Multi-level automated sub-zoning of water distribution systems

Perelman, L.S. , Allen, M. , Preis, A. , Iqbal, M. and Whittle, A.J.

iEMSs 2014 Conference

Bold Visions for Environmental Modeling

Proceedings | Volume 3 | Pages 1214-1816

7th International Congress on Environmental Modeling and Software (iEMSs)

June 15–19, 2014

San Diego, California, USA

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

Multi-level automated sub-zoning of water distribution systems

Lina Sela Perelman [a], **Michael Allen** [b], **Ami Preis**[c], **Mudasser Iqbal** [c], **and Andrew J. Whittle**[a]

[a] *Department of Civil and Environmental Engineering, MIT, Cambridge, MA, USA, (linasela@mit.edu, ajwhittl@mit.edu)*

[b] *Faculty of Engineering and Computing, Coventry University, UK (michael.allen@coventry.ac.uk)*

[c] *Singapore-MIT Alliance for Research and Technology, Singapore (amipreis @smart.mit.edu, mudasser@smart.mit.edu)*

Abstract: Water distribution systems (WDS) are complex pipe networks with looped and branching topologies that often comprise of thousands of links and nodes. This work presents a generic framework for improved analysis and management of WDS by partitioning the system into smaller (almost) independent sub-systems with balanced loads and minimal number of interconnections. This paper compares the performance of three classes of unsupervised learning algorithms from graph theory for practical sub-zoning of WDS: (1) Graph clustering – a bottom-up algorithm for clustering n objects with respect to a similarity function, (2) Community structure – a bottom-up algorithm based on network modularity property, which is a measure of the quality of network partition to clusters versus randomly generated graph with respect to the same nodal degree, and (3) Graph partitioning – a flat partitioning algorithm for dividing a network with n nodes into k clusters, such that the total weight of edges crossing between clusters is minimized and the loads of all the clusters are balanced. The algorithms are adapted to WDS to provide a decision support tool for water utilities. The proposed methods are applied and results are demonstrated for a large-scale water distribution system serving heavily populated areas in Singapore.

Keywords: Water distribution systems; community structure; graph clustering and partitioning

1 INTRODUCTION

Network sub-zoning is one of the tools for leakage and pressure management for water loss control. The requirement of sub-zoning is to define the properties of the sub-zones within a network (e.g. size limit, total demand), to identify their boundaries (i.e. pipes or valves), and to monitor these boundaries for leakage and/or pressure control (with a limited number of meters). For example, the management of district metered areas (DMAs), has proven highly successful for leakage management [Thornton et al., 2008; Kunkel, 2003]. The layout of WDS is typically looped having multiple flow paths from the water sources to consumers. The looped layout of WDS, which provides a high level of reliability to the system supply in the event of mechanical failures (e.g. pipe breaks, valves malfunctions), imposes difficulties on water loss control. Due to the complexity of WDS, the re-design of an existing network can impair water supply, system reliability, and water quality [Grayman et al., 2009]. A number of methods for re-designing existing WDS into independent areas, by the closure of existing valves or disconnection of pipes, have been suggested. These vary from manual trial and error approaches, involving identification of water mains, manual division into districts, and hydraulic simulations [Murray et al., 2010], to highly sophisticated automated tools integrating network analysis, graph theory and optimization methods. The partition of the network is typically achieved by using graph algorithms, e.g. breadth first search and depth first search [Deuerlein, 2008; Perelman and Ostfeld, 2011; Ferrari et al., 2013; Di Nardo et al., 2013a], multilevel partitioning [Di Nardo et al., 2013b], community structure [Diao et al., 2013], and spectral approach [Herrera et al., 2010]. The selection of pipes that need to be disconnected is found by iterative procedures [Ferrari et al., 2013; Diao et al., 2013] or genetic algorithms [Di Nardo et al., 2013a, b].

This work presents a generic framework for simplifying the full-scale WDS by partitioning the system into smaller, balanced sub-zones with a minimum number of inter-connecting pipes/valves without the need to re-design the system. This study compares three types of unsupervised learning algorithms: clustering – representing a more naive approach given limited information about the WDS, community structure – adopted from social studies with similar previous application to WDS sub-zoning [Diao et al., 2013], and network partitioning – adopted from distributed computed and previous similar application [Di Nardo et al., 2013b]. The three methods were applied and tested on a large-scale water distribution system serving heavily populated areas in Singapore and their performance was compared based on different qualitative and quantitative measures.

2 METHODS

Many of the processes in physical, cyber, and social systems are described by complex networks or graphs. Clustering, community structure, and partitioning are closely related methods for understanding and analyzing complex systems, which have been extensively studied by a broad interdisciplinary research community over the past few years [Schaeffer, 2007; Fortunato, 2010]. Generally, given a data set, the goal of these methods is to divide the data set into clusters such that the elements assigned to a particular cluster are similar or connected in some predefined sense.

2.1 Graph clustering

Global clustering is one of the traditional algorithms for clustering n objects with respect to a similarity function. It produces a multi-level or an hierarchical structure of the graph, where each level of the clustering hierarchy defines a different subset and each top-level cluster is composed of sub-clusters. A bottom-up hierarchical algorithm starts with each node forming a unique cluster, followed by a sequential grouping of the two most similar clusters and computation of the centroid of the newly formed cluster. This procedure is repeated until all nodes are grouped into a single cluster. The basic similarity measure of nodes in a physical network is their geographical position. More details can be found in Hastie et al. [2009].

In water distribution systems, distant nodes are not expected to be connected, hence the Euclidean distance between a pair of nodes can be used as a measure of their similarity, i.e. similar nodes will be close to each other. In application to WDS, the number of clusters in which to group the nodes is not known priori, hence knowing the entire hierarchy of the network can be very informative. However, an additional procedure is required to decide how to partition the network. The attained hierarchical clustering of the graph is traversed in a top-down direction. The size (or load) of each top-level cluster is compared to a desired upper bound. If the size of the cluster does not satisfy the size constraint, the traverse continues to attain smaller sub-clusters. Additionally, since the Euclidean distance measure does not consider the connectivity of nodes, the intra and inter-connectivity of each cluster is verified. Finally, to satisfy the lower bound constraint on cluster size, small connected clusters are grouped together.

2.2 Community structure

Community structure is also a bottom-up hierarchical algorithm exploiting the network modularity property as the quality measure of the partition. Modularity, a very popular [Fortunato, 2010] measure of the quality of network partition into clusters, was first introduced by Newman [2004]. It is based on comparing the density of edges in the underlying sub-graphs to the density of edges in a random sub-graph with respect to the same nodal degree (i.e. number of incident edges). Since a random graph is not expected to have a cluster structure, a good community structure would have a higher modularity value. Modularity is always less than one (and can have negative value). Modularity can be computed according to:

$$Q(G,C) = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(c_i, c_j) \tag{1}$$

where m is the number of edges of the graph, A is the adjacency matrix, A_{ij} and P_{ij} is the actual and the expected number of links between nodes i and j, respectively, and $\delta(c_i = c_j) = 1, \delta(c_i \neq c_j) = 0$ indicating whether nodes i and j belong to the same cluster c (i.e. Kronecker delta). The expected number of edges in a random graph between nodes i and j with respect to the same node degrees, k_i and k_j, respectively, is $P_{ij} = k_i k_j / 2m$.

A greedy algorithm [Newman, 2004] for maximizing modularity involves successive merging of two clusters that result in the highest increase in modularity until all nodes are grouped into one cluster. The main steps of the algorithm can be found in Newman [2004]; Clauset et al. [2004]. As in graph clustering, community structure method results in a hierarchical clustering of the network. The exact partition of the graph is again selected by traversing the hierarchical structure from top to bottom and sequentially checking the upper bounds of the created clusters.

2.3 Graph partitioning

The problem of graph partitioning consists of dividing n nodes of the graph into a predefined number k of roughly equal sized clusters such that the number of edges connecting the clusters is minimal and typically it is desired that the cluster have equal size. Graph partitioning is a fundamental approach used in parallel computing, for allocating tasks to multiple processors so as to minimize the communications and equally distribute the computational burden among them. A multi-level graph partitioning approach generalized by Karypis and Kumar [1998] is used in the current work. The problem is solved by performing three main steps: (1) Coarsening – the original graph is reduced into a sequence of smaller graphs by aggregating its nodes and edges based on heavy edge matching. (2) Partitioning – a sequence of bisections of the network until a k-way partition of the graph is attained. (3) Recovering and refining – the original graph is recovered from the k-way partition. During each recovery level, a local refinement heuristics is used to improve the partition by iteratively swapping nodes between two clusters that reduce the weight of the cut edges. The main steps of the graph partition method can be found in more detail in Karypis and Kumar [1998]. The graph partitioning algorithm results in a single partition of the WDS with balanced sub-zones connected by a minimal number of links between the sub-zones. The implementation of the partitioning algorithm to WDS requires the definition of network graph, weights for nodes and links of the graph, and the number of desired sub-zones. The number of sub-zones can be inferred from the desired size of the sub-zones.

2.4 Quality measures

Several qualitative and quantitative measures exist to evaluate the quality of the clustering [Schaeffer, 2007]. The measures for evaluating the sub-zoning of WDS used herein are:

1. Adjacency matrix – visualization of the adjacency matrix is a graphical measure for evaluating the quality of the clustering. When the nodes of a graph are ordered randomly, there is no apparent structure in the adjacency matrix. Re-ordering of the nodes according to their clusters should reveal a tight block-diagonal structure of the adjacency matrix.
2. Cluster diagram – the layout of the clustering of the original graph can qualitatively assist in the evaluation of the clustering of the WDS.
3. Total cut-size – the total number of links connecting the different clusters implies the number of links that need to be monitored for water loss control i.e. this constraint defines the number (and cost) of sensors that need to be installed across the network. Naturally, this number grows with the number of desired sub-zones and should be minimized.

4. Worst cut-size – this measure amounts the total number of links that need to be monitored for a specific cluster. This number should also be minimized to limit the dependencies between the different zones of the WDS.

5. Cluster size – for better control of the WDS it is desired that the load of each sub-zone will be roughly equal. The load is ultimately specified by the water utility and can be measured in terms of the estimated demand, population served, and/or number of connections.

6. Recurrence of inter-cluster edges – this measure can be used to evaluate the suitability of the clustering for investment strategy, for example, a long-term flexible design versus here-and-now design.

3 Application

The three classes of graph clustering algorithms described above were applied to a real large scale network in Singapore. The network consists of 2440 nodes, 1932 pipes, 592 valves, one reservoir, one tank, six pumps, and serves the population of approximately 120,000 people. The required input is network topology, geographic coordinates, weights of nodes and links and sub-zones size constraints. The network was partitioned according to six demand loading constraints for each sub-zone, i.e. 20, 10, 8, 4, 3, 2% of the total daily demand of the network. The result of the clustering of the network are presented in a tabular and graphical schemes, providing statistics for each sub-zone, e.g., the number of nodes, number of intra and inter cluster edges, and the daily demand.

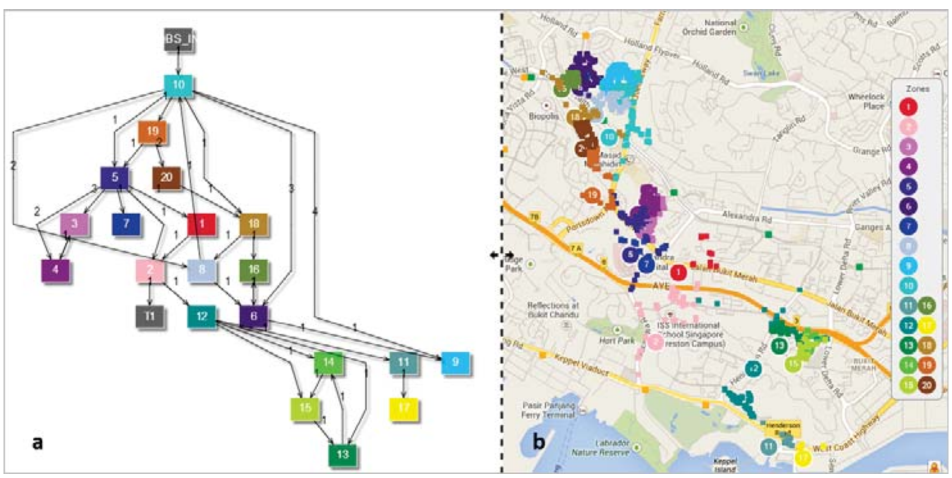

Figure 1. Network partitioning to 20 sub-zones: (a) Block diagram and (b) Map view

Next we compare and evaluate quality measures of the performance of the graph partitioning algorithms. Figure 1(a) demonstrates the structure of the network after division to 20 sub-zones, using the graph partitioning algorithm, and the connections between the different zones and the network sources. The number on the edges shows the number of inter-cluster connecting links and the direction is shown for a representative daily flow pattern of the distribution system. Figure 1(b) graphically shows map view of the network.

Figure 2 shows the adjacency matrices for the network – (a) original network and (b) – (d) network divided into 5, 10, and 25 sub-zones, respectively. The columns and rows of the matrix are reordered corresponding to the sub-zones represented by the blocks of the matrix. A clear cluster structure of the network can be observed from Figure 2(b)-(d) compared to the original structure Figure 2(a).

Figure 3 demonstrates the performance of the three algorithms: graph clustering (blue), community structure (red), and graph partitioning (black), for six different divisions based on four suggested quan-

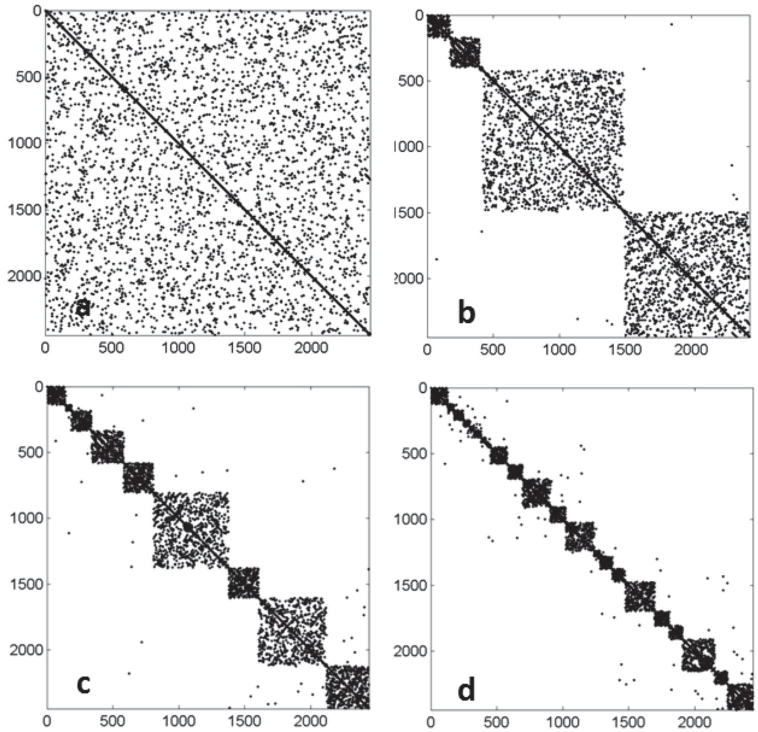

Figure 2. Connectivity matrix: (a) original network and (b)-(d) network divided into 5, 10, and 25 sub-zones, respectively.

titative measures. From the results it can be seen, that, as expected, the total number of inter-cluster connecting links grows with the number of sub-zones (Figure 3(a)). The maximum number of inter-cluster connecting links for a single sub-zone varies around 11, 9, and 8, for the graph clustering, community structure, and graph partitioning methods, respectively (Figure 3(b)). As the number of sub-zones grows, the demand load of each sub-zone decreases (Figure 3(c)). Figure 3(d) shows the fraction of inter-cluster edges that appear more than once during different sub-zoning levels. For example, approximately 16 % of all inter-cluster edges appeared more than once in divisions to 10, 15, 25, and 35 sub-zones based on clustering and community structure approches. The recurrence of inter-cluster connecting edges is similar and higher for the hierarchical methods, i.e. clustering and community structure, compared to the flat partitioning approach. This behavior remains similar for all the partitions.

Figure 4 compares the total cut-size and worst case cut-size for graph partitioning algorithm when pipes and valves are treated: (1) similarly (black), i.e. pipes and valves are allowed on boundary edges of the sub-zones, which is represented by uniform weights of network links, and (2) differently (blue), i.e. only valves are allowed on boundary edges of the sub-zones, which is represented by a adjusted weights on network links. It can be seen, that when considering the actual location of valves, the number of inter-connecting edges is slightly higher, as valves are not installed on every pipe.

4 Conclusions

The partition of water distribution systems into sub-zones is an important tool for leakage and pressure management and for water loss control. This work explores the application of graph-theory approach to the WDS sub-zoning problem. Three classes of algorithms were explored in this work – graph clustering, community structure, and graph partitioning. It was shown that the methods are compatible and applicable for large-scale WDS. The community structure and graph partitioning methods were shown to be more flexible than the graph clustering method, in terms of adaptivity to design constraint by incorporating connectivity of the network and associated weights. The suggested methods can provide a

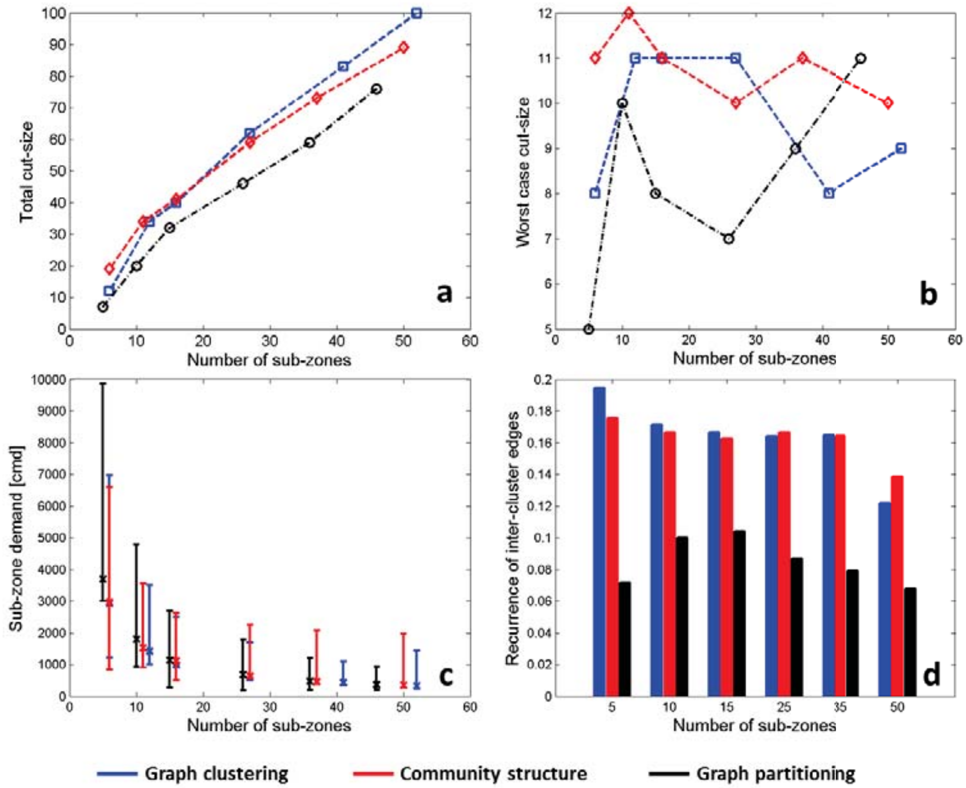

Figure 3. Quality measures: (a) Total cut-size, (b) Worst case cut-size, (c) Sub-zone demand, and (4) Recurrence of inter-cluster edges. Graph clustering, community structure, and graph partitioning methods are represented by blue, red, and black colors, respectively

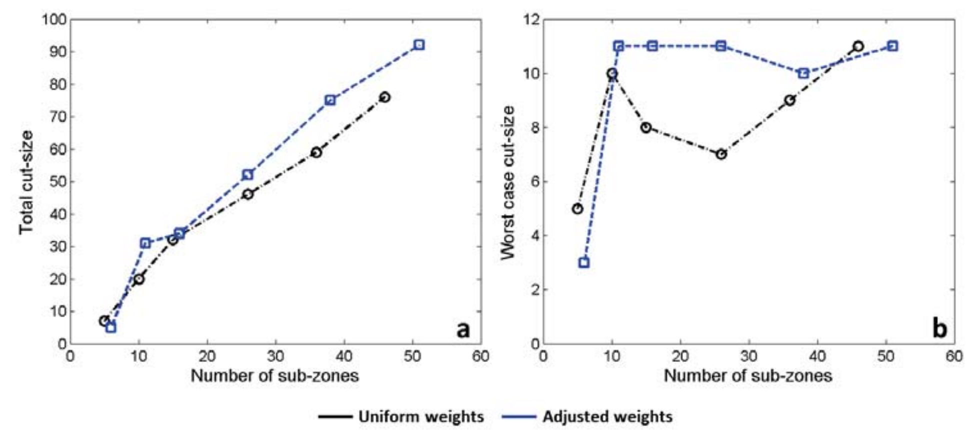

Figure 4. Quality measures for uniform (black) and adjusted (blue) weights of network links: (a) Total cut-size and (b) Worst case cut-size

decision support tool to water utilities for network sectorization and the sub-zone design will depend on investment strategies for monitoring and controlling the WDS.

Acknowledgments

The lead Author is grateful for fellowship support from the MIT-Technion program. The contributing Authors were funded by the National Research Council through the Singapore-MIT Alliance for Research and Technology (SMART) and carried out through a collaborative agreement between the Center for Environmental Sensing and Modeling (CENSAM) and the Public Utilities Board (PUB).

References

Clauset, A., Newman, M. E. J., and Moore, C. (2004). Finding community structure in very large networks. *Phys. Rev. E*, 70(6):066111.

Deuerlein, J. W. (2008). Decomposition model of a general water supply network graph. *J. Hydraul. Eng.*, 134(6):822–832.

Di Nardo, A., Di Natale, M., Santonastaso, G., Tzatchkov, V., and Alcocer-Yamanaka, V. (2013a). Water network sectorization based on graph theory and energy performance indices. *J. Water Resour. Plann. Manage.*, 140(5):620–629.

Di Nardo, A., Di Natale, M., Santonastaso, G. F., and Venticinque, S. (2013b). An automated tool for smart water network partitioning. *Water Resour. Manage.*, 27(13):4493–4508.

Diao, K., Zhou, Y., and Rauch, W. (2013). Automated creation of district metered area boundaries in water distribution systems. *J. Water Resour. Plann. Manage.*, 139(2):184–190.

Ferrari, G., Savic, D., and Becciu, G. (2013). A graph theoretic approach and sound engineering principles for design of district metered areas. *J. Water Resour. Plann. Manage.*, (Published ahead of print).

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, (486):75–174.

Grayman, W. M., Murray, R., and Savic, D. A. (2009). Effects of redesign of water systems for security and water quality factors. In *Proc. World Environ. and Water Resourc. Cong.*, pages 504–514. ASCE, Reston, VA.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*. Springer-Verlag.

Herrera, M., Canu, S., Karatzoglou, A., Perez-Garca, R., and Izquierdo, J. (2010). An approach to water supply clusters by semi-supervised learning. In *Proc., Int. Congress on Environmental Modelling and Software*. IEMSs.

Karypis, G. and Kumar, V. (1998). Multilevel k-way partitioning scheme for irregular graphs. *J. Parall. Distr. Comp.*, 48:96–129.

Kunkel, G. (2003). Committee report: Applying worldwide bmps in water loss control. *J. Am. Water Works Assoc.*, 95(8):65–79.

Murray, R., Grayman, W., Savic, D., and Farmani, R. (2010). *Effects of DMA redesign on water distribution system performance*. Boxall and Maksimovi'c (eds), Taylor and Francis Group, London.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69(6):066133.

Perelman, L. and Ostfeld, A. (2011). Topological clustering for water distribution systems analysis. *Environ. Modell. Software*, 26(7):969 – 972.

Schaeffer, S. E. (2007). Graph clustering. *Computer Science Review*, 1(1):27 – 64.

Thornton, J., Sturm, R., and Kunkel, G. (2008). *Water Loss Control*. McGraw-Hill Companies, Inc.

7th International Congress on Environmental Modelling and Software (iEMSs)
June 15–19, 2014
San Diego, California, USA