing the statistical entropy of the water distribution system with respect to *fixed* pipe flow directions that were *pre-specified*. An inherent weakness of the approach in Tanyimboh and Kalungi (2008, 2009) is that the flow

directions and the pipe sizes are mutually inter-dependent. Therefore, the optimal flow directions are not known in advance. Tanyimboh and Kalungi (2008, 2009) used a single-objective optimization model with linear programming that yields a single solution. By contrast, the present approach has two objectives and, consequently, provides multiple non-dominated solutions that have equal merit in principle. Frequently, there will be many non-dominated solutions. The ease with which the statistical entropy of a water distribution system can be calculated given the pipe flow rates provides an effective method to identify the subset of solutions among the non-dominated solutions that merit further analysis using additional criteria such as hydraulic reliability. In this way, the comparisons of the non-dominated solutions, carried out after completing the optimization proper, were able to identify a solution that is better than the Tanyimboh and Kalungi (2008, 2009) solution in terms of the overall cost and hydraulic performance. The integrated model for whole-life costing of water distribution systems that Tanyimboh and Kalungi (2008, 2009) proposed was utilised in the present work for long-term rehabilitation and upgrading and is thus summarised briefly in Section 2 for completeness.

The second aim of this paper is to assess a multi-objective evolutionary optimization algorithm introduced recently by Siew and Tanyimboh (2012a) that is based on the Non-dominated Sorting Genetic Algorithm II (Deb et al. 2002). Evolutionary optimization algorithms for water distribution systems often use penalties and/or selection procedures (e.g. tournaments) that may involve pair-wise comparisons of the candidate solutions to assess the merits of infeasible solutions when solving optimization problems that have constraints. By contrast, the penalty-free multi-objective evolutionary algorithm proposed by Siew and Tanyimboh (2012a) addresses the node pressure constraints seamlessly, as an integral part of the hydraulic analysis. The hydraulic analysis model that it uses is an enhanced version of EPANET 2 (Rossman 2002) called EPANET-PDX (pressure-dependent extension) that simulates water distribution systems with insufficient flow and/or pressure more realistically (Siew and Tanyimboh 2012b). The proposed genetic algorithm has been discussed previously in terms of the least-cost solution and the smallest number of function evaluations achieved on some standard test problems (Siew and Tanyimboh 2010, 2011, 2012a). However, operators used in genetic algorithms (e.g. mutation) are probabilistic in nature. Thus, a statistically more robust assessment is used here that reflects the stochastic nature of the algorithm. New solutions that are hydraulically feasible and cheaper than the current best solution in the literature were found, for the Kadu et al. (2008) network that represents one of the challenging benchmark problems in the literature. An improvement in cost of almost 4.5 % was achieved.

The remainder of this article consists of brief summaries of the costs considered in the optimization model (in Section 2) and the hydraulic simulation model used (in Section 3). The optimization model is described briefly in Section 4. Two test problems from the literature were considered for this article. The results achieved are discussed in Section 5 followed by concluding remarks in Section 6.

2 Overview of the Integrated Whole-Life Costing Model

The costs used in the optimization model are summarised briefly here. The overall design horizon for the optimization is taken as 20 years that are divided into two phases. A two-phase strategy is more economical; it helps to shorten the periods during which there will be excess capacity in the water distribution system and provides added flexibility to address any uncertainties and other changes that may arise during the first phase. The first phase involves optimizing the design of a new network while the second phase rehabilitates and upgrades the network. The upgrading

options considered are replacement and paralleling of pipes (for consistency with the original specifications of the test problem considered here). Other rehabilitation options such as cleaning and relining can also be implemented in the formulation. For an existing water distribution system, the initial design phase (i.e. Phase I) does not apply. In such a situation only the rehabilitation and upgrading phase (i.e. Phase II) is deployed.

The whole-life costing model (Tanyimboh and Kalungi 2008, 2009; and references therein) is quite complex and thus only the main functions are summarized here. The overall cost can be formulated as:

$$Cost = \sum_{\tau=1}^2 \beta_{\tau} C_{\tau}(s_{\tau}, r_{\tau})(1 + b)^{(d-v)} \tag{1}$$

$$C_{\tau}(s_{\tau}, r_{\tau}) = f1 + f2 + f3 \tag{2}$$

in which $C_{\tau}(s_{\tau}, r_{\tau})$ is the cost of adding capacity r_{τ} in each design phase τ . This cost depends on the added capacity and the existing capacity s_{τ} at the beginning of the relevant design phase. $f1$ represents the cost of pipelines including pipe installation, paralleling, replacement and repair costs. $f2$ represents the indirect cost of setting up construction plant and machinery and is assumed to be incurred at the start of each phase. $f3$ is for the costs that vary depending on the magnitude of the capacity installed. The term $(1+b)^{(d-v)}$ is the compound factor, in which $v=0$ when $\tau=1$; $v=TI, \dots, T2$ when $\tau=2$; TI and $T2$ are, respectively, the minimum and maximum durations (in years) for Phase I; b is the annual compound interest rate for the capital borrowed that has to be paid back after d years. β_{τ} is the product of a discount factor $(1+r)^{-v}$ and price increase factor $(1+c)^v$ where r and c are the discount and the inflation rates in construction cost, respectively; r and c were assumed to be equal.

The pipeline costs can be represented as

$$f1 = f1_a + f1_b + f1_c \tag{3}$$

where $f1_a$ and $f1_b$ represent the costs of new and parallel pipelines, respectively, and are assumed to be equal. $f1_c$ is the cost of replacing pipes and is assumed to be approximately 5 % more than paralleling.

$$f1_a = f1_b = \sum_{ij \in IJ} (\gamma_p * \exp(c_p * D_{ij}) * l_{ij} + REP_{ij}) \tag{4}$$

$$f1_c = \sum_{ij \in IJ} (\gamma_r * \exp(c_r * D_{ij}) * l_{ij} + REP_{ij}) \tag{5}$$

where D_{ij} and l_{ij} are the diameter and length of pipe ij , respectively. IJ represents the set of pipes in the water distribution system. γ_p, γ_r, c_p and c_r are user-specified empirical coefficients; REP_{ij} are the repair costs of the new pipes which can be expressed as

$$REP_{ij} = \sum_{t=tb}^{tr} \frac{J_{ij}(t) * CB_{ij} * FCF(LU_{ij}) * l_{ij}}{(1 + r)^{t-ts+1}}; \forall ij \tag{6}$$

where r is the discount rate, ts and tr are the first and the last year of a given design phase, respectively; tb is the time from which a pipe starts to incur repair costs following the expiry of any warranty that may apply. $FCF(LU_{ij})$ is the failure cost factor for land use, LU_{ij} , for pipe ij .

The failure cost factors reflect the indirect costs due to pipe failures, e.g. disruption to traffic and costs incurred by third parties. CB_{ij} is the repair cost per break and is taken as

$$CB_{ij} = \gamma_{br}(D_{ij} * 1000)^\Phi; \quad \forall ij \tag{7}$$

where γ_{br} and Φ are user-specified empirical coefficients. $J(t)_{ij}$ is the break rate (breaks/km/year) in year t . The break rate was taken as

$$J_{ij}(t) = 0.001974 * \exp(-0.00974 * D_{ij}) * age_{ij}^{1.808}); \quad \forall ij \tag{8}$$

where age_{ij} is the number of years since installation of pipe ij .

Other miscellaneous costs that may be associated with the volume of water supplied—for example, expansion of the sewerage system—may be included if appropriate as follows.

$$f3 = VC * Q_{inst}^{VE} \tag{9}$$

where Q_{inst} is the installed capacity in a design phase in l/s; VC and VE are user-specified empirical coefficients. Details regarding the formulations in Eqs. 1–9 and further references are available in Tanyimboh and Kalungi (2008, 2009).

3 Main Hydraulic Design Equations

We used an external hydraulic analysis model based on EPANET 2 called EPANET-PDX (pressure-dependent extension) (Siew and Tanyimboh 2012b) that, intrinsically, ensures the conservation of mass and energy equations are satisfied. The Hazen-Williams formula for the head loss in a pipe is

$$h_{ij} = \omega l_{ij} \left(\frac{Qp_{ij}}{C_{ij}} \right)^\alpha \frac{1}{D_{ij}^\beta} \tag{10}$$

in which h_{ij} , l_{ij} , Qp_{ij} , C_{ij} and D_{ij} represent the head loss, length, volume flow rate, roughness coefficient and internal diameter for pipe ij , respectively. The adverse effect of ageing on the flow-carrying capacity of pipes was modelled as in Sharp and Walski (1988).

$$C_{ij}(t) = 18.0 - 37.2 \log \left[\frac{e_{0ij} + a_{ij} * age_{ij}}{D_{ij}} \right]; \quad \forall ij \tag{11}$$

where $C_{ij}(t)$ is the Hazen-Williams roughness coefficient in year t , e_{0ij} is the initial roughness (mm) i.e. at the time the pipe was installed and a_{ij} is the roughness growth rate (mm/year).

The nodal demand value Qn_j^{req} is the demand at the end of the relevant design phase.

$$Qn_j^{req} = Q_{0j}^{req} (1 + DGR/100)^t; \quad \forall j \tag{12}$$

where Q_{0j}^{req} is the demand for node j at the start of the relevant design phase, DGR is the (percentage) annual rate of increase in the base demand and t is the number of years.

3.1 Bias-Free Efficient Procedure for Addressing the Minimum Node-Pressure Constraints

The residual pressures at the demand nodes of a water distribution system should be high enough to deliver the quantity of water that is prescribed in the relevant standards for drinking water supply (Twort et al. 2000). Unfortunately, evolutionary algorithms by nature usually

generate both feasible and infeasible solutions. To address node-pressure constraints, penalty methods have been applied widely. For example, Dridi et al. (2008) used constraint-violation penalties in NSGA II (Non-dominated Sorting Genetic Algorithm) and NPGA-2 (Niche Pareto Genetic Algorithm). The only rehabilitation option considered was the replacement of old pipes with new pipes of the same diameter, based on a rather short planning horizon of 5 years. A major disadvantage of the penalty-based approach is that additional case-specific parameters are introduced whose calibration is generally challenging. 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. A review of the methods that have been proposed for handling constraints in evolutionary algorithms in general is available in Coello Coello (2002).

In an attempt to alleviate these difficulties Deb (2000) proposed a constraint-violation dominance concept with the following properties. (a) Any solution with no constraint violation dominates all solutions with constraint violations. (b) Any solution with a constraint violation dominates all solutions with larger constraint violations. This method of handling constraints has the disadvantage that it rates *dominated feasible* solutions more highly than *non-dominated infeasible* solutions. Secondly, by using only the amount of constraint violation exclusively, regardless of any other relevant criteria, comparisons between solutions that have constraint violations ignore the Pareto-optimality condition that is axiomatic in optimization problems with multiple objectives. In the case of water distribution systems, it results in the preferential propagation of *uneconomical* solutions that have small constraint violations at the expense of more economical solutions with larger constraint violations. If, from one generation to the next, the rate at which solutions with constraint violations are removed from the population is excessive, essential genetic material (e.g. small pipe sizes) may become extinct and cause the algorithm to slow down, plateau or converge prematurely.

The approach adopted here allows all the feasible and infeasible solutions generated to compete in a way that is fundamentally bias-free with respect to constraint violations. The proposed penalty-free multi-objective evolutionary algorithm uses pressure-dependent analysis to assess each individual in the population of solutions. Unlike the conventional approach known as demand-driven analysis, pressure-dependent analysis takes proper account of the relationship between the flow and pressure at a node. By definition, feasible solutions satisfy all nodal demands in full. Conversely, infeasible solutions do not and the shortfall in the water they supply represents a real measure of the infeasibility of the water distribution system. In this way, pressure-dependent analysis addresses the node pressure constraints as an integral part of the hydraulic analysis. We used an enhanced version of EPANET 2 (Rossman 2002) called EPANET-PDX (pressure-dependent extension) that carries out pressure-dependent analysis seamlessly (Siew and Tanyimboh 2012b; Seyoum and Tanyimboh 2013).

4 Optimization Problem Formulation and Solution

The proposed approach involves two primary objectives. The first objective is to minimise the overall cost for Phases I and II. The second objective is to ensure all nodal demands are satisfied. The two objective functions F_1 and F_2 are defined as follows.

$$\text{Minimise : } F_1 = CR^2 \quad (13)$$

$$\text{Maximise : } F_2 = DSR^4 \quad (14)$$

CR is the ratio of the cost of a particular solution to the cost of the most costly solution in the entire population within a single generation.

$$CR = \frac{Cost}{Cost^{max}} \quad (15)$$

where $Cost$ is the cost of a particular solution and $Cost^{max}$ refers to the largest cost among all the solutions in the same generation. The demand satisfaction ratio DSR is

$$DSR = \frac{Qn}{Qn^{req}} \quad (16)$$

where Qn is the actual flow supplied based on the available pressure and Qn^{req} is the flow required i.e. the demand. DSR is thus the ratio of the available flow to the required flow and takes values between 0 and 1. A solution that has a DSR that is less than 1 cannot satisfy the demands in full and, inherently, violates at least one minimum node pressure constraint. It is, therefore, infeasible. This is the means by which the proposed algorithm distinguishes between feasible and infeasible solutions. We used the DSR value of the worst-performing demand node in Eq. 16 based on the evidence in Siew and Tanyimboh (2012a). The objective functions in Eqs. 13 and 14 were designed to favour economical solutions that are just feasible or marginally infeasible, in each generation of the genetic algorithm. The exponents (i.e. 2 and 4, respectively, in Eqs. 13 and 14) are default values that have proved satisfactory so far (Siew and Tanyimboh 2012a).

Figure 1 provides a diagrammatic overview of the proposed whole-life optimization approach. The Non-dominated Sorting Genetic Algorithm II (Deb et al. 2002) that is used extensively in many diverse fields was chosen for the computational solution of the optimization problem. One of the advantages of NSGA II is that the number of parameters the user must specify is small (Dridi et al. 2008). A description of NSGA II is not provided here for brevity. We wrote a basic NSGA II computer program in C++ and coupled it directly with the hydraulic analysis model EPANET-PDX (pressure-dependent extension) to form the proposed penalty-free multi-objective evolutionary algorithm. We used binary coding and simple operators namely single-bit mutation, single-point crossover and a tournament with two solutions chosen at random to identify the solutions that participate in crossover. The crowding distance operator in NSGA II (Deb et al. 2002) was amended to permit a greater concentration of feasible solutions near the boundary between the feasible and infeasible regions. Also, this reduces the risk of losing the best solutions in the current generation. This can occur if the crowding distance operator is applied without restriction in the objective space (Siew and Tanyimboh 2012a). By default, the crossover probability used is $p_c=1$, i.e. the number of offspring created through crossover is the same as the population size. The mutation rate is $p_m=Nm/Np$, where Nm is the number of offspring mutated and Np is the population size. The mutation operator used swaps one bit that is selected at random in the solution that is mutated.

5 Results and Discussion

5.1 Whole-Life Design Optimization

The sample network considered is the Wobulenzi water distribution system shown in the Appendix. The data used here are from Tanyimboh and Kalungi (2008) that has the full specifications. The network is partially looped and consists of 1 reservoir, 16 demand nodes,

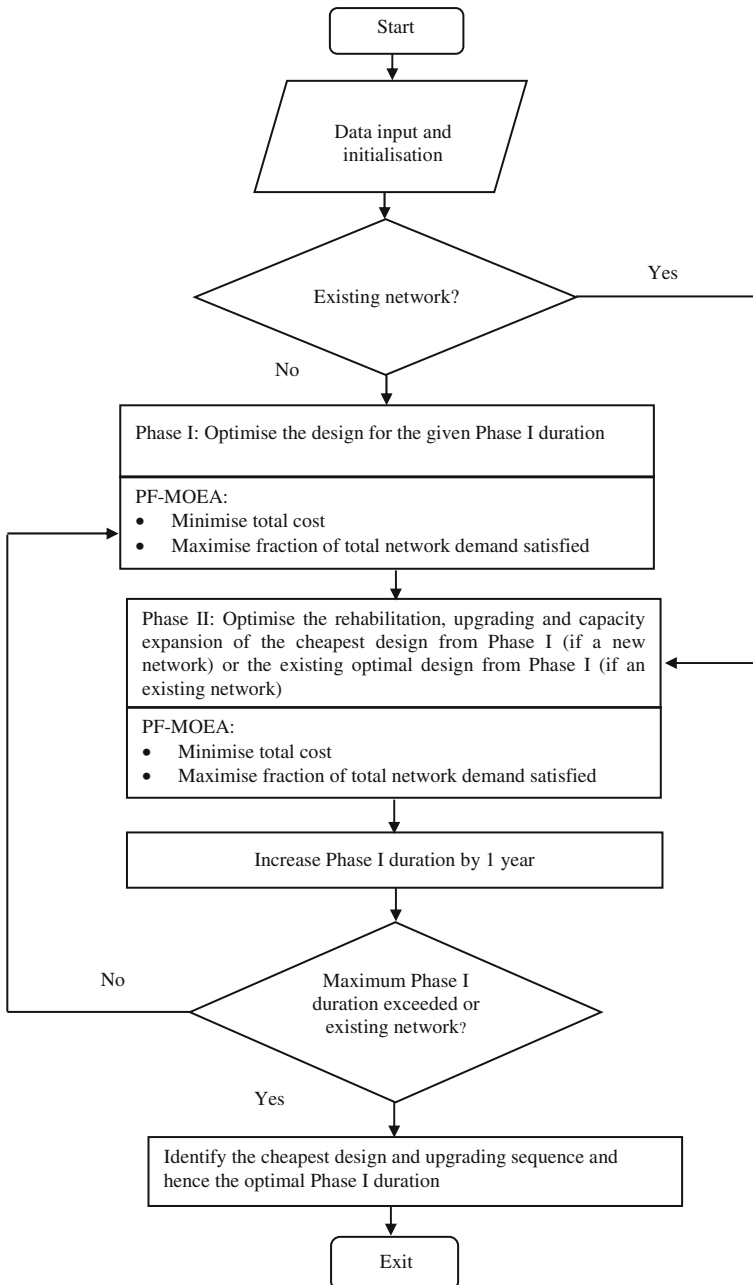


Fig. 1 Flow diagram for the overall design and upgrading methodology

21 pipes and 5 loops. The minimum residual head for full demand satisfaction=15 m; demand growth rate DGR=4 % per annum; peak hour factor=2.0; fire-fighting demand *applied at node 4 only*=25 % of node 4 demand; compound interest rate $b=8$ %; discount rate $r=8$ %; lower and upper limits for the end of Phase I are $T1=7$ years and $T2=14$ years, respectively,

based on a 20-year planning horizon; pipe cost coefficients are $\gamma_p = 32.093$, $c_p = c_r = 3.7$, $\gamma_r = 33.928$, $\gamma_{br} = 108.87$, $\Phi = 0.6067$; pipe warranty period $tb = 6$ years; setting-up cost $f_2 = \$100,000$; installed capacity coefficients are $VC = 130$, $VE = 1.6$; initial pipe roughness $e_{0ij} = 0.0021$ mm; roughness growth rate $a_{ij} = 0.025$ mm/year; the Hazen-Williams head loss coefficients are $C = 130$, $\omega = 10.67$; $\alpha = 1.852$ and $\beta = 4.87$.

In Phase I the decision variables are (a) the duration of Phase I and (b) the pipe diameters. In Phase II the decision variables are the diameters of (a) the existing pipes and (b) any parallel pipes introduced in Phase II. There are 8 pipe diameters to consider: 80, 100, 150, 200, 250, 300, 350 and 400 mm. Therefore, with 21 pipes and 8 pipe sizes, there are $8^{21} \approx 9.2 \times 10^{18}$ feasible and infeasible solutions in Phase I, for each possible Phase I duration of 1 to 20 years. The total number of solutions in Phase I is, therefore, $20 \times 8^{21} \approx 1.8 \times 10^{20}$. The rehabilitation and upgrading options in Phase II are pipe replacement and/or paralleling. There are 9^{21} pipe paralleling options and 8^{21} pipe replacement options; the pipe paralleling options for each pipe include the no-paralleling option. Thus the number of solutions in Phase II is $(8 \times 9)^{21} \approx 1.01 \times 10^{39}$. This constrained optimization problem is nonlinear and has discrete decision variables. There are (a) 16 minimum node-pressure constraints; (b) 16 conservation of mass constraints; and (c) 5 conservation of energy constraints. The hydraulic analysis of the designs ensures constraints (b) and (c) are satisfied. Constraints (a) are addressed through the objective function F_2 (Section 4, Eq. 14).

This is a challenging problem for there are $(20 \times 8^{21})(8 \times 9)^{21} = 1.9 \times 10^{59}$ feasible and infeasible solutions. With reference to Fig. 1, the solution strategy proposed here samples only 8 Phase I durations of 7–14 years. Accordingly, the Phase I optimal design problem is solved, in turn, for each of the 8 alternative durations. For each Phase I duration, the best solution for Phase I is used as the starting point for the rehabilitation problem in Phase II. In this way, the Phase I-plus-Phase II sequence with the least cost is identified as the optimal solution.

The overall efficiency of the proposed solution approach is due to (a) the above-mentioned solution space reduction scheme and (b) the sequential optimization procedure which reduces the computational complexity of the optimization problem. In this way, both the number of decision variables and the size of the solution space under consideration at any given moment are reduced considerably. It is worth re-stating that the duration of Phase I is one of the key decision variables also. *The main aim* of this example is to show that the optimization algorithm proposed can find optimal and near-optimal solutions *quickly*. In addition to its inherent complexity, this example was selected because the combination of linear programming and entropy maximization (used previously by Tanyimboh and Kalungi 2008) provides a design that is considered economical and reliable and, consequently, is not easy to surpass.

Due to the reliability and efficiency of the proposed optimization algorithm (Siew and Tanyimboh 2012a), 10 optimization runs (with different sets of initial populations that were generated randomly) proved to be sufficient. For each optimization run, 10,000 function evaluations or hydraulic simulations (for a population of size 100 and 100 generations) were specified for each design phase. This is equivalent to 20,000 function evaluations (i.e. hydraulic simulations) per Phase I-and-Phase II sequence. Thus there are 160,000 function evaluations in total per optimization run (i.e. 10,000 function evaluations per design phase \times 2 design phases \times 8 alternative Phase I durations); the eight Phase I durations are: 7, 8, 9, ..., 14 years. The probability of crossover and mutation were $p_c = 1$ and $p_m = 0.005$, respectively, for all the 10 optimization runs. An Intel single core personal computer (CPU: 3.2 Hz, RAM: 2 GB) was used. The average central processing unit (CPU) time for one optimization run (consisting of 160,000 function evaluations) was 2.58 h. The maximum CPU time among the 10 optimization runs was 3.07 h while the standard deviation was 0.23 h. The variations in the CPU time are due to differences in the number of iterations that the hydraulic analysis of different solutions requires.

5.1.1 Make-up of the Optimized Costs

Table 1 presents the overall construction, maintenance and failure costs for both the present and previous models. For the present model abbreviated PF-MOEA, the *whole-life* cost attains a minimum for a Phase I duration of 9 years. The solutions from all 10 optimization runs had similar trends. This suggests that the chosen range of Phase I durations (i.e. 7 to 14 years) achieves the objective of bracketing the cheapest solution. All the solutions presented here are fully feasible, i.e. nodal demands and pressures are satisfied in full. For the previous model abbreviated ME-LP (maximum entropy-linear programming) by Tanyimboh and Kalungi (2008), the minimum total cost corresponds to a Phase I duration of 11 years. Figure 2 shows a detailed breakdown of the total design, rehabilitation and upgrading costs, for the alternative Phase I durations considered. The repair costs from Phases I and II contribute the smallest fractions of the total cost. The new network design cost and the various capacity-related costs β_3 in Phase I are the major contributors to the total cost for all the alternative Phase I durations. These costs, along with the repair cost, increase significantly with any delay in rehabilitation and upgrading from a total of approximately 61 % for a Phase I duration of 7 years to 80 % for a Phase I duration of 14 years.

The cheapest solution obtained by the present approach costs \$3,814,298 (Table 1) and has a Phase I duration of 9 years. This is approximately 3.5 % cheaper than the cheapest previous solution of \$3,953,663 with a Phase I duration of 11 years by Tanyimboh and Kalungi (2008). In general, the linear programming approach in Tanyimboh and Kalungi (2008) yields solutions that have segmental pipes with more than one diameter and tend to be cheaper than conventional designs with single-diameter pipes. However, it can be observed in Table 1 that apart from the solution with a Phase I duration of 14 years, the present approach generated solutions with smaller overall costs than the previous model. The main reason is that the previous solutions are maximum entropy-constrained solutions. In general, maximum-entropy solutions are more reliable and, consequently, more expensive than conventional least-cost solutions (Tanyimboh and Templeman 2000). As mentioned previously in Subsection 5.1, the combination of linear programming and entropy maximization that was used in the previous model by Tanyimboh and Kalungi (2008) often yields solutions that are not easily surpassed. Table 1 shows that the cheapest present solutions for all the various Phase I durations have smaller hydraulic reliability values (see the next subsection) than the cheapest previous

Table 1 Cost and hydraulic reliability for the cheapest solutions for the Wobulenzi network

Phase I duration (years)	PF-MOEA cost (\$ Million)			PF-MOEA reliability	ME-LP Cost ^b (\$ Million)		
	Phase I	Phase II	Total		Phase I	Phase II	Total
7	2.862	0.998	3.860	0.999010	2.907	1.386	4.293
8	2.950	0.877	3.827	0.999116	3.006	1.071	4.077
9	3.047	0.768	3.814 ^a	0.999017	3.084	0.966	4.050
10	3.148	0.689	3.837	0.999106	3.200	0.789	3.989
11	3.281	0.593	3.873	0.998867	3.315	0.639	3.954 ^{a,b}
12	3.399	0.504	3.902	0.998869	3.414	0.544	3.958
13	3.565	0.415	3.980	0.998919	3.523	0.461	3.984
14	3.725	0.334	4.059	0.998834	3.631	0.409	4.040

^a Identifies the cheapest solution

^b From Tanyimboh and Kalungi 2008 (reliability value for the cheapest ME-LP solution=0.999197)

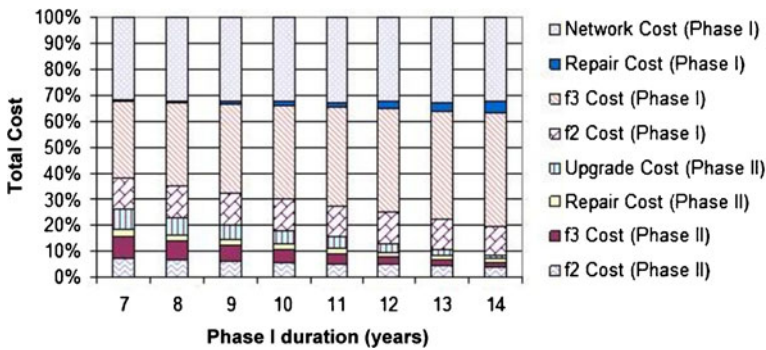


Fig. 2 Breakdown of costs for the cheapest solutions found for the Wobulenzi network

solution. All the hydraulic reliability values in this article relate to the network and operating conditions at the end of the planning horizon in Year 20.

5.1.2 Trade-off Between Cost and Reliability

A temporary pipe closure due to failure or maintenance reduces the capacity of a water distribution system. Therefore, pressure-dependent analysis (Siew and Tanyimboh 2012b) was used to simulate the pipe closures. Pipes were closed individually. In practice, however, the actual locations of isolation valves would be taken into account. A probabilistic pipe failure model (Cullinane et al. 1992) was used to estimate the pipe availability values. The hydraulic reliability, failure tolerance and entropy values (see Section 1) were calculated as defined in Tanyimboh and Templeman (1998, 2000).

Each of the 10 optimization runs generated 8 different sets of non-dominated solutions, i.e. one set for each alternative Phase I duration of 7 to 14 years. Therefore, the 10 optimization runs provided 80 sets of non-dominated solutions in total. The cheapest feasible solution from each of the 80 sets of non-dominated solutions was selected for an assessment of the trade-off between cost and reliability; 80 solutions were thus selected for further analysis. Previous research (Tanyimboh and Sheahan 2002; Tanyimboh and Setiadi 2008) has shown that only a small fraction of the solutions are non-dominated if the selection criteria are (a) cost; (b) entropy; (c) hydraulic reliability; and (d) failure tolerance. This may be because the cost, entropy, hydraulic reliability and failure tolerance are strongly correlated. Therefore, in concert, these criteria reduce drastically the number of non-dominated solutions. Only 5 solutions out of 80 were non-dominated based on cost and entropy. Reliability and failure tolerance values were thus calculated for the 5 solutions. Only 3 solutions out of 5 were non-dominated based on cost, hydraulic reliability and failure tolerance (Table 2). Details of the

Table 2 Performance indicators for the best solutions for the Wobulenzi network

	ME-LP solution	PF-MOEA solutions		
		1	2	3
Phase I duration (years)	11	9	9	10
Cost (\$ Million)	3.954	3.814	3.824	3.864
Hydraulic reliability	0.999197	0.999017	0.999062	0.999275
Failure tolerance	0.924534	0.920000	0.925883	0.935727

three solutions are in the [Appendix](#). The present Solutions 1 to 3 (Table 2) are cheaper than the previous solution. Also, the present solutions have virtually the same hydraulic reliability and failure tolerance values as the previous solution. It can be seen that the present Solution 3 dominates the previous solution with respect to cost, reliability and failure tolerance. Additional details on the various trade-offs are available in Siew (2011). The analysis also revealed that, for the designs under consideration, statistical entropy was a considerably better indicator of hydraulic reliability and failure tolerance than resilience index (Todini 2000).

5.2 Optimization of the Initial Construction Cost Only

This is the second example considered and it involves only the initial construction cost. Therefore, it represents the simplest form of the Phase I optimization problem in that multiple Phase I durations are not considered. The network was taken from Kadu et al. (2008) that has the details of the optimization problem. The network is fully looped and consists of 24 demand nodes, 34 pipes and 9 loops. There are 2 reservoirs with constant water levels of 100 m and 95 m, respectively. The network's layout is shown in the [Appendix](#). Fourteen candidate pipe sizes are available for this network. Therefore, with 34 pipes and 14 pipe sizes, there are $14^{34} \approx 9.3 \times 10^{38}$ feasible and infeasible solutions. The coefficients for the Hazen-Williams formula are $C=130$, $\alpha=1.85$, $\beta=4.87$ and $\omega=10.68$.

Extensive investigation of the effectiveness of proposed optimization algorithm was conducted in this example as summarised in Table 3. The initial populations were generated randomly. The maximum number of function evaluations (i.e. hydraulic simulations) permitted per optimization run was 500,000 and the crossover probability was $p_c=1$ in all cases. Other parameters, including the population size, mutation rate and number of optimization runs were as shown in Table 3. A total of *eight* cases were considered. One case (out of 8) had the default values of the coefficients of the Hazen-Williams formula in EPANET 2 (i.e. $\alpha=1.852$, $\beta=4.871$, $\omega=10.667$) for completeness. Summarised results are shown in Table 3 based on sample sizes (i.e. the total number of optimization runs) of 100 (in 6 cases out of 8) and 30 (in 2 cases out of 8) as the initial results suggested the smaller sample size might also be statistically satisfactory.

The cheapest solution obtained was 125,460,980 Rupees (i.e. with $N_p=500$, $p_m=0.05$ in Table 3), within 436,000 function evaluations. Other researchers have not found this solution previously and it is the cheapest hitherto. Also, the *minimum-cost* solutions found in the eight cases considered were close to the *smallest* minimum-cost. The *means* of the minimum-cost differ from the *smallest* minimum-cost achieved here by only 2.31–4.15 %. The average number of function evaluations to obtain convergence (within the specified maximum of 500,000 function evaluations) ranged from 354,160 to 397,083. The average CPU time to achieve convergence was 1.11–1.25 h. To complete a single optimisation run consisting of 500,000 function evaluations, due to differences in the number of iterations per hydraulic simulation, the average CPU time required was 1.57 h and the standard deviation was 0.10 h on a personal computer (Intel Core 2 Duo with 2.5 GHz CPU and 1.95 GB RAM).

Further limited sensitivity analysis was also conducted to check the influence of the mutation rate. Nine different additional mutation rates spread approximately evenly in the range $p_m=[0.01, 0.7]$ were used; $p_m=0.05$ and $p_m=0.07$ that feature in Table 3 were excluded. The population size was $N_p=200$ and the maximum number of function evaluations permitted was 500,000. Only five optimization runs were executed for each mutation rate. Therefore, with only five trials per mutation rate for the limited sensitivity analysis, a fixed set of five different initial populations (each with $N_p=200$) that were generated randomly was used for all the mutation rates considered. The *minimum-cost* feasible solution for the nine mutation rates ranged from 126,035,000 to 129,832,000 Rupees. For the nine mutation rates, the *average*

Table 3 Computational characteristics of the proposed algorithm for the Kadu network

	Full solution space			Reduced solution space		
Maximum FEs allowed	500×10^3					
Number of GA runs ^d	100 ^a		30 ^{b,d}	100 ^a		
Population size	200	500		100	200	
Mutation rate	0.005			0.005		0.05
Initial construction cost (Rupees $\times 10^6$)		0.05	0.07	0.05		
Minimum	126.740 (1.02) ^c	126.067 (0.48)	125.820 (0.29)	125.551 (0.00)	125.919 (0.36)	125.826 (0.29)
Maximum	134.243 (7.00)	133.005 (6.01)	131.797 (5.05)	128.222 (2.13)	130.868 (4.31)	127.281 (1.45)
Mean	130.662 (4.15)	129.479 (3.20)	128.623 (2.52)	127.019 (1.17)	127.435 (1.57)	126.182 (0.57)
Median	130.904 (4.34)	129.391 (3.13)	128.588 (2.49)	127.054 (1.20)	127.244 (1.42)	126.112 (0.52)
Standard deviation	1.707	1.689	1.430	0.716	1.035	0.328
Number of better solutions from the best run than previous solutions						
Cheaper than Haghghi	19	31	29	16	30	31
Cheaper than Kadu	20	32	30	17	30	31
Number of function evaluations (FEs) to achieve convergence						
Minimum	140,000	198,500	195,500	204,500	32,000	43,400
Maximum	497,800	498,500	497,500	500,000	497,100	499,000
Mean	354,160	396,505	397,083	370,483	278,145	285,998
Median	364,500	413,000	402,750	402,250	293,000	293,600
Standard deviation	101,000	80,000	82,000	98,000	130,000	124,000

^aHazen Williams parameters: $\alpha=1.85$; $\beta=4.87$; and $\omega=10.68$

^bBased on the default EPANET 2 Hazen Williams parameters: $\alpha=1.852$; $\beta=4.871$; and $\omega=10.667$

^cThe deviations from the least cost achieved for the full solution space are shown in the parentheses (%)

^d30 optimization runs were carried out in two tests after six preceding tests with 100 optimization runs revealed highly consistent results

minimum-cost (based on 5 optimization runs per mutation rate) ranged from 127,615,000 to 131,353,800 Rupees. Based on the results in Table 3, it can be expected that a population size of $N_p=500$ would provide even better results. Overall, for the mutation rates and population sizes attempted, the performance of the optimization algorithm was consistently reliable and satisfactory.

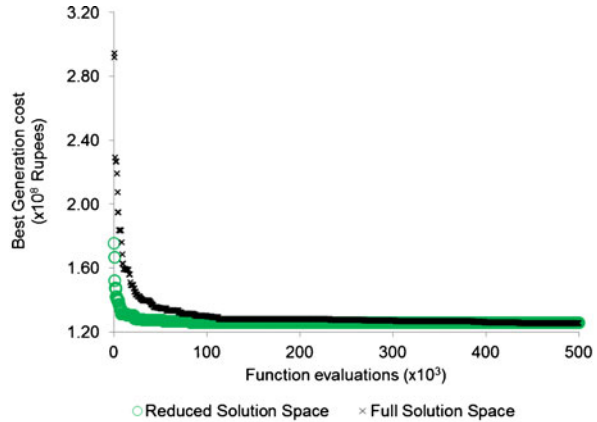
Kadu et al. (2008) proposed a critical path concept to reduce the number of candidate diameters for each pipe. In this way, they reduced the solution space to 8.65×10^{20} feasible and infeasible solutions. We also tested the proposed optimization algorithm as summarised in Table 3 using the same reduced solution space of size 8.65×10^{20} solutions as in Kadu et al. (2008). The *minimum cost* achieved for a feasible solution was 125,826,425 Rupees within 82,400 function evaluations. This is only 0.29 % more costly than the smallest cost we achieved for the full solution space of size $14^{34} \approx 9.3 \times 10^{38}$ solutions. Overall, the values of the cost and function evaluations are improved, on average, by reducing the size of the solution space. On average, approximately 27 % fewer function evaluations, i.e. hydraulic simulations, were required to find a near-optimal solution when the solution space was reduced, in comparison to the full solution space.

Haghighi et al. (2011) also solved the same optimization problem using a hybrid approach consisting of a genetic algorithm and integer linear programming. A comparison between the present approach and the best results reported previously in the literature is shown in the Appendix. For the full solution space, the solution obtained by Kadu et al. (2008) was 131,678,935 Rupees, within 120,000 function evaluations. This is 4.96 % more expensive than the new best solution of 125,460,980 Rupees. Haghighi et al. (2011) achieved 131,312,815 Rupees, within 4,440 function evaluations. This is 4.66 % more expensive than the new best solution. For the reduced solution space, Kadu et al. (2008) obtained a solution of 126,368,865 Rupees, within 25,200 function evaluations. This is 0.72 % more expensive than the new best solution.

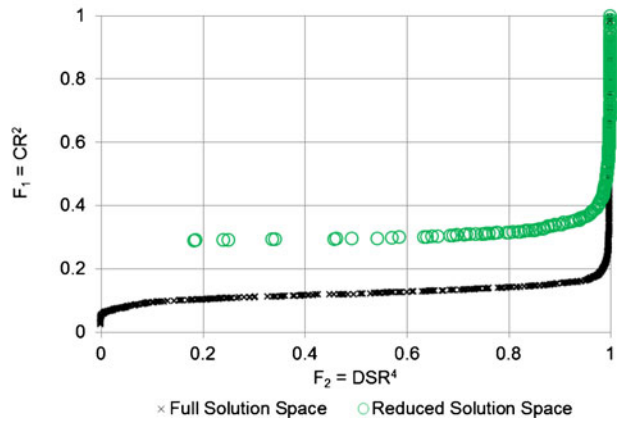
However, the feasibility of the Kadu et al. (2008) and Haghighi et al. (2011) solutions is questionable. Based on EPANET 2, the Kadu et al. (2008) and Haghighi et al. (2011) solutions were deemed infeasible (as shown in the Appendix). The Kadu et al. (2008) solutions violate the minimum node-pressure requirement at Nodes 12, 24 and 25 (for the full solution space) and Node 26 (for the reduced solution space). Similarly, the Haghighi et al. (2011) solution violates the minimum node-pressure requirement at Nodes 13, 24 and 25. By contrast, the new solutions in this article are all feasible.

Figure 3(a) illustrates the progress of the proposed optimization algorithm for the best solutions achieved for both the full and reduced solution spaces. For the full solution space, a cost reduction from 294,152,000 Rupees at the start of the optimization to (131,003,000 Rupees and 73,500 function evaluations) was achieved. The algorithm converged at (125,460,980 Rupees and 436,000 function evaluations). Also, fast reductions in cost from 175,535,000 Rupees at the start, i.e. at zero function evaluation, to (131,184,000 Rupees and 8,000 function evaluations) and then (127,850,000 Rupees and 25,000 function evaluations) were achieved for the reduced solution space. The algorithm finally converged at (125,826,425 Rupees and 82,400 function evaluations). It is worth emphasizing, also, that the algorithm found *dozens* of feasible solutions that are cheaper than the solutions found by Kadu et al. (2008) and Haghighi et al. (2011) as summarised in Table 3. For example, for the full solution space and a population size of $N_p=500$, each of the mutation rates i.e. $p_m=0.005$ and $p_m=0.05$ achieved *30 solutions* that are cheaper than Kadu et al. (2008) and Haghighi et al. (2011) in a *single* optimization run. For the total 705 optimization runs executed, the proposed algorithm discovered more than 3,800 individual solutions approximately (with mean, median and standard deviation of Rs128,940,641, Rs129,006,500 and Rs1,424,767, respectively) that

Fig. 3 Progress graphs (a) and Pareto-optimal fronts (b) for the best optimization runs for the Kadu network



(a) Evolution of the cost of the cheapest feasible solution



(b) Frontier-optimal solutions

are feasible and cheaper than the previous best solution of Rs131,312,815 by Haghghi et al. (2011). The distribution of these new solutions is shown in the Appendix. These results illustrate clearly the *high evolutionary sampling efficiency* of the proposed algorithm. In other words, the number of solutions evolved and analysed on average before finding a near-optimal solution is small in comparison to the size of the solution space. Also, the small distance between the graphs in Fig. 3(a) for the full and reduced solution spaces is worth a mention, considering that the reduced solution space is approximately a factor of 10^{18} smaller than the full solution space.

The pareto-optimal fronts for the best optimization runs are shown in Fig. 3(b) for both the full and reduced solution spaces. In the full solution space, all permissible pipe sizes were included. However, in the reduced solution space, for each pipe, the pipe sizes that are unlikely to be feasible and/or competitive were excluded, leaving only 3–5 options in each case. Consequently, for the low-cost solutions, the cost-ratio values for the full solution space are smaller than the reduced solution space. It can be seen in Fig. 3(b) that the front for the full solution space (in which $Np=500$) has a higher density of solutions than the front for the reduced solution space (in which $Np=200$), and some solutions with the smallest cost-ratio values are missing in the front for the reduced solution space.

6 Conclusions

Application of the proposed optimization algorithm, so far, has been relatively straightforward. In total, 705 optimization runs were executed for the Kadu et al. (2008) network for which there were 352.5 million hydraulic simulations. The results suggest that both the optimization and hydraulic simulation algorithms are efficient and reliable. The penalty-free multi-objective evolutionary algorithm proposed uses pressure-dependent analysis. This accounts for the pressure dependency of the nodal flows and obviates the need for penalties or tournament selection procedures to address violations of the nodal pressure constraints. It is encouraging, also, that the algorithm seems reasonably stable with respect to the mutation rate. This suggests that fine tuning of the mutation rate may not be essential.

The whole-life design, rehabilitation and upgrading model developed takes into consideration the deterioration over time of both the structural integrity and hydraulic capacity of every pipe. Both direct and indirect failure costs are included. The upgrading options of paralleling and replacement of pipes are considered along with the timing. The hydraulic reliability and the overall cost are considered when choosing the best few design options to recommend. For the whole-life design optimization problem considered (Tanyimboh and Kalungi 2008), a solution was found that is both cheaper and more reliable than the previous solution in the literature based on linear programming and entropy maximization. The results achieved are consistent with the previous results (Tanyimboh and Kalungi 2008) and demonstrate the benefits of the whole-life design optimization approach. The timing of the rehabilitation and upgrading is important. Therefore it should be optimized along with the pipe diameters. It would be useful, also, to extend the whole-life design optimisation problem to address a water supply system that involves other components such as pumps, valves and multiple storage facilities.

For the benchmark optimization problem concerned with the initial construction cost only (Kadu et al. 2008), *thousands of solutions that are both fully feasible and cheaper* than the best known solutions in the literature were found. These results provide encouragement to extend the proposed approach to even more challenging optimization problems involving, for example, much larger water distribution networks in the real world and/or more complex aspects such as optimal pump scheduling and tank operation that require extended period simulation. Finally, results for the reduced solution space demonstrated a significant reduction in the number of function evaluations needed to find optimal solutions. This strongly suggests a need for further research to develop an efficient solution space reduction technique that is capable of selecting appropriate candidate pipe sizes dynamically.

Acknowledgments This research was funded in part by the UK Engineering and Physical Sciences Research Council (EPSRC Grant EP/G055564/1), the British Government (i.e. Universities UK's Overseas Research Students' Award Scheme) and the University of Strathclyde. The authors acknowledge the financial support with thanks.

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

- Awumah K, Goulter IC, Bhatt SK (1990) Assessment of reliability in water distribution networks using entropy-based measures. *Stoch Hydrol Hydraul* 4(4):325–336
- AWWA (2012) Buried no longer: confronting America's water infrastructure challenge

- 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
- Cullinane MJ, Lansey KE, Mays LW (1992) Optimization-availability-based design of water distribution networks. *J Hydraul Eng* 118(3):420–441
- Dandy GC, Engelhardt MO (2001) Optimal scheduling of water pipe replacement using genetic algorithms. *J Water Resour Plan Manag* 127(4):214–223
- Deb K (2000) An efficient constraint handling method for genetic algorithms. *Comput Methods Appl Mech Eng* 186(2):311–338
- 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. *J Comput Aided Civ Infrastruct Eng* 23(8):625–635
- Haghighi A, Samani HMV, Samani ZMV (2011) GA-ILP method for optimisation of water distribution networks. *J Water Resour Manag* 25(7):1791–1808
- Kadu MS, Gupta R, Bhave R (2008) Optimal design of water networks using a modified genetic algorithm with reduction in search space. *J Water Resour Plan Manag* 134(2):147–160
- Rossman LA (2002) EPANET 2 user's manual. Water Supply and Water Resources Division, National Risk Management Research Laboratory, Cincinnati, OH45268
- Saleh S, Barlow E, Tanyimboh T (2012) Unbiased and accurate assessment of surrogate measures of hydraulic reliability of water distribution systems. 14th WDSA Conference, Adelaide, Australia, 24–27 September 2012
- Setiadi Y, Tanyimboh TT, Templeman AB (2005) Modelling errors, entropy and the reliability of water distribution systems. *Adv Eng Softw* 36(11–12):780–788
- Seyoum AG, Tanyimboh TT (2013) Pressure dependent network water quality modelling. Proceedings of ICE: Water Management. doi:10.1680/wama.12.00118
- Shannon C (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–428
- Sharp WW, Walski TM (1988) Predicting internal roughness in water mains. *J AWWA* 80(9):34–40
- Siew C (2011) A penalty-free multiobjective evolutionary optimization approach for the design and rehabilitation of water distribution systems. PhD thesis, Department of Civil Engineering, University of Strathclyde Glasgow, UK
- Siew C, Tanyimboh TT (2010) Penalty-free multi-objective evolutionary optimization of water distribution systems. 12th WDSA Conference, Tucson, Arizona, September 2010
- Siew C, Tanyimboh TT (2011) Penalty-free evolutionary algorithm optimization for the long-term rehabilitation and upgrading of water distribution systems. World Env. & Water Resources Congress, Palm Springs, California, ISBN 978-0-7844-1173-5, pp 214–223
- Siew C, Tanyimboh TT (2012a) Penalty-free feasibility boundary-convergent multi-objective evolutionary algorithm for the optimization of water distribution systems. *Water Resour Manag* 26(15):4485–4507. doi:10.1007/s11269-012-0158-2
- Siew C, Tanyimboh TT (2012b) Pressure-dependent EPANET extension. *Water Resour Manag* 26(6):1477–1498
- Tanyimboh TT, Kalungi P (2008) Optimal long term design, rehabilitation and upgrading of water distribution networks. *Eng Optim* 40(7):637–654
- Tanyimboh TT, Kalungi P (2009) Multicriteria assessment of optimal design, rehabilitation and upgrading schemes for water distribution networks. *Civ Eng Environ Syst* 26(2):117–140
- Tanyimboh TT, Setiadi Y (2008) Joint layout, pipe size and hydraulic reliability optimization of water distribution systems. *Eng Optim* 40(8):729–747
- Tanyimboh TT, Sheahan C (2002) A maximum entropy based approach to the layout optimization of water distribution systems. *Civ Eng Environ Syst* 19(3):223–253
- Tanyimboh TT, Templeman AB (1993) Calculating maximum entropy flows in networks. *J Oper Res Soc* 44(4):383–396
- Tanyimboh TT, Templeman AB (1998) Calculating the reliability of single source networks by source head method. *Adv Eng Softw* 29(7–9):449–505
- Tanyimboh TT, Templeman AB (2000) A quantified assessment of the relationship between the reliability and entropy of water distribution systems. *Eng Optim* 33(2):179–199
- Todini E (2000) Looped water distribution networks design using a resilience index based heuristic approach. *Urban Water J* 2(3):115–122
- Twort AC, Ratnayaka DD, Brandt MJ (2000) Water supply. Arnold, London
- Wagner JM, Shamir U, Marks DH (1988) Water distribution reliability. *J Water Resour Plan Manag* 114(3):253–294