Automatic generation of water distribution systems based on GIS data

Robert Sitzenfrei*, Michael Möderl, Wolfgang Rauch

Unit of Environmental Engineering, Institute of Infrastructure Engineering, University of Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria

ABSTRACT

In the field of water distribution system (WDS) analysis, case study research is needed for testing or benchmarking optimisation strategies and newly developed software. However, data availability for the investigation of real cases is limited due to time and cost needed for data collection and model setup. We present a new algorithm that addresses this problem by generating WDSs from GIS using population density, housing density and elevation as input data. We show that the resulting WDSs are comparable to actual systems in terms of network properties and hydraulic performance. For example, comparing the pressure heads for an actual and a generated WDS results in pressure head differences of ±4 m or less for 75% of the supply area. Although elements like valves and pumps are not included, the new methodology can provide water distribution systems of varying levels of complexity (e.g., network layouts, connectivity, etc.) to allow testing design/optimisation algorithms on a large number of networks. The new approach can be used to estimate the construction costs of planned WDSs aimed at addressing population growth or at comparisons of different expansion strategies in growth corridors.

Software availability

Name of software: WDS Designer (Water Distribution System Designer)
Contact address: Robert Sitzenfrei, Unit of Environmental Engineering, University of Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria. Tel: ++43 (512) 507-6695; Fax: ++43 (512) 507-2911
Year first available: 2010
Software required: MCRInstaller.exe, version 7.13 or Matlab (R2010a) Version 7.10.0.499
Programming language: Matlab
Program size: ∼16 MB
Availability: Contact the authors to obtain this software and user manual
Cost: Free

1. Introduction

In the field of water distribution system (WDS) analysis, case study research is a well-known instrument for testing or benchmarking optimisation strategies, evaluating newly developed technologies, proving hypotheses, testing new software and analysing uncertainties in models. With a limited number of case studies for evaluations, investigations are case specific and it is difficult to transfer the obtained results to other boundary conditions. In addition, very differing case studies make it difficult to identify driving factors. However, data availability for real case research is limited due to the time and cost of data collection. In addition, the data on WDSs that are available are often incomplete or of insufficient quality. Fear of terrorist attacks has led to restricted access to sensitive WDS data (Brumbelow et al., 2007). However, such data are crucial for research tasks. Past studies on optimisation problems (di Pierro et al., 2009; Ostfeld and Tubaltzev, 2008; Behzadian et al., 2009; Giustolisi and Savic, 2010; Savić et al., 2011), water quality analyses or assessments (Rodriguez and Serodes, 1999; May et al., 2008; Tamminen et al., 2008; Wu et al., 2009a), performance analysis (Ramos et al., 2009) and demand forecasting (Alvisi et al., 2007; Zhou et al., 2002) had to use one or at most a few WDS case studies as application examples. To make e.g. optimisation studies comparable and also to tackle the limited data availability, benchmark models have been introduced in the

* Corresponding author. Tel.: +43 512 507 6695; fax: +43 512 507 2911.
 E-mail address: robert.sitzenfrei@uibk.ac.at (R. Sitzenfrei).

literature, including the Two Loop System (Alperovits and Shamir, 1977; Todini, 2000; Kalungi and Tanyimboh, 2003), New York City Tunnels (Schaake and Lai, 1969; Olsson et al., 2009), the Hanoi network (Fujiwara and Khang, 1990; Savic and Walters, 1997; Dandy et al., 1996) and the "Anytown" network (Walski et al., 1987; Walters et al., 1999). For incomplete data sets, Wang (2008) presented a Bayesian method for multi-site stochastic data generation. To tackle the problem of data availability, Brumbelow et al. (2007) presented two single virtual cities, denoted Micropolis and Mesopolis, that are fully described both in GIS and as an Epanet2 model (Rossman, 2000) for further analysis (e.g., Torres et al., 2009). However, the manual process for creating such a virtual case study is tedious, and investigations are very case specific.

In Möderl et al. (2011) a set of 2280 virtual water distribution systems (vWDSs) were generated by means of a stochastic approach. The set is based on variations in the layout and size of the vWDS. For this set, a constant elevation for all vWDS junctions is assumed, and the elevation of the reservoirs is kept constant. The demand per junction is determined by a stochastic variation. The layout characteristics of each virtual system are determined by the generation process (variation of size and expansion) and do not meet any boundary conditions or GIS data requirements. Hence, the generated vWDSs are conceptual and not fully comparable with real world WDSs. However, analyses using that set of vWDSs are based on a large set of different case studies thus providing general applicable findings instead of case specific and difficult to generalize results (Möderl et al., 2011).

The objective of this work is to generate WDSs benchmark configurations based on given GIS data in order to test different design approaches, such as based on the use of optimisation algorithms. The generation process of the WDSs is carried out using a newly developed graph concatenation approach (GCA) that is introduced in this paper. With the GCA, default blocks representing reoccurring network WDS motifs are selected and combined to form an entire WDS (network layout). The blocks are chosen to meet the requirements of given GIS data (housing and population densities) and follow a layout design strategy. Compared to existing approaches such as that presented in Möderl et al. (2011), more variations in the generation process can be performed. For example, the level of connectivity in the system can be varied from completely looped to branched but with user-defined intermediate levels. In addition, the number of sources, the pipe-sizing procedure, and other factors can easily be varied. Although elements like valves and pumps are neglected, the generated vWDSs can form the basis for further comprehensive investigations (e.g., Möderl et al., 2010; Hellbach et al., 2011). The method is applied to the GIS data from the real-world case study of Wolf-Codera Ranch (WCR), and the generated WDS data are compared with the real WDS data from the WCR.

2. Methods and materials

In this section, the GCA methodology developed for WDS generation is described. The approach is implemented in a standalone Matlab tool with a graphical user interface (GUI), denoted WDS Designer. The WDS Designer GUI is presented in Sitzenfrei et al. (2010). The results of the WDS generation process are discussed and compared with data from a real-world WDS (Fig. 1).

The key method of generating the layout of WDSs in this paper is the newly developed graph concatenation approach (GCA). Within the GCA, the modular design system (MDS, Möderl et al., 2011) is used for graph theory-based data administration and provides an interface to the hydraulic solver Epanet2 (Rossman, 2000). The choice to use Epanet2 is made for convenience because it is a widely applied software tool. Any other solver (including pressure-driven solvers such as those as used by, e.g., Todini, 2003; Wu et al., 2009b; Giustolisi and Laucelli, 2011) could be applied in the GCA as well without changing either the method or the results (the implemented diameter design and performance assessment are performed under standard conditions). The innovative approach concatenates different blocks from a database while meeting the requirements of the underlying GIS data

and following layout and reservoir design processes. For diameter design, an auto design algorithm for large systems that is based on economic flow velocity and the consideration of either future demands (Alvisi et al., 2007) or peak demands (Diao et al., 2010) is used. All of the parameters for these design strategies are determined by input parameters through the GUI of the WDS Designer. The parameters for the cost estimations that enable cost design, performance evaluation (e.g., sufficient supply pressure, water age and overall performance) and graph properties of the WDS can be defined via the GUI, and the results can be visualised. In addition, the GUI allows the import of Epanet2 case study data and the export of generated or redesigned networks to Epanet2.

2.1. Modular design system (MDS)

The generated WDSs are described using the graph theory-based MDS (Modular design system) approach presented in Möderl et al. (2011). Following the idea of describing complex networks with "network motifs" (Milo et al., 2002), the MDS can be used to describe and administrate entire network structures (e.g., energy supply networks, district heating networks or water distribution networks) by means of graph theory. For MDS, the demand junctions are points on a grid, and the number of connected pipes to the junctions are binary coded (Möderl et al., 2011). Hence, all of the pipes and junctions are represented by an explicit MDS graph matrix. In the MDS approach, an assumption is made that a maximum of four pipes can be connected to any single junction. For the application for WDSs presented in this work that assumption is appropriate (see Fig. 5, maximum 4 connections to a single junction). The coordinates of the junctions in the MDS can be freely defined (e.g., distances between the junctions or angles between the pipes). In this study, a quadratic arrangement of the grid points is used. Therefore, the four connections to a single junction are oriented in the four cardinal directions. By summing the binary values for the existing connections, the elements in the MDS graph matrix as shown in Fig. 2 range from 0 to 15 (no connections: $0 = 0 \cdot 2^0$ (east) $+ 0 \cdot 2^1$ (north) $+ 0 \cdot 2^2$ (west) $+ 0 \cdot 2^3$ (south); four connections: $15 = 1 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 1 \cdot 2^3$). All system attributes, such as demand per junction, elevation of each junction, pipe diameters, and pipe roughness, are organised within the MDS. The MDS provides an interface with the chosen hydraulic solver by writing text based input files. In this study, the well-known and freely available hydraulic solver Epanet2 is used because the hydraulic evaluations for the WDSs are performed under standard conditions (i.e., no source failures, pipe breakages, etc.). Other conditions might require the use of more advanced solution routines, which would not pose a problem because any other hydraulic solver that is either based on text files as input or available as an open source can be easily used with this method.

2.2. Graph concatenation approach (GCA)

With the GCA subsets of the water network, designated blocks can be concatenated from a database to form an entire WDS. The blocks in the database as well as the entire WDS are represented by graph matrices using the MDS. The implemented default blocks are based on a 5×5 grid raster and therefore a 5×5 graph matrix. Hence, a default block consists of a maximum of 25 junctions (see Fig. 2).

For the generated WDS, the decision of a 5×5 grid raster was made on the one hand to have small, generic network motifs and on the other hand in order to have a manifold database of network motifs. The default blocks have differing connectivities (see Fig. 3), i.e., number of dead end branches, differing layout structures, numbers of junctions in a block or numbers of redundant connections. The GCA presented here uses four sets of default blocks. These sets are denoted as O, L, H and U according to the appearance of their basic structure (see Fig. 3). The O set consists of blocks representing network motifs of looped network structures (high connectivity), while the L, H and U blocks represent blocks of branched network structures (low connectivity) (see Fig. 3). Using the MDS for data administration, the user can also design new blocks and add them to the GCA.

Because each element of the raster has its own interdependent coordinates, these blocks can be freely defined; all of the blocks can be scaled, warped and concatenated to bigger blocks through matrix concatenation. In this study, the GCA uses data that are rasterised to a cell size corresponding to the size of the default blocks. Depending on the level of detail of the obtained network, the cell size has to be chosen. For this study, Fig. 7 shows how the cell size of 300 m was determined. If no data is available, a cell size of 300 m is recommended.

The required GIS data (raster data) for population density, water demand and housing density are therefore reclassified according to the block raster ($i = 1...N$ raster cells with a housing density above 0). The housing density in this context describes the dwelling unit density, which can be determined based on dwelling sizes and floor space required per capita. Based on that, the number of junctions per block area is determined.

In the following section, the GCA flowchart is described and illustrated using an example for one cell (see Fig. 4). The input data for the GCA include both the GIS raster data and other input parameters such as the connectivity parameter F_{Layout} that specifies which block sets of the database can be used in the GCA. Based on the housing density for each raster cell $H(i)$ (for all cells $i = 1...N$) the GCA then determines which blocks meet the requirements of the GIS data and defined parameters ($F_{Layout} = 100\%$ specifies that the branched sets (L, H, U set) are used). For
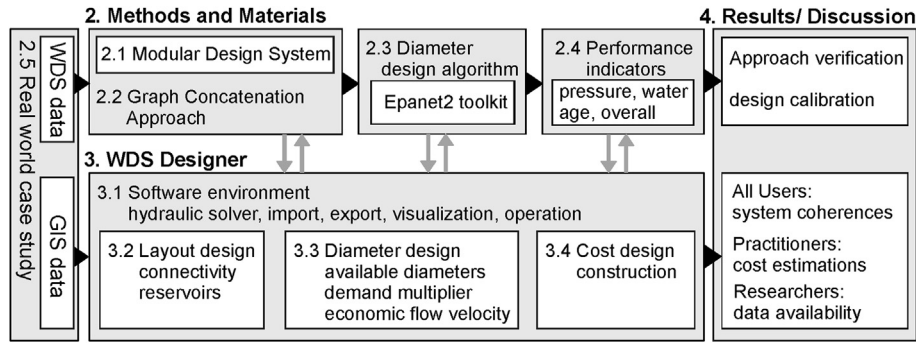
Fig. 1. Development and application of the presented approach.

example, a current housing density of $H(i) = 11$ requires 11 junctions in a block. In the database, there is a set of appropriate blocks {K} that meet these requirements. From among the appropriate blocks {K}, one block $k \in \{K\}$ is randomly selected. A random rotation angle $\alpha \in \{0°, 90°, 180°, 270°\}$ is applied to the block k, which results in block k_{rot}. Block k_{rot} is placed at position x,y by concatenating the graph matrix of block k_{rot} to the graph matrix of the entire WDS. Through the algorithmic application of this procedure to each cell and considering the variation of the selection process, an entire WDS can be composed. The water demand for each junction is distributed according to the population densities P(i) within the cells, and the elevation is determined by projecting the junctions on digital elevation map E. Finally, water sources are needed for the WDS. Reservoirs can be added to the WDS following a reservoir design (see Section 3.2.2).

2.3. Diameter design algorithm and construction costs

After the network layout is set, the system, and in particular the pipe diameters, must be designed based on the maximum demand. The generated WDS can be extensive, with several thousand nodes and pipes. For the optimisation of the WDS in terms of hydraulic performance, pipe diameters, pumping costs or CO_2-emissions, among other factors, various algorithms and approaches are available (Ostfeld and Tubaltzev, 2008; Maier et al., 2003; di Pierro et al., 2009; Gupta et al., 1999; Wu et al., 2010). The aim of the automatic pipe sizing algorithm used in this effort is not to obtain an optimised WDS but to design a WDS following state-of-the-art design rules in minimal computation time. Therefore, a single objective approach is implemented in this study and it is assumed that the diameter design is totally free (no limitations were e.g. large diameters are installed). Nevertheless, the generated WDS can be used to investigate the performance of optimisation algorithms via the evaluation of different generated WDSs with various layout characteristics.

For the diameter design used in this work, the available pipe diameters and corresponding economic flow velocities can be user-defined and input via the GUI of the WDS Designer. The by default implemented diameters and economic flow velocities are shown in Table 1. The implemented approach is based on an iterative calculation of flow velocities with Epanet2. Starting with the smallest available pipe diameter for all of the pipes (e.g. 80 mm), the flow velocities are determined using the Epanet2 hydraulic solver (first iteration). If the calculated flow velocity in a pipe exceeds the corresponding economic flow velocity for the current diameter, before the next iteration that diameter is incrementally increased to the next available diameter (see Table 1, e.g. from 80 mm to 100 mm). After incrementally increasing all of the pipe diameters with uneconomic flow velocities, a hydraulic recalculation (next iteration) is performed. This procedure is repeated until either the computed velocity in each pipe is below the economic flow velocity or the highest available pipe diameter is obtained. Therefore, for the 11 available diameters according to Table 1 first line, maximum 11 iterations are performed. By defining customised pipe construction costs for each pipe diameter or using the already implemented default values, a rough estimation of construction costs of the generated WDS is obtained.

2.4. Performance indicators

To estimate the hydraulic performance and water quality of the investigated real and generated WDSs, performance indicators for sufficient hydraulic pressure and
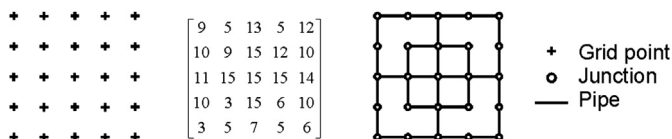
water age (i.e., residence time) based on Möderl et al. (2011) are used. The performance indicator for sufficient hydraulic pressure (PI1) is determined by

$$PI1 = \frac{\sum_{j=1}^{J} Q_{del,j}(p)}{\sum_{j=1}^{J} Q_{req,j}} \quad Q_{del}(p) : p \in (a,b);$$

In this equation, the sum of water $Q_{del}(p)$ that is delivered with sufficient pressure over all J junctions is divided by the sum of the required amount of water Q_{req}. The limits a and b for PI1 are the lower and upper hydraulic pressure heads, respectively. Varying values for pressure limits can be obtained from the literature and from national guidelines. Trifunovic (2006) reports minimum pressure limits for several cities ranging between 20 and 40 m per water column. Chung et al. (2009) report that 14 m is a sufficient water pressure. The maximum pressure limits may also vary. For example, for WDSs with high water losses and a high risk of pipe breakages, a low maximum pressure limit may be applied. On the other hand, for an alpine case study with large topographic height differences, maximum pressure limits up to 100 m and more are used (Möderl et al., 2012). Therefore, the pressure limits for evaluation are kept variable and can be defined by the user. By default, a lower pressure limit of $a = 40$ m and an upper pressure limit of $b = 100$ m are used. The sufficient water age can be determined by the equation for PI2:

$$PI2 = \frac{\sum_{j=1}^{J} Q_{del,j}(Age)}{\sum_{j=1}^{J} Q_{req,j}} \quad Q_{del}(Age) : Age \in (c,d)$$

In this equation, the sum of water $Q_{del}(Age)$ that is delivered with a sufficiently low water age over all J junctions is divided by the sum of the required amount of water Q_{req}. The lower limit for the PI2 (default $c = 0$ h) and the upper limit d are variable. The US-EPA investigated more than 800 water supply utilities, resulting in average residence times of 1.3–3 days (US-EPA, 2002). The default value for the upper limit of the water age is thus assumed to be $d = 24$ h, but it can also be defined differently by the user. PI3 is the product of PI1 and PI2, which represents a combined performance indicator:

$$PI3 = \frac{\sum_{j=1}^{J} Q_{del,j}(p) \cdot Q_{del}(Age)}{\sum_{j=1}^{J} Q_{req,j}} \quad Q_{del}(p) : p \in (a,b); \ Q_{del}(Age) : Age \in (c,d)$$

These performance indicators can be applied to real-world WDSs and generated WDSs, taking into account all nodes and pipes. Therefore, with these performance indicators, very different systems can be compared using these normalised single values PI1, PI2 and PI3.

2.5. Real-world case study data

To demonstrate the applicability of the GCA, a real-world case study consisting of an Epanet2 model of a WDS and GIS data is used. The GCA can be applied to any appropriate GIS data. In this paper, one case study data is used as input. The GIS data are derived from the Epanet2 model of the Wolf-Codera Ranch (WCR) (Lippai, 2005). The WCR is a region of expected future developments of the Colorado Springs utilities.

The given Epanet2 model provides data regarding the elevation and water demand for each junction in the WDS. From the water demand of the Epanet2 model, GIS data for population density as well as housing density can be estimated by evaluating the number of WDS junctions in a cell. The WCR model contains approximately 2000 pipes and 1800 junctions. The extension of the supply area is
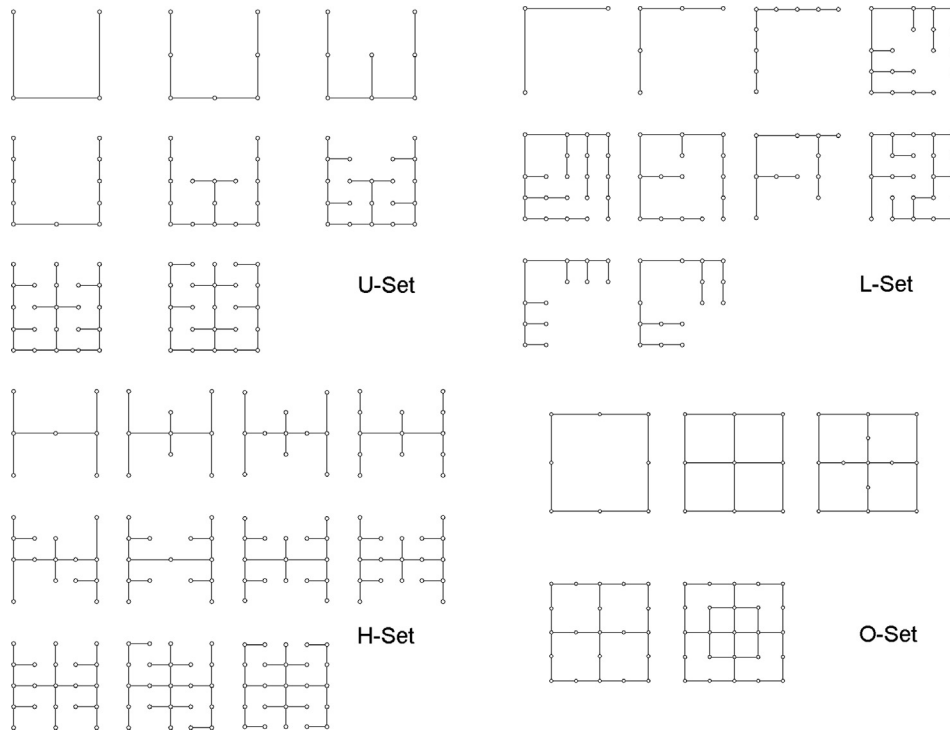

Fig. 2. Left side: basic block grid, middle: graph matrix of an O-Block, right side: O-Block layout.

Fig. 3. Default sets: O, L, H and U blocks with increasing numbers of junctions.

approximately 10 km² with an estimated average demand of 160 l/s. Nearly 90% of the pipes are DN200 (diameter 200 mm or 8 inch) or lower because of the utility's standards that require minimum diameters for distribution lines (DN200) and DN150 (diameters 150 mm or 6 inch) for dead ends (Lippai, 2005). A maximum DN600 (diameter 600 mm or 24 inch) is used in the planned WDS.

In Fig. 6, the digital elevation model of the WCR model is shown. The ground elevation of the service area is between 2.100 m and 2230 m. The elevation is only available at the junctions. Between the junctions, the grid is interpolated (Delaunay Triangulation, de Berg et al., 1997), and the grid is extrapolated to defined boundary points. The junctions of the water distribution system are plotted according to their elevation, as shown in Fig. 6.

In the WCR model, 36% of the nodes are without demand or connecting tasks (Sitzenfrei, 2010). Over 90% of these dead end pipes have lengths less than 10 m. Following the principles of skeletonisation (Walski et al., 2003), these dead ends without demand are removed for the analysis in this work. Therefore, 1150 junctions and approximately 1350 pipes remain in the system; these are further used to

determine the appropriate cell size. In the block database, blocks that contain from 3 to 25 junctions in a cell are available (see Fig. 3 and blue area in Fig. 7). The ways in which different cell sizes for rasterising housing density result in different distributions of required blocks are investigated. Cell sizes ranging from 50 m up to 500 m (in 50 m steps) are investigated. In Fig. 7, each cumulative distribution function (CDF) represents one cell size and specifies the distribution of the required blocks in the demand area for that specific cell size. With a cell size of 300 m (black CDF in Figure 7), 89% of the junction densities can be represented with available blocks (black line within blue area). Larger cell sizes (CDFs to the right of the black line) result in more than 25 junctions in a cell. Smaller cell sizes (CDFs to the left of the black line) mean that some of the available blocks are not used, which also results in short pipe lengths (see Fig. 11). For a cell size of 300 m, the cells that are not within the range of the database are those that require one or two junctions. It is assumed that these cells can be approximated with blocks with 3 or 4 junctions.

Therefore, in this work, the housing density is reclassified to a cell size of 300 m to obtain the highest range of junction densities that are covered with the block database and to approximate the pipe length distribution in the WCR model (see also Fig. 11). A detailed description of how the optimal cell size is determined can be obtained in Sitzenfrei (2010).

With the presented methodology, WDSs with different characteristics can be generated that, for example, have differing pipe sizing specifications. To ensure that the generated WDS and the WCR have similar layouts, a *wolf mode* is implemented in the approach. By applying this mode, the positions of the reservoirs, the available pipe diameters and the economic flow velocities for the pipe-sizing process are kept the same as those in the WCR or are set in such a way that similar characteristics can be obtained (see Section 3.3).

3. GCA implementation – WDS designer

3.1. Software environment

The presented GCA approach is implemented in WDS Designer. This tool is programmed using the Matlab environment and is operated by a GUI. It uses the open source software Epanet2 (Rossman, 2000) to run the hydraulic simulations for pipe diameter design and performance evaluation. The GUI functions as an operator's panel and allows visualisation of the network structure, parameter selection for the WDS generation process, and graph and pipe diameter distribution evaluation. All of the parameters in the following sections can be set via the GUI. Both the GUI and the
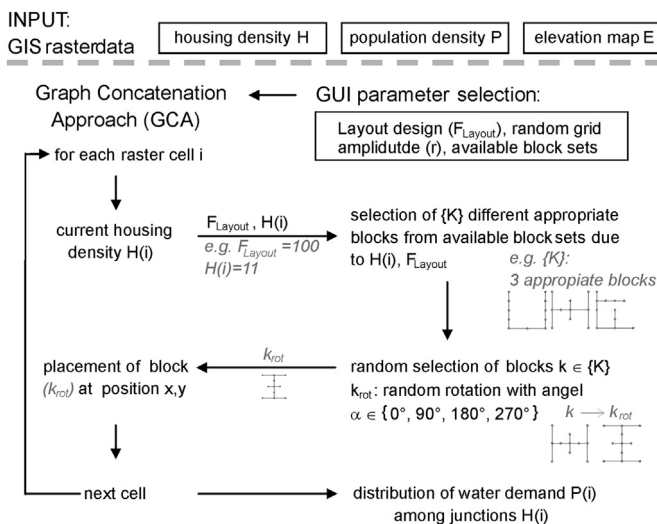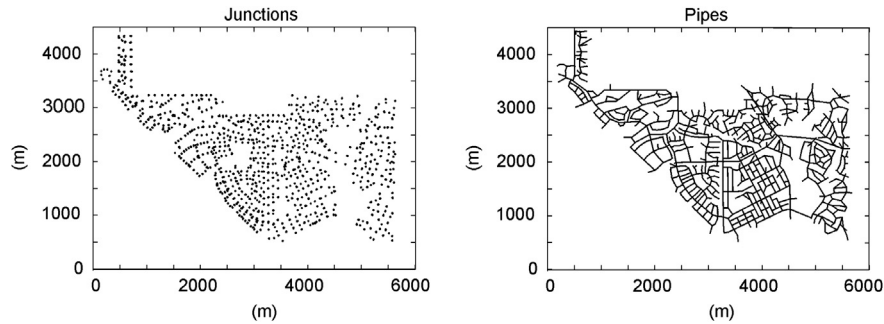

Fig. 4. GCA flowchart.

Fig. 5. Left side: demand and junction density, right side: layout of input data.

kernel are compiled using standalone application for Windows. This standalone application requires either Matlab 2010a or higher or MCRInstaller.exe, version 7.13. The MCRInstaller.exe is freely available on request for users of WDS Designer.

3.2. Layout design

The parameters for layout design control the WDS generation process. The generated WDSs are based on the MDS quadratic grid. To vary this constant grid, a stochastic deviation of the position can be applied at each grid point. The maximum deviation can be defined by the random grid amplitude (r) as a percentage of the grid size. The value for the random grid amplitude can be specified via the GUI of the WDS Designer.

3.2.1. WDS connectivity

The connectivity is determined by the factor F_{Layout}, which can be set via the GUI. The probability of selecting a loop or branched block for the generation process is thereby defined. Setting $F_{Layout} = 0\%$ results in the use of only O-Set blocks, which generate completely looped water distribution systems (high connectivity). If $F_{Layout} = 100\%$, only L, U or H set blocks are applied, which results in a branched water distribution system. By default, the branched set consists of the L, U and H blocks (low connectivity). A new user-defined block set can also be designed by writing the graph matrix in a text file. These user-defined sets can be added into the GUI and further selected in the generation process. Furthermore, the branched set or the looped set can be modified. For example, a branch system can be specified to use only L blocks.

3.2.2. Reservoirs for WDS

The reservoir parameter determines the number of reservoirs set in the generation process. The reservoirs are placed on either specified or random positions within the area of the WDS. To use random positions may result in poor hydraulic performance (Möderl et al., 2010). Therefore, it is recommended to use specified positions. At least one reservoir must be placed, and at most 10 reservoirs can be set in the generation process (the number 10 results from the implemented numerical routines only). The

reservoir height parameter defines the elevation above the surface at the current position.

In Fig. 8, on the right side, the positions of the reservoirs denoted as A, B1 and C in the WCR model are marked with red circles. In the WCR, there is no reservoir at B2, but there is connection with B1 via a pipe with a diameter of 600 mm. This connection cannot be included in the GCA. The Epanet2 model for the Wolf-Codera Ranch determines the system inflows at points B1 and B2 under standard conditions. This means that two reservoirs, B1 and B2, are used in the generated WDS instead of the single reservoir, B1, in the WCR. Furthermore, the hydraulic simulation of the WCR model reveals that reservoir C does not contribute to the water supply under standard conditions. Therefore, for the generated WDS, reservoir C is disregarded because hydraulic analyses are only performed under standard conditions in this study. If the *wolf mode* is activated, the heights and positions of the sources of the real-world data are used, but in the simplified manner described above (Fig. 8).

3.3. Diameter design

With the diameter design properties, the parameters for the auto pipe-sizing algorithm (available pipe diameters (see Section 3.3.1), demand multiplier and economic flow velocity) are defined. The aim of the auto pipe-sizing process is to maintain the defined economic flow velocities within the WDS.

3.3.1. Available diameters

Pipe diameters that are applicable in the auto pipe-sizing process are selected from an input data set denoted as the pipe vector. By default, pipe vector one (PV1, see Table 1) is used. Optionally, the diameters used in the real world data can be selected with pipe vector two (PV2, see Table 1). Additional user-defined pipe vectors with corresponding design velocities can be specified via the GUI.

3.3.2. Demand multiplier

With the demand multiplier (−)prospective future demand (Zhou et al., 2002; Alvisi et al., 2007; Ghiassi et al., 2008) as well as
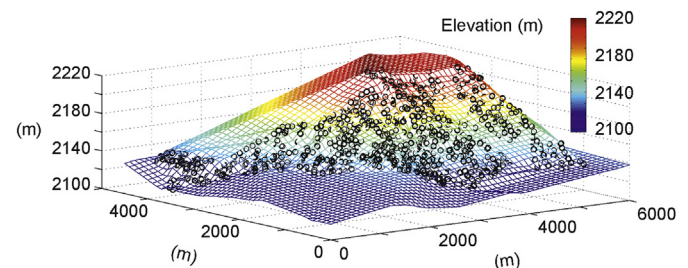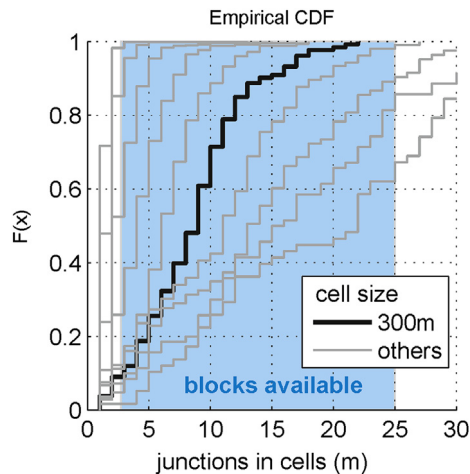
Table 1
Pipe vectors (pv1 and pv2) and related velocities and costs.

PV1 DN (mm)	80	100	125	150	200	250	300	350	400	500	600
Dim+ mode v (m/s)	1	1	1	1.5	1.5	1.75	1.75	2	2	2	2
PV2 DN (mm)				150	200		300		400		600
Wolf mode v (m/s)				0.015	0.015		0.7		1.2		1.2
Default costs (€/m)	57	63	77	91	121	161	196	246	296	391	500

Fig. 6. Digital elevation model of the input data, including the location of demand junctions.

Fig. 7. Blocks available (blue area) and junctions in cells with different cell sizes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

demand peaks (Diao et al., 2010) can be considered in the auto pipe-sizing process. The actual demand is therefore multiplied by a factor, usually in a range from 0.5 to 4(−). By default, the demand multiplier is set to 1.5(−) according to Trifunovic (2006).

3.3.3. Economic flow velocity

With the economic flow velocity (m/s), the maximum tolerable velocity for the auto pipe-sizing process is defined. The value range is defined from 0.5 m/s to 3.0 m/s.

The upper limit of 3.0 m/s is set to avoid water hammer problems (Walski et al., 2003). The defined constant flow velocity is considered for every pipe diameter. By default, the design velocity is set to 1.0 m/s for all diameters (Trifunovic, 2006). With the *dim+* mode provided by the GUI, the values for the velocities for different diameters are set according to Gujer (2002). If the *dim+* mode is activated for the auto pipe-sizing process, PV1 is selected with varying velocities for the diameters according to Table 1 (PV1 and *dim+* mode). The *wolf mode* provided by the GUI follows a strategy of generating a WDS with a similar diameter distribution as the WCR model. If the *wolf mode* in the auto design section is activated, PV2 is selected because only these parameters appear in the real world WDS of the WCR. The economic flow velocities for the diameters are set, as shown in Table 1 (PV2 and *wolf mode*). The values for the pipe velocities are derived from a calibration of the auto pipe-sizing algorithm to attain a comparable diameter distribution for the real world diameter data and the redesigned diameter data (see Fig. 11 right hand side).

3.4. Costs design

The cost design (total construction cost only − no cost of operation) is based on the assumption of costs per metre. The investment costs for pipes are estimated to represent 70% of the total construction costs. Table 1 shows the costs according to Mutschmann and Stimmelmayr (1999) multiplied by the factor 1/0.7. Optionally, a user-defined pipe vector with corresponding economic flow velocities and specific costs can be defined via the GUI. The estimation of costs in connection with the hydraulic performance allows for the economic design of WDSs.

4. Results and discussion

The results for several applications of the GCA are presented, and the generated WDSs are analysed. First, the WCR network is redesigned using the implemented auto pipe sizing and compared with the real-world data. This process gives a clear indication of the limits and potential of the proposed methodology. Second, generated WDSs with varying layouts are presented and investigated with regard to hydraulic performance, water quality and costs. Finally, the properties of the generated WDSs are compared with the real-world WDS.

4.1. WCR model

The WDS layout of the WCR model is shown in Fig. 9. The colours and the thicknesses of the pipes indicate their diameters. On the left side, the original data are shown, and on the right side, a redesigned WCR is shown. For the redesign process, PV1 is used for the available pipe diameters with an economic flow velocity of 0.5 m/s for all diameters and a demand multiplier of 2.5. For the cost estimations, the default costs in Table 1 are used. Fig. 9 reveals that the redesigned WDS has a similar spatial distribution for diameters equal to or higher than DN400 (400 mm or 16 inch). However, some redistribution of flow capacities can be observed. For example, the redesigned WDS has no transport pipe connections (diameters larger DN300 300 mm or 12 inch) between reservoir A and the parts of the right side of the WDS. In the real world, such additional connections, which are unnecessary under standard conditions, are built to lower the vulnerability of the distribution system (Todini, 2000; Kalungi and Tanyimboh, 2003). A visual comparison of the real and redesigned system regarding the smallest diameters reveals that the chosen parameter for auto pipe-sizing results in smaller diameters (DN80 80 mm). Therefore, the smallest diameters in the WCR are over-designed to cope with a higher prospective demand and to meet the utility standards of a DN150/DN200 minimum diameter.

Utilities' standard requirements result in higher costs in relation to the redesigned WDS (cost estimation original data: €17.2
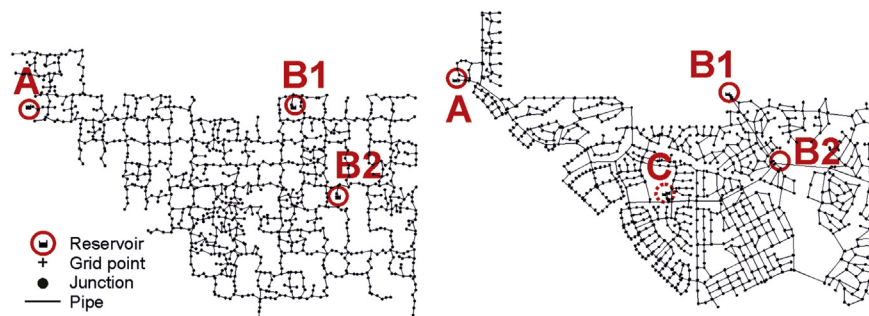
Fig. 8. Left: simplified positions of the reservoirs generated WDS; right: positions in real world.
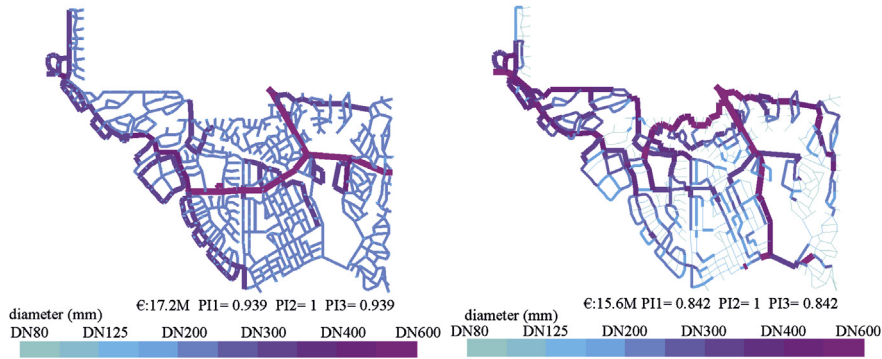
Fig. 9. Real-world WDS. Left: real-world diameter distribution, right: redesigned diameter distribution.

million; redesigned WDS: €15.6 million). The performance indicator PI1 for sufficient pressure is also different (see Fig. 9; pressure limits 40 and 100 m; original data: PI1 = 0.939; redesigned WDS: PI1 = 0.842). With limits for water age of 0 and 24 h, PI2 is sufficient for all of the junctions in both the original WDS and the redesigned WDS.

4.2. Generated WDS

In this second application, WDSs are created based on the GIS data of the WRC model but with varying system properties regarding connectivity, reservoir placement, and other variables. Fig. 10 shows two WDSs that are generated with reservoirs corresponding to the real world. On the left side, a connectivity factor of F_{Layout} = 0% is applied. As a result, a completely looped WDS is generated. For the auto pipe-sizing process for the looped WDS, economic flow velocities according to *wolf mode* are used, and a demand multiplier for prospective demand of 1.5 is applied. The PV2 data are used for available diameters. The right side of Fig. 10 shows a WDS that is generated with F_{layout} = 100%, and therefore, only branched blocks are used for the GCA. For the auto pipe-sizing process, an economic flow velocity of 1 m/s is considered, and a demand multiplier for prospective demand of 1.3 is applied to demonstrate diversity in the design process.

The PI1 for the completely looped system PI1 (with pressure limits of 20 and 100 m) is marginally lower (PI1 = 0.900) than that for the branched layout (PI1 = 0.929). The difference is 0.029 in PI1, which means that compared to the branched system, 2.9% of the demand weighted pressure heads at the junctions of the looped system (see definition of PI1) are outside of the specified pressure limits. From the detailed hydraulic evaluation (see Section 4.4), it is observed that the looped system has lower friction losses due to its high redundancy, resulting in a 2.9% shift beyond the upper

pressure limit (above 100 m). For both of the generated WDSs, sufficient water qualities are obtained (PI2 = 1). The cost for the looped WDS is €15.6 million, which is higher than that for the branched WDS because PV2 (DN150 for the smallest available diameter) is used and the connection redundancy and therefore the number of pipes are high. To highlight possible variations for the entirely looped system, a demand multiplier of 1.5 is applied, and for the branched system, a demand multiplier of 1.3 is applied. The costs for the looped system are lower than those in the real world. This difference can be attributed to the pipe-sizing in the WCR model and the redundancy of large-diameter back bone pipes. For the WDS that is generated using only branched blocks, the costs are €10.5 million, which is significantly lower than for the real WDS. The costs are lower because of the lower number of connections, the use of PV1 for the auto pipe-sizing process and the demand multiplier of 1.3.

4.3. Pipe lengths and diameter distributions

On the left side of Fig. 11, the pipe lengths of the systems are analysed by means of cumulative distribution functions (CDF). If no random grid amplitudes are considered ($r = 0$), then the pipe lengths of a WDS are a constant 75 m (Fig. 11 left hand side WDS r_0, black solid line). A cell size of 75 m is derived from the rasterisation of the housing density map to maintain a maximum junction density of 25 (junctions/cell), which corresponds with the maximum junction density of the blocks in the GCA. The median value of the WCR model corresponds to the cell size and therefore approximately to the median values of all of the generated WDSs ($r = 0$; 33; 67; 100). With an increasing deviation r, there is an increasing variation in the length distribution.

Comparing the small pipe lengths (below the cell size of 75 m), there are more short pipes in the WCR than in the generated WDS.
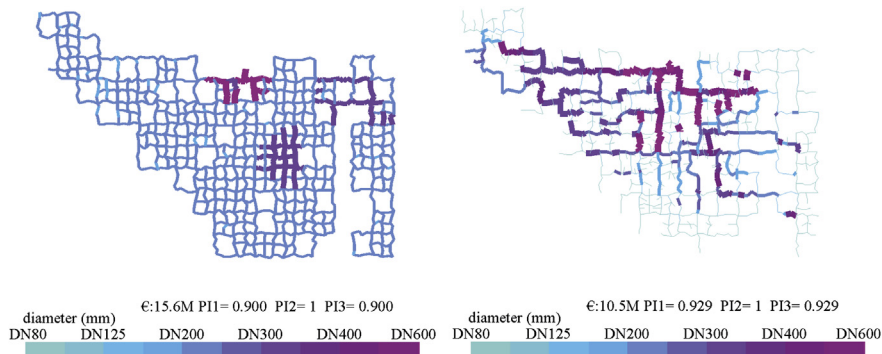

Fig. 10. Generated WDS. Left: diameter distribution looped layout, right: diameter distribution branched layout.
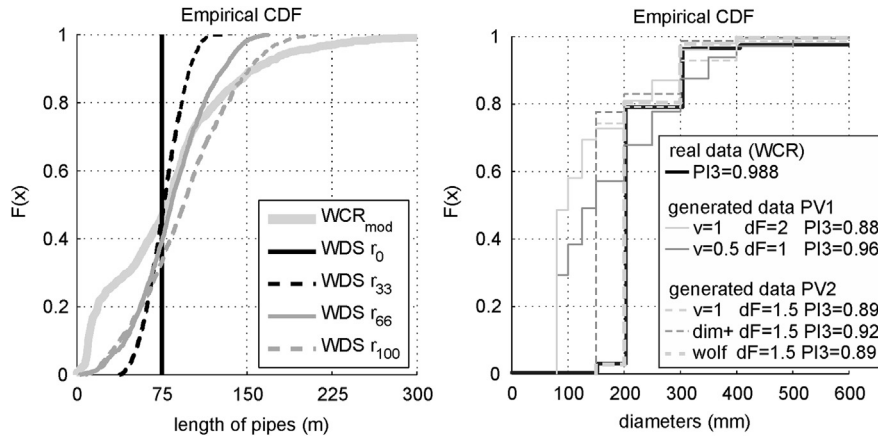
Fig. 11. Pipe length distributions of generated and real world WDS (WCR).

Table 2
Comparison of pressure heads.

Description	Maximum (m)	Minimum (m)	Median (m)
Real-world WDS	109.7	10.2	68.6
Generated WDS completely looped	113.1	10.5	76.2
Generated WDS completely branched	102.0	10.6	68.4
Generated WDS $F_{Layout} = 50\%$	105.9	10.1	68.3

The reason for this difference is that these short pipes cannot be generated by means of the GCA in its current form. This issue could be addressed by taking into account short dead-end pipes in the design of the blocks. Approximately 50% of the pipe lengths in the WCR can be directly reproduced with the generated WDS. By taking into account that in the generated WDS, longer pipes are implicitly included in the fractional pipe mean, approximately 75% of the cumulative distribution of the pipe lengths in the generated WDS is in good agreement with the distribution in the WCR.

The right side of Fig. 11 shows the CDF of diameters of a generated WDS ($F_{Layout} = 50\%$) that has different connectivities (pipe vectors PV1 or PV2, velocity v, demand multiplier dF) than the WCR (Fig. 11 right hand side, black solid line). The performance indicators PI3 for all of the WDSs are within a narrow range of 0.88–0.99 (for water pressure the limits 20 and 100 m were used and for water age 0 and 24 h). By applying the design diameter vector PV1 (grey solid lines), at least 55% of the pipe diameters are below DN200. For the three generated WDS on the right side of Fig. 11, PV2 is used and a demand multiplier of 1.5 is applied (grey dotted lines). For these three generated WDSs, the economic flow velocities are varied. In comparison with the WCR (black solid line),

a constant economic flow velocity of 1 m/s (light grey solid line) leads to more small-diameter pipes (DN150) for the generated WDS (75% are DN150). By considering the economic flow velocities according to *dim+ mode* (dark grey solid line), even more small diameters occur. The wolf mode utilises economic flow velocities and a demand multiplier to obtain a similar diameter distribution for the generated WDS (thick grey dotted line) and the WCR. There is a good fit between the real-world data and the generated WDS design with the *wolf mode*.

4.4. Hydraulic performance

For a comparison of hydraulic performance, hydraulic pressure heads of the WCR and the generated WDS are investigated. Table 2 shows the maximum, minimum and median pressure heads for different systems. The layouts of the completely looped and completely branched generated systems are shown in Fig. 10. In addition, a generated system with $F_{Layout} = 50\%$ is investigated. The minimum values for the pressure heads for all of the generated systems are in good agreement with the real-world WDS. In addition, the median values are in good agreement, except in the completely looped system. This lack of agreement can be traced to the overcapacity and high redundancy in that system. This redundancy also applies to the maximum pressure for the completely looped system. For the completely branched system, there are more head losses due to the small amount of redundancy, and therefore, this system has the largest difference in maximum pressures heads compared to the real-world WDS.

On the left side of Fig. 12, the spatial distribution of the pressure head differences is shown. For both, the real world system and the
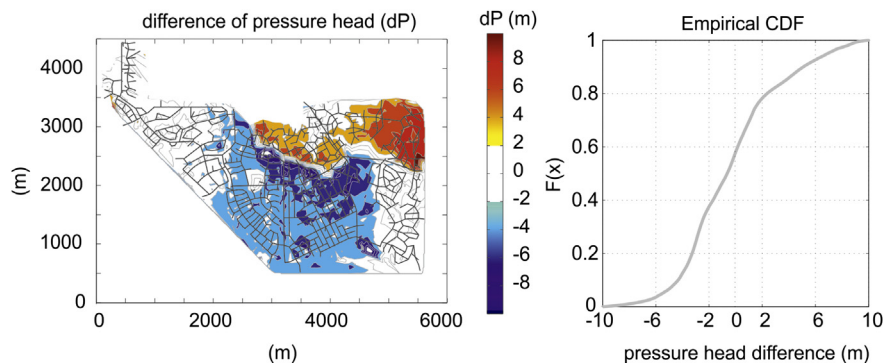

Fig. 12. Difference of pressure head of WCR and generated system.

Table 3
Sum of construction costs for WDS and generated WDS.

Description	Figure	Construction costs 2005 (millions €)
Real-world WDS	Fig. 9 left side	17.2*1.13 = 19.4
Real-world WDS redesigned	Fig. 9 right side	15.6*1.13 = 17.6
Generated WDS completely looped	Fig. 10 left side	15.6*1.13 = 17.6
Generated WDS completely branched	Fig. 10 right side	10.5*1.13 = 11.9

generated system with $F_{Layout} = 50\%$, the surfaces of the hydraulic pressure heads are calculated. This calculation is necessary because the specific locations of the junctions in the two systems do not coincide exactly. The hydraulic pressure head surface of the generated system is then subtracted from the hydraulic pressure head surface of the real-world system. On the left side of Fig. 12, the white areas are where the absolute value of the difference in the hydraulic pressure head is less than 2 m. This area is approximately 40% of the entire supply area (see also the CDF in Fig. 12 right). The light blue (in the web version) areas and the orange (in the web version) areas indicate that the difference is between −2 and −4 m, and between 2 and 4 m, respectively. These areas account for approximately 35% of the supply area. The absolute value of the pressure head difference is more than 8 m for only 3% of the supply area, and the entire supply area is within ±10 m.

On the left side of Fig. 12, there are no connections between the upper and the lower parts of the real-world system at the black dashed line. However, there are connections in the generated system; these connections exist because the gap between the two parts in comparison to the cell size is rather small and is therefore not detected in the generation process. In the spatial pressure distribution, there is a distinct difference (orange area next to blue area). The difference in pressure head in that area is less than ±6 m. By contrast, on the lower right side of the system, the gap in the real-world system was detected in the generation process, and therefore, there is a good agreement with regard to hydraulic pressure heads in that area (differences less than ±2 m).

4.5. Construction costs estimation

The construction costs for the four different WDSs are summarised in Table 3. Because the cost estimation for the original WCR model is based on Lippai (2005), all cost estimations are discounted to the year 2005 based on inflation. Considering the cost estimation generated, the WDS Designer for the original WDS of €17.2 million (based on costs from the year 1999) and an inflation rate of 13% for the period 1999–2005 results in estimated 2005 construction costs of €19.4 million (based on German experience). This is in the same value range as the cost estimation of Lippai (2005), which was $28.8 million (utility experience in United States: €21.3 million, exchange rate as of 03.01.2005). The estimated costs for the original WCR model are higher than for the redesigned system as well as for the two generated WDSs because the auto diameter design is based on the economic flow velocity under regular conditions but does not include safety factors that would be included in the real world. Therefore, the implemented pipe-sizing algorithm could be enhanced in future analyses by including critical conditions. However, even without those safety factor considerations, the values for the cost estimations are very close to each other.

5. Conclusion and outlook

In this paper, the graph concatenation approach (GCA) for the algorithmic generation of water distribution systems (WDSs) based on GIS data is developed and presented. It is shown that the geometrical properties of the generated WDS are comparable to real-world WDSs because the design algorithm provides distribution functions for pipe length and diameter that are equivalent to real world WDSs. It is also shown that the hydraulic performance of the generated and the real-world WDS are only marginally different in terms of the minimum, maximum and median pressures. Additionally, the spatial distribution of the hydraulic pressure head differences between the generated and the real-world WDS shows that sufficient agreement is achieved. A difference in hydraulic pressure heads of less than ±4 m is observed for 75% of the supply area, and for 40% of the supply area, the difference is less than ±2 m. Furthermore, the robustness of the developed procedure is demonstrated by generating, investigating and comparing different virtual WDSs. For example, the impact of different network connectivity on hydraulic performance and costs is demonstrated.

The strength of the presented approach is that it provides complex WDS with varying properties based on GIS data that are valuable for testing models and algorithms as well as for other purposes. The GCA is implemented in the Matlab tool WDS Designer. Therefore, for practitioners and researchers the GCA can be utilised and can help them in identifying system properties and interactions. For practitioners, the GCA helps to roughly assess the construction costs of WDSs based on GIS data for population, housing density and elevation and to compare different expansion strategies for growth corridors. However, the WDS generated with the GCA cannot substitute for a detailed real-world planning process.

In future developments, an interface to GIS data will be implemented to apply the graph concatenation approach to any boundary conditions. In implementing the GIS interface, the cell size of the rasterisation can correspond to the scaling of the used blocks and can therefore be defined by the user to obtain the desired pipe lengths. An additional future feature is the use of a generic grid that is not restricted to a quadratic shape. This can be achieved by implementing a grid based on potential theory following, for example, a river course, costal lines or main transport routes. Furthermore, a connection of the generated WDS to construction of new developments will be integrated into the system, and a definition of the positions of reservoirs should be included to provide improved construction cost estimations.

References

Alperovits, E., Shamir, U., 1977. Design of optimal water distribution-systems. Water Resources Research 13 (6), 885–900.

Alvisi, S., Franchini, M., Marinelli, A., 2007. A short-term, pattern-based model for water-demand forecasting. Journal of Hydroinformatics 9 (1), 39–50.

Behzadian, K., Kapelan, Z., Savic, D., Ardeshir, A., 2009. Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. Environmental Modelling & Software 24 (4), 530–541.

Brumbelow, K., Torres, J., Guikema, S., Bristow, E., Kanta, L., 2007. Virtual Cities for Water Distribution and Infrastructure System Research. World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat, 15–19 May 2007, Tampa. Florida.

Chung, G., Lansey, K., Bayraksan, G., 2009. Reliable water supply system design under uncertainty. Environmental Modelling & Software 24 (4), 449–462.

Dandy, G.C., Simpson, A.R., Murphy, L.J., 1996. An improved genetic algorithm for pipe network optimization. Water Resources Research 32 (2), 449–458.

de Berg, M., van Kreveld, M., Overmars, M., Schwarzkopf, O., 1997. Computational Geometry: Algorithms and Applications. Springer, ISBN 354061270X.

di Pierro, F., Khu, S.T., Savic, D., Berardi, L., 2009. Efficient multi-objective optimal design of water distribution networks on a budget of simulations using hybrid algorithms. Environmental Modelling & Software 24 (2), 202–213.

Diao, K., Barjenbruch, M., Bracklow, U., 2010. Study on the impact of peaking factors on a water distribution in Germany. Water Science and Technology: Water Supply 10 (2).

Fujiwara, O., Khang, D.B., 1990. A two-phase decomposition method for optimal design of looped water distribution networks. Water Resources Research 26 (4), 539–549.

Ghiassi, M., Zimbra, D.K., Saidane, H., 2008. Urban water demand forecasting with a dynamic artificial neural network model. Journal of Water Resources Planning and Management-ASCE 134 (2), 138–146.

Giustolisi, O., Laucelli, D., 2011. Water distribution network pressure-driven analysis using the Enhanced Global Gradient Algorithm (EGGA). Journal of Water Resources Planning and Management-ASCE 137 (6), 498–510.

Giustolisi, O., Savic, D., 2010. Identification of segments and optimal isolation valve system design in water distribution networks. Urban Water Journal 7 (1), 1–15.

Gujer, W., 2002. Siedlungswasserwirtschaft. Springer-Verlag, Berlin, ISBN 3-540-43404-6.

Gupta, I., Gupta, A., Khanna, P., 1999. Genetic algorithm for optimization of water distribution systems. Environmental Modelling & Software 14 (5), 437–446.

Hellbach, C., Moderl, M., Sitzenfrei, R., Rauch, W., 2011. Influence of Network Properties and Model Purpose on the Level of Skeletonization. World Environmental & Water Resources Congress, May 22-26, 2011, Palm Springs, California, USA.

Kalungi, P., Tanyimboh, T.T., 2003. Redundancy model for water distribution systems. Reliability Engineering & System Safety 82 (3), 275–286.

Lippai, I., 2005. Water System Design by Optimization: Colorado Springs Utilities Case Studies. Online Resource: http://centres.exeter.ac.uk/cws/benchmarks/expansion/42-wolf-cordera-ranch.

Maier, H.R., Simpson, A.R., Zecchin, A.C., Foong, W.K., Phang, K.Y., Seah, H.Y., Tan, C.L., 2003. Ant colony optimization distribution for design of water systems. Journal of Water Resources Planning and Management-ASCE 129 (3), 200–209.

May, R.J., Dandy, G.C., Maier, H.R., Nixon, J.B., 2008. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. Environmental Modelling & Software 23 (10–11), 1289–1299.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: simple building blocks of complex networks. Science 298 (5594), 824–827.

Möderl, M., Sitzenfrei, R., Rauch, W., 2010. How Many Network Sources Are Enough? World Environmental & Water Resources Congress EWRI, May 16-20, 2010, Rhode Island, USA.

Möderl, M., Sitzenfrei, R., Fetz, T., Fleischhacker, E., Rauch, W., 2011. Systematic generation of virtual networks for water supply. Water Resources Research 47.

Möderl, M., Sitzenfrei, R., Jarosch, H., Rauch, W., 2012. Identifying Hydropower Potential in Water Distribution Systems of Alpine Regions. World Environmental & Water Resources Congress, May 20–24, 2012, Albuquerque, New Mexico.

Mutschmann, J., Stimmelmayr, F., 1999. Taschenbuch der Wasserversorgung. Vieweg, Braunschweig.

Olsson, R.J., Kapelan, Z., Savic, D.A., 2009. Probabilistic building block identification for the optimal design and rehabilitation of water distribution systems. Journal of Hydroinformatics 11 (2), 89–105.

Ostfeld, A., Tubaltzev, A., 2008. Ant colony optimization for least-cost design and operation of pumping water distribution systems. Journal of Water Resources Planning and Management-ASCE 134 (2), 107–118.

Ramos, H., Tamminen, S., Covas, D., 2009. Water supply system performance for different pipe materials part II: sensitivity analysis to pressure variation. Water Resources Management 23 (2), 367–393.

Rodriguez, M.J., Serodes, J.B., 1999. Assessing empirical linear and non-linear modelling of residual chlorine in urban drinking water systems. Environmental Modelling & Software 14 (1), 93–102.

Rossman, L.A., 2000. EPANET 2 User Manual. National Risk Management Research Laboratory – U.S. Environmental Protection Agency, Cincinnati, Ohio.

Savic, D.A., Walters, G.A., 1997. Genetic algorithms for least-cost design of water distribution networks. Journal of Water Resources Planning and Management-ASCE 123 (2), 67–77.

Savić, D.A., Bicik, J., Morley, M.S., 2011. A DSS generator for multiobjective optimisation of spreadsheet-based models. Environmental Modelling & Software 26 (5), 551–561.

Schaake, J., Lai, D., 1969. Linear Programming and Dynamic Programming Applications to Water Distribution Network Design. No. 116. Department of Civil Engineering, Massachusetts Institute of Technology, USA.

Sitzenfrei, R., 2010. Stochastic Generation of Urban Water Systems for Case Study Analysis. PhD thesis. Unit of Environmental Engineering, University of Innsbruck. Free available at: http://www.uibk.ac.at/umwelttechnik/teaching/phd/diss_sitzenfrei.pdf.

Sitzenfrei, R., Möderl, M., Rauch, W., 2010. WDS Designer – a Tool Algorithmic Generation of Water Distribution Systems Based on GIS Data. World Environmental and Water Resources Congress, 4193–4204. http://dx.doi.org/10.1061/41114(371)427.

Tamminen, S., Ramos, H., Covas, D., 2008. Water supply system performance for different pipe materials part I: water quality analysis. Water Resources Management 22 (11), 1579–1607.

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

Todini, E., 2003. A more realistic approach to the "extended period simulation" of water distribution networks. Advances in Water Supply Management, 173–183.

Torres, J.M., Brumbelow, K., Guikema, S.D., 2009. Risk classification and uncertainty propagation for virtual water distribution systems. Reliability Engineering & System Safety 94 (8), 1259–1273.

Trifunovic, N., 2006. Introduction to Urban Water Distribution. UNESCO-IHE Lecture Note Series. Taylor & Francis Group, London, ISBN 0-415-39517-8.

US-EPA, 2002. Effects of Water Age on Distribution System Water Quality. United States Environmental Protection Agency, Washington.

Walski, T.M., Downey Brill, E., Gessler, J., Goulter, I.C., Jeppson, R.M., Lansey, K., Lee, H.-L., Liebman, J.C., Mays, L., Morgan, D.R., Ormsbee, L., 1987. Battle of the network models: epilogue. Journal of Water Resources Planning and Management 113 (2), 191–203.

Walski, T.M., Chase, D.V., Savic, D.A., Grayman, W.M., Beckwith, S., Koelle, E., 2003. Advanced Water Distribution Modeling and Management. ISBN: 978193449301.

Walters, G.A., Halhal, D., Savic, D., Ouazar, D., 1999. Improved design of "Anytown" distribution network using structured messy genetic algorithms. Urban Water 1 (1), 23–38.

Wang, Q.J., 2008. A Bayesian method for multi-site stochastic data generation: dealing with non-concurrent and missing data, variable transformation and parameter uncertainty. Environmental Modelling & Software 23 (4), 412–421.

Wu, J., Lu, J., Wang, J., 2009a. Application of chaos and fractal models to water quality time series prediction. Environmental Modelling & Software 24 (5), 632–636.

Wu, Z.Y., Wang, R.H., Walski, T.M., Yang, S.Y., Bowdler, D., Baggett, C.C., 2009b. Extended global-gradient algorithm for pressure-dependent water distribution analysis. Journal of Water Resources Planning and Management-ASCE 135 (1), 13–22.

Wu, W.Y., Simpson, A.R., Maier, H.R., 2010. Accounting for greenhouse gas emissions in multiobjective genetic algorithm optimization of water distribution systems. Journal of Water Resources Planning and Management-ASCE 136 (2), 146–155.

Zhou, S.L., McMahon, T.A., Walton, A., Lewis, J., 2002. Forecasting operational demand for an urban water supply zone. Journal of Hydrology 259 (1–4), 189–202.