In recent years, epidemiologic studies and vulnerability assessments have come to rely on water distribution system models for refined estimates of street-level exposure scenarios in conducting health studies and assessing potential contamination occurrences. A water distribution system network analysis at the street level can result in thousands to tens of thousands of hydraulic appurtenances such as pipelines, hydrants, and valves, making model calibration, verification, and system operation simulations time-consuming and complex. Methods of aggregation, skeletonization, and network simplification have been used to reduce the complexity of the networks being analyzed. The issue then is whether network aggregation based on hydraulic characteristics and controls conserves mass so that the simplified network provides reliable results for the water quality analyses needed for epidemiologic studies and contamination assessment. In this study, a network aggregation methodology based on both hydraulic and water quality aggregation of an all-pipes network was applied to a complex all-pipes network described in the literature and used for an epidemiologic study. Simulation results obtained using the quality aggregation method, closely resembled both hydraulic (i.e., pressures) and water quality (i.e., concentrations) behavior of the original full-sized (all-pipes) system, while reducing system size by almost half.

BY LINA PERELMAN, MORRIS L. MASLIA, AVI OSTFELD, AND JASON B. SAUTNER

Using aggregation/skeletonization network models for water quality simulations in epidemiologic studies

Urban and municipal water distribution systems can consist of thousands to tens of thousands of hydraulic appurtenances such as pipelines, valves, tanks, hydrants, and pumps. With the capacity and capability of today's computers and database management software (such as geographic information systems), it is possible to develop network representations of all of the hydraulic appurtenances and characteristics of a water distribution system and represent them using an all-pipes model. However, these all-inclusive models have the disadvantage of requiring significantly greater effort for calibration, verification, and postsimulation analysis. In addition, planning and conducting field tests to gather systemwide data to characterize the network by necessity require an understanding of a less complex system that generally responds in the same manner as the more complicated all-pipes network.

Historically, reduced-pipe, aggregated, or skeletonized network models were relied on as a means of characterizing and understanding the hydraulics of a water distribution system when computer speed and storage requirements were limited. Because desktop computers now rival or exceed the computational capacity of their predecessor mainframe equivalents, it would seem that analysts would tend to conduct all-pipes model network analyses on a routine basis. Even with today's high-speed and high-capacity computers, however, some degree of network simplification or aggregation may be necessary to obtain results and assessment estimates within reasonable time frames and limited budgets.

2008 © American Water Works Association

With the recent focus on intentional and nonintentional contamination (Guan et al, 2006) and optimal sensor placement (Ostfeld & Solomons, 2004), skeletonization issues are under scrutiny because of the potential need for rapid, near real-time assessment capabilities. Another consideration that has arisen is the use of water distribution system modeling to assist with historical reconstruction of contamination occurrences for use in epidemiological health studies and exposure assessments (Maslia et al, 2001; 2000). In these analyses, decades-long simulation times may need to be undertaken so that historically synthesized information can be useful to epidemiologists. Again the question arises, can network skeletonization methods be effectively developed and properly applied so that robust and consistent solutions are derived for both hydraulic and water quality analyses?

To address issues associated with network skeletonization, the authors developed an aggregation method based on both hydraulic and water quality considerations. The methodology was then applied to a complex all-pipes network previously described in the literature and used for an epidemiologic study (Maslia et al, 2000), and results for the aggregation and two skeletonization methods were compared with results for the all-pipes network.

This elevated storage tank is part of the Dover Township (N.J.) water distribution system. An "all-pipes" representation of the system was described in an epidemiological study and was used in this article to provide comparisons with the aggregated and skeletonized models.

LITERATURE REVIEW

The need for network simplification through skeletonization and aggregation dates back to the days of manual methods such as the Hardy Cross procedure for balancing network flows (Cross, 1936). Because of the extensive hand calculations that were required, the method could realistically handle only a limited number of pipelines. With the advent and availability of high-speed computers with large memory capacity, data-handling procedures and computer programs were developed for representing large distribution system networks, as described elsewhere (Rossman, 1994; Sarikelle & Chuang, 1989; Grayman & Males, 1988; Grayman et al, 1988). With these tools available and even faster computers being developed, complete network characterization using all-pipes water distribution system models might be expected to become routine practice. However, depending on the problem being analyzed, the availability of field data to properly characterize and describe network features, and time and budget constraints, system complexity simplification through the application of aggregation techniques is a viable and possibly necessary approach for analysis of water distribution systems.

Historically, skeletonization techniques have been developed based on hydraulic criteria that required the skeletonized or simplified system to resemble the configuration or connectivity of the original all-pipes network and perform similarly in terms of hydraulic variables, i.e., pressures, water levels, pump operations, and demands (Ulanicki et al, 1996; Anderson & Al-Jamal, 1995; Hamburg & Shamir, 1988). In fact, many of today's distribution system modeling and analysis packages contain some type of skeletonization module or routine for rapid system simplification. Although published literature describing the effects of network simplification on modeling and water distribution system analysis accuracy has been limited, some quantitative analyses are available (Grayman & Rhee, 2004; Grayman et al, 1991; Characklis, 1988; Eggener & Polkowski, 1976). The results of these quantitative analyses appear to indicate that water quality analyses can be more sensitive to network skeletonization than standard hydraulic analyses. Thus the ability to compute accurate water quality concentrations in a distribution system under unsteady hydraulic and water quality conditions is of paramount importance, given the current objectives of many modeling analyses.

METHOD OVERVIEW AND DEVELOPMENT

Water quality aggregation. The method proposed here is capable of aggregating the all-pipes model to a skeletonized model containing fewer nodes and links, yet preserves the hydraulic and water quality behavior of the full system. The aggregation algorithm allows the user to select a set of nodes (e.g., monitoring station locations) within the system to remain in the aggregated model, which calculates a new topology and component characteristics of the skeletonized system. The final result

2008 © American Water Works Association

of the aggregation model is a simplified system that preserves pressures and concentrations of the full network at the skeletonized system.

The aggregation method incorporates the EPANET hydraulic and water quality simulation model (Rossman, 2000; 1994) to perform extended period simulation of hydraulic and water quality behavior within pressurized pipe networks. The changes in constituent concentration along pipes are predicted using a one-dimensional advective–reactive transport equation. The constituent mixing at nodes is assumed to be complete and instantaneous. The concentration leaving the node is computed as a flow-weighted average of the incoming pipe concentrations. Water is assumed to be completely mixed in tanks and reservoirs. Details of the governing equations and solution algorithms that form the basis of the EPANET simulation platform are described elsewhere (Rossman, 2000; 1994).

The water quality aggregation algorithm extends the hydraulic aggregation algorithm (Ulanicki et al, 1996) and combines two different search algorithms. The main stages of the algorithm are shown in Figure 1 and are briefly described in subsequent sections of this article. A full description of the algorithm can be found in Perelman and Ostfeld (2006).

Directed graph. A water distribution system can be modeled as a graph, *G*, with a set of nodes, *V*, representing the sources and consumer nodes and a set of edges, *E*, representing the connecting pipes, pumping units, and valves. The first step of the algorithm is to represent the water distribution system as a directed graph. The graph representing the distribution system is formulated on the basis of the topological structure of the system, and the directions of the edges are defined on the basis of hydraulic simulation results under operational pattern history. In the directed graph, there are two types of edges in which flow can and cannot reverse.

Strongly connected components. The next step is to reduce the original directed graph to its strongly connected components. A strongly connected component is a set of nodes such that any two nodes are connected by at least one directed path. This implies that the concentration at a node depends on the concentration of any other node within its strongly connected component. The directed graph (i.e., the outcome of the first step) is reduced to its strongly connected components using the depth-first search (DFS) algorithm (Tarjan, 1972).

All-inclusive models have the disadvantage of requiring significantly greater effort for calibration, verification, and postsimulation analysis.

The DFS is an algorithm for searching a graph. Initially all the nodes of the graph are unexplored. The search begins at a node of the graph, moving in the direction of an outgoing edge from that node and discovering a new unvisited node. The search explores as far as possible along the edges before backtracking to the previously reached node that still has unexplored edges. The search terminates when all the nodes of the graph have been visited. The DFS algorithm is highly efficient, with a linear running time of the number of nodes and links of the system.

Directed acyclic metagraph. After all strongly connected components of the directed graph have been identified, all nodes of a strongly connected component are joined together into a single metanode, yielding a directed acyclic graph (DAG), i.e., a graph without cycles. The resulting network consists only of links in which flow does not reverse, i.e., the network does not contain any circular paths, and its nodes comprise the strongly connected components. The DAG represents hierarchical relationships among all the nodes of the system. In other

TABLE 1 Summary of water distribution system networks used in the application examples

Network Identification*	Number of Nodes	Number of Links	Total Pipe Length—*mi*	Reduction in Nodes From All-pipes Network—%	Reduction in Links From All-pipes Network—%	Reduction in Pipe Length From All-pipes Network—%
All-pipes	14,945	16,056	484.6	NA	NA	NA
Aggregated	8,173	8,995	289.1	45.3	44.0	40.3
Skeletonized A	7,032	8,064	399.7	46.0	56.2	17.5
Skeletonized B	3,076	3,984	383.4	73.3	80.8	20.8

NA—not applicable

*The source for the all-pipes network was Maslia et al (2001; 2000). The aggregated network was derived from the quality aggregation method proposed by Perelman and Ostfeld (2006). The skeletonized networks A and B were derived from skeletonization software described in MWH Soft Inc. (2003).

2008 © American Water Works Association

words, a constituent measured at critical nodes (e.g., monitoring stations) has to pass through all its upstream nodes. The designation of a critical node is explained in detail and an illustrative example is provided in Perelman and Ostfeld (2006).

Remaining nodes. In the next step, the metagraph is topologically sorted and all upstream nodes that can contribute to the concentration of some critical nodes are identified. This is achieved by reversing the directions of all links in the metagraph and searching the reverse DAG using the breadth-first search (BFS) algorithm.

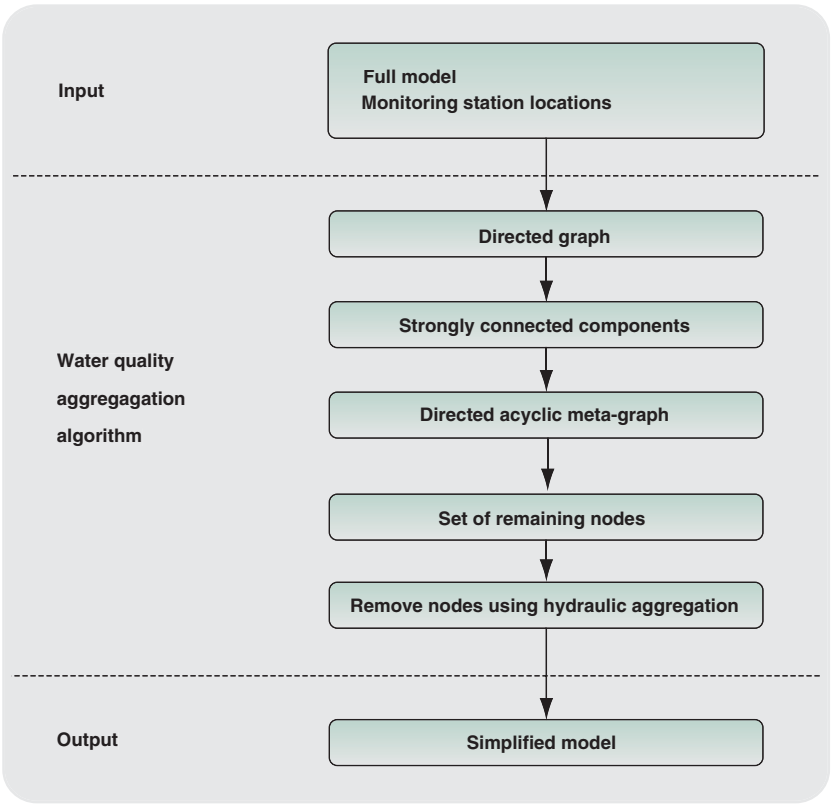
Like the DFS, the BFS is an algorithm for searching a graph. The search starts at a root node (i.e., critical node) and visits all the nodes that are adjacent to the current node before moving to the nodes that are farther away. The search is not restarted in other connected components of the graph; nodes not reachable from the root node are ignored. The overall running time of BFS is again linear in the size of its input, i.e., number of nodes and links of the system. The nodes that are not identified by the BFS do not have direct influence on the concentration of the critical nodes (e.g., nodes with monitoring stations) and can be removed from the network.

Hydraulic aggregation. The network is skeletonized by eliminating network nodes and links using the hydraulic aggregation algorithm (Ulanicki et al, 1996). The method used by Ulanicki and colleagues is based on the reduction of the linearized model of the full nonlinear system by eliminating some of the variables using Gauss elimination (Hammerlin & Hoffman, 1991). The process involves the following four stages.

- Create a full nonlinear model by formulating mass conservation equations for all nodes of the network (Kirchoff's law 1) and energy balance equations for all fundamental loops of the network (Kirchoff's law 2).
- Linearize the established model. A linearized approximation of the nonlinear system describes the relationship between small changes in nodal heads and demands around an operating point.
- Reduce the linear model using the Gauss elimination procedure. For each removed node, one step of Gauss elimination is executed. The original demand assigned to the removed node is redistributed among the connected nodes, the connecting links are removed, and new characteristics are calculated for the remaining elements of the system.
- Recover the skeletonized nonlinear model from the skeletonized linear model.

Aggregation and skeletonization terminology. Two example problems are used to demonstrate the proposed aggregation methodology and compare results with alternative network skeletonization techniques. For the purposes of these example problems, the following terminology is applied to the aggregated or skeletonized water distribution system networks. The all-pipes network is the water distribution system shown in Figure 2. This system, detailed in other research (Maslia et al, 2001; 2000), is characterized by 16,056 links and 14,945 nodes (Table 1). The network was used as part of an epidemiological case-control study of the Dover Township (N.J.) water distribution system (NJDHSS, 2003). The aggregated system network is the skeletonized water distribution system derived from the application of the quality aggregation method and is characterized by 8,995 links and 8,173 nodes. The skeletonized system A network is the water distribution system network derived by applying a commercially available hydraulic skeletonizer[1] to the all-pipes

FIGURE 1 Main stages of the aggregation model



2008 © American Water Works Association

network and is characterized by 8,064 links and 7,032 nodes. The skeletonized system B network represents a further simplification of skeletonized system A and is characterized by 3,984 links and 3,076 nodes. Table 1 lists the number of nodes and links for each of these networks and the percent reduction of nodes, links, and total pipe length, compared with the all-pipes network.

Comparison of the all-pipes and aggregated networks. The water distribution system serving the Dover Township area has been the subject of publicly documented health studies and analyses (Maslia et al, 2001; 2000). The all-pipes network described in the aforementioned references was selected to test the hydraulic and water quality aggregation methods previously described. The EPANET model (Rossman, 2000; 1994) describing the water distribution system (Figure 2) is characterized by 16,056 links, 14,945 junctions, 8 ground-level and elevated storage tanks, 12 high-lift or booster pumps, and 20 water supply wells. The skeletonized network was constructed using the quality aggregation method. The resulting aggregated system contains 8,173 nodes and 8,995 links, which represents approximately a 45% reduction in the number of system nodes and links (Table 1).

Two hydraulic and water quality simulations were conducted to assess the accuracy of the aggregated network. The first scenario simulated intentional contaminant intrusion, and the second simulated the transport of barium—a naturally occurring, conservative groundwater constituent—within the water distribution system.

Scenario 1. Scenario 1 was characterized by a 6-h intrusion of a hypothetical, conservative contaminant with a 100-mg/L concentration at the Parkway (N.J.) well field (Figure 2). The source concentration is described by the pattern factor shown in Figure 3. Simulated pressure and concentration results using the all-pipes and aggregated network models were compared at three hypothetical monitoring station locations in the network—junctions 4564, 14125, and 14429 (Figure 2). These results are shown graphically in Figure 4, part A. A comparison of results for pressures and concentrations of the aggregated model network found good agreement between the all-pipes and the aggregated network results throughout the duration of the 24-h simulation. For example, the root mean square (RMS) of concentration differences between the all-pipes and aggregated networks over a 24-h simulation for junctions 4564, 14125, and 14429 were 9.48, 1.43, and 0.51 mg/L, respectively (Table 2 and Figure 4, part A). The RMS of concentration difference was defined according to Eq 1:

$$\text{RMS} = \sqrt{\frac{\sum_{i=1}^{24} (C_{AP,i} - C_{EN,i})^2}{24}} \quad (1)$$

in which $C_{AP,i}$ is the simulated concentration for all-pipes network for hour i and $C_{EN,i}$ is the simulated concentration for the example network (aggregated, skeletonized A, or skeletonized B) for hour i.

Scenario 2. For scenario 2, a water quality simulation was conducted using barium as a tracer. Barium is a naturally occurring, conservative constituent in the Dover Township source water (groundwater). Details of this simulation are available elsewhere (Maslia et al, 2000). The underlying boundary conditions of the all-pipes network were applied to the aggregated model network described previously.

Results of the scenario 2 simulation, shown graphically in Figure 4, part B, compared simulated pressures and concentrations at three selected monitoring locations—nodes 4564, 5134, and 14125. Measured barium concentrations, obtained from a one-time grab sample taken at 8:00 a.m., are also shown on the concentration graphs (Figure 4, part B). Comparison of results for pressures and concentrations of the aggregated model network indicated good agreement between the all-pipes and the aggregated network results throughout the duration of the 24-h simulation. The RMS of concentration differences between the all-pipes and aggregated networks over a 24-h simulation for junctions 4564, 5134, and

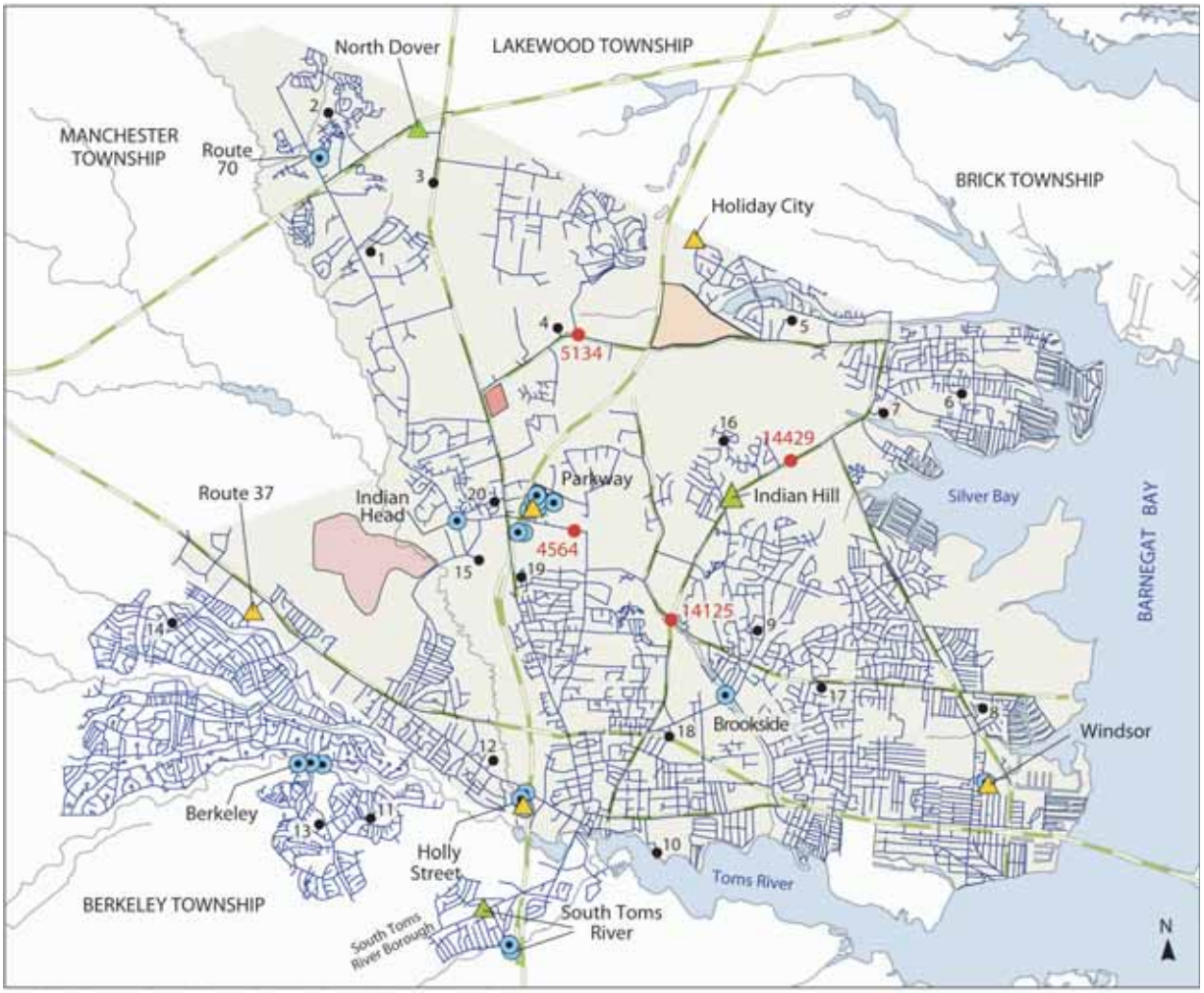
TABLE 2 Comparison of simulated concentration differences for water distribution system networks used in application examples

Example Network	RMS of Concentration Difference Between All-pipes Network and Example Networks for 24-h Simulation—*mg/L*					
	Scenario 1			Scenario 2		
	Node 4564	Node 14125	Node 14429	Node 4564	Node 5134	Node 14125
Aggregated	9.48	1.43	0.51	1.13	2.70	0.98
Skeletonized A	17.51	13.83	12.24	6.94	18.57	3.73
Skeletonized B	18.87	20.01	11.97	5.82	18.18	4.33

RMS—root mean square

2008 © American Water Works Association

FIGURE 2 All-pipes network representation of the 1996 water distribution system serving the Dover Township (N.J.) area



Map Point	Node ID	Map Point	Node ID	Map Point	Node ID	Map Point	Node ID
1	4319	6	9617	11	1829	16	14271
2	5450	7	14749	12	2601	17	13685
3	4409	8	7854	13	1326	18	5866
4	5130	9	8442	14	3581	19	4475
5	8858	10	5621	15	12737	20	12788

Adapted from Maslia et al, 2000

2008 © American Water Works Association

14125 were 1.13, 2.70, and 0.98 mg/L, respectively (Table 2 and Figure 4, part B).

Comparison of all-pipes and skeletonized networks. Network skeletonization techniques are based on hydraulic criteria that require the skeletonized or simplified system to resemble the configuration or connectivity of the original all-pipes network and to have similar performance in terms of hydraulic variables—pressures, water levels, pump operations, and demands. For the current analysis, a commercially available network skeletonizer[1] was applied to the all-pipes water distribution system network shown in Figure 2. The resulting skeletonized networks, referred to as skeletonized system A and skeletonized system B, were used to conduct additional simulations for scenarios 1 and 2. Configurations of skeletonized systems A and B are listed in Table 1.

Results reported for scenario 1 were for EPANET model nodes 4564, 14125, and 14429 (Figure 4, part A). Results reported for scenario 2 were for model nodes 4564, 5134, and 14125 (Figure 4, part B). Model node locations for which results are discussed are shown in Figure 2. In both scenarios, the results of EPANET simulation runs were taken as reference parameters for the hydraulic and quality performance of the system, and the all-pipes network, aggregated system, and skeletonized systems A and B were compared on the basis of these results.

Scenario 1. As described previously, scenario 1 was characterized by a 6-h intrusion of a contaminant with a 100-mg/L concentration at the Parkway well field (Figure 2). Pressure and concentration results for scenario 1 based on EPANET simulations using skeletonized system A and skeletonized system B networks are shown in Figure 4, part A, at three monitoring station locations in the network—nodes 4564, 14125, and 14429.

At node 4564, simulations based on networks of skeletonized systems A and B completely missed the second concentration peak that occurred at simulation hour 14, even though comparisons of simulated pressures between the all-pipes and skeletonized networks were equivalent. By comparison, the aggregated system network duplicated the second concentration peak and was nearly equivalent to the simulated concentration peak of the all-pipes network.

At node 14125, the concentration fronts simulated by skeletonized systems A and B arrived 1–2 h earlier than the front simulated using the all-pipes and the aggregated system networks. Additionally, the maximum concentrations simulated by skeletonized systems A and B were about 40% greater than the maximum concentration simulated using the all-pipes and aggregated system networks.

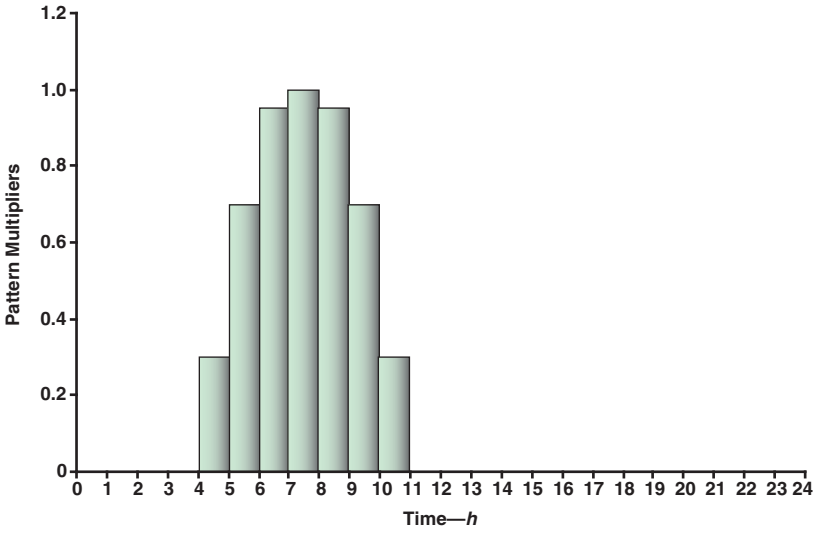
At node 14429, the simulated concentration front from skeletonized systems A and B arrived 1 h earlier than the simulated concentration front derived from the all-pipes and the aggregated system networks. Similar to the results for node 4564, simulated concentrations for skeletonized system A and B poorly reproduced the second concentration peak (occurring during simulation hours 21–23), compared with the all-pipes and aggregated system networks.

Comparison of the RMS of concentration differences listed in Table 2 demonstrated that skeletonized networks A and B produced much larger RMS of concentration differences than did the aggregated network. For example, for nodes 4564, 14125, and 14129, skeletonized network A

TABLE 3 Approximate computational times resulting from a 1,008-h simulation

Network	Scenario 1—s	Scenario 2—s
All-pipes	165	200
Aggregated	76	80
Skeletonized A	70	73
Skeletonized B	34	37

FIGURE 3 EPANET pattern factors assigned to simulation scenario 1



2008 © American Water Works Association

had RMS of concentration differences of 17.51, 13.83, and 12.24 mg/L, respectively, indicating much poorer agreement with the all-pipes network, compared with the aggregated network. Similarly, large RMS of concentration differences were observed between skeletonized network B and the all-pipes network (Table 2).

Scenario 2. In scenario 2, a water quality simulation was conducted to simulate the transport of a naturally occurring conservative constituent, barium. Figure 2 shows the locations of three selected monitoring stations (nodes 4564, 5134, and 14125). As with scenario 1 results, simulated pressures for all skeletonized networks (aggregated, skeletonized system A, and skeletonized system B) were nearly equivalent to pressures simulated by the all-pipes network during the 24-h simulation (Figure 4, part B). At node 4564, the aggregated network concentrations were nearly identical to concentrations simulated by the all-pipes network. Compared with the all-

TABLE 4 Comparison of simulated concentrations of barium at dead-end nodes removed by aggregation and skeletonization process

	All-pipes			Aggregated				Skeletonized A				Skeletonized B			
Map Point Number	Node ID*	Concentration *mg/L*	Pipe Diameter *in.*	Node ID*	Concentration *mg/L*	Pipe Diameter *in.*	Absolute Difference *%*	Node ID*	Concentration *mg/L*	Pipe Diameter *in.*	Absolute Difference *%*	Node ID*	Concentration *mg/L*	Pipe Diameter *in.*	Absolute Difference *%*
1	4319	34.7	4	4323	34.7	8	0.06	4323	1.4	8	96.02	4323	0.0	8	99.88
2	5450	34.7	4	5430	34.7	8	0.00	5431	0.0	8	100.00	5440	0.0	8	100.00
3	4409	34.7	16	4412	34.7	12-16	0.00	4410	0.0	16	100.00	4410	0.0	16	100.00
4	5130	45.1	8	5129	45.0	12	0.04	5129	41.1	12	8.88	5050	40.9	12	9.25
5	8858	42.9	6	8801	43.4	12	0.95	8801	34.4	12	19.98	8796	34.5	12	19.63
6	9617	42.1	2	9572	42.4	8	0.78	9572	38.0	8	9.86	9572	37.7	8	10.48
7	14749	42.1	8	14747	42.2	12	0.12	14746	39.9	12	5.25	9396	39.6	12	6.08
8	7854	39.7	6	7844	40.0	12	0.55	7836	40.0	12	0.68	7836	40.1	12	0.86
9	8442	41.4	2	8431	42.0	6	1.47	8431	42.4	6	2.56	8431	43.4	6	4.78
10	5621	23.7	2	2992	23.6	10	0.59	2992	22.9	10	3.45	2992	24.4	10	2.57
11	1829	12.6	2	2396	12.6	8	0.00	2396	12.6	8	0.00	2396	12.6	8	0.00
12	2601	41.5	8	2599	37.4	12	9.86	2599	38.3	12	7.64	2597	38.7	12	6.72
13	1326	22.7	2	1351	22.7	16	0.00	1327	22.7	12	0.00	1327	22.7	12	0.00
14	3581	22.7	6	822	22.7	12	0.00	822	22.7	12	0.00	822	22.7	12	0.00
15	12737	25.9	8	3801	27.4	12	5.76	3801	28.1	12	8.66	3801	28.0	12	8.08
16	14271	41.8	6	8667	42.0	12	0.53	8420	41.7	12	0.24	8417	41.5	12	0.81
17	13685	42.0	8	6691	41.9	12	0.36	6691	41.2	12	2.00	6691	42.1	12	0.02
18	5866	38.1	6	5852	37.6	6	1.13	5852	36.4	6	4.44	5852	37.3	6	2.10
19	4475	40.4	4	4467	39.9	8	1.41	4467	41.6	8	2.89	4462	41.6	8	2.92
20	12788	42.4	2	4551	42.3	16	0.42	4554	41.5	16	2.21	4554	40.7	16	4.15

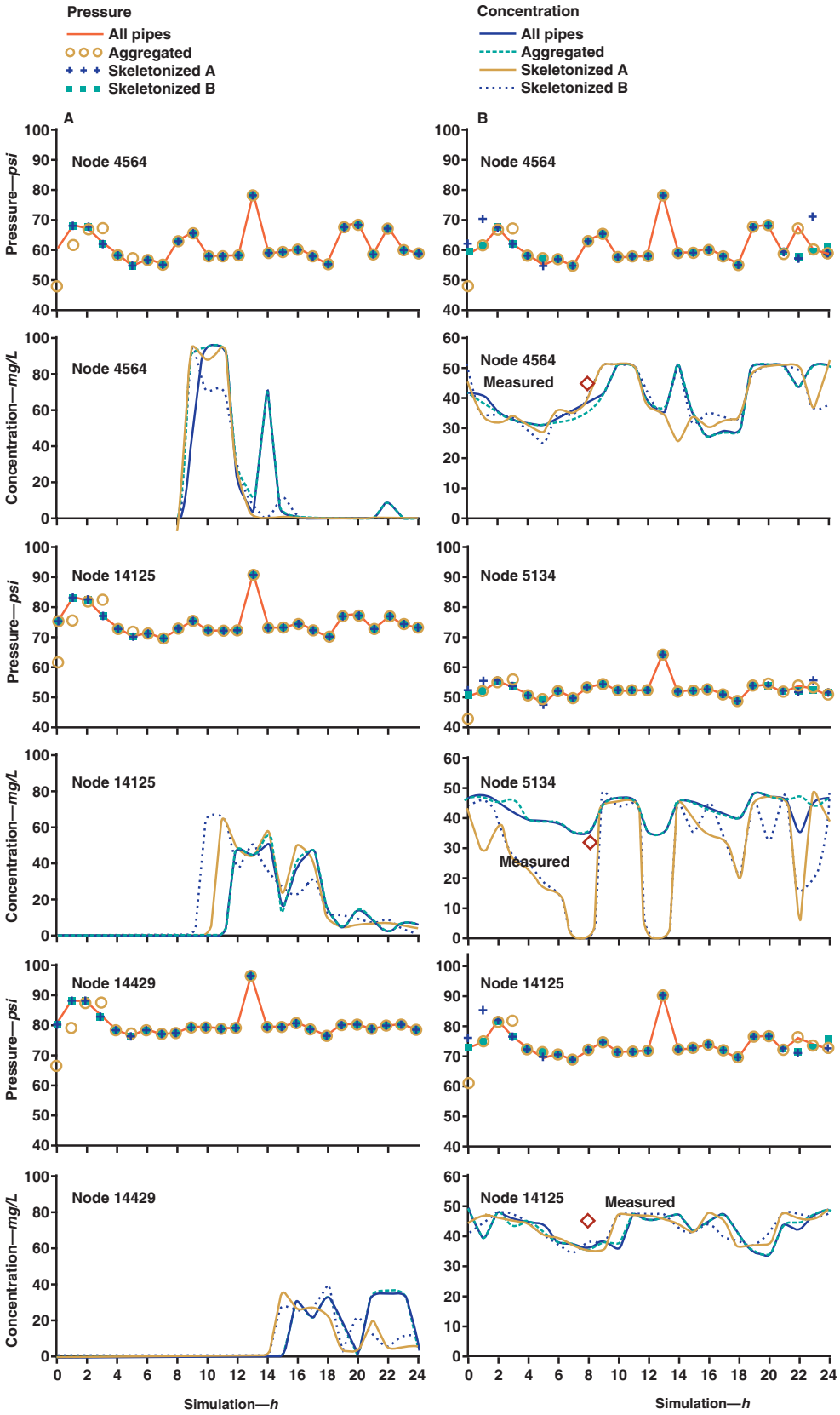
ID—identification

Concentrations represent 24-h average. Refer to Figure 2 for location of all-pipes node; node on map is identified using map point number.

*Node ID of all-pipes model was a dead-end node; node IDs of aggregated, reduced A, and reduced B networks were the closest remaining nodes. For example, node 4319 was a dead-end node in the all-pipes network (Figure 2), and node 4323 was the closest remaining node to the location of node 4319 for the aggregated, reduced A, and reduced B networks.

2008 © American Water Works Association

FIGURE 4 Simulated pressure and concentration results comparing all-pipes, aggregated, skeletonized A, and skeletonized B networks for scenario 1 (A) and scenario 2 (B)



All-pipes network simulation results are from Guan et al (2006); measured data are from Maslia et al (2000).

2008 © American Water Works Association

pipes network, however, concentrations simulated by skeletonized networks A and B generally both underestimated concentrations (hours 5, 14, and 22) and overestimated concentrations (hours 9 and 16).

For node 5134, the aggregated network and the all-pipes network showed simulated concentrations gently varying throughout the 24-h simulation. However, when the all-pipes and aggregated networks were compared with skeletonized networks A and B, significant simulated concentration differences were apparent. Skeletonized networks A and B showed concentrations exhibiting a pulsed behavior resulting in significant declines (hours 7–8, 12–13, 18, and 22) and increases (hours 9, 14, 19, and 23) in concentrations that were not at all similar to the all-pipes network simulated concentrations.

Results for node 14125 again demonstrated that concentrations simulated by the aggregated network were equivalent to the all-pipes network. Simulated concentrations derived from skeletonized networks A and B showed that concentrations arrived earlier (during simulation hours 9–10) when compared with the all-pipes and aggregated network concentrations. Generally, concentrations simulated by skeletonized networks A and B tended to overestimate or underestimate the all-pipes network simulated concentrations, whereas concentrations derived from the aggregated network were equivalent.

As in scenario 1, the computed RMS of concentration differences listed in Table 2 showed much closer agreement between the aggregated network and the all-pipes network than between skeletonized networks A and B and the all-pipes network. For example, for nodes 4564, 5134, and 14125 in scenario 2, the RMS of concentration differences between the aggregated and all-pipes network were 1.13, 2.70, and 0.98 mg/L, respectively. In contrast, the RMS of concentration differences between skeletonized network A and the all-pipes network were 6.94, 18.57, and 3.73 mg/L, respectively.

DISCUSSION

In this research, simulation results were obtained using the aggregated network methodology described by the authors. Pressure and concentration results were compared with results using an all-pipes network and results derived from skeletonized networks A and B using a commercially available skeletonization program. Two scenarios were investigated. For scenario 1, a hypothetical, intentional 100-mg/L contaminant was introduced for 6 h according to the pattern shown in Figure 3. In scenario 1, the contaminant source was characterized by a single source location (one node). For scenario 2, a naturally occurring and conservative groundwater constituent, barium, was traced through the water distribution system. Unlike in scenario 1, the source in scenario 2, i.e., barium, was spatially distributed throughout the network because it entered the water distribution system at water supply well locations (Figure 2).

For each scenario, although simulated pressures were nearly equivalent for each network representation, the aggregated network derived from the methodology presented here provided superior results in terms of matching simulated concentrations with the all-pipes network, compared with results obtained using skeletonized networks A and B. This observation may have important implications relative to assessment and prediction of intentional or nonintentional attacks on a water distribution system as well as implications for epidemiological studies relying on water distribution system modeling for the conduct of case-control studies. Because skeletonized networks A and B relied solely on equivalent hydraulic networks, there was no guarantee that resulting concentrations would be similar to concentrations derived using an all-pipes network. In the scenario 1 example (Figure 4, part A), this caused simulated concentrations using skeletonized networks A and B to

- arrive earlier than concentrations simulated by the all-pipes and aggregated networks for node 14125,
- completely miss a second concentration front at node 4565,
- significantly undersimulate a second concentration front at node 14429, and
- produce much larger RMS of concentration differences over a 24-h simulation period, compared with the aggregated network.

One reason for using simplified network representations of an all-pipes network is to increase the computational speed for which results can be obtained when assessing intentional or unintentional attacks on a water distribution system. At the same time, however, simulation reliability cannot be sacrificed for the sake of computational speed. Table 3 compares the approximate computational times for each of the four networks (all-pipes, aggregated, skeletonized A, and skeletonized B) for a 1,008-h simulation. Computational times were determined using a computer workstation[2] with dual central processing units, 32-bit operating system, and 8 MB of random access memory.

Compared with the all-pipes network, the aggregated methodology reduced computational time by more than half, and the skeletonized networks reduced the computational time even further. However, because the skeletonized networks did not provide accurate results for reliable water quality solutions, removing additional nodes and links from the network and thus further reducing computational costs represented no advantage. The results presented here demonstrated that the aggregation methodology reduced network computations and still provided reliability in terms of simulated concentrations.

For epidemiological and health studies, reliability of simulated concentrations is paramount in order to avoid study bias or misclassification in terms of exposure con-

2008 © American Water Works Association

centrations. As demonstrated in scenario 2 (Figure 4, part B), significant concentration differences and RMS of concentration differences (Table 2) were observed using skeletonized networks A and B, compared with the aggregated network. In the case of node 5134, for example, this could result in significant underestimation of an exposure concentration, leading to a bias or misclassification of an exposed population. Again, simulation reliability would be compromised for the sake of increased computational efficiency if researchers were to use a network-reduction scheme based solely on hydraulic criteria rather than the graph approach based on both hydraulic and water quality criteria used by the aggregated network methodology proposed by the authors.

In applying network aggregation and reduction methods to water distribution system analyses, an important consideration in epidemiologic studies is how dead ends lost through skeletonization affect the reliability of simulation results. In other words, how well do the closest remaining nodes predict water quality in the deleted nodes? Although the results presented here cannot completely answer this question, they offer some insight.

Using the scenario 2 results as an example, an analysis was conducted by comparing the concentrations at dead-end nodes from the all-pipes network to the closest remaining nodes from the aggregated, skeletonized system A, and skeletonized system B networks. Results are shown in Table 4 for 20 selected nodes, whose spatial distributions are shown in Figure 2. Table 4 also lists the pipe diameters corresponding to each of the selected nodes of the networks (all-pipes and simplified). As an example, for node 4319 in the all-pipes network (Table 4), the closest remaining node from the aggregated, skeletonized A, and skeletonized B networks was node 4323. The simulated 24-h averaged concentration at node 4319 was 34.7 mg/L for the all-pipes network, compared with concentrations of 34.7 mg/L for the aggregated network, 1.4 mg/L for skeletonized network A, and 0.0 mg/L for network B. At this location, therefore, the aggregated network methodology provided identical water quality results to the all-pipes network, whereas skeletonized networks A and B provided poor comparisons.

REFERENCES

Anderson, E.J. & Al-Jamal, K.H., 1995. Hydraulic Network Simplification. *ASCE Jour. Water Res. Planning & Mgmt.,* 121:3:235.

Characklis, W.G., 1988. Bacterial Regrowth in Distribution Systems. AwwaRF, Denver.

Cross, H., 1936. Analysis of Flow in Networks of Conduits or Conductors. University of Illinois Engineering Experiment Station Bull. 286, Urbana, Ill.

Eggener, C.L. & Polkowski, L.B., 1976. Network Models and the Impact of Modeling Assumptions. *Jour. AWWA,* 68:4:189.

Grayman, W.M.; Clark, R.M.; & Males, R.M., 1988. Modeling Distribution System Water Quality: Dynamic Approach. *ASCE Jour. Water Res. Planning & Mgmt.,* 114:3:295.

Grayman, W.M.; Males, R.M.; & Clark, R.M., 1991. Effects of Skeletonization in Distribution System Modeling. Proc. AWWA Seminar on Computers in the Water Industry, Houston.

Grayman, W.M. & Males, R.M., 1988. Database Establishment Procedures. Rept., University of Michigan, EPA Cooperative Agreement CR-81 4324-01, Ann Arbor, Mich.

Grayman, R.M. & Rhee, H., 2004. Assessment of Skeletonization in Network Models. Proc. Water Resources 2000, Minneapolis.

Guan, J.; Aral, M.M.; Maslia, M.L.; & Grayman, Wm., 2006. Identification of Contaminant Sources in Water Distribution Systems Using Simulation–Optimization Method: Case Study. *ASCE Jour. Water Res. Planning & Mgmt.,* 132:4:252.

Hamburg, D. & Shamir, U., 1988. Schematic Models for Distribution Systems Design I: Combinatorial Concept. *ASCE Jour. Water Res. Planning & Mgmt.,* 114:2:129.

Hammerlin, G. & Hoffman K.H., 1991. *Numerical Mathematics.* Springer-Verlag, New York.

Maslia, M.L.; Sautner, J.B.; Aral, M.M.; Gillig, R.E.; Reyes, J.J.; & Williams, R.C., 2001. Historical Reconstruction of the Water Distribution System Serving the Dover Township Area, New Jersey: January 1962–December 1996. Rept., Agcy. for Toxic Substances and Disease Registry, Atlanta.

Maslia, M.L.; Sautner, J.B.; Aral, M.M.; Reyes, J.J.; Abraham, J.E.; & Williams, R.C., 2000. Using Water Distribution System Modeling to Assist Epidemiologic Investigations. *ASCE Jour. Water Res. Planning & Mgmt.,* 126:4:180.

MWH Soft Inc., 2003. H_2OMAP Water Users Manual. MWH Soft Inc., Broomfield, Colo.

NJDHSS (New Jersey Department of Health and Senior Services), 2003. Case Control Study of Childhood Cancers in Dover Township (Ocean County), New Jersey. Rept., NJDHSS, Trenton, N.J.

Ostfeld, A. & Solomons, E., 2004. A Stochastic Early Warning Detection System Model for Drinking Water Distribution System Security. Proc. EWRI World Water and Environmental Resources Congr. Reston, Va.

Perelman, L. & Ostfeld, A., 2006. Aggregation of Water Distribution Systems for Contamination Detection. Proc. 8th Annual Intl. Water Distribution Systems Analysis Symp., Cincinnati.

Rossman, L.A., 2000. EPANET 2 Users Manual. National Risk Management Research Laboratory, US Environmental Protection Agency, Cincinnati.

Rossman, L.A., 1994. EPANET Users Manual. Risk Reduction Engineering Laboratory, US Environmental Protection Agency, Cincinnati.

Sarikelle, S. & Chuang, Y., 1989. Analysis of Water Distribution Systems on a Supercomputer. Proc. AWWA Computer Specialty Conf., Denver.

Tarjan R., 1972. Depth-First Search and Linear Graph Algorithms. *SIAM Jour. Computing,* 1:2:146.

Ulanicki, B.; Zehnpfund, A.; & Martinez, F., 1996. Simplification of Water Distribution Network Models. Proc. 2nd Intl. Conf. on Hydroinformatics, Zürich.

2008 © American Water Works Association

Table 4 shows the percent absolute difference between concentrations simulated by the all-pipes network and the simplified networks. Absolute difference in percent was computed as in Eqs 2, 3, and 4 for the aggregated system, skeletonized A system, and skeletonized B system, respectively:

$$\frac{|(C_{AP} - C_{Aq})|}{C_{AP}} \times 100\% \quad (2)$$

$$\frac{|(C_{AP} - C_{SA})|}{C_{AP}} \times 100\% \quad (3)$$

$$\frac{|(C_{AP} - C_{SB})|}{C_{AP}} \times 100\% \quad (4)$$

in which C_{AP}, C_{Ag}, C_{SA}, and C_{SB} are simulated concentrations of all-pipes, aggregated, skeletonized system A, and skeletonized system B networks, respectively. For the aggregated network, the absolute percent difference ranged between 0 and < 10% with a mean absolute difference of about 1%. For skeletonized networks A and B, the absolute differences ranged between 0 and 100% with a mean absolute difference of ~ 20%. The results shown in Table 4 provide additional confirmation that, for the water distribution system network used in an epidemiological case-control study (NJDHHS, 2003) and tested in example applications discussed here, the aggregation network methodology provided highly reliable results in terms of water quality simulations.

The aggregated network was created primarily to ensure that the pressures and concentrations were maintained at selected nodes (monitoring stations); the nature of the algorithm allows the network to sustain pressures and concentrations at all remaining nodes of the system. An important issue, however, is the evaluation of pressures and concentrations at nodes that were eliminated from the original (all-pipes) network. This issue requires additional investigation and development of quantification metrics so that objective comparisons can be made between the all-pipes and skeletonized systems using a variety of network-reduction algorithms.

CONCLUSION

This article introduced a network skeletonization methodology based on both hydraulic and water quality aggregation of an all-pipes network. The methodology was then applied to a complex all-pipes network that had been previously described in the literature and used for an epidemiologic study. Results of two simulations comparing previously published results for an all-pipes network simulation indicated that the aggregated network provided highly reliable pressures and concentrations at network nodes when compared with pressures and concentrations simulated by the all-pipes network. In this study, the aggregation method was capable of reducing the system size by almost half (Table 1), while still preserving system characteristics in terms of reliably simulating pressure and concentrations (Tables 2 and 4). Additional research is being conducted to explore the possibility of further reducing the system size, thereby further increasing computational efficiency while preserving simulation reliability.

ACKNOWLEDGMENT

The research described in this article was funded by the Technion Grand Water Research Institute in Haifa, Israel; the North Atlantic Treaty Organization; Science for Peace (Toronto, Ont.) project CBD.MD.SFP 9814556; and the Agency for Toxic Substances and Disease Registry, US Department of Health and Human Services, Atlanta, Ga. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the Agency for Toxic Substances and Disease Registry or the US Department of Health and Human Services.

ABOUT THE AUTHORS

Lina Perelman is a doctoral candidate at Technion—Israel Institute of Technology in Haifa, Israel, where she received bachelor's and master's of science degrees. For more than two years, she has been developing models for aggregation of water distribution systems, focusing on hydraulic and quality system behavior. Morris L. Maslia (to whom correspondence should be addressed) is a research environmental engineer with the Agency for Toxic Substances and Disease Registry (ATSDR), 4770 Buford Hwy., Mail Stop F59, Atlanta, GA 30341-3717; e-mail mmaslia@cdc.gov. Avi Ostfeld is a senior lecturer in the Department of Civil and Environmental Engineering at Technion. Jason Sautner is an environmental health scientist at ATSDR.

Date of submission: 07/18/07
Date of acceptance: 02/04/08

FOOTNOTES

[1]H_2OMAP Water, MWH Soft Inc., Broomfield, Colo.
[2]Dell 640 Precision, Dell Inc., Round Rock, Texas.

If you have a comment about this article, please contact us at journal@awwa.org.

2008 © American Water Works Association