H$_2$Open Journal

H$_2$Open Journal Vol 6 No 3, 343 doi: 10.2166/h2oj.2023.029

Optimization of urban water pipe network design using fast-messy genetic algorithms (fmGA)

Berhanu Fanta Alemaw * and Tshepo E. Jankie
Department of Geology, University of Botswana, Gaborone, Botswana
*Corresponding author. E-mail: alemaw@ub.ac.bw; bfalemaw@gmail.com

BFA, 0000-0002-2588-9016

ABSTRACT

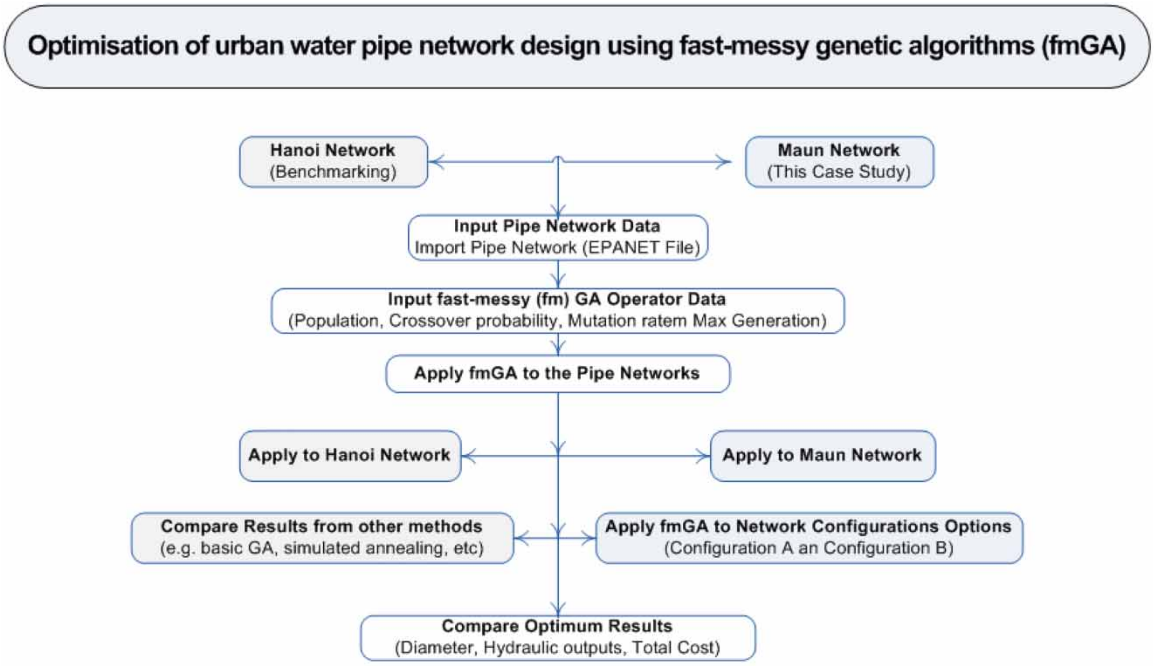
To have an efficient water distribution network, optimal design alternatives need to be identified and analysed using combined hydraulic modelling and optimization. This paper reports on the application of the fast-messy genetic algorithm (fmGA) coupled with hydraulic modelling tool EPANET to assess and select an optimum design and operational alternative for a water distribution pipe network. The sole objective of the optimization modelling was to minimize network costs subject to hydraulic and design constraints. The fmGA was first tested using the benchmark case study of the Hanoi network and then applied to the real network of Maun, Botswana which is considered as the case study. We have compared our results of the fmGA model application with other optimization techniques applied to the Hanoi network. The findings of the test revealed that the fmGA is superior to other popular metaheuristic optimization methods in terms of processing speed with comparable accuracy and pressure constraints fulfilled for all nodes. It also provides the best solution. For the water distribution network of Maun, the best pipeline configuration and route were determined, in which Configuration B is found to be the best and least cost solution to improve the water supply situation in Maun.

Key words: EPANET, fast-messy genetic algorithm, hydraulic modelling, optimization, water distribution network, water supply

HIGHLIGHTS

- A combined hydraulic modelling and optimization modelling has been employed to enable the design of an efficient water distribution network, where optimal design alternatives are identified and analysed.
- This paper employs the fast-messy genetic algorithm (fmGA) coupled with the hydraulic modelling tool EPANET to assess and select an optimum design and operational alternative for a water distribution pipe network.

GRAPHICAL ABSTRACT



HIGHLIGHTS

The proposed model improves on existing optimization models, which was used to minimise network cost as a basis for pipe network hydraulic design and network optimization.

The proposed fmGA Model was first tested using the well-know benchmark system of the Hanoi pipe Network, and it is found to be superior to other popular metaheuristic methods (basic GA, simulated annealing, shuffled frog leaping algorithm, shuffled complex evolution) in terms of processing speed while meeting all design constraints.

The model was then applied to the real network of Maun, Botswana to make decisions on conflicting situations of alternative pipe network and pumping configurations are to be evaluated.

The developed fmGA model was used to explore opportunities for pipe networks with pumping and alternative pipe configurations that was used to find the best trunk network pipeline route with least cost.

INTRODUCTION

Over the years, cities and urban centres in Botswana have water distribution systems that have expanded and face challenges due to unprecedented growth of the city dimensions, population and water demand brought by urbanization and economic activities (Government of Botswana 2016). The water systems have been faced with major problems of high water losses due to leakage, recurrent water shortages due to drought and the need for infrastructure upgrades to augment water supply shortfalls.

Over the past three decades, numerous studies have been undertaken to integrate optimization techniques with hydraulic models in an effort to optimize the design and operation of water distribution networks. Techniques that have been used include deterministic methods, that involve linear programming methods (e.g. Cisty 2010), non-linear programming methods (e.g. Fujiwara & Khang 1990) and dynamic programming methods (e.g. He & Hui 2006; Martínez-Bahena *et al.* 2018; Ramírez *et al.* 2021). Other techniques also involve stochastic or metaheuristic methods including, among others, simulated annealing (e.g. Kirkpatrick *et al.* 1983; Cunha & Sousa 1999, 2001), tabu search (e.g. Fanni *et al.* 2000) and genetic algorithms (GAs) (e.g. Savic & Walters 1997; Wu 1998; Van Dijk *et al.* 2008).

Today, it is widely accepted that GAs are the best option in water distribution network optimization due to their reliability, robustness and rapid speed of search. The literature review shows that a lot of research has focused on improving methods of optimization and testing such new techniques on well-researched example problems. There are also notable studies that attempted to solve optimization problems of real-world water distribution networks. Water distribution pipe network problems were earlier solved as linear problems (Alperovits & Shamir 1977). According to computational complexity theory, the water distribution design problem is classified as a complex one because the computational effort grows exponentially as the instance size increases. To solve these kinds of problems, it is necessary to use optimization methods such as metaheuristics (Savic & Walters 1997; Liberatore & Sechi 2009; Cruz-Chávez *et al.* 2014).

Through a combination of different concepts derived from artificial intelligence, biological evolution and statistical mechanisms, metaheuristics provide a general framework for creating new hybrid algorithms (Savic & Walters 1997; Liberatore & Sechi 2009). Reca & Martínez (2006) used a GA in conjunction with the EPANET (EPA 2021) hydraulic network solver to optimize a real complex irrigation network and obtained satisfactory results. Meanwhile, Wu & Sage (2006) showed that GA optimization can also be used for water loss detection through hydraulic model calibration. Prasad & Park (2004) applied a GA to widely researched water distribution design problems and produced better results than previous works. GAs have been applied extensively to optimize water distribution systems, despite the high computational intensity involved (Savic & Walters 1997; Vairavamoorthy & Ali 2000; Keedwell & Khu 2005; Shau *et al.* 2005; van Dijk *et al.* 2008). Today, it is widely accepted that GAs are the best option for water distribution network optimization due to their reliability, robustness and rapid speed of search.

The messy GA to investigate whether a GA is able to identify and recombine building blocks to form optimal solutions which were developed by Goldberg *et al.* (1989). The messy GA encodes individuals (parameters) as a vector of pairs, where each pair specifies the location and value of a bit. Special rules deal with over-specified (duplicate) or underspecified (missing) bits forming an individual (parameter). The fast-messy genetic algorithm (fmGA) improves upon the messy GA by reducing the computational efforts in the early phases of the messy GA (Goldberg *et al.* 1989). The gene expression messy GA enhances the messy GA to actively search for linkage relationships within a search space (Kargupta 1996; Wu & Garibay 2002). A fmGA is a special clone of a common simple GA. This type represents a new, more powerful kind in the GA branch. It resists premature local-minimum fall and solves problems in a shorter time (Goldberg *et al.* 1993). The fmGA is assessed as an optimization tool and then used to optimize the real trunk network adopted in this study. The current study employing the fmGA was evaluated against other algorithms developed and applied to similar optimization.

The Shuffled Frog Leaping Algorithm (SFLA) is a metaheuristic for solving discrete optimization problems that were recently applied to determine optimal discrete pipe sizes for new pipe networks and for network expansions (Eusuff & Lansey 2003). The shuffled complex evolution (SCE) algorithm is based on a GA that combines the best features of multiple complex shuffling and competitive evolution based on a simple search method (Atiquzzaman & Liong 2004). The SCE algorithm has been applied to the optimization of water distribution networks (Liong & Atiquzzaman 2004).

Cunha & Sousa (2001) applied and recommended a simulated annealing method that can be used to find the least-cost design for a gravitational looped water distribution network. Simulating annealing (SA) algorithm is one of the most preferred heuristic methods for solving optimization problems. SA was introduced by inspiring the annealing procedure of metalworking (Kirkpatrick *et al.* 1983).

Non-dominated Sorting Genetic Algorithm II (NSGA-II) alleviates all shortcomings of multi-objective evolutionary algorithms (MOEAs) that use non-dominated sorting and sharing. The MOEAs have been criticized mainly for three main difficulties related to (1) computational complexity due to a large number of objectives and population size; (2) their non-elitism approach and (3) the need to specify a sharing parameter; all these challenges are well addressed in NSGA-II, as noted in Deb *et al.* (2002).

With increasing computing power, there have been advancements in the field of water supply design which evolved the emergence of a number of software packages for hydraulic analysis and modelling of water distribution networks. They include both public domain and commercial software including EPANET (EPA 2021), Branch (USGS 2021), H2Onet (Water Simulation 2021), HydrauliCAD (Pipeflow 2021), WaterCAD (Bentley 2021a), HYDROFLO3 (CESD 2021), Pipe2014 (KYPipe 2021), Synergi Water (DNV 2021), WaterGEMS (Bentley 2021b), among others. A review of various software available for designing and modelling water distribution networks is noted in Sonaje & Joshi (2015).

In this article, EPANET was used to develop two alternative simple hydraulic models of the Maun trunk water distribution network. The models were built from existing network GIS data. The viability of the algorithm was first tested using the benchmark case study of the Hanoi network and then it was applied to the real trunk network of Maun of the water supply scheme, located in Northern Botswana. Detailed information and hydraulic data on the Hanoi water distribution network are provided in Fujiwara & Khang (1990). The Hanoi Network offers simplicity in that it was first optimized by Fujiwara & Khang (1990) using a two-phase non-linear programming method, which then has subsequently been used as a case study for a number of optimization techniques including GAs.

This paper finally demonstrates how fmGA can be used to optimize a hydraulic system coupled with the public domain hydraulic software, EPANET and then applied to assess and select an optimum design alternative. Fortran codes were used to iteratively prepare inputs and simulate hydraulic results using EPANET. EPANET is not just a competitor to other software but is also the engine for many current ones offered commercially. Its open-source nature and versatile functionality have made it a popular choice among developers and users alike for hydraulic and water quality modelling of water distribution systems. Model calibration and sensitivity analysis were performed to determine the sensitivity and determined parameters that provided accurate and fast convergence in determining the minimum cost of a benchmark Hanoi network.

METHODS AND DATA

Methodology

The fmGA was used in this study to achieve network optimization of proposed pipe network configurations. The fmGA was first introduced by Goldberg *et al.* (1993) as a solution to the 'initialization bottleneck' problem experienced with the standard messy genetic algorithm (mGA). Messy GAs are a type of GA that are capable of processing variable-length strings (trial solutions) that may be under- or over-specified with respect to the problem being solved (Goldberg *et al.* 1989; Wu *et al.* 2012). The original GA only processes fixed-length and fixed-coded strings, which makes it unlikely to solve NP-hard and 'deceptive' problems (such as complex water distribution networks). The messy GA can solve such complex problems; however, they are hindered by the aforementioned problem. The steps and flow chart of the messy GA are presented in Figure 1.

Problem formulation

The fmGA divides the optimization process into two phases; the initial primordial phase and the juxtapositional phase. It is the misappropriation of processing time between these processes that brings about the infamous 'initialization problem'. This can be described as a primordial phase.

During the primordial phase, the mGA initializes by adopting enumeration for all strings of length, k, to create an initial population of size (n) given as $n = 2^k \binom{l}{k}$, where l is the problem size (determined by the number of decision variables) and k is the string length. Consequently, the number of evaluations of the initial population increases exponentially as the problem length l and/or the order k of the building blocks increases. This gives an initialization processing time of $O(l^k)$ while the juxtapositional processing time is $O(l \log l)$. This is a significant difference in processing time causing a bottleneck at the initialization phase (Goldberg *et al.* 1993).

The fmGA overcomes the initialization problem by using probabilistically complete initialization and a gene (building block) filter procedure. Both techniques effectively speed up the search process of the original messy GA (mGA). While the original mGA enumerates all order-k building blocks in a deterministic manner, probabilistically complete initialization allows the same order-k building blocks to be defined using a smaller number of strings. The probabilistically complete initialization technique generates an initial population of strings with length of $l - k$. Building blocks of order-k are filtered out by a gradual reduction of string length through the random deletion of genes. This process of selection and random deletion at specified intervals is called building block filtering (Wu 1998).

Least-cost optimization of water distribution networks is based on finding a combination of pipe sizes which yield minimum cost for a given network layout and demand pattern. The model is subject to constraints of (1) the continuity equation, (2) the energy equation and (3) minimum/maximum pressure at nodes. In any optimization problem, the objective function needs to be first formulated. The only decision variable in the design problem is pipe size (diameter), therefore the objective function becomes a cost function of pipe diameters and lengths.

$$C_T = \sum_{i=1}^{N} c_i(D_i) \cdot L_i \tag{1}$$

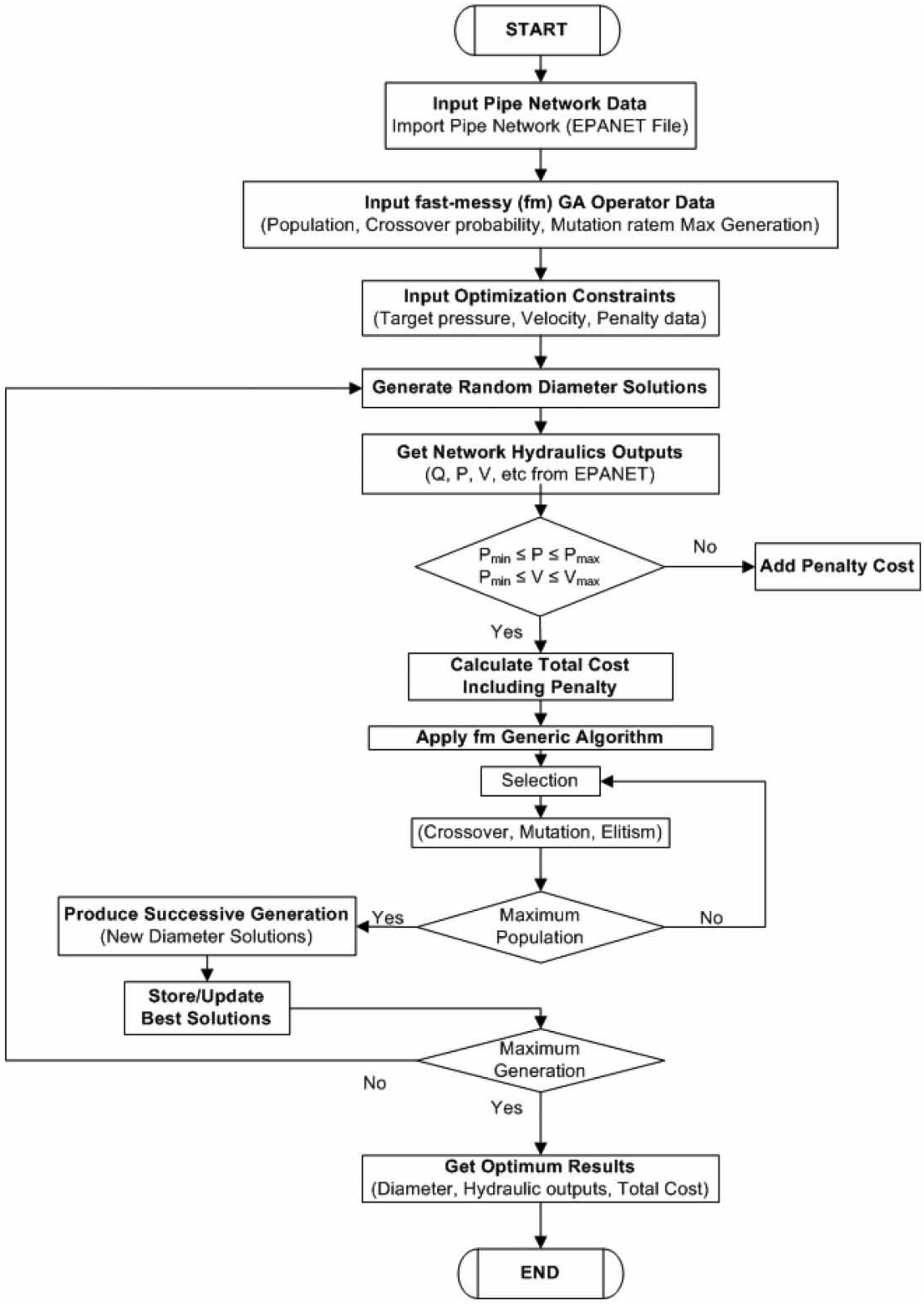


Figure 1 | The steps and flowchart of the messy GA optimization process.

In which C_T is the total network cost, N is the total number of pipes, $c_i(D_i)$ is the cost per unit length of pipe i with diameter D_i and L_i is the length of pipe i.

A hydraulic model such as WaterGEMS or EPANET can simulate the flow rate in pipes, pressure at nodes and height of water in the tanks. The hydraulic models apply network mass conservation (Equation (2)) and energy conservation (Equation (3)) to compute the flow rates and pressures, respectively, in the network.

The continuity equation, also known as the law of conservation of mass, dictates that the difference between inflow and outflow at a node must equal the demand and/or storage at that particular node. Continuity equations are solved at every node yielding a set of algebraic equations in terms of flow.

$$\sum Q_{\text{in}} - \sum Q_{\text{out}} = \sum Q_{\text{ext}} \tag{2}$$

where Q_{in} is the flow entering node; Q_{out} is the flow exiting node and Q_{ext} is the external demand or supply at node.

The law of conservation of energy, also known as the Bernoulli principle, has to be satisfied for each link (pipe). It dictates that difference in energy at any two points connected in a network must be equal to the energy gains from pumps and energy losses caused by friction in pipes and fittings that occur in the path between them. Most importantly, this means that the sum of the head losses around a loop must equal zero. This can be expressed as follows:

$$H_1 + \sum h_p = H_2 + \sum h_L + \sum h_M \tag{3}$$

where H_1 and H_2 is total head at point 1 and 2, respectively; h_p is the head gains due to pumping; h_L is head losses due to pipe friction (h_f) and h_M is minor head losses due to valves and fittings.

The friction head loss can be calculated using the Hazen–Williams (Equation (4)):

$$h_f = \frac{6.8193}{D^{1.167}}\left(\frac{V}{C}\right)^{1.1852} \text{ or } h_f = \frac{w\,L\,Q^{1.1852}}{C^{1.1852}D^{4.871}} \tag{4}$$

where L is the length (m); D is the diameter (m); V is the velocity (m/s); g is the gravitational acceleration (m/s²); C is the Hazen–Williams friction factor and ω is the numerical conversion constant ($\omega = 10.667$ is adopted in this paper).

The minimum pressure head constraint at each node is given as follows:

$$H_i \geq H_{i\min}, \quad i = 1, 2, 3, \ldots, \mathrm{N} \tag{5}$$

where H_i is head at node i, $H_{i\min}$ is minimum required head at node i and N is the number of nodes.

The network solution technique used is the Gradient Method. The method was first proposed by Todini & Pilati (1987) and is based on an algorithm that is capable of solving complex and large water networks. This method essentially employs the Newton-Raphson technique, however, with the added ability of simultaneously taking into account both nodal heads and pipe flows as equations of mass and energy are solved (Paluszczyszyn 2015). The gradient method is trusted as a robust and computationally effective solution technique such that it forms the basis of the widely used freeware modelling software, EPANET (Rossman 2000), as well as several commercial grade softwares including WaterCAD and WaterGEMS. A brief presentation of the gradient algorithm solution method is given below (Mays 2000).

First, the non-linear energy equations are linearized using previous flow estimates to yield the full expression of the network response in matrix form:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix}\begin{bmatrix} Q \\ h \end{bmatrix} = \begin{bmatrix} -A_{10}h_0 \\ q \end{bmatrix} \tag{6}$$

where A_{11} is the diagonal matrix containing the linearization coefficients; A_{12} ($=A_{21}{}^T$) is the incidence matrix of unknown head nodes; A_{10} is the incidence matrix of fixed head nodes; Q is pipe flow; h is nodal head; q is nodal demand and h_0 is fixed nodal head.

Differentiating both sides of the equation yields:

$$\begin{bmatrix} \beta A_{11} & A_{12} \\ A_{21} & 0 \end{bmatrix}\begin{bmatrix} dQ \\ dh \end{bmatrix} = \begin{bmatrix} dE \\ dq \end{bmatrix} \tag{7}$$

where dE and dq are residuals of pipe equations to be solved iteratively for flow and head; and βA_{11} is the diagonal matrix of the exponents of the pipe equations.

The equation above is a set of linear equations in terms of dQ and dh. Once solved, Q and h are updated by the following equations:

$$Q_{k+1} = Q_k + dQ \tag{8}$$

$$h_{k+1} = h_k + dh \tag{9}$$

Subsequent updates continue until residuals of dE and dq are reduced to zero, thereby reaching convergence.

Benchmarking the model

The developed fmGA optimization model was tested against benchmark problems in order to establish its functionality and efficiency. Interactive Fortran code was developed and used to test common networks for which many optimizations have been performed; these include traditional and heuristic methods (Savic & Walters 1997). Two case study systems were tested: Hanoi (Case study 1) and Maun network (Case study 2) detailed below are tested against and discussed below.

In both these case studies, the widely used Hazen–Williams equation, embedded within EPANET, is utilized to determine the friction loss in a pipe link between two nodes. Previous researchers have investigated the Hanoi system (Case study 1) extensively and obtained numerous solutions that met the defined fitness function of minimum cost based on the constraint of required pressure and demand at every node. Different researchers obtained different solutions mainly due to different interpretations of the Hazen–Williams equation (Savic & Walters 1997) which made direct comparison not always possible as noted in Van Dijk *et al.* (2008).

Data used

The fmGA optimization was first tested using the benchmark case study of the Hanoi network and then applied to the real network of Maun, Botswana which is considered as the case study.

Case study 1: Hanoi network

The Hanoi water distribution network is shown in Figure 2. Data for the Hanoi network was retrieved from Fujiwara & Khang (1990). The network consists of 1 reservoir (node 1), 31 demand nodes and 34 pipes. The minimum pressure head required at each node is 30 m. Pipe lengths range between 100 and 3,500 m and the demand at each node diameter ranges 60–1,345 m^3/h. The set of commercially available pipe diameters and their unit cost

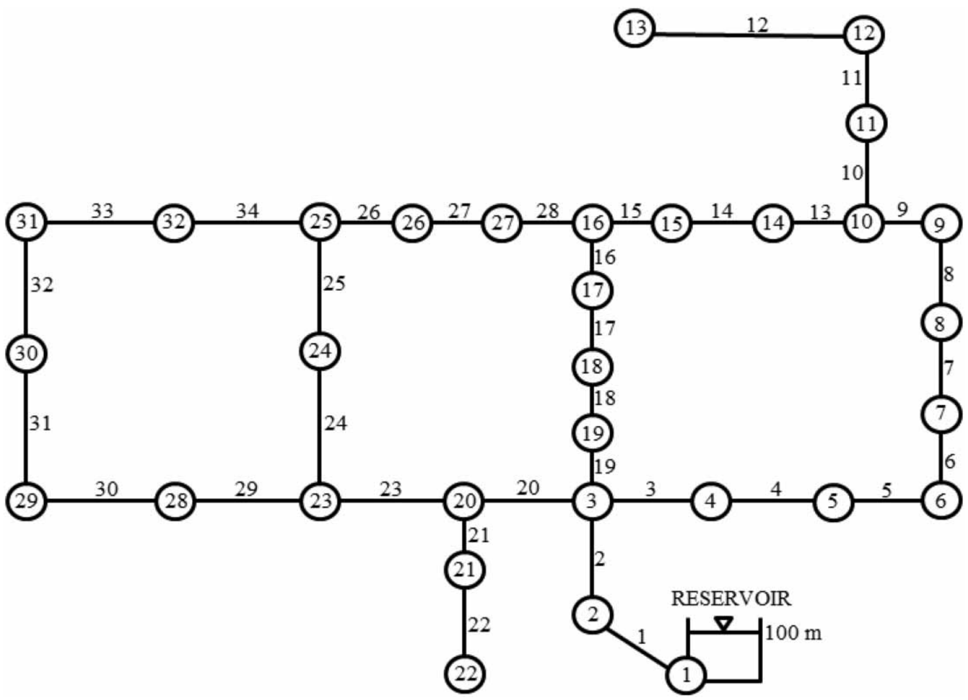


Figure 2 | Hanoi trunk network.

is given in Table 1. The total search space for the network is 6^{34} different possible network designs. The Hazen–Williams coefficient for all links is 130. Pipe cost is calculated as $C = 1.1(aD)^{1.5}$, where C is the cost per metre length (cost unit) and D is the pipe diameter in inches where $a = 1$, and in mm where $a = 0.039$.

Case study 2: Maun network

The Maun water distribution system (Figure 3) has two main conveyance networks with pipe and pump configurations that were subject to evaluation in this study, shown in Figures 4 and 5. The entire Maun water distribution network consists of 2,943 pipes, 3 distribution reservoirs, 7 pump stations (housing 18 pumps) and 9 storage tanks. The total pipe length is 413 km. After treatment, both surface water (river) and groundwater are pumped into the network. In 2014–2015, total water abstraction was estimated at 10,137 m^3/day, with 80% attributed to groundwater sources (Government of Botswana 2016). Demand is estimated at 8,319 m^3/day. Recently, the network has been struggling to meet demand due to low river levels, inefficient treatment facilities and malfunctioning boreholes.

Each zone in Maun is supposed to be supplied by zonal tanks to meet the respective zonal demands as shown in Table 2. The nodal level zonal demands in Configurations A and B are summarized in Tables 3 and 4,

Table 1 | Available pipe diameters and associated cost

Diameter	(Inches)	12	16	20	24	30	40
	(mm)	300	400	500	600	700	1,000
Unit cost	(cost unit/metre)	45.73	70.40	98.39	129.33	180.75	278.28

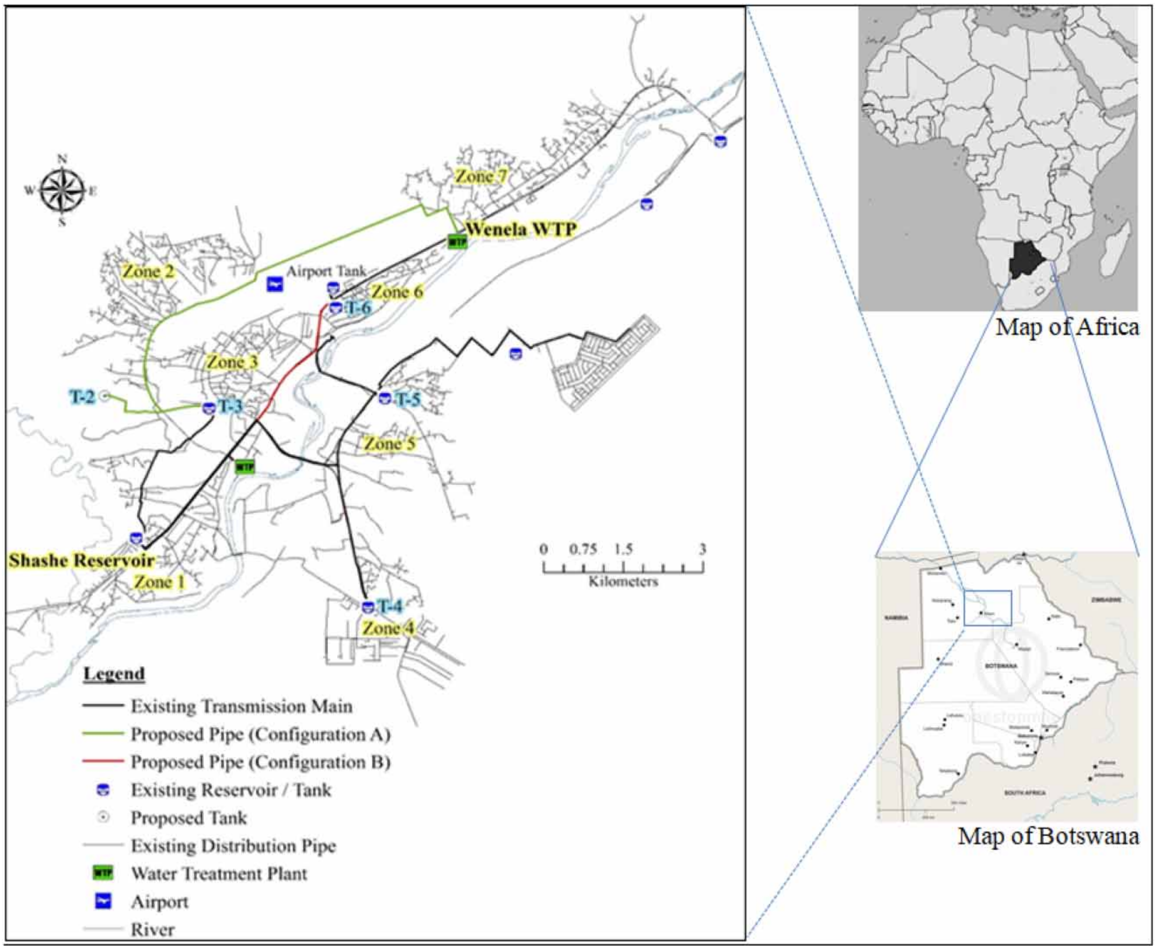


Figure 3 | Maun water distribution network and proposed pipelines.

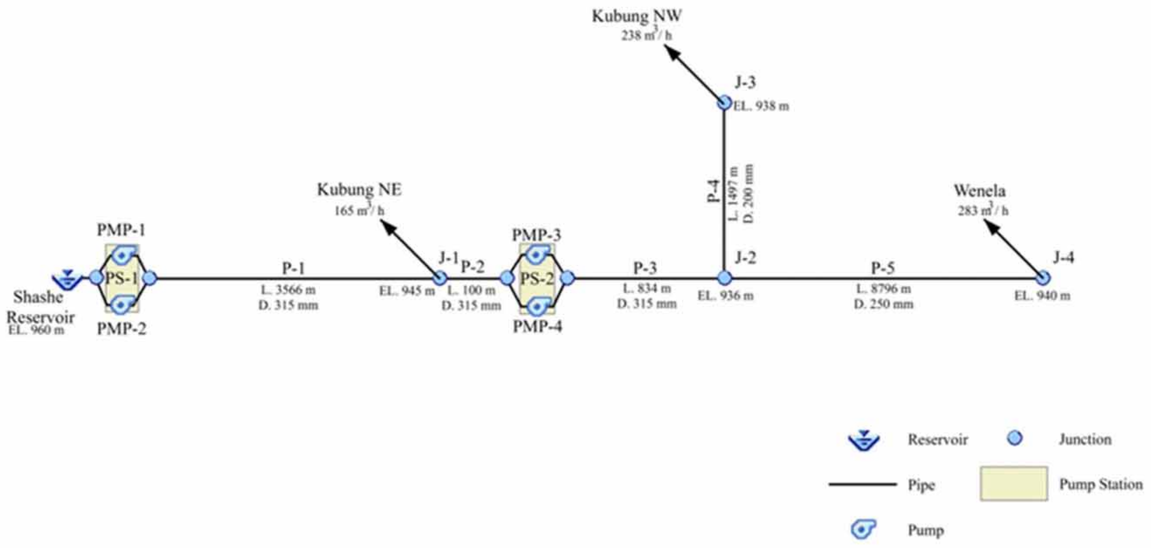


Figure 4 | Profile of design Configuration A with pipe numbers, nodes and supply sub-areas with demands.

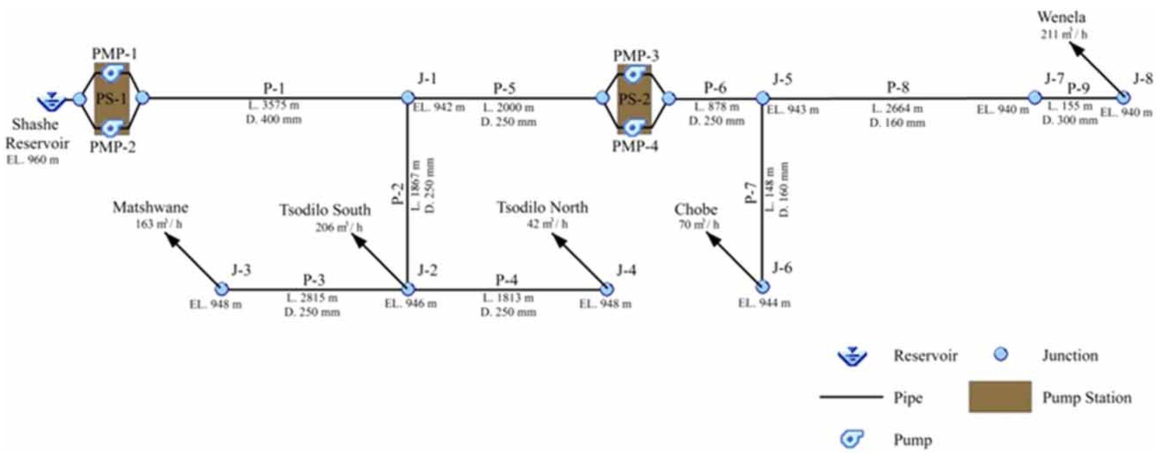


Figure 5 | Profile of design Configuration B with pipe numbers, nodes and supply sub-areas with demands.

Table 2 | Current and future zonal demand estimates

Zone	Supply tank	Baseline (m^3/h)	2020 (m^3/h)	2025 (m^3/h)
2	T-2	211.70	222.57	237.61
3	T-3	146.84	154.43	164.92
4	T-4	145.11	152.61	162.98
5	T-5	221.21	232.65	248.46
6	T-6	62.58	65.85	70.38
7	WTP	188.19	197.87	211.27

respectively. Pipe of pressure classes of 160 m (16 bar) was considered with a minimum pressure 25 m that is also required to meet local standards. Pumps are used to achieve flow deliveries in the network to overcome higher elevations at zonal tanks and demand sites/nodes, without violating maximum pressure classes of pipes. The performance curve of the pump stations used in each configuration of Maun is shown in Table 5.

Table 3 | Demand patterns for Configuration A

	Flow rate (m^3/h) for design event		
Node	**[1] Baseline**	**[2] 2020**	**[3] 2025**
3	146.84	154.43	164.92
5	0.00	0.00	0.00
6	211.70	222.57	237.61
7	250.77[a]	263.72[a]	281.65[a]

[a]Combined demand for zones 6 and 7.

Table 4 | Demand patterns for Configuration B

	PHD (m^3/h) for design event		
Node	**[1] Baseline**	**[2] 2020**	**[3] 2025**
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	145.11	152.61	162.98
6	221.21	232.65	248.46
8	0.00	0.00	0.00
9	62.58	65.85	70.38
10	188.19	197.87	211.27

Table 5 | Pump station performance curve of PS1 and PS2

Pump station 1 (PS1)		Pump station 2 (PS2)	
Flow (m^3/h)	**Head (m)**	**Flow (m^3/h)**	**Head (m)**
500	90	294	104
600	85	400	95
700	80	440	90
800	72	500	82
898	66		
900	3		

To alleviate production challenges at the Wenela water treatment plant (WTP), this study focuses on finding an optimum trunk mains pipeline route from the Shashe Wellfield reservoir to Wenela WTP. This involves designing two alternate routes (Configuration A, Figure 4 and Configuration B, Figure 5) for choosing the most economically viable route. Both routes were selected based on topography and existing street plans.

For the Maun network, Table 6 shows the available commercial pipe diameters and assumed unit costs. This applies to both network configurations. A Hazen–Williams C-factor of 130 is assumed for all pipes.

For the Maun network, the optimization criteria were based on (1) satisfying the demand growth from 2016 to 2025 and (2) satisfying the minimum and maximum junction pressure head of 25 and 160 m, respectively.

RESULTS AND DISCUSSION

Case study 1: Hanoi network

The Hanoi network example has been tested using various network optimization algorithms and schemes including the GA scheme (Savic & Walters 1997; Van Dijk *et al.* 2008), SA scheme (Cunha & Sousa 1999), SFLA

Table 6 | Available pipe diameters and associated cost

Diameter (mm)	Unit cost (cost unit per metre length)
100.0	35.00
150.0	41.00
200.0	49.00
250.0	59.00
300.0	72.00
350.0	88.00
400.0	107.00

scheme (Eusuff & Lansey 2003) and SCE scheme (Liong & Atiquzzaman 2004). Results from these scheme applications have been used to compare results and draw conclusions from this study.

For the case of the Hanoi network, the ideal parameters adopted were based on values shown in Table 7. The objective of the optimization problem was to determine the required pipe diameters that would yield the least total cost while still supplying the demand and adhering to the system constraint of minimum pressure at each node. The maximum number of trials and non-improvement solutions was set to 3,000,000 and 20,000, respectively. The ideal parameters (Table 7) were determined after approximately 15 iterative optimization runs.

EPANET hydraulic network model was used interactively with FORTRAN coded program in this study which employs the standard Hazen–Williams equation (Equation (4)). Only studies that used discrete-diameter methods similar to this work were considered for comparison. Other studies in the past have employed continuous-diameter methods (e.g. Fujiwara & Khang 1990) and split-pipe methods (e.g. Eiger *et al.* 1994). A comparison of the solutions of this study with those of the other authors is summarized in Table 8.

The distribution of pressure heads for the optimal network of Hanoi is shown in Figure 6. The solution was simulated with EPANET and none of the nodes violated the pressure constraint. Statistical values of pressure heads obtained in the study in comparison with other published studies using the Hanoi network (Model 1 to Model 6) are presented in Table 9. In summary, an optimal network cost unit of 6,109,049.00 was found by this study. The fmGA required approximately 1,300,000 function evaluations and 2 minutes and 46 seconds of CPU time.

Table 7 | Key GA parameters used in the Hanoi network

GA parameter	Value	GA parameter	Value
Population size	500	Mutation probability	0.010
Cut probability	0.017	Random seed	0.500
Splice probability	0.600	Penalty factor	100,000,000

Table 8 | Comparison of solution with other authors

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Author	Savic & Walters (1997)	Savic & Walters (1997)	Cunha & Sousa (1999)	Eusuff & Lansey (2003)	Liong & Atiquzzaman (2004)	Van Dijk *et al.* (2008)	This Study
Algorithm	GA	GA	SA	SFLA	SCE	GA	fmGA
ω	10.5088	10.9031				10.667	10.667
Cost (unit $\times 10^6$)	6.073	6.195	6.056	6.073	6.220	6.110	6.109
Function evaluations	10^6	10^6	53, 000	26, 987	25, 504	495	1.3×10^6
Run time	3 h	3 h	2 h	/	11 min	6.4 min	166 s

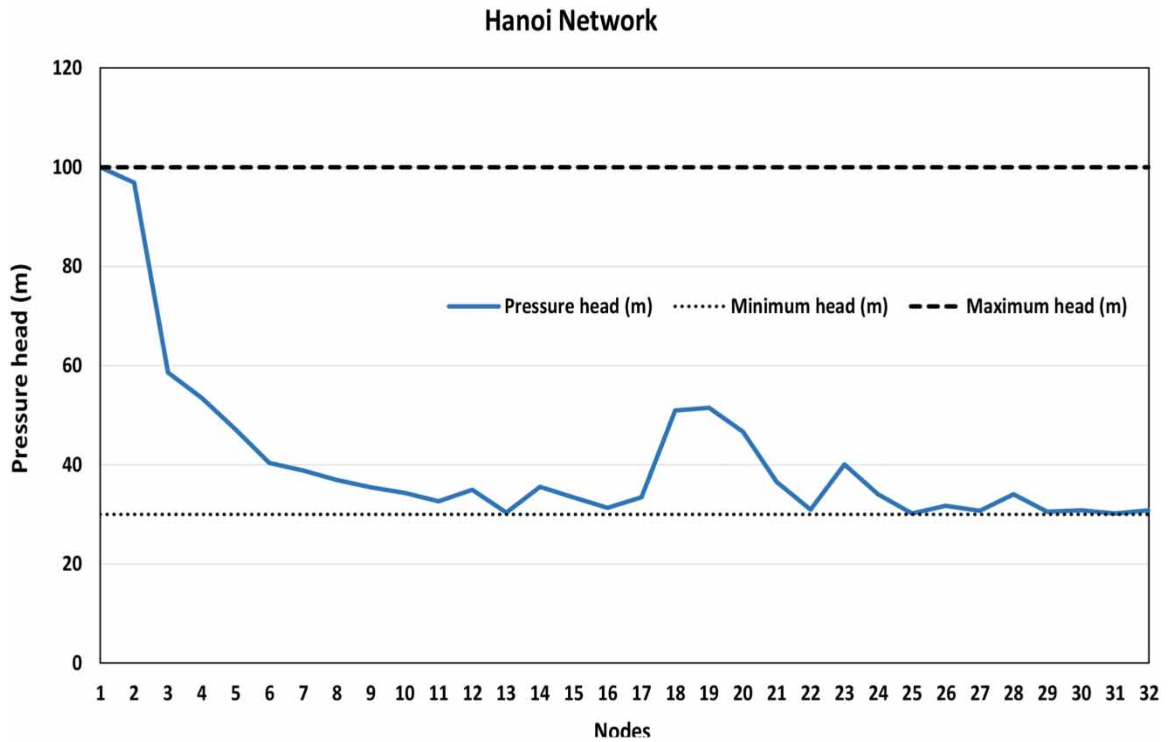


Figure 6 | Pressure head of Hanoi network calculated using fmGA.

Table 9 | Resultant pressure heads (metres) for the Hanoi network problem

Statistics	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Min head (m)	30.11	30.54	30.03	30.11	30.05	30.16	30.26
25%ile (m)	32.74	33.02	34.86	32.74	33.60	32.59	31.75
Mean head (m)	43.68	44.20	45.23	43.68	44.31	43.80	43.58
75%ile (m)	46.12	49.59	50.49	46.12	51.48	49.21	45.47
Max head (m)	100.00	100.00	100.00	100.00	100.00	100.00	100.00

There is no record of actual pressure heads at the various nodes of the Hanoi network and it was not possible to report on model accuracy except that of satisfaction of continuity of flow at each node and conservation of energy in each loop of the network, which is achieved in all the modelling cases. Of the seven models, three had computed hydraulic models with infeasible solutions of heads less than 30, namely Models 1, 3 and 4. To evaluate the performance of the proposed fmGA model (Model 7), results of Models 2, 5 and 6 were used. Two random sets of hydraulic heads at 20 nodes and 12 nodes were selected using simple random selection and were used as calibration and validation modes to present the results of Model 7 (fmGA) against the other metaheuristics GA models those which did not results at all nodes did not show infeasible solutions of pressure heads of $H_i < 30$ m, namely Models 1, 3 and 4. The result of model performance in terms of correlation coefficient and Nash–Sutcliffe model efficiency criteria is summarized in Table 10. Furthermore, the correlation plot of hydraulic heads at 20 calibration and 12 validation nodes is shown graphically in Figure 7, with a 45 degree line for the proposed fmGA model (Model 7) against Models 2, 5 and 6. All the models satisfy the minimum and maximum heads of 30 and 100 m, respectively.

Sensitivity analysis

Optimal solutions are not guaranteed in evolutionary algorithms; hence, maximizing 'near optimal' solutions is essential as noted in Mora-Melia *et al.* (2013). These parameters interact in a complex way, best parameter

Table 10 | Model performance on hydraulic heads of the Hanoi network problem

	fmGA vs. GA (Model 2)		fmGA vs. GA-SCE (Model 5)		fmGA vs. GA (Model 6)	
Mode	r	R^2	r	R^2	r	R^2
Calibration	0.997	99.3%	0.986	97.2%	0.993	98.6%
Validation	0.994	98.9%	0.809	65.4%	0.975	95.1%

combinations are searched to obtain better solutions with faster convergence (Eiben *et al.* 2007). A sensitivity analysis was carried out to determine the most influential parameters and to obtain the best parameter combinations for effective execution of the algorithm using input parameters that significantly affect the overall performance and speed of the Hanoi network. The parameters tested with increasing order of influence on minimum cost function are: Population Size; Crossover/Splice Probability; Mutation Rate and Cut Rate Probability. From literature, the most common parameters of the penalty function were set to 100×10^6, and the head-loss coefficients of the Hazen–Williams formula in the energy conservation equation were set to a fixed value of $\omega = 10.667$. Function evaluations were set between 1.0 to 150×10^4.

The sensitivity analysis was conducted to determine the influence a set of parameters had on predicting cost for the optimal network configuration for Hanoi water distribution. Figure 8 shows the sensitivity of the hydraulic network cost to variations of different parameters. Figure 8(a) shows the variation of the network cost ($) with the number of iterations. Values of the probability of mutation, crossover and cut rate were also considered for the number of iterations which are selected as 1, 5, 50, 100, 120, 140 and 150×10^4. It can be noted that the cost approaches a fixed value after 1.3×10^6 iterations. In addition, the cost for all populations follows the closer values especially after 0.5×10^6 iterations; therefore, it may be concluded that the population parameter has little influence on this network. Here, the minimum cost is $\$6.109 \times 10^6$.

The crossover values of 50, 60, 70 and 80% were analysed, and the results of the four cases are shown in Figure 8(b), where minimum value cost is achieved with a crossover. It is seen that the network cost became a minimum after 200,000 iterations with a crossover value of 60% or splice probability of 0.060. Mutation is the most critical parameter of GAs. In Figure 8(c), the probability of mutation is increased from 0.5 to 2.0%, and its influence on cost is investigated. This parameter is critical to the performance of GA since changing its value makes a significant change in the cost. For a mutation probability of 1.0%, Figure 8(c) shows that the minimum network cost of $\$6.109 \times 10^6$ is reached after 100,000 iterations. In Figure 8(d), the influence of cutting probability in the range of 0.5 to 2.5% on the network cost is investigated. This figure shows that the value of cutting probability of 1.7% leads to the desired minimum cost. The results indicate a broad decreasing cost of the network after 100,000 iterations.

Case study 2: Maun network

For the case of the design and analysis study of case 2 of the Maun network, the ideal parameters adopted were based on values shown in Table 11. There was no pressure constraint violation at any of the junctions for all design events. Table 12 shows the solutions for Configurations A and B. It can be noted that Configuration B with longer and several pipes of smaller diameters offers the cheapest option for water conveyance, as shown in Table 12. Pipe pressure classes of 16 m were considered in both configurations and a minimum pressure of 25 m is required to be maintained in water distribution networks to meet local standards. Thus, the maximum pressure of 160 m and a minimum pressure of 25 m were used as system constraints (Figure 9).

Both pipeline routes, Configuration A and Configuration B, were optimized for minimum pipeline cost with respect to baseline and future peak hour demands. A difference in cost of 261,491.62 units between the two pipeline options. As shown in Table 12, Configuration B is the least-cost pipeline design configuration for water conveyance in the study area. With additional factors considered such as pumping initial and operating costs, further decisions will be done for the best overall water supply solution. This study assumed a network with consistently sufficient water supply, which is currently not the case now. The authors recommend that a future modelling study based on head-dependent analysis would be ideal to assess the actual supply shortfall in the network and employ optimization techniques based on that data.

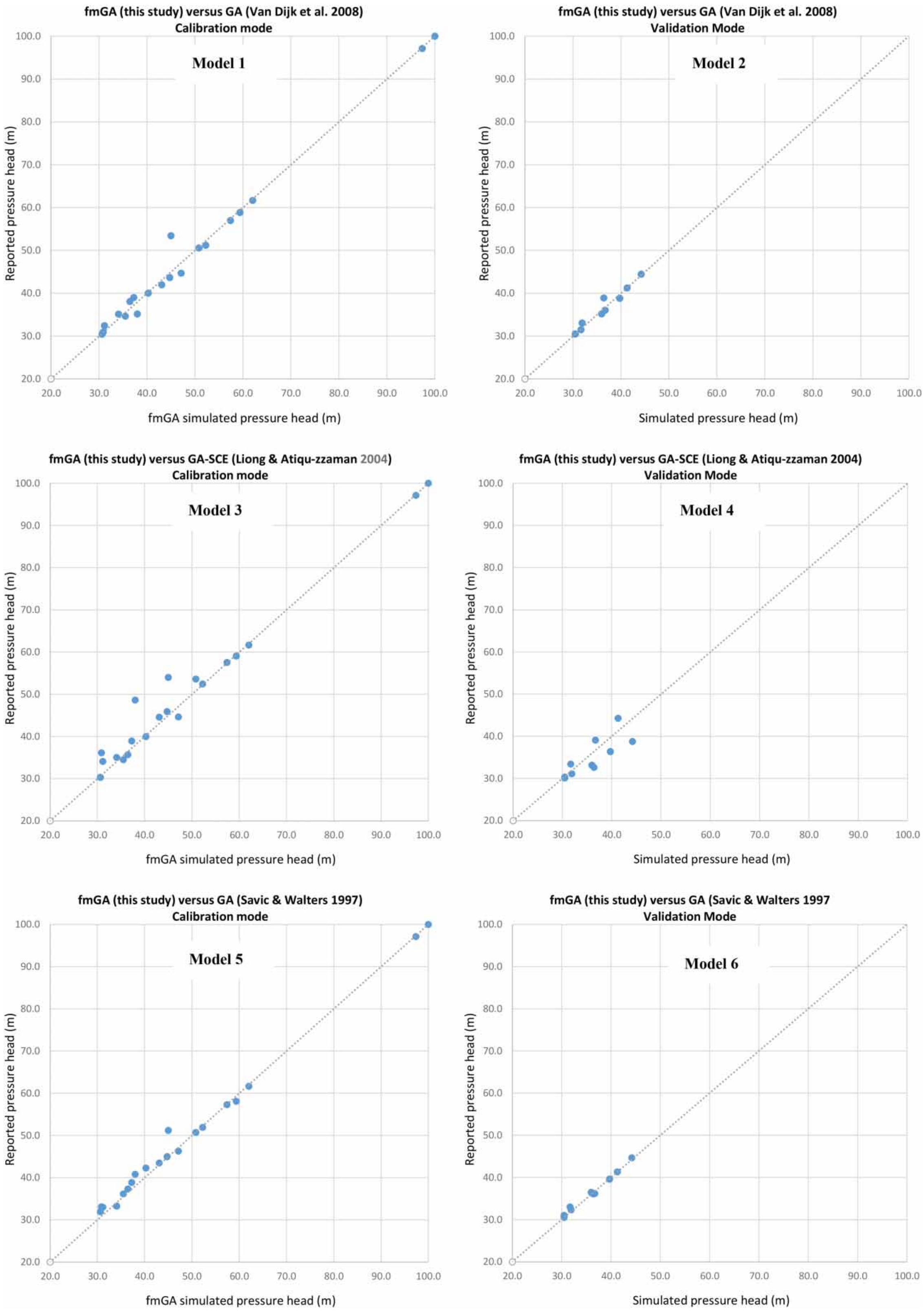


Figure 7 | Correlation plot of the hydraulic heads at 20 nodes (calibration model) and 12 nodes (validation model) simulated with the proposed fmGA model in relation to three other GA models, Models 2, 5 and 6.

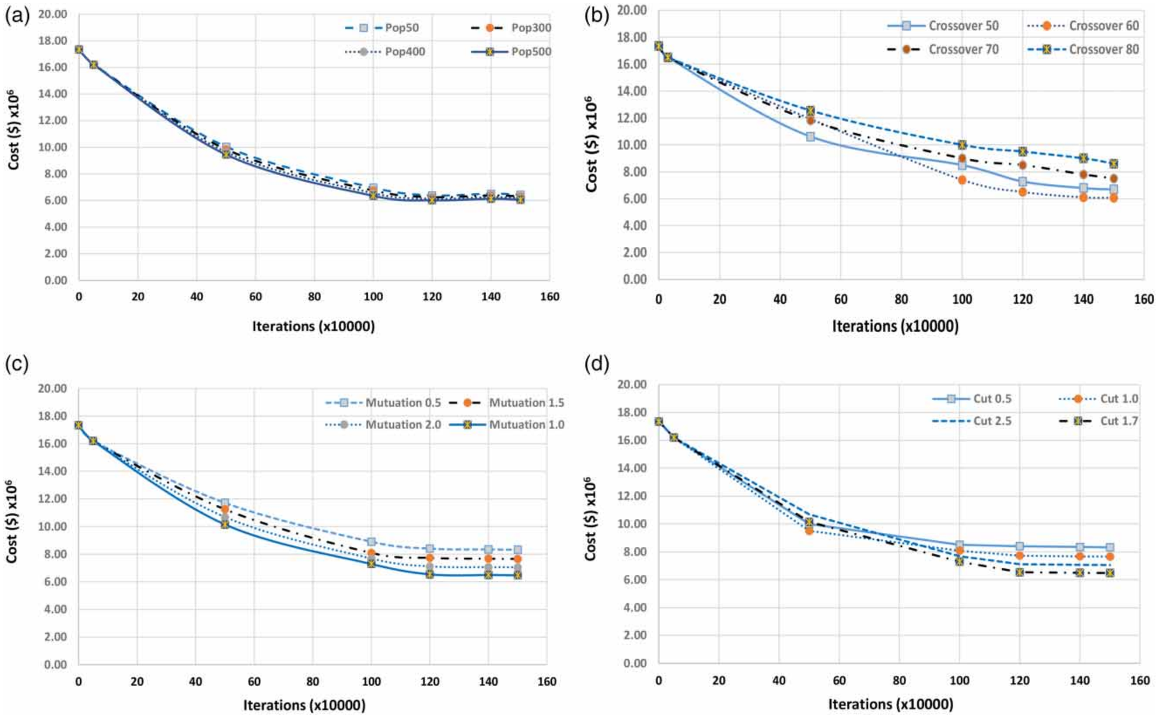


Figure 8 | Sensitivity analysis of the parameters of the genetic algorithm: (a) population, (b) crossover, (b) mutations and (d) cutting.

Table 11 | GA parameters used in the case study of the Maun network

GA optimization parameters	Value	GA optimization parameters	Value
Maximum era number	6	Random seed	0.5
Era generation number	150	Penalty factor	1,000,000
Population size	50	Stopping criteria:	
Cut probability	1.7	Maximum trials	250,000
Splice probability	60.0	Non-improvement generations	500
Mutation probability	1.5		

Table 12 | Solution for the Maun network Configurations A and B

Solution for Configuration A			Solution for Configuration B		
Pipe	**Diameter (mm)**	**Cost (units)**	**Pipe**	**Diameter (mm)**	**Cost (units)**
P-1	300	256,752.10	P-1	300	257,400.92
P-2	150	4,100.70	P-2	350	164,296.70
P-3	200	40,866.17	P-3	250	166,085.21
P-4	250	88,323.71	P-4	300	130,536.93
P-5	400	941,172.42	P-5	300	144,000.68
			P-6	150	35,916.00
			P-7	100	5,167.40
			P-8	250	166,321.00
	Total cost	**1,331,215.09**		**Total cost**	**1,069,722.16**

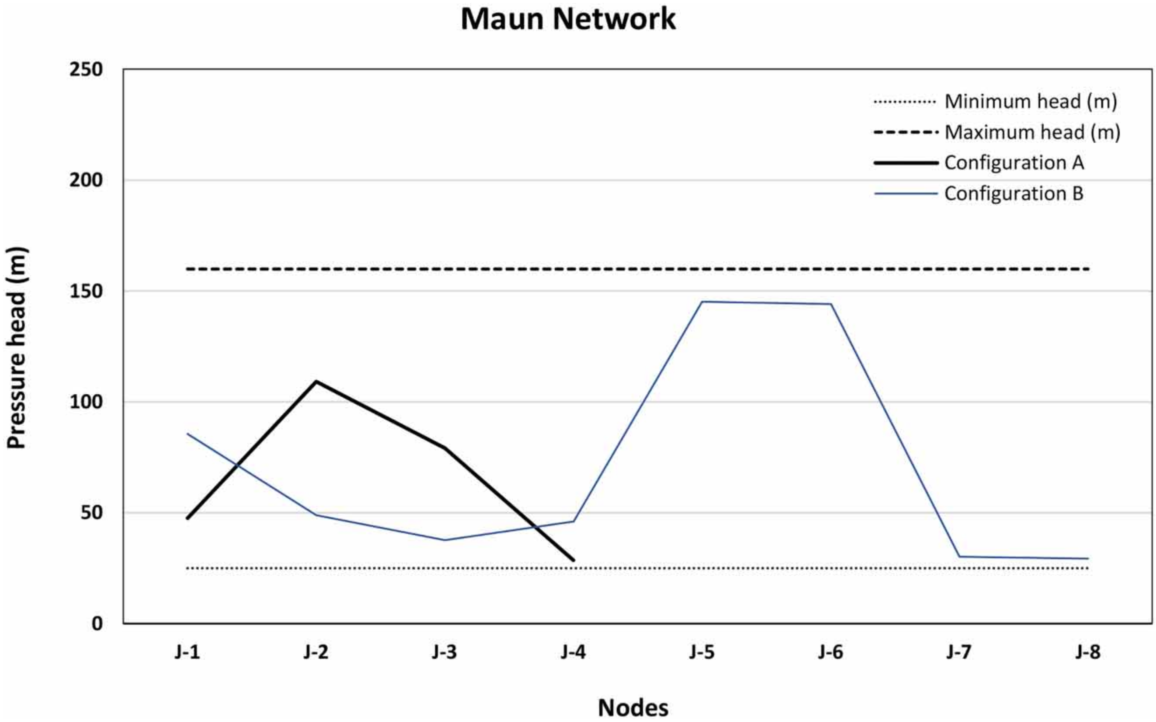


Figure 9 | Pressure head of Maun network calculated using fmGA.

CONCLUSIONS AND RECOMMENDATIONS

The effectiveness of an fmGA optimization model was assessed using a benchmark network in Hanoi. The results demonstrated that this approach yielded accurate outcomes and achieved rapid convergence within a limited number of generations compared to other GA-based methods. The fast-messy algorithm exhibited accurate results and displayed faster initial convergence. By utilizing a FORTRAN program in conjunction with EPANET, an interactive hydraulic analysis tool is developed that can aid in the design and analysis of both new and existing water distribution networks. Sensitivity analysis was conducted to determine the parameter combinations that result in the minimum cost for the benchmark Hanoi network. Additionally, the model was applied to evaluate and optimize alternative pipeline configurations for a new system in the Maun network located in Botswana. EPANET, which is widely recognized and extensively tested, serves as the hydraulic modelling software, while the developed FORTRAN program handles the interactive inputting of parameters for the optimization of the water distribution networks. This approach proves valuable as a tool for optimizing the design, rehabilitation and operation of water distribution systems in both urban and rural supply networks.

Further research can be a cost-effective tool for optimization applications and design of water supply systems that meet local design and hydraulic requirements. The approach can be used to provide optimal design solutions for more complex water distribution networks. Generally, this approach has the potential to be applied in designing more intricate water distribution networks.

ACKNOWLEDGEMENTS

The authors acknowledge the information by the Ministry of Land Management, Water and Sanitation Services, availed for public works contracts which inspired research to Master's degree training of the second author at the University of Botswana.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alperovits, E. & Shamir, U. 1977 Design of optimal water distribution systems. *Water Resour. Res.* **13**, 885–900.

Atiquzzaman, M. D. & Liong, S. Y. 2004 Application of shuffled complex evolution to water distribution network rehabilitation. In: *Hydroinformatics*, pp. 882–889. doi:10.1142/9789812702838_0109.

BENTLEY 2021a *Water Distribution Modeling and Analysis Software-WaterCAD*. Available from: https://www.bentley.com/en/watercad.

BENTLEY 2021b *Water Distribution Analysis and Design Software-WaterGEMS*. Available from: https://www.bentley.com⟩product-line⟩watergems.

CESD 2021 *Civil Enginieering Software Database*. Tahoe Design Software. Available from: https://www.cesdb.com/tahoe-design-software.htm (retrieved 10 June 2021).

Cisty, M. 2010 Hybrid genetic algorithm and linear programming method for least-cost design of water distribution systems. *Water Res. Manag.* **24**, 1–24.

Cruz-Chávez, M. A., Ávila-Melgar, É. Y., Cruz-Rosales, M. H., Martínez-Bahena, B., Flores-Sánchez, G., 2014 Search space analysis for the combined mathematical model (linear and nonlinear) of the water distribution network design problem. In: *Artificial Intelligence and Soft Computing*, Vol. 2014(8467) (Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L. A. & Zurada, J. M., eds). ICAISC 2014; Lecture Notes in Computer Science; Springer International Publishing, Zug, Switzerland, pp. 347–359.

Cunha, M. D. C. & Sousa, J. 1999 Water distribution network design optimization: simulated annealing approach. *J. Water Resour. Plan. Manag.* **125** (4), 215–221.

Cunha, M. A. C. & Sousa, J. 2001 Hydraulic infrastructures design using simulated annealing. *J. Infrastruct. Syst.* **125** (4), 31–39. doi:10.1061/(ASCE)1076-0342(2001)7:1(32).

Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A. M. T. 2002 A fast and elitist multiobjective genetic algorithm NSGA-II. *Evolut. Comput.* **6** (2), 182–197. doi:10.1109/4235.996017.

DNV 2021 *Advancing Smart Water Networks*. Available from: https://www.dnv.com/services/synergi-water-modules-4123 (retrieved 20 June 2021).

Eiben, A. E., Michalewicz, Z., Schoenauer, M. & Smith, J. E. 2007 Parameter control in evolutionary algorithms. In: *Parameter Setting in Evolutionary Algorithms*. (F. G. Lobo, C. F. Lima & Z. Michalewicz, eds.) Springer: Berlin/Heidelberg, Germany, pp. 19–46.

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

EPA 2021 *EPANET Application for Modeling Drinking Water Distribution Systems*. Available from: https://www.epa.gov/water-research/epanet (retrieved 15 August 2021).

Eusuff, M. M. & Lansey, K. E. 2003 Optimization of water distribution network design using the shuffled frog leaping algorithm. *J. Water Resour. Plan. Manag.* **129** (3), 210–225.

Fanni, A., Liberatore, S., Sechi, G. M., Soro, M. & Zuddas, P. 2000 Optimization of water distribution systems by a tabu search metaheuristic. In: *Computing Tools for Modeling, Optimization and Simulation* (M. Laguna & J.-L. González-Velarde, eds.). Springer, Cham, Switzerland, pp. 279–298.

Fujiwara, O. & Khang, D. B. 1990 A two-phase decomposition method for optimal design of looped water distribution networks. *Water Resour. Res.* **26** (4), 539–549. doi:10.1029/WR026i004p00539.

Goldberg, D., Deb, K. & Korb, B. 1989 Messy genetic algorithms: motivation, analysis, and first results. *Complex Syst.* **3**, 493–530.

Goldberg, D., Deb, K., Karugupta, H. & Harik, G., 1993 Rapid accurate optimization of difficult problems using fast messy genetic algorithms. In: *Proc. 5th Int. Conf. Genetic Algorithms* (S. Forrest, ed.), Morgan Kaufmann Publishers, San Mateo, CA, pp. 56–64.

Government of Botswana. 2016 *Botswana Water Accounting Report 2014/15*. WAVES, CAR, and MMEWR, Gaborone.

He, Y. & Hui, C.-W. 2006 Dynamic rule-based genetic algorithm for large-size single-stage batch scheduling. *Comput-Aided Chem. Eng.* **21**, 1911–1916.

Kargupta, H. 1996 The gene expression messy genetic algorithm. In *Proc. IEEE Int. Conf. Evolutionary Computation*, pp. 814–819. IEEE Press.

Keedwell, E. & Khu, S. T. 2005 A hybrid genetic algorithm for the design of water distribution networks. *Eng. Appl. Artificial Intelligence* **18**, 461–472.

Kirkpatrick, S., Gelatt Jr, C. D. & Vecchi, M. P. 1983 Optimization by simulated annealing. *Science* **220** (4598), 671–680.

KYPIPE. 2021 *Pipe2014: KYPipe (Steady State Analysis)*. Available from: https://kypipe.com/kypipe2/ (retrieved 20 June 2021).

Liberatore, S. & Sechi, G. M. 2009 Location and calibration of valves in water distribution networks using a scatter-search meta-heuristic approach. *Water Resour. Manag.* **23**, 1479–1495.

Liong, S. Y. & Atiquzzaman, M. 2004 Optimal design of water distribution network using shuffled complex evolution. *J. Inst. Engineers, Singapore* **44** (1), 93–107.

Martínez-Bahena, B., Cruz-Chávez, M. A., Ávila-Melgar, E. Y., Cruz-Rosales, M. H. & Rivera-Lopez, R. 2018 Using a genetic algorithm with a mathematical programming solver to optimize a real water distribution system. *Water* **10** (10), 1318. https://doi.org/10.3390/w10101318.

Mays, L. W. 2000 *Water Distribution System Handbook*. McGraw-Hill Professional Publishing, New York, NY, USA.

Mora-Melia, D., Iglesias-Rey, P. L., Martinez-Solano, F. J. & Fuertes-Miquel, V. S. 2013 Design of water distribution networks using a pseudo-genetic algorithm and sensitivity of genetic operators. *Water Resour. Manag.* **2013** (27), 4149–4162.

Paluszczyszyn, D. 2015 *Advanced Modelling and Simulation of Water Distribution Systems with Discontinuous Control Elements*. *PhD*, De Montfort University, Leicester.

PIPEFLOW. 2021 *Pipe Flow Software – Flow Rate & Pressure Drop Calculator Software*. Available from: https://www.pipeflow.co.uk/ (retrieved 15 July 2021).

Prasad, T. D. & Park, N. S. 2004 Multiobjective genetic algorithms for design of water distribution networks. *J. Water Resour. Plan. Manag.* **130** (1), 73–82.

Ramírez, R. M., Juárez, M. L. A., Mora, R. D., Morales, L. D. P., Mariles, Ó. A. F., Reséndiz, A. M., Elizondo, E. C. & Paredes, R. B. C. 2021 Operation policies through dynamic programming and genetic algorithms, for a reservoir with irrigation and water supply uses. *Water Res. Management* **35**, 1573–1586.

Reca, J. & Martínez, J. 2006 Genetic algorithms for the design of looped irrigation water distribution networks. *Water Resour. Res.* **42** (5), 1–9. http://dx.doi.org/10.1029/2005WR004383.

Rossman, L. A. 2000 *EPANET 2: User's Manual*. National Risk Management Research Laboratory Office of Research and Development, US Environmental Protection Agency, Cincinnati, OH, USA.

Savic, D. A. & Walters, G. A. 1997 Genetic algorithms for least-cost design of water distribution networks. *J. Water Res. Plan. Manag.* **123** (2), 67–77.

Shau, H. M., Lin, B. L. & Huang, W. C. 2005 Genetic algorithms for design of pipe network systems. *J. Marine Sci. Tech.* **13** (2), 116–124.

Sonaje, N. P. & Joshi, M. G. 2015 A review of modeling and application of water distribution networks (WDN) softwares. *Int. J. Tech. Res. Appl.* **3** (5), 174–178.

Todini, E. & Pilati, S. 1987 A gradient algorithm for the analysis of pipe networks. *Paper presented at the Computer Applications in Water Supply: Vol. 1 – Systems Analysis and Simulation*.

USGS 2021 *The Branch-Network Dynamic Flow Model – BRANCH*. Available from: https://water.usgs.gov/cgi-bin/man_wrdapp?branch(1) (retrieved 15 June 2021).

Vairavamoorthy, K. & Ali, M. 2000 Optimal design of water distribution systems using genetic algorithms. *Comput.-Aided Civ. Infrastruct. Eng.* **15**, 374–382.

Van Dijk, M., Van Vuuren, S. & Van Zyl, J. 2008 Optimising water distribution systems using a weighted penalty in a genetic algorithm. *Water SA* **34** (5), 537–548.

WATER SIMULATION. 2021 *H_2Onet*. Available from: https://www.http://www.water-simulation.com/wsp/2004/12/24/h2onet/ (retrieved 14 July 2021).

Wu, Z. Y. 1998 *Messy Genetic Algorithms for Optimisation of Water Distribution Systems Including Water Hammer*. University of Adelaide, Adelaide.

Wu, A. S. & Garibay, I. 2002 The proportional genetic algorithm: gene expression in a genetic algorithm. *Genet. Program. Evolvable Mach.* **3**, 157–192.

Wu, Z. Y. & Sage, P. 2006 Water loss detection via genetic algorithm optimization-based model calibration. In: *ASCE 8th Annual International*, p. 5.

Wu, Z. Y., Wang, Q., Butala, S., Mi, T. & Song, Y. 2012 *Darwin Optimization Framework User Manual*. Bentley Systems, Incorporated, Watertown, CT, USA.

First received 30 January 2023; accepted in revised form 9 June 2023. Available online 23 June 2023