Multiobjective contaminant response modeling for water distribution systems security

Ami Preis and Avi Ostfeld

ABSTRACT

Following the events of 9/11/2001 in the US, the world public awareness to possible terrorist attacks on water supply systems has increased significantly. The security of drinking water distribution systems has become a foremost concern around the globe. Water distribution systems are spatially diverse and thus are inherently vulnerable to intentional contamination intrusions. In this study, a multiobjective optimization evolutionary model for enhancing the response against deliberate contamination intrusions into water distribution systems is developed and demonstrated. Two conflicting objectives are explored: (1) minimization of the contaminant mass consumed following detection, versus (2) minimization of the number of operational activities required to contain and flush the contaminant out of the system (i.e. number of valves closure and hydrants opening). Such a model is aimed at directing quantitative response actions in opposition to the conservative approach of entire shutdown of the system until flushing and cleaning is completed. The developed model employs the multiobjective Non-Dominated Sorted Genetic Algorithm–II (NSGA-II) scheme, and is demonstrated using two example applications.

Key words | algorithm, EPANET, multiobjective, optimization, water distribution systems, water security

Ami Preis
Avi Ostfeld (corresponding author)
Faculty of Civil and Environmental Engineering,
Technion-I.I.T, Haifa 32000,
Israel
E-mail: *ostfeld@tx.technion.ac.il*

NOMENCLATURE

EPS	Extended Period Simulation
MOGA	Multiobjective Genetic Algorithm
NSGA-II	Non-Dominated Sorted Genetic Algorithm – II
SA	Sensitivity Analysis
SCADA	Supervisory Control and Data Acquisition
SPEA II	Strength Pareto Evolutionary Algorithm II
$C_i(t)$	contaminant concentration at node i at time t
F_1	total contamination mass consumed following detection
F_2	total number of field operations (i.e. valves closure and hydrants opening)
HY	total number of hydrants
I	node index
j	hydrant index
k	valves index
M	number of objective functions
N	total number of consumer nodes
t	elapsed time from first detection time t_d
VA	total number of valves
$V_i(t)$	volume of consumed water at node i at time t
$x^{(1)}, x^{(2)}$	possible two solutions

doi: 10.2166/hydro.2008.061

INTRODUCTION

A water distribution system is typically comprised of tanks, pipes and pumps delivering treated water from treatment plants to consumers. Even a moderate system can contain hundreds of kilometers of pipes and numerous delivery points, making such a system inherently vulnerable.

The threats on water distribution systems can be partitioned into three major groups according to the resulted means of their enhanced security: (1) a direct attack on main infrastructure (dams, treatment plants, storage reservoirs, pipelines, etc.); (2) a cyber attack disabling the functionality of the water utility Supervisory Control and Data Acquisition (SCADA) system, taking over control of key system components; and (3) a deliberate chemical or biological contaminant injection at one of the system nodes.

Of the above threats, a deliberate chemical or biological contaminant injection is the most difficult to address, because of the uncertainty of both the type of injected contaminant and its consequences and of the location and injection time. Principally, a pollutant can be injected at any water distribution system connection (node) using a pump or a mobile pressurized tank. Although backflow preventers provide an obstacle to such actions, they do not exist at all connections and might not be functional at some.

Obviously, if all the system nodes could be reliably monitored then the maximum level of safety would be gained. This is clearly not the case, and to cope with this constraint various methodologies for monitoring stations layout design were developed (e.g. Kessler *et al.* 1998; Al-Zahrani & Moied 2001; Woo *et al.* 2001; Ostfeld & Salomons 2004; Berry *et al.* 2005, 2006; Propato 2006).

Once monitoring stations are in place, a complimentary model should provide the ability to solve the contamination source identification problem of revealing the characteristics of a contaminant intrusion: its location, starting time, injection duration and mass rate. This is essential for implementing a response model for recovering the system after a contamination intrusion, which is the subject of this manuscript.

LITERATURE REVIEW

Shang *et al.* (2002) suggested an input-output model which provides information about the relationships between water quality at input and output locations by tracking water parcels and moving them simultaneously along their paths. Laird *et al.* (2005) presented an origin-tracking algorithm for solving the inverse problem of contamination source identification based on a nonlinear programming framework. Laird *et al.* (2006) addressed the non-uniqueness difficulty of the outcome of the nonlinear model of Laird *et al.* (2005) by incorporating a mixed-integer quadratic program to refine the solutions provided by the nonlinear formulation. Preis & Ostfeld (2006*a*,*b*) introduced methods for solving the contamination source identification problem: (1) a hybrid approach using a coupled Model Trees–Linear Programming scheme (Preis & Ostfeld 2006*a*) and (2) a genetic algorithm (Holland 1975) EPANET (USEPA 2002) framework (Preis & Ostfeld 2006*b*).

Response modeling for optimizing activities after a contamination event is at its infancy. Poulin *et al.* (2006) introduced an algorithm based on a set of heuristic operational and safety rules to isolate contaminated zones, to (1) minimize the risk that contaminated water is consumed; (2) identify network valves to be closed to safely contain the contaminated water; and (3) define of a set of operations to efficiently flush the contaminated water from the network to return to normal operation. The model was demonstrated on the Town of Valcourt water distribution system.

PROBLEM FORMULATION

The effort in this study is to quantitatively model the series of actions required after a contamination intrusion detection for recovering a water distribution system to normal operation. Such actions are aimed at both minimizing the contamination exposure to public and the number of the required field operations (i.e. valves closure and hydrants flushing). These two objectives are provided below.

Consumed contamination mass F_1

F_1 is defined as the total contamination mass consumed following detection:

$$F_1 = \sum_{i=1}^{N} \sum_{t=t_d}^{\mathrm{EPS}} C_i(t) \times V_i(t) \qquad (1)$$

where i is node index, N is total number of consumer nodes, t is elapsed time from first detection time t_d, EPS (Extended Period Simulation) is overall simulation duration, $c_i(t)$ is contaminant concentration at node i at time t and $V_i(t)$ is volume of consumed water at node i at time t.

Number of field operations F_2

F_2 is the total number of field operations (i.e. valves closure and hydrants opening) needed for isolation and flushing of the contamination out of the system:

$$F_2 = \sum_{k=1}^{\mathrm{VA}} \mathrm{VA}_k + \sum_{j=1}^{\mathrm{HY}} \mathrm{HY}_j \tag{2}$$

where k is valves index, VA is total number of valves (VA_k has the value of 1 if the kth valve was closed for isolation and 0 otherwise), j is hydrant index, HY is total number of hydrants (HY_j has the value of 1 if the jth hydrant was opened for flushing and 0 otherwise). Obviously as F_2 increases, F_1 decreases; F_2 and F_1 therefore compete.

MULTIOBJECTIVE OPTIMIZATION

Multiobjective optimization deals with finding the vector of decision variables which satisfy a set of constraints and optimizes a vector function whose elements represent the objective functions. Consequently, a multiobjective optimization problem can be formalized as follows.

Optimize

$$F(x) = (f_1(x), f_2(x), \ldots, f_M(x))^T \tag{3}$$

subject to

$$g_i(x) > 0, \quad i = 1, 2, \ldots, k \tag{4}$$

$$e_j(x) = 0, \quad j = 1, 2, \ldots, l \tag{5}$$

where k and l are inequality and equality constraints, respectively, and $x = (x_1, x_2, \ldots, x_n)^T$ is the vector of decision variables.

The goal in multiobjective optimization is to find from all the sets of solutions which satisfy Equations (4) and (5) the set of solutions which yield optimal values with respect to all the objective functions. This set of solutions is called the Pareto optimal solution set or the non-dominated solution set. Each solution x in the Pareto optimal set is optimal in the sense that it is not possible to improve one objective without making at least one of the others worse.

Any two solutions $x^{(1)}$ and $x^{(2)}$ are compared based on domination, where a solution $x^{(1)}$ is said to dominate $x^{(2)}$ if the following conditions hold:

1. $x^{(1)}$ is no worse than $x^{(2)}$ in all objectives:

$$f_j(x^{(1)}) \ntrianglerighteq f_j(x^{(2)}) \quad \forall j, \quad j = 1, \ldots, M \tag{6}$$

2. $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective:

$$f_j(x^{(1)}) \triangleleft f_j(x^{(2)}) \quad \exists j, \quad j = 1, \ldots, M \tag{7}$$

where $\triangleleft$ indicates a better performance evaluation of an objective function and M is the number of objective functions.

There are two interconnected conceptual goals in multiobjective optimization: to find a set of solutions as close as possible to the Pareto optimal set and to guarantee that the set of solutions is as diverse as possible.

In recent years, several methods have been developed for multiobjective optimization with the emphasis on extending single evolutionary optimization methods such as genetic algorithms to multiobjective evolutionary methodologies. Three of the most popular implemented algorithms are: multiobjective genetic algorithm (MOGA) (Fonseca & Fleming 1995), the non-dominated sorting genetic algorithm II (NSGA II) (Deb *et al.* 2000), and the strength Pareto evolutionary algorithm II (SPEA II) (Zitzler *et al.* 2001).

In water distribution systems, management multiobjective evolutionary optimization techniques have been intensively employed over the last five years. Prasad & Park (2004) presented a multiobjective genetic algorithm approach to the optimal design of a water distribution network by minimizing the network cost versus maximizing the network resilience. Network resilience is defined as a reliability surrogate measure, taking into consideration excess pressure heads at the network nodes and loops with practicable pipe diameters. Prasad *et al.* (2004) used a multiobjective genetic algorithm approach to minimize the total disinfectant dose of booster chlorination stations versus the maximization of the volumetric demand within specified residual chlorine limits. Farmani *et al.* (2005) compared three evolutionary multiobjective optimization algorithms for water distribution system design, by visualizing the resulting non-dominated fronts of each of the methods and by using two performance indicators. Vamvakeridou-Lyroudia *et al.* (2005) employed a genetic algorithm multiobjective scheme to tradeoff the least cost to maximum benefits of a water distribution system design problem, with the benefits evaluated using fuzzy logic reasoning.

The solution scheme implemented in this study is the NSGA-II (Deb *et al.* 2000) which was found to be robust and reliable in previous water distribution systems management applications. The algorithm (Figure 1) trades off the exposure of contamination to the public (F_1) with the number of field operations (F_2).

Initial population generation (set of random strings)

↓

Evaluation of each string's fitness for all objectives

↓

Sorting strings into Pareto fronts: string$_i$ dominates string$_j$ if the following two conditions are met:

i. The fitness of string$_i$ is as good as the fitness of string$_j$ for all objectives
ii. The fitness of string$_i$ is better than the fitness of string$_j$ for at least one objective

↓

Crowding distance sorting within each Pareto front: each front is sorted from the highest crowding distance to the lowest to guarantee the spreading of solutions at the Pareto fronts, where the *crowding distance of string$_i$* is the average distance between the fitness vector of all objectives of string$_i$ and the fitness vector of all objectives of the other members of front$_k$

↓

Generation of a new population using the GA operators: selection, crossover, and mutation. Post the two-step sorting procedure; the best (in the multiobjective sense) strings are selected for reproduction

↓

Stopping condition (for this NSGA-II application): the process is repeated for a predefined number of iterations or until no increase in the number of non-dominated chromosomes through ten generations is accomplished

Figure 1 | Methodology flowchart.

EXAMPLE APPLICATIONS

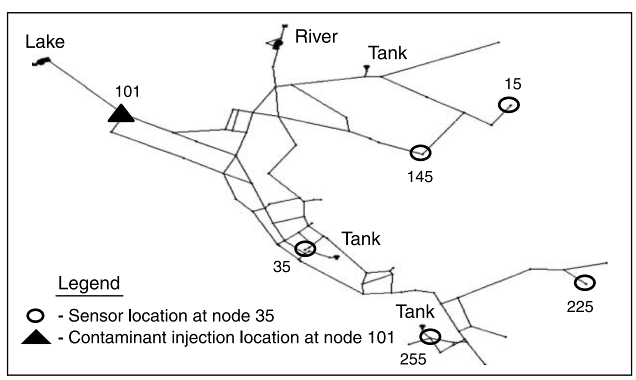
The methodology is demonstrated on two example applications: (1) EPANET Example 3 and (2) Battle of the Water Sensor Networks Example 1. Network layouts and data can both be downloaded from USEPA (2002) and Ostfeld *et al.* (2006), respectively.

Example 1

The system's layout, the valves (assuming two valves are present on each link) and hydrant (nodes number) locations, the sensor layout and a contaminant intrusion point are shown in Figure 2. The system consists of two constant head sources: a lake and a river, three elevated storage tanks, 120 pipes, 94 nodes (consumers and internal nodes) and two pumping stations. It is assumed that the discharge of each hydrant is 0.003154 m^3/sec and that the system is subject to a demand flow pattern of 24 h.

The early warning detection system consists of five sensors at nodes 15, 35, 145, 225 and 255. It is assumed that the sensors were located using a design methodology to maximize the detection likelihood of a random intrusion (e.g. Ostfeld & Salomons 2004).

The pollution scenario imposed for testing the methodology is an intrusion at node 101 at 08:00 of a contaminant with a mass rate of 0.00467 kg/sec for a duration of 6 h. As a



Valves location (links number)	Hydrants location (nodes number)
111, 175, 105, 116, 177, 215, 204, 237, 269, 173, 123, 107, 229, 311, 155, 309, 221, 231, 317, 301	40, 50, 60, 601, 61, 120, 129, 164, 169, 173, 179, 181, 183, 184, 187, 195, 204, 206, 208, 241, 249, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275

Figure 2 | Example 1 layout (EPANET Example 3).

result of the intrusion, the sensor located at node 35 detected the contaminant presence at 10:55. After the first detection, it is assumed that the system response time for initiating actions for handling the event is 1 h. Therefore, at 11:55 protective response actions began, including: (1) contamination source identification; (2) isolation and containment by valves closure; (3) flushing by hydrants opening; and (4) public notification.

Following previous applications of NSGA-II for water distribution systems optimization, the probabilities of crossover and mutation were set equal to 0.75 and 0.07, respectively. The NSGA-II population consisted of 24 chromosomes, and the total number of generations was set to 30. The average running time on an IBM 3.6 GHz, 1 GB of RAM was about 10 min. The overall computational time is composed of the EPANET running time for each multiobjective GA chromosome (i.e. simulation time of a possible operational response to a pollution event) multiplied by the number of chromosomes at the population and by the number of the GA iterations. As the running time of the EPANET simulation model is almost similar for each chromosome (i.e. the pollution event is constant throughout the model implementation), the total running time of each optimized solution is also constant.

The Pareto optimal front of the consumed contamination mass (F_1) versus the number of field operations (F_2) is shown in Figure 3.

It can be seen from Figure 3 that a total of 242 kg of the contaminant, up to the total EPS time of 24 h, would have been consumed if no operations were performed. Compare this with 19 kg if 20 field operations were activated. It can also be seen that the improvement in minimizing the consumed contamination decreases exponentially with regard to the number of field operations. Selecting $F_1 = 25$ kg and $F_2 = 10$ as a compromised solution yields valves closure and hydrants flushing as shown in Figure 4. It can be seen from Figure 4 that flushing is concentrated at the vicinity of sensor 35 (the sensor which first detected the contamination presence) and downstream at node 208. Links 105, 107 which are directly connected to the contaminant intrusion location, and links 177, 204, 221 located downstream of the intrusion, are recommended to be closed.

Four sensitivity analyses (SA) were conducted for this example, each altering one of the genetic algorithm parameters. In SA1, the population was increased to 48 (24 at the base run). As a result, a slightly better Pareto front was obtained but the computational time was increased relative to the increase in the number of generations. The same result was obtained for SA2 at which the number of generations was increased to 60 (30 at the base run). In SA3, the crossover probability was altered to 0.85 (0.75 at the base run), and in SA4 the mutation probability was increased to 0.15 (0.07 at the base run). The outcomes for both SA3 and SA4 have not shown a substantial influence on the algorithm performance. From this, it can be concluded that the most dominant GA parameters were the population size and the number of generations with the crossover and mutation probabilities affecting the algorithm convergence less. However, the increase in both population size and generation number, which increased the computational time significantly, did not result in improved Pareto fronts.

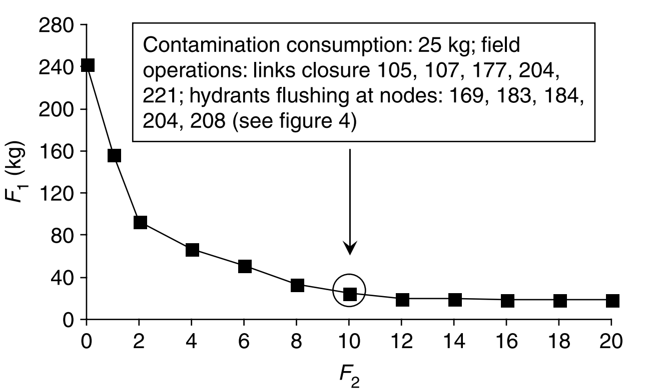


Figure 3 | Optimal Pareto front for Example 1: consumed contamination mass F_1 (kg) versus number of field operations F_2.

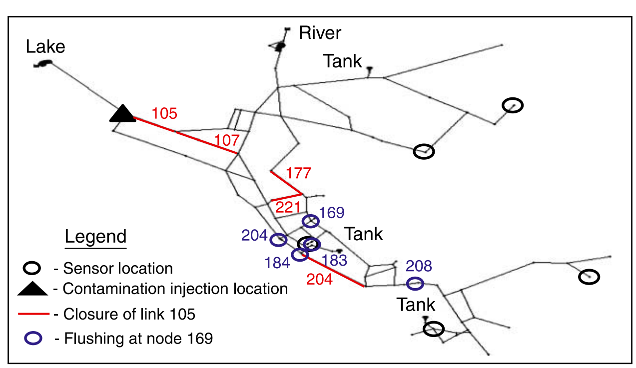


Figure 4 | Optimal 10 field operations ($F_2 = 10$) for Example 1.

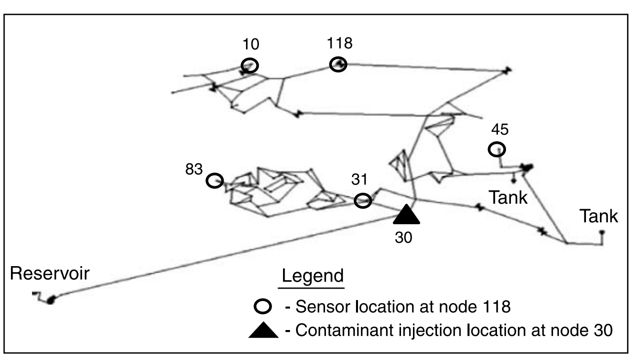
Example 2

The second example application consists of a more complicated hydraulic regime and is aimed at emphasizing the robustness of the model. The system layout, valves (assuming two valves are located on each link) and hydrants (nodes number) locations, sensors layout and a contaminant intrusion point are shown in Figure 5. The system (Battle of the Water Sensor Networks Example 1, Ostfeld *et al.* 2006) consists of one constant head source, two elevated storage tanks, 170 pipes, 129 nodes (consumers and internal nodes) and two pumping stations. The system is subject to a demand flow pattern of 48 h. The hydrants flushing discharge is assumed to be 0.003154 m^3/sec.

The early warning detection system consists of five sensors at nodes 10, 31, 45, 83 and 118. The pollution scenario imposed for testing the methodology is an intrusion at node 30 at 08:00 of a contaminant with a mass rate of 0.00567 kg/sec for a duration of 8 h.

As a result of the intrusion, the sensor located at node 31 detected the contaminant existence at 08:24. After the first detection, it is assumed that the system response time for initiating actions for handling the event is 1 h. Protective response actions therefore began at 09:24.

The probabilities of crossover and mutation were set equal to 0.75 and 0.07, respectively. The NSGA-II population consisted of 24 chromosomes, and the total number of generation was set to 30. Average running time on an IBM 3.6 GHz, 1 GB of RAM was about 15 min.



Valves location (links number)	Hydrants location (nodes number)
0, 1, 11, 33, 40, 45, 109, 143, 66, 35, 42, 160, 46, 111, 34, 164, 31, 94, 53, 73, 51, 71, 70, 69, 68, 67, 121, 11, 115, 113	1, 7, 9, 57, 14, 16, 21, 24, 26, 29, 36, 38, 47, 48, 56, 62, 79, 86, 91, 68

Figure 5 | Example 2 layout (Battle of the Water Sensor Networks Example 1).

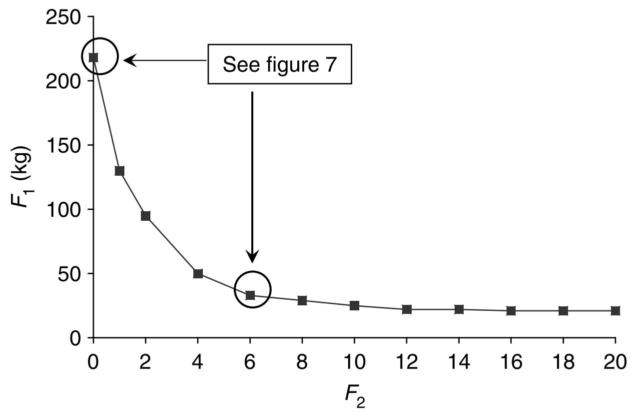


Figure 6 | Optimal Pareto front for Example 2: consumed contamination mass F_1 (kg) versus number of field operations F_2.

The Pareto optimal front of the consumed contamination mass (F_1) versus the number of field operations (F_2) is shown in Figure 6. A total of 216 kg of the contaminant up to the total EPS time of 48 h would have been consumed if no response actions were performed, compared to 45 kg if 20 field operations were activated.

Figure 7 presents the maximum exposure envelopes with no field operations (upper part of Figure 7, marked A)

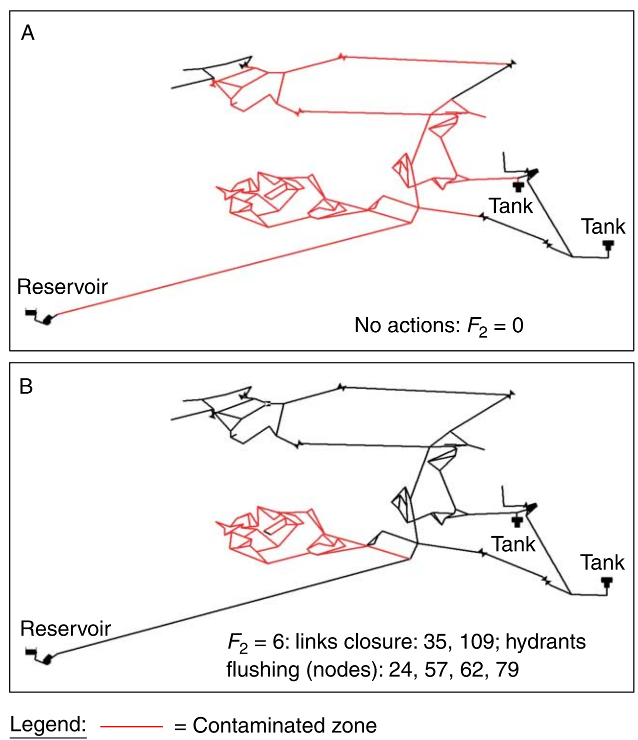


Figure 7 | Maximum exposure envelopes for Example 2 for no actions ($F_2 = 0$) (upper part, marked A) and for six optimal operations ($F_2 = 6$) (lower part, marked B).

and with six optimal operations (lower part of Figure 7, marked B). It can be seen from Figure 7 that with $F_2 = 6$ the maximum contamination spread is substantially reduced compared to no actions ($F_2 = 0$).

CONCLUSIONS

Response modeling for contamination intrusions is in its early research stages. This study developed and demonstrated a multiobjective model for enhancing the response to a contamination event. The NSGA-II was implemented to tradeoff the contamination exposure to public versus the number of field operations required to recover the system to its normal state. Such a model can provide a tool to allow decision-makers to react quantitatively to a contamination event, as opposed to immediately shutting down the entire system.

Research is being conducted into implementation of the methodology to larger water distribution systems, the reduction of the computational time required for solving the problem and into incorporation of uncertainty in both the ability of sensors to detect contaminants and in the system data (e.g. distribution of flows, pressures).

ACKNOWLEDGEMENTS

This research was supported by the Henri Gutwirth Fund for the Promotion of Research at the Technion, by the Institute for Future Defense Technologies Research Named for The Medvedi, Shwartzman and Gensler families, by the Technion Grand Water Research Institute (GWRI) and by NATO (Science for Peace (SfP) project no. CBD.MD.SFP 981456).

REFERENCES

Al-Zahrani, M. & Moied, K. 2001 Locating optimum water quality monitoring stations in water distribution system. In *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges*, Proceedings of the EWRI-ASCE annual conference on Water Resources Planning and Management, published on CD.

Berry, J. W., Fleischer, L., Hart, W. E., Phillips, C. A. & Watson, J. P. 2005 Sensor placement in municipal water networks. *Water Resour. Plann. Manage. Div.* **131** (3), 237–244.

Berry, J. W., Hart, W. E., Phillips, C. A., Uber, J. G. & Watson, J. P. 2006 Sensor placement in municipal water networks with temporal integer programming models. *Water Resour. Plann. Manage. Div.* **132** (4), 218–224.

Deb, K., Agrawal, S., Pratap, A. & Meyarivan, T. 2000 A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In *Proceedings of the Parallel Problem Solving from Nature VI Conference*, Paris, France. pp. 849–858.

Farmani, R., Savic, D. A. & Walters, G. A. 2005 Evolutionary multi-objective optimization in water distribution network design. *Eng. Optim.* **37** (2), 167–183.

Fonseca, C. M. & Fleming, P. J. 1995 An overview of evolutionary algorithms in multiobjective optimization. *Evol. Comp.* **3** (1), 1–16.

Holland, J. H. 1975 *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor.

Kessler, A., Ostfeld, A. & Sinai, G. 1998 Detecting accidental contaminations in municipal water networks. *Water Resour. Plann. Manage. Div.* **124** (4), 192–198.

Laird, C. D., Biegler, L. T., Bloemen Waanders, B. G. & Bartlet, R. A. 2005 Contamination source determination for water networks. *Water Resour. Plann. Manage. Div.* **131** (3), 125–134.

Laird, C. D., Biegler, L. T. & van Bloemen Waanders, B. G. 2006 Mixed-integer approach for obtaining unique solutions in source inversion of water networks. *Water Resour. Plann. Manage. Div.* **132** (4), 242–251.

Ostfeld, A. & Salomons, E. 2004 Optimal layout of early warning detection stations for water distribution systems security. *Water Resour. Plann. Manage. Div.* **130** (5), 377–385.

Ostfeld, A., Uber, J.G. & Salomons, E. 2006 Battle of the Water Sensor Networks (BWSN): A Desgin Challenge for Engineers and Algorithms. In *Proceedings of 8th Annual International Symposium on Water Distribution Systems Analysis*, Cincinnati, Ohio, USA, published on CD. Available online at: www.projects.ex.ac.uk/cws/ (accessed 21 March 2008).

Poulin, A., Mailhot, A., Grondin, P., Delorme, L. & Villeneuve, J.-P. 2006 Optimization of operational response to contamination in water networks. In *Proceedings of 8th Annual International Symposium on Water Distribution Systems Analysis*, Cincinnati, Ohio, USA, published on CD.

Prasad, T. D. & Park, N.-S. 2004 Multiobjective genetic algorithms for design of water distribution networks. *Water Resour. Plann. Manage. Div.* **130** (1), 73–82.

Prasad, T. D., Walters, G. A. & Savic, D. A. 2004 Booster disinfection of water supply networks: multiobjective approach. *Water Resour. Plann. Manage. Div.* **130** (5), 367–376.

Preis, A. & Ostfeld, A. 2006 Contamination source identification in water systems: a hybrid model trees–linear programming scheme. *Water Resour. Plann. Manage. Div.* **132** (4), 263–273.

Preis, A. & Ostfeld, A. 2006*b* Optimal sensors layout for contamination source identification in water distribution systems. In *Proceedings of 8th Annual International Symposium on Water Distribution Systems Analysis*, Cincinnati, Ohio, USA, published on CD.

Propato, M. 2006 Contamination warning in water networks: general mixed-integer linear models for sensor location design. *Water Resour. Plann. Manage. Div.* **132** (4), 225–233.

Shang, F., Uber, J. G. & Polycarpou, M. M. 2002 Practical back-tracking algorithm for water distribution system analysis. *Environ. Eng.* **128** (5), 441–450.

USEPA 2002 EPANET 2.0. Available online at: www.epa.gov/ORD/NRMRL/wswrd/epanet.html (accessed 21 March 2008).

Vamvakeridou-Lyroudia, L. S., Walters, G. A. & Savic, D. A. 2005 Fuzzy Multiobjective Optimization of Water Distribution Networks. *Water Resour. Plann. Manage. Div.* **131** (6), 467–476.

Woo, H. M., Yoon, J. H. & Choi, D. Y. 2001 Optimal monitoring sites based on water quality and quantity in water distribution systems. In *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges*, Proceedings of the EWRI-ASCE annual conference on Water Resources Planning and Management, published on CD.

Zitzler, E., Laumanns, M. & Thiele, L. 2001 SPEA2: improving the strength Pareto evolutionary algorithm for multiobjective optimization. Evolutionary Methods for Design, Optimisation and Control. In *Proceedings of the EUROGEN2001 Conference*, Athens, Greece (CIMNE: Barcelona), pp. 95–100.

First received 15 November 2007; accepted in revised form 12 April 2008