Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 70 (2014) 1055 – 1062
Procedia Engineering
www.elsevier.com/locate/procedia

12th International Conference on Computing and Control for the Water Industry, CCWI2013

Improving incomplete water distribution system data

M. Mair[a]*, W. Rauch[a], R. Sitzenfrei[a]

[a]*Unit of environmental engineering/University Innsbruck, Technikerstr. 13, 6020 Innsbruck, Austria*

Abstract

Data for water distribution systems (WDS) are often not available or of poor quality. One promising approach for improving these is collecting data sets with strong coherences to the WDS and reconstructing possible WDS by using these data sets. An example for such a strongly correlated data set is the street network which can easily be accessed (e.g. open street map). The aim of this paper is to systematically analyze the impact of data improvement from alternative sources for creating WDS models. Investigations showed that hydraulic WDS models with a mean pressure error of three meters can be created by knowing 30% of pipes with a diameter ≥250 mm.

Keywords: Improving data; water supply system modeling; stochastic data reconstruction; street network data; hydraulic modeling

1. Introduction

For water distribution system analysis, data from case studies is crucial. But depending on the modeling aim and therefore the required accuracy of the model output, different levels of data quality are required (Hellbach et al., 2011). Contrary to practical applications where the aim is to describe a specific system sufficiently and very accurately, for research purposes it is often more important to gather information on many different water distribution systems with different characteristics in order to obtain case unspecific results from evaluations (Sitzenfrei et al., 2013). Therefore, recently researchers started to create few virtual water distribution systems

* Corresponding author. Tel.: +43-512-507-62129.
E-mail address: michael.mair@uibk.ac.at
The presentation of this paper, delivered by Robert Sitzenfrei, received the *CCWI 2013 Early Career Award*

doi:10.1016/j.proeng.2014.02.117

manually (Torres, 2006) or even automatically in great number (e.g. Möderl et al., 2007; Sitzenfrei et al., 2010b; Möderl et al., 2011; Trifunovic et al., 2012; Muranho et al., 2012). All these approaches are capable to create different kinds of water distribution systems. But a lack of all these approaches is that the generated network layouts are only a simplified representation of real systems. E.g. the approach presented in Möderl et al., 2011) has the limitation that only four pipes can be connected to one junction in the cardinal directions. This results in a rectangular grid of junctions in which all pipes have the same lengths. Likewise only reservoirs and junctions are regarded, therefore only gravity driven water supply can be represented in those models. Hence, more accurate approaches which mimic the layout characteristics of real world systems are required. More recently, researchers also investigated the evolvement of water distributions systems over time (e.g. Yazdani et al., 2011; Chang et al., 2012). For these research tasks like identifying and modeling optimal and robust future expansion strategies for water distribution systems the issue of data availability is even more challenging. One way to address the lack of data in this context is also to come back to data of virtual water distribution systems. Sitzenfrei et al., 2012) presented an approach for modeling dynamic expansion of water distribution systems for new urban developments. In that work, only virtual data was regarded because for calibration and validation of such an approach, historical data is required. But collecting historical network data for analyzing the dynamics of a network is a complex and time consuming task (Sitzenfrei et al., 2010a). In order to address the issue of simplified representation of real systems for the automatic generation of virtual water distribution systems and also to gain knowledge and sufficient data on the historical development of water distribution systems, new approaches are required. One promising approach for improving insufficient data sets is 1) collecting easy to access data sets with strong coherences to the water distribution system and 2) reconstructing possible water infrastructure data with a stochastic approach from these data sets. Sitzenfrei et al., 2013) presented an approach in which water distribution systems can be generated based on GIS data of elevation, population and housing densities. With that approach also different types of networks can be generated (e.g. looped or branched networks) and promising results were obtained for the pressure distribution in the investigated supply area. For sewer model creation, Blumensaat et al., 2012) presented an approach for generating possible representations of sewer models (Rossman, 2004) which is based on the street layout. In Mair et al., 2012) coherences in capacity, design and layout of water infrastructures are analyzed and described by comparing street, water supply and sewer network data. According to this study, there is a strong coherence for the layout between street network and water distribution network.

The aim of this paper is to analyze the impact of data improvement from other available data sources (e.g. street network data, population density) for creating water distribution models. Different data sets with different quality are compared with regard to the hydraulic performance, layout and asset costs between model results provided by a) improved incomplete data (e.g. only poor knowledge on spatial layout of the water distribution system improved with street network data) and b) the complete water distribution network model. By systematically altering data sets of incomplete water distribution network and water demand data and with a successive comparison of the improved incomplete network model with the complete network model, the impact on e.g. hydraulic performance can be quantified. All results presented in this paper are based on a detailed case study in the Alps with approximately 120 thousand inhabitants.

Nomenclature

DEM	(m) Digital elevation map
DN	(mm) diameter
J	(-) Set of supply junctions in a water distribution model
#J	(-) number of elements in a set J
p	(m) hydraulic pressure
PI	performance indicator
PI_1	(-) Performance indicator for the relative pressure difference between models
PI_2	(m) Performance indicator for mean absolute pressure difference between models
Q	(m^3/s) water demand

2. Material and methods

The principal idea of this work is use easy to access data like street layout, elevation, and population densities to create semi-virtual hydraulic models. The presented generation procedure is repeated by using different amount of the available data for creation of the semi-virtual hydraulic models (Rossman, 2000). In a next step, the results of the hydraulic modeling of each created model are compared with traditional assembled and calibrated hydraulic water supply models (which is assumed to be the exact representation of the system).

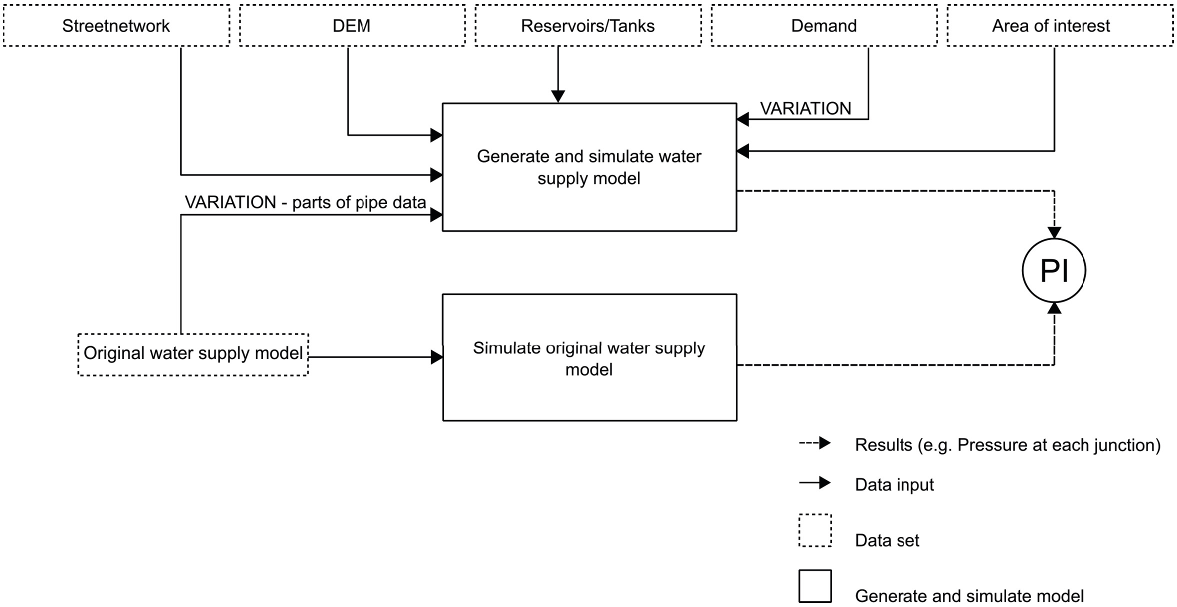

Fig. 1. Workflow setup for benchmarking artificial generated water supply models.

For that, a benchmark system was developed to automatically generate water supply models (EPANET2 Rossman, 2000) and evaluating their performance according to two different performance indicators (PI). The whole benchmark system was set up in a scientific workflow engine called DynaMind (Urich et al., 2012). Fig. 1 shows the general concept of the benchmark system. In the context of this work we have identified five different data sets (Fig.1 top row boxes), which are at least needed to automatically generate and simulate hydraulic models. The accuracy of the automatic generated model is determined by comparing simulation result of the generated model with simulation results from an already existing exact model. To determine the impact of additional data (e.g. location, diameter, roughness of real system pipes or demand with spatial resolution) the benchmark simulations are repeated by successively increasing the amount of additional data derived from the exact model.

2.1. Water supply network layout and pipe sizing

In Mair et al., 2012) it was demonstrated that up to 78 percent of a street network are containing up to 81 percent of the total length of a water supply network. According to that fact it is obvious to use street network data during the modeling process because of its easy accessibility (e.g. open street map, Google maps). This data is used to determine the layout of a water supply network by means of graph theory, which is the starting point for automatically generating a water supply model. The basic steps of the algorithm to generate a water supply network model in this work are:

- Trim the street network data set according to the area of interest
- Join additional pipe data set with the street data set
- Generate a minimum spanning tree out of the joined data set (Kruskal, 1956)
- Distribute demand points over the spanning tree
- Remove all leafs which have no demand points (A leaf is a vertex with only one connected edge)
- Generate loops
- Connect all reservoirs and/or tanks to the looped graph
- Pipe sizing of the generated graph based on a simple pipe sizing algorithm (Sitzenfrei et al., 2013)

A detailed description of the graph theoretical generation procedure can be obtained in a future publication. The output of this algorithm is an EPANET2 model which can be simulated and compared with the original EPANET2 model within the area of interest.

2.2. Performance indicators – PI

For evaluating the accuracy of the generated water supply models two different performance indicators (Eq.(1) and Eq.(2)) are used. Both of them are comparing the pressure at different locations between the generated and original EPANET2 model. Table 1 shows the exact meaning and definition of each variable and function occurring in both equations (Eq.(1) and Eq.(2)).

Table 1. Definition of variables and functions within both performance indicators.

Function or variable name	Set of all junctions with a demand greater zero in the original water supply model
J	Number of elements in the set of junctions
#*J*	Pressure at junction *i* within the original model
p_i	Pressure at coordinate x and y of p_i (pressure within original model) projected in the area of interest of the generated model. This projection is network layout unspecific and only depending on the real location within the area of interest.
$\text{proj}(p_i)$	Set of all junctions with a demand greater zero in the original water supply model

Eq.(1) (PI_1) is an indicator for the relative pressure difference between both models. Values smaller than one indicate the prediction of too low pressure values in the generated model. Values greater than one are indicating an over estimate of pressure. In contrast to that Eq.(2) (PI_2) indicates the mean absolute pressure difference.

$$PI\_1 = \frac{\sum_0^{i=\#J} \frac{\text{proj}(p_i)}{p_i}}{\#J} \text{ [-]} \tag{1}$$

$$PI\_2 = \sqrt{\frac{\sum_0^{i=\#J} (\text{proj}(p_i) - p_i)^2}{\#J}} \text{ [m]} \tag{2}$$

2.3. Case study – Innsbruck

The presented method is applied to the case study Innsbruck. It is a city in the Alps with 120 thousands inhabitants. For the area of interest a significant amount of data is available. In particular, a street network data set extracted from open street map, a digital elevation map in a resolution of 10x10 meter, the total demand Q of the water supply network with 0.7 m^3/s (detailed spatial resolution, uniform distributed and normal distributed over the area of interest) and the position of all reservoirs. These data sets (Initial data set) are used as input for the model generation algorithm, which generates an EPANET2 model. The pressure distribution of that model is shown in Fig. 2.

Fig. 2. Pressure distribution within the water supply network of the case study Innsbruck.

2.4. Generation and testing procedure

The generated models are compared with an already existing and well tested EPANET2 model of the case study containing detailed information of all components (e.g. pipes, demand with spatial resolution, etc.). The network generation algorithm is tested by appending additional pipe information (location, diameter, roughness) of the detailed EPANET2 model to the initial data set. In detail, five different sets of additional pipe information are appended to the initial data set, which are all pipes of the original EPANET2 model with a diameter greater or equal than 500mm, 250mm, 150mm or 0mm, respectively. Fig. 3. shows the fraction of the total network length by choosing all pipes greater or equal a certain diameter (DN). For example the total length of all pipes with a diameter greater or equal 250mm describes 30 percent of the overall system. In the generation procedure, the rest of the pipe network is created by means of the street network and graph analysis.

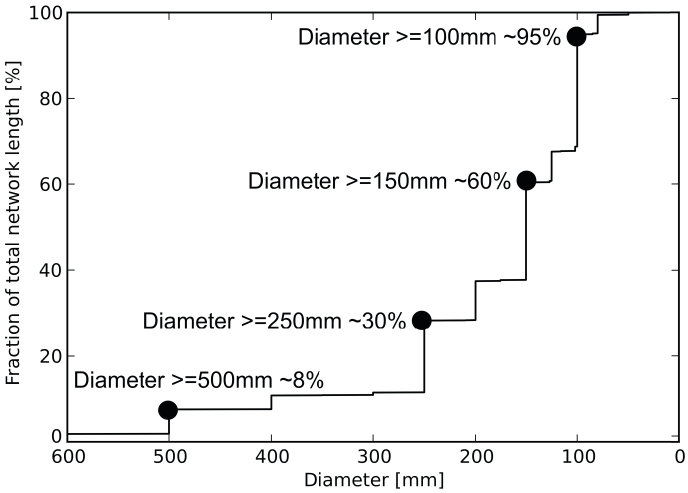

Fig. 3. Fraction of total network length of pipes equal or greater than a certain diameter.

3. Results

Fig. 4 shows the result for both performance indicators (PI_1 – left and PI_2 - right) by using the developed model generation algorithm. The x-axes for both plots are describing additional pipe information data sets which are appended to the initial data set (see also Fig. 4). The initial data set is defined by street network data set, DEM, Reservoirs/Tanks, Area of interest plus a variation of the demand data set: normal distributed, uniform distributed and exact demand.

By using the initial data set combined with additional pipe information with a diameter greater or equal to 500mm (eight percent of the total system) the generated water supply model is predicting too low pressure heads - between five and ten percent compared to the detailed model (Fig. 4. – PI_1 (>500)). This is an absolute pressure difference of between ten and seven meters (Fig. 4 – PI_2 (>500)). Compared to automatic generated models by using all pipes with a diameter greater or equal to 250mm (30 percent of total system) performance according to PI_1 is near to one, which is a pressure difference of between two and three meters according to PI_2 (Fig. 4 – PI_2 (>250)). By using all pipes of the original EPANET2 model as additional information nearly same result can be observed as compared to the pipe data set with diameters greater or equal to 250mm. The reason in this case for PI_1 not is being equal to exactly one and PI_2 not being equal to exactly zero is that the model generation algorithm uses a digital elevation map as input of a resolution of 10x10m for mapping the elevation to each junctions within the system. A fine grained DEM (e.g. 1x1 meter) would result in a lower PI_2 value.

Both diagrams are containing three different curves which is the variation of using exact distributed (original demand), uniform distributed and normal distributed demand within the initial data set. Both plots show that it is not necessary to know the exact position of all demand points within a system. For the scope and aim of this work models generated with uniform distribution of the total demand (0.7 m^3/s) within the area of interest have the same performance as knowing the real demand points of the system.

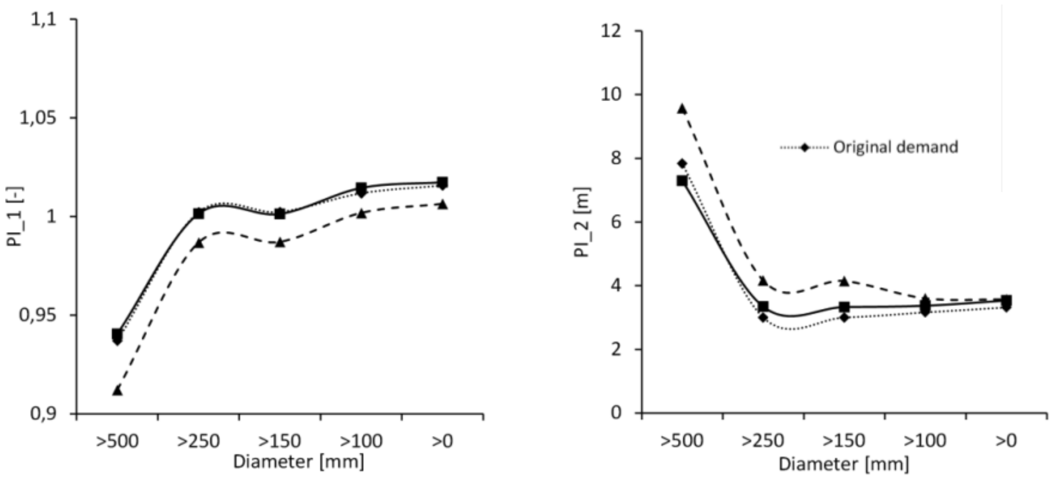

Fig. 4. Data amount for the case study Innsbruck.

4. Conclusion

In this manuscript the minimum data requirements are investigated which are required for developing a valid water supply model with the aim of generating the network layout, automatically determining pipe sizes and simulating the pressure within the system. The presented model generation algorithm uses easy accessible data sets (e.g. street network – Open Street network data) as input, combined with additional well known data of the real water supply network (e.g. data set of water supply pipes with a diameter greater or equal to a certain diameter). The simulation results of automatically generated water supply models are compared with simulation results of an exact model of the case study by using two different performance indicators. Both of them are indicating the mean pressure difference between the generated and detailed model, where the first one is a relative performance indicator and the second an absolute indicator. The whole benchmark setup was applied to the case study Innsbruck, a city within the Alps with 120 thousand inhabitants. Applying the model generation algorithm on a digital elevation map (resolution 10x10 m), reservoirs/tanks, street network, uniform distributing demand of 0.7 m^3/s and area of interest data sets plus all pipes with a diameter equal or greater than 250mm (30 percent of total length of all pipes within the real system) as additional information results in a relative performance value (PI_1) near to one and a mean absolute pressure difference of approximately three meters (PI_2).

Both performance indicators show only little difference if using exact spatial information of demand nodes as compared to uniform distributing the total demand within the area of interest for modeling the water supply system.

Acknowledgements

This work was funded by the Austrian Science Fund (**FWF**) in the project **DynaVIBe P23250-N24**. The authors gratefully acknowledge the financial support.

References

Blumensaat, F.; Wolfram, M. and Krebs, P. (2012): Sewer model development under minimum data requirements. In: Environmental Earth Sciences, Vol. 65(5), p. 1427-1437.

Chang, N.-B.; Qi, C. and Yang, Y.J. (2012): Optimal expansion of a drinking water infrastructure system with respect to carbon footprint, cost-effectiveness and water demand. In: Journal of Environmental Management, Vol. 110(0), p. 194-206.

Hellbach, C.; Moderl, M.; Sitzenfrei, R. and Rauch, W. (2011): Influence of Network Properties and Model Purpose on the Level of Skeletonization. In: Proceedings of the World Environmental & Water Resources Congress, Palm Springs, California, USA, May 22 - 26, 2011, Vol. 414, p. 10.

Kruskal, J.B. (1956): On the shortest spanning subtree of a graph and the traveling salesman problem. In: Proceedings of the American Mathematical society, Vol. 7(1), p. 48-50.

Mair, M.; Sitzenfrei, R.; Möderl, M. and Rauch, W. (2012): Identifying multi utility network similarities. In: Proceedings of the World Environmental and Water Resources Congress, Albuquerque, New Mexico, 20.05.2012-24.05.2012, p. 3147-3153.

Möderl, M.; Fetz, T. and Rauch, W. (2007): Stochastic approach for performance evaluation regarding water distribution systems. In: Water Science and Technology, Vol. 56(9), p. 29-36.

Möderl, M.; Sitzenfrei, R.; Fetz, T.; Fleischhacker, E. and Rauch, W. (2011): Systematic generation of virtual networks for water supply. In: Water Resources Research, Vol. 47.

Muranho, J.; Ferreira, A.; Sousa, J.; Gomes, A. and Marques, A.S. (2012): WaterNetGen: an EPANET extension for automatic water distribution network models generation and pipe sizing. In: Water Science & Technology: Water Supply, Vol. 12(1), p. 117-123.

Rossman, L.A. (2000): EPANET 2 user manual. National Risk Management Research Laboratory - U.S. Environmental Protection Agency.

Rossman, L.A. (2004): Storm water management model - User's manual Version 5.0. National Risk Management Research Laboratory - U.S. Environmental Protection Agency.

Sitzenfrei, R.; Fach, S.; Kleidorfer, M.; Urich, C. and Rauch, W. (2010a): Dynamic virtual infrastructure benchmarking: DynaVIBe. In: Water Science and Technology: Water Supply, Vol. 10(4), p. 600.

Sitzenfrei, R.; Moderl, M.; Mair, M. and Rauch, W. (2012): Modeling Dynamic Expansion of Water Distribution Systems for New Urban Developments. In: Proceedings of the World Environmental & Water Resources Congress, Albuquerque, New Mexico, USA, May 20 - 24, 2012.

Sitzenfrei, R.; Möderl, M. and Rauch, W. (2010b): Graph-based approach for generating virtual water distribution systems in the software ViBe. In: Water Science and Technology: Water Supply, Vol. 10(6), p. 923-932.

Sitzenfrei, R.; Möderl, M. and Rauch, W. (2013): Automatic generation of water distribution systems based on GIS data. In: Environmental Modelling & Software, Vol. 47, p. 138-147.

Torres, J. (2006): Micropolis: A Virtual City for Water Distribution System Research Applications. Texas A&M University, Zachry Department of Civil Engineering.

Trifunovic, N.; Maharhan, B. and Vairavamoorthy, K. (2012): Spatial network generation tool for water distribution network design and performance analysis. In: Water Science and Technology: Water Supply, Vol. 13(1).

Urich, C.; Burger, G.; Mair, M. and Rauch, W. (2012): DynaMind–A Softwaretool for Integrated Modelling of Urban Environments and their Infrastructure. In: Proceedings of the 10th International Conference on Hydroinformatics, Vol. 14, p. 18.

Yazdani, A.; Otoo, R.A. and Jeffrey, P. (2011): Resilience enhancing expansion strategies for water distribution systems: A network theory approach. In: Environmental Modelling & Software, Vol. 26(12), p. 1574-1582.