Water Resour Manage (2014) 28:333–350
DOI 10.1007/s11269-013-0485-y

HydroGen: an Artificial Water Distribution Network Generator

Annelies De Corte · Kenneth Sörensen

Received: 14 October 2013 / Accepted: 25 November 2013 /
Published online: 13 December 2013

Abstract Many (metaheuristic) techniques for water distribution network (WDN) design optimisation already have been developed. Despite of the aforementioned scientific attention, only few, high-quality benchmark networks are available for algorithm testing, which, in turn, hinders profound algorithm testing, sensitivity analysis and comparison of the developed techniques. This absence of high-quality benchmark networks motivated us to develop a tool to algorithmically generate close-to-reality virtual WDNs. The tool, called HydroGen, can generate WDNs of arbitrary size and varying characteristics in EPANET or GraphML format. The generated WDNs are compared to (and shown to closely resemble) real WDNs in an analysis based on graph-theoretical indices. HydroGen is used to generate an extensive library of realistic test networks on which (metaheuristic) methods for the optimisation of WDN design can be tested, allowing researchers in this area to run sensitivity analyses and to draw conclusions on the robustness and performance of their methods.

Keywords Water distribution networks · Network generation · HydroGen ·
Water distribution network design

1 Introduction

Water distribution network design (WDND) has been a popular subject of study over the past thirty years. As the name suggests, this research field supports decisions related to the design of water distribution networks, such as network expansion, pipe replacement, or network redesign due to changing demand.

A. De Corte (✉) · K. Sörensen
Departement Engineering Management, Universiteit Antwerpen,
Prinsstraat 13 – B.510, 2000 Antwerp, Belgium
e-mail: annelies.decorte@uantwerpen.be

Water distribution network design problems can generally be formulated as NP-hard, non-linear constrained combinatorial optimisation problems. In Section 2 we discuss the basic variant of WDND optimisation problems. Many optimisation algorithms have been developed for this basic WDND optimisation problem, as well as for its extensions. Since all variants are NP-hard (Yates et al. 1984), and the evaluation of the constraints requires the solution of a highly non-linear set of differential equations, research has primarily focused on the development of metaheuristic techniques to find good solutions in reasonable time. To profoundly test, evaluate, and compare the developed algorithms would require an extensive library of easily available test instances, varied in size and characteristics. As we show in Section 3, however, only a very limited number of such test instances are available, and most of these are not very realistic. In Section 4 we provide an overview of existing methods to generate artificial WDNs.

To remedy this shortcoming and advance future research on WDND, a software tool is developed in this paper (see Section 5), that is able to generate realistic artificial WDNs. By adjusting the parameters of the tool, instances can be generated of arbitrary size and characteristics. The tool is called HydroGen and is freely available. Additionally, HydroGen is used to generate a large database of instances, that is also available for testing purposes.

To ensure the realism of the instances generated by Hydrogen, an analysis of real WDNs is performed in Section 6. This analysis compares the networks generated by HydroGen on the basis of several graph-theoretical indices. Finally, Section 7 concludes and gives applications of HydroGen.

2 The WDND Optimisation Problem

In this section, we discuss the basic WDND optimisation problem, that is the subject of a large number of papers. The aim of this problem is to determine the optimal diameter and type for each pipe in a given network layout. The basic problem is a simplification of reality in that demand for water is considered to be static, and no pumps are assumed to be present in the network (for this reason, these networks are called "gravity-fed").

The objective of the basic WDND optimisation problem is to minimise the total cost of the network design. The cost of an individual pipe depends on the *type* t that is chosen for this pipe from a list of commercially available types T. The type of a pipe determines both its diameter and the material of which it is made, which in turn determine its hydraulic properties. If the cost per meter of a pipe of type t is represented by C_t and the length of pipe p is represented as L_p, the objective function of the WDND optimisation problem can be written as:

$$\text{Minimise total cost} = \sum_{p \in P} \sum_{t \in T} x_{tp}.C_t.L_p \tag{1}$$

where x_{tp} is a binary decision variable that determines whether pipe p is chosen from type t ($x_{tp} = 1$) or not ($x_{tp} = 0$).

The objective function is conditioned by physical mass and energy conservation laws, and by minimum head requirements in the demand nodes.

The *mass conservation law* must be satisfied for each node $n \in N$. This law states that the volume of water flowing into a node in the network per unit of time must be equal to the volume of water flowing out of this node. Let Q_{ij} represent the water flowing from node i

to node j, and let S_n be the supply and D_n the demand of node n (all expressed in m^3s^{-1}) then the following should hold:

$$\sum_{i \in N \setminus n} Q_{in} - \sum_{j \in N \setminus n} Q_{nj} = D_n - S_n \qquad \forall n \in N \tag{2}$$

Furthermore, for each closed loop $l \in O$, the *energy conservation law* must be satisfied. This law states that the sum of pressure drops in a closed loop is zero. Pressure drops (also called head losses) in piping systems are caused by wall shear in pipes and friction caused by piping components such as junctions, valves, and bends. In past research, only the first type of friction losses (in the pipes) are taken into account. Let ΔH_p represent the head loss in a pipe p (in m) that connects nodes n_1 and n_2, then:

$$\Delta H_p = H_{n_1} - H_{n_2}$$

For the closed loop l, the energy conservation law can therefore be stated as:

$$\sum_{p \in \text{loop } l} \Delta H_p = 0 \qquad \forall l \in O \tag{3}$$

Approximating the head losses in the pipes of the network is usually done using the Hazen–Williams approximation:

$$\Delta H_p = w \frac{Q^a L_p}{C^a D^b} \tag{4}$$

In this equation, Q is the water flow rate in (m^3s^{-1}), L_p is the pipe length (in m), C is the Hazen–Williams roughness coefficient (unitless), D is the pipe diameter (in m), and a, b, and w are unitless parameters. Parameters D and C are determined by the type of a pipe and are assumed given for each available type. Finally, *minimum pressure head requirements* exist for every (demand) node $n \in N$. Let H_n be the pressure head in node n (in m) and H_n^{min} the minimum pressure head in node n (in m). This constraint therefore can be represented as:

$$H_n \geq H_n^{min} \qquad \forall n \in N \tag{5}$$

Several variants of the basic WDND optimisation problem have been proposed. These extensions are more realistic in that the water demand is dynamic (changes during the course of the day), that tanks (with finite capacity) are used in addition to reservoirs for the water supply and that pumps are added to the formerly exclusively gravity-fed networks. Another extension is the formulation of the above single-objective optimisation problem as a multi-objective one (Montalvo et al. 2010; Jung et al. 2013; Siew and Tanyimboh 2012). As stated in Section 5, HydroGen is not only able to generate test networks for the basic WDND optimisation problem, but also for the more complex extensions containing demand patterns, tanks and pumps.

3 Need for Realistic Networks

As stated earlier, a lot of research has been done in the field of optimisation of WDND over the past thirty years. Due to the flexibility and ability to find close-to-optimal solutions, many different metaheuristics frameworks have been applied to WDND optimisation problems. An overview of the applied metaheuristic techniques is given in Table 1. A more detailed description of the applied methods and their results on the most common benchmark networks is provided in De Corte and Sörensen (2013).

Table 1 Overview of the metaheuristic techniques applied to the WDND optimisation problem

Local search metaheuristics	
Simulated Annealing	(Loganathan et al. 1995; Cunha and Sousa 1999, 2001)
Tabu Search	(Cunha and Ribeiro 2004)
Population-based metaheuristics	
Genetic Algorithm	(Murphy and Simpson 1992; Simpson et al. 1994)
Improved Genetic Algorithm	(Dandy et al. 1996; Savic and Walters 1997)
Real Coding Genetic Algorithm	(Gupta et al. 1999; Vairavamoorthy and Ali 2000; Reca et al. 2008)
Differential Evolution	(Vasan and Simonovic 2010)
Memetic Algorithm	(Baños et al. 2007; Baños et al. 2010)
Cross-Entropy	(Perelman and Ostfeld 2007)
Scatter Search	(Lin et al. 2007)
Immune Algorithm	(Chu et al. 2008)
Improved Immune Algorithm	(Chu et al. 2008)
Shuffled Frog Leaping Algorithm	(Eusuff and Lansey 2003)
Constructive metaheuristics	
Ant Colony optimisation	(Maier et al. 2003)
Ant Systems	(Zecchin et al. 2005; Zecchin et al. 2006)
Max-Min Ant System	(Zecchin et al. 2006)
Particle Swarm Harmony Search	(Geem 2009)

A criticism with respect to the current state of the art, made by De Corte and Sörensen (2013) and Sitzenfrei (2010), is that the metaheuristics developed in this field are not adequately tested and thus, no strong conclusions can be drawn on their performance.

A proper evaluation of optimisation methods for WDND is hindered by two factors: (1) the number of available benchmark networks is limited, and (2) the available networks do not resemble real WDNs. An overview of currently available benchmark networks is given in Table 2.

A direct consequence of the limited test data availability is that metaheuristics for the WDND optimisation problem tend to be over fitted to these few instances, and therefore most likely not very robust.

The limited availability of realistic benchmark networks has been recognised by numerous authors: (Brumbelow et al. 2007; Yazdani and Jeffrey 2011; Möderl et al. 2011; Cunha and Ribeiro 2004), etc. Two reasons for the limited availability of realistic WDND benchmark instances can be distinguished. First, creating such instances from real-life WDNs is a time-consuming and expensive process of data collection, data conversion, data digitalisation, calibration and validation. Secondly, for various reasons water distribution companies are not eager to share their data and the dissemination of this data, even for research ends, is often prohibited.

A second problem is that the scarce available benchmark networks do not resemble realistic WDNs in terms of size and characteristics. The problem size of a WDND optimisation problem instance, a measure of the difficulty of finding good solutions for it, is determined by the size of the solution space and the size of the set of equations that needs to be solved. The size of the set of equations is determined by the number of junctions and loops in the

Table 2 Dimensions of benchmark and real networks

Network	nodes	pipes	loops	pipe types	equations	solution space
Benchmark networks						
Twoloop network (Alperovits and Shamir 1977)	7	8	2	14	18	$14^8 = 1.5 \times 10^9$
Gessler network (Gessler 1985)	12	14	3	8	30	$8^{14} = 4.4 \times 10^{12}$
NYCT network (Schaake and Lai 1969)	20	21	2	16	44	$16^{21} = 1.9 \times 10^{25}$
Hanoi network (Fujiwara and Khang 1990)	32	34	3	6	70	$6^{34} = 2.9 \times 10^{26}$
Balerma network (Reca and Martínez 2006)	447	454	8	10	910	$10^{454} = 1.0 \times 10^{454}$
Real networks						
Small district: 217 dwellings (Belgium)	45	53	9	39*	108	$39^{53} = 2.1 \times 10^{84}$
Blacksburg (USA) (Bragalli et al. 2008)	31	35	5	39*	72	$39^{35} = 4.9 \times 10^{55}$
Fossolo (Italy) (Bragalli et al. 2008)	37	58	22	39*	118	$39^{58} = 1.9 \times 10^{92}$
Pescara (Italy) (Bragalli et al. 2008)	71	99	29	39*	200	$39^{99} = 3.3 \times 10^{157}$
Modena (Italy) (Bragalli et al. 2008)	272	317	46	39*	636	$39^{317} = 2.3 \times 10^{504}$
Small town: 495 dwellings (Belgium)	469	495	27	39*	992	$39^{495} = 3.8 \times 10^{787}$
East-Mersea (UK) (Yazdani and Jeffrey 2011)	755	769	15	39*	1,540	$39^{769} = 3.4 \times 10^{1,223}$
Richmond (USA) (Yazdani and Jeffrey 2011)	872	957	86	39*	1,916	$39^{957} = 4.5 \times 10^{1,522}$
Colorado Springs (USA) (Yazdani and Jeffrey 2011)	1,786	1,994	209	39*	3,990	$39^{1,994} = 3.8 \times 10^{3,172}$
Kumasi (Ghana) (Yazdani and Jeffrey 2011)	2,799	3,065	267	39*	6,132	$39^{3,065} = 4.1 \times 10^{4,876}$

Note: Networks for which the number of pipe types was not available, were assumed to have 39 possible pipe types (13 possible diameters and 3 possible materials), based on data of a Flemish water distribution company

network: the total number of equations N_e is determined by the number of pipes m (head loss equations), junctions n (mass conservation equations), independent loops l (energy conservation equations) and cost (objective) functions cf: $N_e = m + n + l + cf$. From Euler's law for finite, connected planar graphs can be derived that $n + l - m = 1$ holds. Usually, the number of cost functions $cf = 1$. Therefore, it follows that $N_e = 2m + 2$. In other words, the number of equations that need to be solved to evaluate a given solution is roughly equal to twice the number of pipes in the network.

The size of the solution space, i.e., the number of potential solutions, is determined by the number of pipes m and the number of possible pipe types t, determined by diameter and material. If every pipe can take every available configuration, the number of possible solutions is t^m. Table 2 compares the problem size of some of the benchmark networks in the WDND optimisation literature to the size of some real-life instances.

As stated earlier, and as can be derived from Table 2, these few benchmark networks have dimensions that are much smaller than those of realistic networks. In De Corte and Sörensen (2013) is shown that, due to this not-too-challenging structure of these benchmarks, almost every developed metaheuristic technique finds the reported optimal solution. Table 3 shows the results for the techniques applied to the New York City Tunnels problem. Results for the other benchmarks are summarised in De Corte and Sörensen (2013).

Besides being much smaller than realistic instances, these benchmark networks miss some characteristics that typify real networks. In reality, we distinguish varying types of networks, such as: gravity-fed versus pump-fed networks; tree-like versus looped network structures; densely populated versus rural areas; domestic versus industrial demand nodes; ...

Concluding it can be said that it is of crucial importance that researchers can test their developed methods extensively on a wide range of realistic networks for numerous reasons: to avoid over fitting, to be able to verify a methods' performance in realistic circumstances, to evaluate its performance in comparison to other methods that are tested on the same set of test networks and to be able to perform sensitivity analyses to verify the methods robustness to varying input parameters. Such extensive testing will foster further development of this research area.

4 Existing Methods for the Generation of Artificial WDNs

Some attempts to (manually or automatically) construct realistic artificial WDNs can be found in the literature. This section provides an overview of the different methods, visual examples of their results can be found in Fig. 1. In Section 6 a comparison of these networks based on several graph-theoretical indices is given.

4.1 Manual Construction

Several artificial WDNs have been manually constructed. The **EXNET** network is set up by the Centre for Water Systems of Exeter University and described in Farmani et al. (2004). This distribution network serves approximately 400,000 inhabitants. Brumbelow et al. (2007) developed two virtual cities, called **Micropolis** and **Mesopolis**. Micropolis is based on the development pattern of a small city of 5,000 inhabitants, whereas Mesopolis represents a larger city of about 100,000 residents. Both cities resemble a very realistic network setting, because the authors mimic the development pattern of a city over a timespan of 130 years, which is reflected in the pipe materials and the network topology.

Although manually constructed artificial WDNs may in theory be very realistic, the shortcomings of the manual construction process are legion. Sitzenfrei (2010) emphasizes the time-consuming process, which in turn prohibits the development of an extensive library of test networks.

Table 3 Overview of results: New York City Tunnels problem by Schaake and Lai (1969)

Method	Author(s)	w	F	TC
Improved Genetic Algorithm	(Dandy et al. 1996)	–		38.80
Real Coded Genetic Algorithm	(Vairavamoorthy and Ali 2000)	na	IF	37.09
Particle Swarm optimisation	(Montalvo et al. 2008)	na	F	38.64
Scatter Search	(Lin et al. 2007)	10.5088	IF	36.68
Simulated Annealing 1	(Cunha and Sousa 2001)	10.5088	IF	37.10
Tabu Search 1	(Cunha and Ribeiro 2004)	10.5088	IF	37.13
Tabu Search 2	(Cunha and Ribeiro 2004)	10.5088	IF	37.13
Immune Algorithm	(Chu et al. 2008)	10.5088	IF	37.13
Modified Immune Algorithm	(Chu et al. 2008)	10.5088	IF	37.13
Genetic Algorithm	(Savic and Walters 1997)	10.5088	IF	37.13
Genetic Algorithm	(Lippai et al. 1999)	10.5088	IF	38.13
Ant Colony optimisation	(Maier et al. 2003)	10.6668	F	38.64
Shuffled Frog Leaping Algorithm	(Eusuff and Lansey 2003)	10.6688	F	38.80
Ant System	(Zecchin et al. 2005)	10.6688	F	38.64
Max-Min Ant System	(Zecchin et al. 2006)	10.6688	F	38.64
Harmony Search	(Geem 2006)	10.6688	F	38.64
Particle Swarm Harmony Search	(Geem 2009)	10.6688	F	38.64
Differential Evolution	(Vasan and Simonovic 2010)	10.6668	F	38.64
Scatter Search	(Lin et al. 2007)	10.675	F	38.64
Shuffled Frog Leaping Algorithm	(Eusuff and Lansey 2003)	10.6688	IF	38.13
Scatter Search	(Lin et al. 2007)	10.6668	IF	38.13
Simulated Annealing 2	(Cunha and Sousa 2001)	10.6792	IF	38.80
Genetic Algorithm	(Savic and Walters 1997)	10.9031	F	40.42
Immune Algorithm	(Chu et al. 2008)	10.9031	F	40.42
Modified Immune Algorithm	(Chu et al. 2008)	10.9031	F	40.42
Scatter Search	(Lin et al. 2007)	10.9031	F	40.42
Simulated Annealing 1	(Cunha and Sousa 2001)	10.9031	IF	40.40

Note: w = hydraulic coefficient , F = feasibility under EPANET2.0, TC = total cost in monetary units

4.2 Algorithmic Generation

Modular Design System (MDS) by Möderl et al. (2011) deliver a noteworthy contribution by explicitly addressing the issue of poor availability of realistic test networks in the field of WDNs. Moreover, the authors generate 2,280 artificial water distribution networks by concatenating smaller water supply blocks of looped and branched layout. As a consequence, all pipes have equal lengths, as can bee seen in Fig. 1, which leads to poor resemblance to reality for this property.

Moreover, only junctions, reservoirs and pipes are taken into consideration. Concluding, it can be stated that the main advantage of this technique is that this systematic generation leads to a wide range of test networks. Main shortcoming of this technique is that it makes significant simplifications and assumptions: no pumps, tanks nor valves are present and the network layout includes some unrealistic symmetries. These simplifications do not correspond with reality.

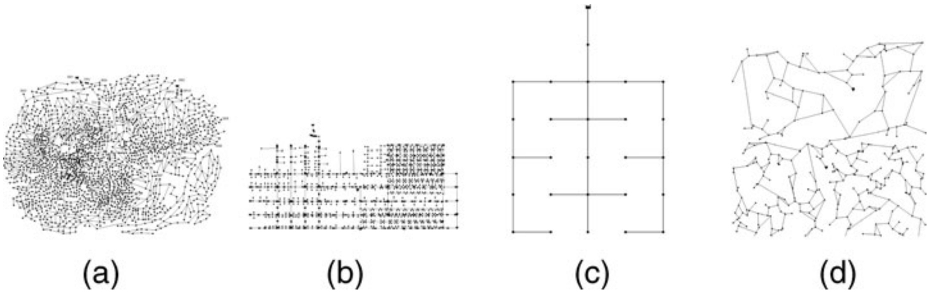

Fig. 1 (**a**) Exnet; (**b**) Micropolis; (**c**) Modular Design System; (**d**) WaterNetGen

A more recent work on the generation of WDN is due to Muranho et al. (2012). They developed **WaterNetGen**, which is an EPANET extension for automatic WDN generation and pipe sizing. This method has a lot of similarities with our generation procedure. One disadvantage of WaterNetGen is that valves, pumps and reservoirs are not added in the algorithmic generation. Moreover, a completely random allocation of demand nodes in a rectangular plane is a less realistic representation, compared to the circular layout applied in HydroGen.

5 HydroGen

The developed WDN generation method, HydroGen, attempts to address the shortcomings of the aforementioned methods. The tool generates realistic WDNs of arbitrary size and characteristics algorithmically, which are available in EPANET input format, since this is the most frequently used hydraulic solver, and GraphML, an XML-based graph exchange file format. A database is available online (http://www.antor.ua.ac.be/hydrogen).

The generation method is divided into six phases:

1. Generate clusters.
2. Generate tree structure.
3. Add water reservoirs, tanks and pumps.
4. Generate loops using intra-cluster pipes.
5. Generate inter-cluster pipes.
6. Assign base demand and demand load patterns.

A visualisation of the different steps in the generation process is shown in Fig 2.

5.1 Generation of Clusters

In a first step, random cluster centres are generated in a two dimensional plane. The water demand nodes are constructed in a circular layout around these cluster centres, with a uniform random polar angle and a radius for which holds: $P[R_{min} \leq radius \leq \alpha^d(R_{max} - R_{min})] = \alpha$, with α being the probability that a generated point will lay within a certain distance from the center and d an adjustable parameter, visual examples can be seen in Fig. 3. The generated water demand nodes are not equal to the drinking water needs of individual

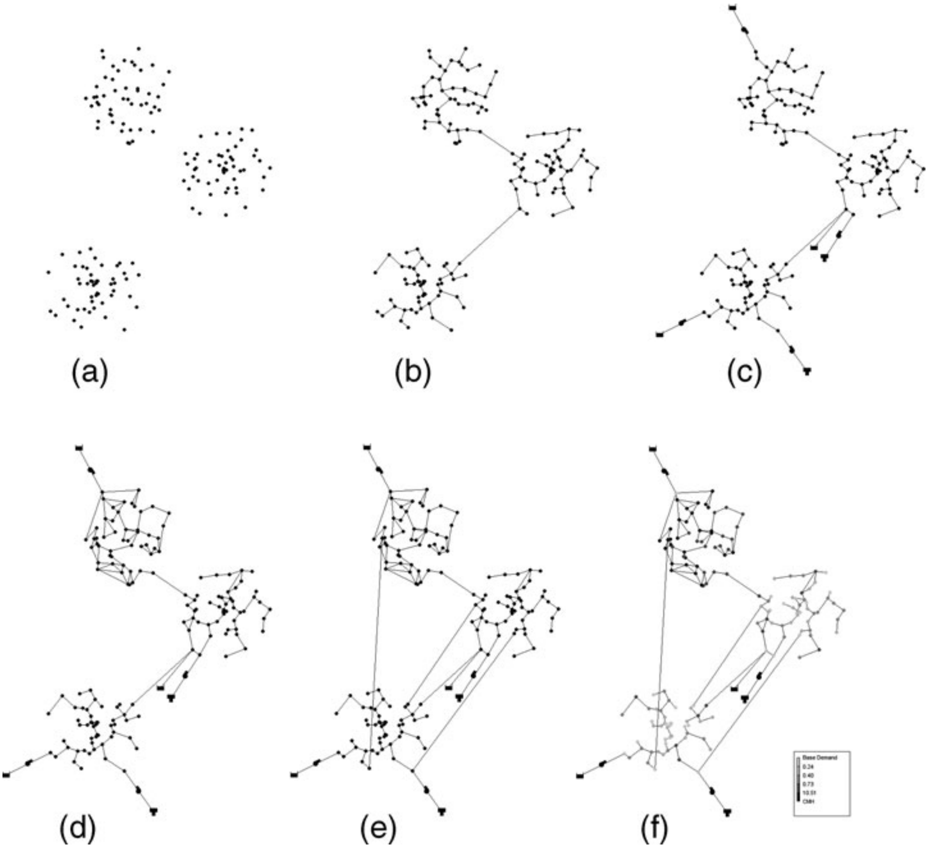

Fig. 2 Stepwise generation of (**a**) clusters; (**b**) tree structure, (**c**) reservoirs, tanks, pumps; (**d**) intra-cluster pipes; (**e**) inter-cluster pipes; (**f**) base load and demand patterns

households, but to clustered demands of households, restaurants, hotels, etc. This is further explained in Section 5.6.

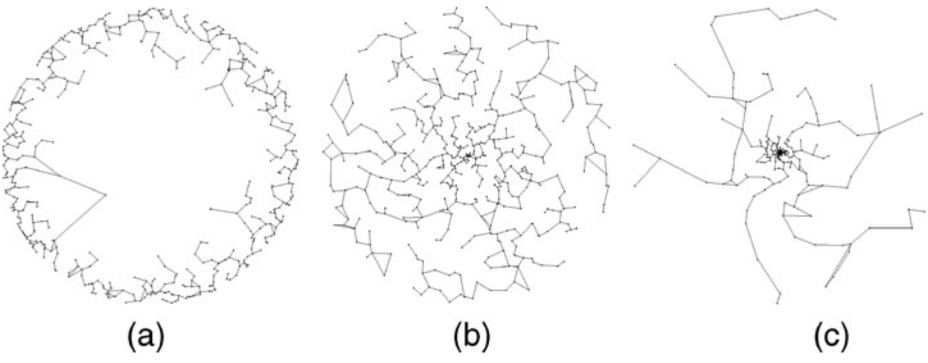

Fig. 3 HydroGen examples of a network of 500 nodes with cluster type 3 and varying d-factor: (**a**) $d = 0.1$; (**b**) $d = 1$; (**c**) $d = 10$

5.2 Generation of Tree Structure

A minimum spanning tree, connecting all of the demand nodes in the plane, is drawn, using Prim's algorithm (Prim 1957). These connections represent the water distribution pipes. Every pipe has a begin-node, end-node, length, diameter and roughness coefficient. Begin-node, end-node and length are defined by the generation itself: nodes are assigned while drawing the spanning tree and the edge weights or pipe lengths are the Euclidean distances. Pipes implicitly contain shut-off valves with their initial status set on open.

5.3 Addition of Reservoirs, Tanks and Pumps

For every reservoir or tank that has to be added to the WDN, a random demand node at the outside of the cluster is selected, and the water supply is connected to this demand node with a new pipe or pump. Reservoirs are infinite external sources of water to the network. Tanks have a limited water storage capacity and operate within their minimum and maximum water levels.

5.4 Generation of Loops with Intra-Cluster Pipes

Although a tree structure could efficiently provide every demand node with sufficient drinking water, no real-life WDN is designed as a tree, as stated under Section 6. Loops are added to increase water delivery reliability. Moreover, huge pressure changes are avoided by an interwoven net. The user can predefine a cluster type for each cluster. Each cluster type corresponds with a specific meshedness coefficient M (see also Section 6), which in turn is related to the number of intra-cluster pipes that is added. Depending on the cluster type, some clusters will have more loops than others, which can be seen in Fig. 6. Duplication of pipes and the crossing of pipes are avoided by a preliminary check. Moreover, in reality, the number of neighbouring pipes (or the degree of that node) is limited to four, which is also taken into account in the generation procedure. There are three pre-defined cluster types:

1. rural area ($M = 0$): no intra-cluster pipes, retaining minimum spanning tree configuration
2. very densely populated area ($0 << M$): generation of a complete planar graph (via triangulation), followed by adjustments so that the degrees of the nodes are ≤ 4.
3. urban area ($M = 0.16$): generation of loops, avoiding duplication and crossing of pipes and maintaining degrees of the nodes ≤ 4.

Three example networks, with an equal number of nodes but a different cluster type are shown in Fig. 4, in Table 5, their parameters are shown. It is clear that a higher meshedness coefficient implies a lower diameter and a lower average path length, for an equal number of nodes. This shows that more looped networks have a higher efficiency in terms of water delivery, which means that the pressure drops related to the supplying of water are lower.

5.5 Generation of Inter-Cluster Pipes

Additionally, inter-cluster pipes can be added to the WDN. The user can specify the number of inter-cluster pipes that are added randomly, connecting different clusters. In contrast to the intra-cluster pipes, inter-cluster pipes can cross other pipes.

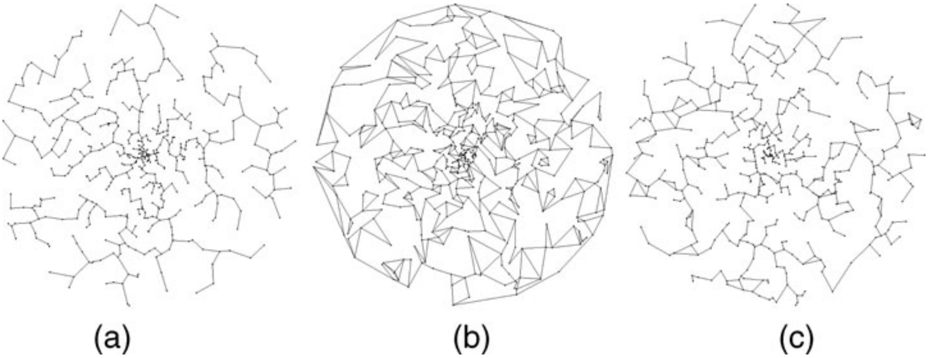

Fig. 4 HydroGen examples of (**a**) rural area - type 1; (**b**) very densely populated area = type 2; (**c**) urban area = type 3

5.6 Assignment of Base Load and Demand Patterns

In a final step, both the demand pattern and the base load must be specified for every demand node. Five different user categories with corresponding base loads (Van Tomme and De Sutter 2004) and demand patterns are defined:

- residential (households): 0.01 m^3/h
- industry (chemical industry, production plants, factories,...): 0.17 m^3/h
- commercial (restaurants, cafs, hotels, ...): 0.20 m^3/h
- public services (hospitals, universities, ...): 0.48 m^3/h
- energy (refinery, electricity and gas production): 121.68 m^3/h

The water demand of residential users, commercial users and public services is assigned to the pipelines and subsequently clustered to the connected demand nodes, based on the average number of customers per kilometre of pipeline in Flanders (Belgium), which is 43 customers per kilometre of pipeline for residential users (VMW 2013), 20 commercial users per kilometre and 5 public services buildings per kilometre. The water demand of industrial users and energy plants is assigned to the nodes directly.

An abundant number of realistic network settings can be built by adjusting HydroGen's parameters. By fine-tuning these parameters, one can obtain networks that have a high resemblance to real networks in terms of the indices stated in Section 6. This is in sharp contrast to earlier used benchmark networks and earlier developed network generation methods, which can be derived from Table 2.

6 Graph-Theoretical Comparison to Real WDNs

Networks, or graphs, are a popular object of study in the field of physics, mathematics and computer science. Abundant literature has been written about the analysis of biological networks, social networks, etc. However, WDNs have not received much research attention yet, as noted by Newman (2010), which could be explained by the fact that it is hard to obtain data on real WDNs, as stated under Section 3.

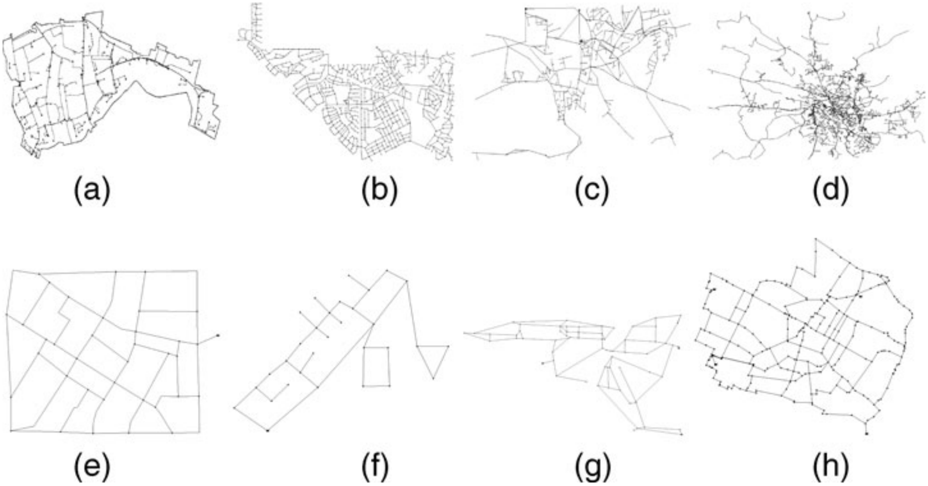

Fig. 5 (**a**) East-Mersea; (**b**) Colorado Springs; (**c**) Richmond; (**d**) Kumasi; (**e**) Fossolo; (**f**) Blacksburg; (**g**) Pescara; (**h**) Modena

An analysis of WDNs can be found in Yazdani and Jeffrey (2011) and Yazdani and Jeffrey (2012). The first contribution provides more general insights in the analysis of water distribution systems and shows how the network structure influences operational efficiency, reliability and robustness. The latter uses principles of graph theory and network analysis to assess the structural vulnerability and robustness of WDNs.

In this section, we analyse several real WDNs on a number of graph-theoretical indices, explained below. The limited number of available instances rules out an analysis with statistically significant results. Therefore, we limit ourselves to an informal comparison. These real networks are then compared to both the networks generated by HydroGen and to the other available benchmark networks in the literature.

Ten real networks are used in this analysis: a small district and a medium-sized town (Belgium); East-Mersea (UK), Richmond (UK), Colorado Springs (USA) and Kumasi (Ghana) (Yazdani and Jeffrey 2011), Balcksburg (USA), Fossolo (Italy), Pescara (Italy) and Modena (Italy) (Bragalli et al. 2008). These networks are shown graphically in Fig. 5.

A WDN can be represented as a graph $G = (N, E)$, containing a set N of n nodes (i.e., water demand and supply nodes) and a set E of m edges (i.e., water distribution pipes). Since the direction of the water flow in the pipes is subject to occasional changes (e.g., pressure changes, pump activation, change of status of shut-off valves, etc.), WDNs are represented as undirected graphs. Moreover, under normal circumstances, there exists at least one path between every pair of nodes and therefore, WDNs are connected graphs. Furthermore, a WDN can be represented as a graph where no edges cut one another (except at their nodes). This planarity could be violated in a WDN due to pipes crossing without intersecting, but this situation is uncommon in reality, therefore, planarity can be considered as a good approximation for WDNs.

For a WDN with n nodes, m edges and l independent loops, Euler's formula for finite connected planar graphs establishes that $n - m + l = 1$. In this paper, sources (water reservoirs or water tanks) are counted as nodes and pipes connecting the sources to the

Table 4 Column headings and symbols used in Table 5

n	number of nodes
m	number of edges/pipes
l	number of independent loops
T	method can generate tanks
P	method can generate pumps
DP	method can generate dynamic demand patterns
L_{avg}	average pipe length
M	meshedness coefficient
m/n	edge-to-node ratio
k_{max}	maximal degree
k_{avg}	average degree
d	diameter
apl	average path length

network are counted as edges. A graph with n nodes and m edges can be described by its $n \times n$ adjacency matrix A, which is defined as:

$$\text{for } i \neq j A_{ij} = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$
$$\text{for } i = j : A_{ij} = 0.$$

The *degree* k_i of a node i represents the number of edges incident to this node: $k_i = \sum_{j=1}^{n} A_{ij}$. The *average degree* k_{avg} of a graph is defined as the average of the degrees of all the nodes of that graph: $k_{avg} = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{2m}{n}$. The average degree is a measure of connectivity and gives information on the sparseness of the network: treelike network structures have an average degree of approximately 2, whereas an average degree of about 4 indicates a more complete network with grid pattern. As can be derived from the average degree of the observed real networks ($2 < k_{avg} \ll 4$), WDNs tend to be rather sparse. Another interesting parameter is the *maximum degree* k_{max} of the network. Practical limitations on the way water pipes can be connected mean taht, in real WDNs, this value does not usually exceed 4 ($k_{max} \leq 4$), independently of the size of the network. This is in sharp contrast to many other types of networks (e.g., social networks, information networks or biological networks), for which the maximum degree increases for larger networks.

The *edge-per-node ratio*, m/n, gives an indication on the level of connectivity. A ratio of approximately 1 indicates a tree structure, whereas a ratio very close to 3 indicates a complete planar graph. The *meshedness coefficient*, $M = \frac{l}{2n-5}$, evaluates the number of loops in comparison to the maximum number of loops (Buhl et al. 2006). A tree structure will have a coefficient equal to 0 and a complete planar graph will have a coefficient of 1. Due to the fact that the nodes of WDNs typically have degrees ≤ 4, it is very unlikely that WDN's will be complete planar graphs.

Networks with higher meshedness coefficients and edge-per-node ratios characterise networks with more redundant pipes, which are less sensitive to pressure shocks, pipe breakages, etc. due to the existence of alternative water supply paths.

Tables 4 and 5 shows that, for real networks, $1 \leq m/n \ll 3$ and $10^{-2} \leq M \leq 10^{-1} \ll 1$, which indicates that all real WDNs are looped, but rather sparse networks, with low connectivity. This is explained by the fact that adding additional (redundant) pipes to a WDN carries a high infrastructure cost and is only done if the benefits in terms of robustness

Table 5 Dimensions of real and generated networks

Network	n	m	l	T	P	DP	L_{avg}	M	m/n	k_{max}	k_{avg}	d	apl
Real networks													
Blacksburg (USA) (Bragalli et al. 2008)	31	35	5				177	0.09	1.13	4	2.26	9	4
Small district (Belgium)	45	53	9				42	0.11	1.18	3	2.36	12	5
Fossolo (Italy) (Bragalli et al. 2008)	37	58	22				145	0.32	1.57	4	3.14	8	4
Pescara (Italy) (Bragalli et al. 2008)	71	99	29				491	0.21	1.39	5	2.79	20	9
Modena (Italy)(Bragalli et al. 2008)	272	317	46				227	0.09	1.17	5	2.33	38	14
Small town (Belgium)	469	495	27				78	0.03	1.06	3	2.11	94	36
East-Mersea (UK)(Yazdani and Jeffrey 2011)	755	769	15				28	0.01	1.02	4	2.04	97	34
Richmond (UK) (Yazdani and Jeffrey 2011)	872	957	86				633	0.05	1.10	4	2.19	135	51
Colorado Springs (USA) (Yazdani and Jeffrey 2011)	1,786	1,994	209				187	0.06	1.12	4	2.23	69	26
Wolf-Cordera Ranch (USA) (Centre for Water Systems 2013)	2,034	2,241	208				184	0.05	1.10	5	2.20	69	26
Kumasi (Ghana) (Yazdani and Jeffrey 2011)	2,799	3,065	267				316	0.05	1.10	4	2.19	120	34
Generated networks													
EXNET(Farmani et al. 2004)	1,893	2,418	526				241	0.14	1.28	10	2.55	54	21
Micropolis (Brumbelow et al. 2007)	1,577	1,626	50	✓	✓	✓	85	0.02	1.03	6	2.06	98	37
Mesopolis (Brumbelow et al. 2007)	1,595	2,224	630	✓	✓	✓	1,589	0.20	1.39	8	2.79	94	31
MDS: nid1710 (Möderl et al. 2011)	175	238	64				500	0.19	1.36	4	2.72	36	15
MDS: nid1000 (Möderl et al. 2011)	213	292	80				500	0.19	1.37	4	2.74	44	16
MDS: nid1267 (Möderl et al. 2011)	351	352	2				500	0.00	1.00	4	2.01	75	40
MDS: nid140 (Möderl et al. 2011)	713	992	280				500	0.20	1.39	4	2.78	144	49
MDS: nid2144 (Möderl et al. 2011)	1,256	1,832	577				500	0.23	1.46	4	2.92	74	26
WaterNetGen: 3 clusters (Muranho et al. 2012)	300	387	88				248	0.15	1.29	4	2.58	50	21
WaterNetGen: 5 clusters (Muranho et al. 2012)	550	703	154				227	0.14	1.28	4	2.56	66	26
WaterNetGen: 7 clusters (Muranho et al. 2012)	850	1,097	248				124	0.15	1.29	4	2.58	75	28

Table 5 (continued)

Network	n	m	l	T	P	DP	L_{avg}	M	m/n	k_{max}	k_{avg}	d	apl
WaterNetGen: 9 clusters (Muranho et al. 2012)	2,030	2,607	578				210	0.14	1.28	4	2.57	114	47
WaterNetGen: 11 clusters (Muranho et al. 2012)	2,660	3,420	761				96	0.14	1.29	4	2.57	126	53
HydroGen: cluster type 1	500	499	0	√	√	√	69	0.00	1.00	4	2.00	104	41
HydroGen: cluster type 2	500	727	228	√	√	√	651	0.23	1.45	4	2.91	34	16
HydroGen: cluster type 3	500	548	49	√	√	√	172	0.05	1.10	4	2.19	81	30
HydroGen: (a)	50	53	4	√	√	√	115	0.04	1.06	4	2.12	21	9
HydroGen: (b)	150	164	15	√	√	√	169	0.05	1.09	5*	2.19	32	13
HydroGen: (c)	300	329	30	√	√	√	171	0.05	1.10	5*	2.19	66	24
HydroGen: (d)	500	549	50	√	√	√	188	0.05	1.10	5*	2.20	114	36
HydroGen: (e)	750	824	75	√	√	√	173	0.05	1.10	4	2.20	79	32
HydroGen: (f)	1,050	1,154	105	√	√	√	752	0.05	1.10	4	2.20	91	36
HydroGen: (g)	1,400	1,539	140	√	√	√	56	0.05	1.10	4	2.20	92	38
HydroGen: (h)	1,800	1,979	180	√	√	√	68	0.05	1.10	4	2.20	138	49
HydroGen: (I)	2,250	2,474	225	√	√	√	11	0.05	1.10	5*	2.20	107	44
HydroGen: (j)	2,750	3,024	275	√	√	√	14	0.05	1.10	5*	2.20	124	53

Note: * : <1 % of all nodes has a degree of 5. WaterNetGen networks are generated under default settings. The first set of HydroGen networks is generated with varying cluster type and $d = 1$, visual representations can be found in Fig. 4, the second set of HydroGen networks is generated with cluster type 3 and $d = 1$, visual representations can be found in Fig. 6

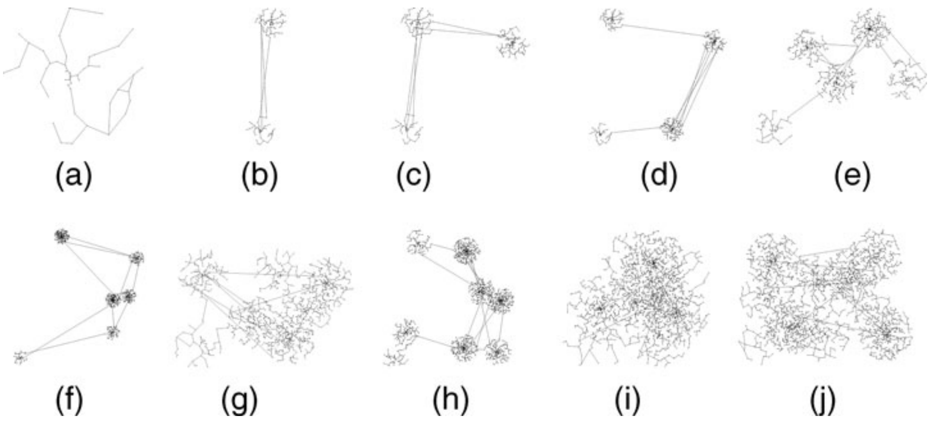

Fig. 6 HydroGen examples with varying distance parameters: (**a**) 1 cluster of 50 nodes; (**b**) 2 clusters of 50 and 100 nodes, (**c**) 3 clusters of 50, 100 and 150 nodes; (**d**) 4 clusters of 50 to 200 nodes; (**e**) 5 clusters of 50 to 250 nodes; (**f**) 6 clusters of 50 to 300 nodes; (**g**) 7 clusters of 50 to 350 nodes; (**h**) 8 clusters of 50 to 400 nodes; (**i**) 9 clusters of 50 to 450 nodes; (**j**) 10 clusters of 50 to 500 nodes

or resilience of the network outweigh the costs. This important trade-off, robustness versus cost, is faced by all WDN companies. It can be seen as a multi-objective WDN design optimisation problem and would be an interesting field of study.

As can be derived from Table 5 and Fig. 6, HydroGen is able to generate more realistic test networks than previously developed methods in terms of network dimensions and pipe length distribution, as well as allowing for extensions such as dynamic demand patterns, tanks and pumps. Moreover, since it is an algorithmic generation using randomness, HydroGen is able to generate an infinite amount of non-identical test networks.

7 Conclusion

It has been shown that many (metaheuristic) techniques for WDND optimisation are poorly tested, which makes it difficult to draw solid conclusions on their performance. This lack of rigorous analyses is mainly caused by the absence of high-quality test networks, which motivated us to develop a tool to algorithmically generate realistic artificial WDNs.

The developed tool, HydroGen, generates WDNs of arbitrary size and varying characteristics in both EPANET and GraphML format. The use of cluster types, the possibility to add dynamic demand profiles, tanks and pumps and parameter fine-tuning enables HydroGen to generate close-to-reality WDNs.

The generated WDNs are compared to real WDNs in an analysis based on graph-theoretical indices. It is shown that the networks generated by HydroGen have a high resemblance to real WDNs, in sharp contrast to the frequently used benchmark networks and most of the previously developed generation methods.

HydroGen is used to generate an extensive library of realistic test networks, which is available via http://antor.ua.ac.be/. The generated networks can be used for metaheuristic testing and improving and allows researchers in this area to perform sensitivity analyses and to draw conclusions on the robustness and performance of their methods. HydroGen networks can support researchers in their development of new metaheuristic methods and therefore, research in the area of water distribution network design optimisation.

Future work in this area could be done by extending the generation procedure to other output formats or by extending the generation to other types of distribution networks such as gas distribution networks, oil distribution networks or electric grids.

Acknowledgments This research was funded by the Research Foundation - Flanders (FWO). The authors would like to thank Jochen Janssens for the useful input on network generation methods. This research has been partially supported by the Interuniversity Attraction Poles (IUAP) Programme initiated by the Belgian Science Policy Office.

References

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

Baños R, Gil C, Agulleiro J, Reca J (2007) A memetic algorithm for water distribution network design. Soft Comput Ind Appl 39:279–289

Baños R, Gil C, Reca J, Montoya F (2010) A memetic algorithm applied to the design of water distribution networks. Appl Soft Comput 10:261–266

Bragalli C, D'Ambrosio C, Lee J, Lodi A, Toth P (2008) Water network design by minlp. Tech. rep., IBM

Brumbelow K, Torres J, Guikema S, Bristow E (2007) Virtual cities for water distribution and infrastructure systems research. In: Proceedings of the world environmental and water resources congress. Tampa, FL, May

Buhl J, Gautrais J, Reeves N, Solé R, Valverde S, Kuntz P, Theraulaz G (2006) Topological patterns in street networks of self-organized urban settlements. Eur Phys J 49:512–522

Centre for Water Systems (2013) Water distribution network benchmarks. emps.exeter.ac.engineering/research/cws/downloads/benchmarks

Chu CW, Lin MD, Liu GF, Sung Y (2008) Application of immune algorithms on solving minimum-cost problem of water distribution network. Math Comput Model 48:1888–1900

Cunha M, Ribeiro L (2004) Tabu search algorithms for water network optimization. Eur J Oper Res 157:746–758

Cunha M, Sousa J (1999) Water distribution network design optimization: simulated annealing approach. J Water Resour Plan Manag:125

Cunha M, Sousa J (2001) Hydraulic infrastructures design using simulated annealing. J Infrastruct Syst 7(1):32–39

Dandy G, Simpson A, Murphy L (1996) An improved genetic algorithm for pipe network optimisation. Water Resour Res 32:449–458

De Corte A, Sörensen K (2013) Optimisation of gravity-fed water distribution network design: a critical review. Eur J Oper Res 228:1–10

Eusuff M, Lansey K (2003) Optimization of water distribution network design using the shuffled frog leaping algorithm. J Water Resour Plan Manag:129

Farmani R, Savic D, Walters G (2004) Exnet benchmark problem for multi-objective optimization of large water systems. In: Proceedings IFAC workshop

Fujiwara O, Khang D (1990) A two-phase decompostion method for optimal design of looped water distribution networks. Water Resour Res 26:539–549

Geem Z (2006) Optimal cost design of water distribution networks using harmony search. Eng Optim 38:259–280

Geem Z (2009) Particle-swarm harmony search for water network design. Eng Optim 41:297–311

Gessler J (1985) Pipe network optimization by enumeration. Comput Appl Water Resour:572–581

Gupta I, Gupta A, Khanna P (1999) Genetic algorithm for optimization of water distribution systems. Environ Model Softw 14:437–446

Jung B, Filion Y, Adams B, Karney B (2013) Multi-objective design optimization of branched pipeline systems with analytical assessment of fire flow failure probability. Water Resour Manag 27:3663–3678

Lin M, Liu Y, Liu G, Chu C (2007) Scatter search heuristic for least-cost design of water distribution networks. Eng Optim 39:857–876

Lippai I, Heaney J, Laguna M (1999) Robust water system design with commercial intelligent search optimizers. J Comput Civ Eng 13:135–143

Loganathan G, Greene J, Ahn T (1995) Design heuristic for globally minimum cost water distribution systems. J Water Resour Plan Manag 121:182–192

Maier H, Simpson A, Zecchin A, Foong W, KYPhang HYSeah, Tan C (2003) Ant colony optimization for design of water distribution systems. J Water Resour Plan Manag 129:200–209

Möderl M, Sitzenfrei R, Fetz T, Fleischhacker E, Rauch W (2011) Systematic generation of virtual networks for water supply. Water Resour Res 47:1–10

Montalvo I, Izquierdo J, Pérez R, Tung M (2008) Particle swarm optimization applied to the design of water supply systems. Comput Math Appl 56:769–776

Montalvo I, Izquierdo J, Schwarze S, Pérez-Garcá R (2010) Multi-objective particle swarm optimization applied to water distribution systems design: an approach with human interaction. Math Comput Model 52:1219–1227

Muranho J, Ferreira A, Sousa J, Gomes A, Marques AS (2012) Waternetgen: an epanet extension for atuomatic water distribution network models generation and pipe sizing. Water Sci Technol Water Supply 12:117–123

Murphy L, Simpson A (1992) Genetic algorithms in pipe network optimisation. Research Report R93, University of Adelaide

Newman MEJ (2010) Networks: an introduction. Oxford University Press

Perelman L, Ostfeld A (2007) An adaptive heuristic cross-entropy algorithm for optimal design of water distribution systems. Eng Optim 39:413–428

Prim R (1957) Shortest connection networks and some generalizations. Bell Sys Tech J 36:1389–1401

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

Reca J, Martínez J, Gil C, Baños R (2008) Application of several meta-heuristic techniques to the optimization of real looped water distribution networks. Water Resour Manag 22:1367–1379

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

Schaake J, Lai D (1969) Linear programming and dynamic programming applications to water distribution network design. Report 116, Department of Civil Engineering. MIT, Cambridge

Siew C, Tanyimboh T (2012) Penalty-free feasibility boundary convergent multi-objective evolutionary algorithm for the optimization of water distribution systems. Water Resour Manag 26:4485–4507

Simpson A, Dandy G, Murphy L (1994) Genetic algorithms compared to other techniques for pipe optimization. J Water Resour Plan Manag 120:423–443

Sitzenfrei R (2010) Stochastic generation of urban water systems for case study analysis. Dissertation, Leopold Franzens Universität Innsbruck, Innsbruck

Vairavamoorthy K, Ali M (2000) Optimal design of water distribution systems using genetic algorithms. J Comput Aided Civ Infrastruct Eng:374–382

Van Tomme I, De Sutter R (2004) Berekening van het watergebruik in 2002 en analyse van het watergebruik in de periode 19991-2002, studie uitgevoerd in opdracht van de vlaamse milieumaatschappij. MIRA MIRA/200X/06:1–57

Vasan A, Simonovic S (2010) Optimization of water distribution network design using differential evolution. J Water Resour Plan Manag 136:279–287

VMW (2013) Watermeter 2012: Drinkwaterproductie en -levering in cijfers. Tech. rep., VMW

Yates D, Templeman A, Boffey T (1984) The computational complexity of the problem of determining least capital cost designs for water supply networks. Eng Optim 7:143–155

Yazdani A, Jeffrey P (2011) Complex network analysis of water distribution systems. Chaos 21:1–10

Yazdani A, Jeffrey P (2012) Applying network theory to quantify the redundancy and structural robustness of water distribution systems. J Water Resour Plan Manag 138:153–161

Zecchin A, Simpson A, Maier H, Nixon J (2005) Parametric study for an ant algorithm applied to water distribution system optimization. IEEE Trans Evol Comput 9:175–191

Zecchin A, Simpson A, Maier H, Leonard M, Roberts A, Berrisford M (2006) Application of two ant colony optimisation algorithms to water distribution system optimisation. Math Comput Model 44:451–468