

Optimal Design of Sensor Placement in Water Distribution Networks

Mustafa M. Aral¹; Jiabao Guan²; and Morris L. Maslia³

Abstract: In this study we provide a methodology for the optimal design of water sensor placement in water distribution networks. The optimization algorithm used is based on a simulation-optimization and a single-objective function approach which incorporates multiple factors used in the design of the system. In this sense the proposed model mimics a multiobjective approach and yields the final design without explicitly specifying a preference among the multiple objectives of the problem. A reliability constraint concept is also introduced into the optimization model such that the minimum number of sensors and their optimal placement can be identified in order to satisfy a prespecified reliability criterion for the network. Progressive genetic algorithm approach is used for the solution of the model. The algorithm works on a subset of the complete set of junctions present in the system and the final solution is obtained through the evolution of subdomain sets. The proposed algorithm is applied to the two test networks to assess the selected design. The results of the proposed solution are discussed comparatively with the outcome of other solutions that were submitted to a water distribution systems analysis symposium. These comparisons indicate that the algorithm proposed here is an effective approach in solving this problem.

DOI: 10.1061/(ASCE)WR.1943-5452.0000001

CE Database subject headings: Water distribution systems; Optimization; Algorithms; Probe instruments.

Author keywords: Water distribution system; Optimization; Genetic algorithms; Water sensor networks.

Introduction

The recent terrorism concerns have caused emergency management personnel and public health officials to devote attention to the security of infrastructure systems in the United States and around the world (Congressional Research Service 2002). Since the water supply infrastructures are not built to withstand an intentional attack, they are vulnerable to external attacks (House Science Committee 2001). Since the occurrence of the recent terrorism events, various preventative measures have been enacted. However, because these measures depend on a lengthy response time, they are sorely inadequate and fail to provide an early warning or instantaneous notification that would be required for an intentional attack. Thus, to be protective of public health, water distribution system security requires faster detection methods and response of a computerized control network—an early warning detection system—to an intentional attack.

Reducing or minimizing the vulnerability of water distribution systems require the implementation of a detection system that can (1) assure that any degradation in water quality is quickly identi-

fied and (2) rapid assessments can be made identifying areas of the distribution system that have been impacted. For this purpose a well-designed water sensor network system is necessary to detect the contamination events in minimum time reducing the risk of prolonged exposure. The design of this water sensor network deals with several factors such as cost, reliability of the designed system, detection time, and the population exposed. A good design should provide an acceptable trade-off among these competing factors.

More recently a special session entitled “Battle of the Water Sensors Network” (BWSN 2006) was dedicated to this issue during the conference of WDSA (2006) held in Cincinnati. The special session included 14 papers addressing this problem and the researchers provided solutions to this problem using different methods and different points of view (BWSN 2006). In this paper we also provide a comparison of the results presented in this study with the outcome provided by the writers of the other studies, as referenced above.

Literature Review

Prior to the recent terrorism events, the scientific literature primarily focused on assessing water quality monitoring networks in terms of efficiency and cost to guarantee a reliable supply of potable water to the public. Ostfeld and Salomons (2004) provided a review of this literature, which will not be repeated here owing to brevity. Recently, the literature has been focused more on the optimal layout of early warning detection systems with respect to intentional biological or chemical attacks and assessment of the risk to population exposure (Bahadur et al. 2001; Bahadur et al. 2003; Ostfeld and Salomons 2004, 2006). For a comprehensive resource on security of water supply systems, the reader is also referred to Mays (2004). More recently, the optimal

¹Director, Multimedia Environmental Simulations Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332 (corresponding author). E-mail: maral@ce.gatech.edu

²Senior Research Engineer, Multimedia Environmental Simulations Laboratory, School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

³Research Environmental Engineer, Agency for Toxic Substances and Disease Registry, 1600 Clifton Rd., Mail Stop E-32, Atlanta, GA 30333.

Note. This manuscript was submitted on August 22, 2008; approved on September 11, 2008; published online on March 18, 2009. Discussion period open until June 1, 2010; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Water Resources Planning and Management*, Vol. 136, No. 1, January 1, 2010. ©ASCE, ISSN 0733-9496/2010/1-5-18/\$25.00.

sensor network design problem was discussed in the BWSN session of the conference WDSA (2006), as referenced above. These studies provided 14 different approaches to the solution of the sensor placement problem. We provide a comparison of the results of these designs in this study using standardized analysis of the results with respect to the criteria identified for the BWSN exercise. However, the reader is referred to the conference proceeding (BWSN 2006) for further details of each design consideration discussed in these 14 applications.

The design of an effective water sensor network may be formulated as an optimization model. A variety of models and solution algorithms, including integer programming methods and combinatorial heuristics, have been developed and reported in the literature. A detailed review of mixed-integer programming models to determine the placement of contaminant sensors within a municipal water network can be found in Berry et al. (2005) and Propato (2006). As the objective function, these models use the concept of minimization of the expected population that is at risk for an attack and solve the problem using intelligent variations of branch and bound approaches. Ostfeld and Salomons (2004) proposed a two-stage approach to determine the optimal layout of an early warning detection stations in the water distribution system. In their methodology a randomized pollution matrix (RPN) was constructed based on random injection events and a genetic algorithm was used to obtain maximum column coverage of the RPN. Although both of these approaches are effective, they may not provide an effective and efficient solution for large-scale water distribution systems because of the heavy computational burden they may require.

In the optimal placement of sensors in a network, the objective function selected directly influences the sensor placement location in the system. In most models that are used in the literature a single objective is selected for this purpose. For example: (1) minimizing total contaminated volume consumed (Kessler et al. 1998); (2) minimizing the time of detection (Kumar et al. 1999); (3) maximizing detection coverage and probability (Ostfeld and Salomons 2004); or (4) minimizing population affected by contamination (Berry et al. 2005). Actually the optimal water sensor placement problem is a multiobjective decision problem. To the best of our knowledge, there are a few reported models in the literature that use a multiobjective solution approach for this application (Watson et al. 2004; Ostfeld and Salomons 2004; Propato 2006). Perhaps the reason lies in the fact that the optimization algorithm may tend to be extremely inefficient for an integer programming solution containing thousands of variables which would be the case for a large water distribution system in some cases and in other cases where multiple feasible solutions are provided, a selection of the final design is most commonly compromised since the final selection would still reflect the preferences or the point of view of the decision maker. One of the goals of the method proposed in this study is to overcome this dilemma through the formulation of a single synthetic objective function that would mimic the performance of a multiobjective approach. Thus, in this study a single-objective optimization model is developed that incorporates the four criteria adopted in the Battle of the Water Sensors Network (BWSN 2006), as described below. A progressive genetic algorithm (PGA) is used to solve the model to overcome the computational inefficiency issues in the study of the large-scale water distribution networks. The outcome, based on our comparisons with other solutions, is a robust methodology which may be useful in solving this problem effectively.

Table 1. Characteristics of the Hypothetical Networks 1 and 2

Parameter	Network 1	Network 2
Junctions	126	12,523
Pipes	168	14,822
Reservoirs	1	2
Tanks	2	2
Pumps	2	4
Valves	8	8

Water Distribution System Characteristics and the Sensor Network Evaluation Criteria

Two hypothetical networks were supplied for the BWSN application (BWSN 2006) and we will use these two water distribution systems to demonstrate the performance of the model proposed here. The characteristics of these two water distribution systems are given in Table 1.

To evaluate the proposed algorithm's efficiency and reliability in detecting contamination events using these two networks, four quantitative design objectives were identified, as described in the BWSN exercise:

Expected Time of Detection Z_1 . For a particular contamination scenario, the time of detection by a sensor is defined as the elapsed time from the start of the contamination event to the first presence of a nonzero contaminant concentration at a sensor. This detection time is denoted as t_i where the subscript i refers to the i th sensor location. The time of detection for the sensor network for a particular random contamination event, t_d , can be identified as the minimum detection time among all sensors present in the design, $t_d = \min t_i$. In this case the objective function to be minimized is the expected value computed over the assumed probability distribution of random contamination events

$$Z_1 = E(t_d) \quad (1)$$

where $E[\cdot]$ =expected value and will be approximated by Monte Carlo simulation for the purpose of comparing alternative designs. The assumed probability distribution of contamination events is discussed below in conjunction with other design assumptions.

Expected Population Affected prior to Detection Z_2 . For a particular contamination scenario, the population affected is measured as a function of the ingested contaminant mass. The ingested contaminant mass in turn depends on the time of detection for the sensor network, as described above. The key assumption here is that no mass will be ingested after detection since it is assumed that an appropriate response, such as the shutdown of the system, will be initiated at that time. For a particular contamination scenario, the mass ingested—prior to detection—by any individual at network node i can be calculated as

$$M_i = \phi \Delta t \sum_{k=1}^N c_{ik} \rho_{ik} \quad (2)$$

where ϕ =mean volumetric ingestion rate (L/day); Δt =time step (days); c_{ik} =contaminant concentration at junction i at time step k (mg/Liter); ρ_{ik} =dose rate multiplier for junction i at time step k and calculated by $\rho_{ik} = q_{ik} / \bar{q}_i$; \bar{q}_i =average water demand at junction i ; and N =number of time steps prior to detection.

A typical dose-response model is used to express the probability that any person ingesting mass M_i will be affected (i.e., become infected or symptomatic)

Table 2. Hydraulic and Water Quality Parameters

Parameter	Value
Mean volumetric ingestion rate	2 L/day
Probit slope parameter	0.34
Dose with 0.5 probability of becoming infected	41 mg/kg
Assumed body weight	70 kg
Hazard concentration threshold	0.3 mg/L
Per capita water consumption rate	300 L/day

$$R_i = \Phi\{\beta \log_{10}[(M_i/W)/D_{50}]\} \quad (3)$$

where R_i =probability [0, 1] that a person who ingests contaminant mass M_i will become infected or symptomatic; β =Probit slope parameter (dimensionless); W =assumed body weight (kg); D_{50} =dose that would result in a 0.5 probability of becoming infected or symptomatic (mg/kg); and Φ =standard normal cumulative distribution function.

Based on this, the population affected for a particular contamination scenario is calculated by

$$P_a = \sum_{i=1}^m R_i P_i \quad (4)$$

where P_i =population assigned to node i ; and m =total number of junctions. For simplicity, the population can be estimated by the total water demand divided by total per capita water consumption rate. During the design stage, the contamination scenarios will be randomly generated and contamination can occur at any junction and at any time step according to a uniform probability distribution. The more scenarios one considers in this approach, the more reliable the designed water sensor network will be. Of course, the use of a large number of scenarios will increase the computation time and the computer memory requirements. According to this criterion, the measure Z_2 is based on the expected value of P_a computed over the assumed probability distribution of contamination events

$$Z_2 = E[P_a] \quad (5)$$

where $E[\cdot]$ again denotes the expected value and will be approximated by Monte Carlo simulation for purposes of comparing alternative sensor placement designs. The assumed probability distribution of contamination events is discussed below in conjunction with other design assumptions. In the scenarios tested, the contaminant intrusion occurs at a junction of the system with injection flow rate of 125 L/h at a contaminant concentration of 230,000 mg/L and the injection duration is 2 h. The hydraulic and water quality parameters used for these two networks are given in Table 2.

Expected Demand of Contaminated Water prior to Detection Z_3 . Z_3 =expected volume of consumed contaminated water prior to detection

$$Z_3 = E[V_d] \quad (6)$$

where V_d denotes the total volumetric water demand that exceeds a predefined hazard concentration and $E[\cdot]$ =expected value, as defined above. As to the expected population affected, the key assumption is that no water is delivered after detection. The objective Z_3 (as Z_2 and Z_1) is to be minimized.

Reliability Z_4 . Given a sensor network design (i.e., number of sensors and sensor locations) the reliability (the probability of detection) is defined as

$$Z_4 = \frac{1}{N} \sum_{i=1}^N d_i \quad (7)$$

where $d_i=1$ if contamination scenario i is detected and zero if otherwise and N =number of the total contamination scenarios considered. Contrary to the objectives Z_1 , Z_2 , and Z_3 , the objective Z_4 is to be maximized.

It is important to note that for all four measures described above, the detection of a contaminant by a sensor is an important part of the computation of the value of the measure. In the optimization protocol proposed in this study, if a contamination scenario is not detected, that event is not ignored but instead the time of detection for that event is set to be the end of the simulation period (i.e., maximum simulation time). The effect of a scenario not being detected is also directly reflected in the reliability measure as well. A good solution should yield lower values of the former three measures and a higher reliability value. As discussed in detail below, the methodology and the procedure developed here strikes a balance between these four measures. It can be stated that in the procedure described here the “trade-off” coefficients of the objective function are defined in terms of the four measures defined above and thus they are optimized as well. This is the most important contribution of the algorithm proposed here.

Mathematical Formulation

The performance of the network designed is evaluated using the four criteria given above. These criteria are all compatible in the following sense. In order to reduce the number of exposed population and contaminated water volume, one would attempt to design the system to minimize the detection time and clearly a higher reliability is necessary for the system to be effective. However, it is important to recognize that emphasizing any one of these criteria over the others will result in a different sensor placement pattern for the water distribution network. One alternative to minimize the effect of this dilemma is to use a multiobjective optimization approach. However, the multiobjective approach, in one way or another, will be eventually associated with a trade-off methodology in the selection of the final design. Thus, we start with the premise that there is no single mathematical formulation that would yield all four criteria to reach their true optimal values, which would be otherwise obtained if these criteria were used as a stand alone criterion. Our goal is to identify an objective function that would strike a balance between these four objectives without artificial controls or one that would eventually reflect the preferences of the decision maker in selecting one design among many feasible optimal solutions that may be provided. For this purpose, in this study a single-objective formulation approach is selected. The goal is to blend the four criteria into one objective function as best as possible. To accomplish this, the following objective function is proposed:

$$F = \underset{X}{\text{minimize}} \left\{ [1 - r(X)] \times E_s \left[\sum_{i=1}^N \sum_{t=t_s^{in}}^{t_s^d(X)} (t - t_s^{in} + 1) V_{is}(t) \right] \right\} \quad (8a)$$

where s =index of contamination scenarios or events; X =decision vector; and $X=[x_1, x_2, \dots, x_N]^T$ where x_i =decision variable, $x_i=\{0, 1\}$. If $x_i=1$, a sensor is placed at junction i , otherwise there is no sensor placement at junction i . $E_s[\cdot]$ =expected value of total water demand that exceeds a predefined threshold

hazard concentration, C_{\min} in the water distribution system; t_s^{in} =index of injection time at which the contaminant is injected into the system for contamination event s ; $V_{is}(t)$ =volume of contaminated water consumed at junction i for contamination event s at time step t ; i =index of junctions in the system; N =total number of junctions in the system; t =index of time steps; and $r(X)$ =reliability measure of event detection (%). If the N_s contamination events are randomly generated with uniform distribution within a specified period, the equation given above can be written as

$$F = \underset{X}{\text{minimize}} \left\{ \frac{[1 - r(X)]}{N_s} \sum_{s=1}^{N_s} \left[\sum_{i=1}^N \sum_{t=t_s^{in}}^{t_s^d(X)} (t - t_s^{in} + 1) V_{is}(t) \right] \right\} \quad (8b)$$

where N_s =number of contamination events. If we identify t_s^d as the time of detection for sensor network for the contamination events t_s^d can be calculated as

$$t_s^d(X) = \min_j \{t_{js}\} \quad (9)$$

where j =index of sensors in solution X and t_{js} =time of detection at sensor j for event s . Then, t_s^d will be an implicit function of X . In the objective function above, Eq. (8a) or (8b), $V_{is}(t)$ is calculated by

$$V_{is}(t) = \begin{cases} 0 & \text{if } C_{is}(t) < C_{\min} \\ q_i(t)\Delta t & \text{if } C_{is}(t) \geq C_{\min} \end{cases} \quad (10)$$

where $q_i(t)$ =water demand at junction i at time step t and Δt =time step interval. The reliability measure $r(X)$ is determined by the solution vector X . During the optimization process, $r(X)$ can be approximately estimated by the reliability, which is defined as

$$r = \frac{1}{N_s} \sum_{s=1}^{N_s} d_s(X) \quad (11)$$

$d_s=1$ if contamination scenario s is detected and $d_s=0$ if otherwise.

This objective function synthetically reflects the four criteria described above in a unified form. As stated earlier, the population affected is a function of water volume consumed. The minimization of water volume contaminated can directly realize the purpose of reducing the population affected. In the model proposed above, the time of detection directly affects the objective function value through two factors. One is associated with the summation terms, the longer the time of detection is, the more terms there will be in the summation. This results in a larger objective function value which we would be minimizing. The other is associated with the term $(t - t_s^{in} + 1)$. This term is included as a penalty coefficient. The larger the value of t , the larger the penalty coefficient will be. However, it is clear that this penalty function is not preassigned but minimized as well, as the solution progresses. Thus, the optimal solution is associated with detection time being small, which is the objective of our model. The term $[1 - r(X)]$ is a coefficient associated with the reliability of the solution. Higher reliability values results in a smaller coefficient. This yields smaller objective function value and vice versa. Again, this coefficient is not a set constant but is also minimized as the optimization process progresses. We anticipate that the optimized results obtained using this objective function may yield a balanced inherent trade-off solution for the four criteria identified above. More important, as emphasized in this study, the bias that

exists in the trade-off coefficients defined in the objective function is also optimized as the solution progresses. Thus, the trade-off is based on the balance that can be found among the four objectives of the problem.

In the water distribution system, each junction node may be selected as a candidate location for the installation of a sensor. Installation of more sensors will increase the reliability of the sensor network, albeit at a higher cost. The purpose of optimal design of water sensor network is to determine the number of sensors that are needed and where to place these sensors while satisfying the specified reliability of the system. Explicitly, the number of sensors needed is a function of the specified reliability. Therefore, the following constraint is also considered in the formulation:

$$\sum_{i=1}^N x_i = M(r_c) \quad (12)$$

$$r = \text{prob}\{\text{detected events}/M = m\} \geq r_c \quad (13)$$

where $M(r_c)$ =number of sensors needed for the specified reliability; m =specified number of sensors; and r_c =predefined reliability.

Eq. (13) indicates that the probability of detected events under the condition of a specified number of sensors must be larger than or equal to a given reliability level of the system for the water distribution system security. If Eq. (13) is not included in the model, the optimization problem may be identified as the case where one seeks optimal sensor placement with a specified number of sensors. Otherwise, the optimization problem also identifies the optimal number of sensors necessary and their optimal placement given the reliability constraint. Both optimization problems will be addressed in the numerical applications discussed below.

Optimal Solution Methodology

The optimization model described above is $\{0,1\}$ integer programming problem. Although the conventional $\{0,1\}$ integer programming methods, such as branch and bound and enumeration and implicit enumeration, can be used to solve this problem, they may tend to be inefficient for large number of candidate junctions. In recent years, the application of heuristic algorithms is also explored in the solution of these problems. A genetic algorithm was used to design the early warning detection systems by Ostfeld and Salomons (2004). In their study a genetic algorithm framework in integration with the network solver EPANET (Rossman 2000) was proposed and the methodology was demonstrated through two example applications. However, we must emphasize that both example applications studied in their paper contain less than 20 candidate junctions. There is no doubt that the proposed approach is efficient in selecting the sensor locations from 20 candidate junctions. However, if the number of candidate junctions is in the order of thousands, this approach may be inefficient because the crossover operation used in the genetic algorithm may lose its functionality. For example, selecting five-sensor locations from 1,000 candidate junctions, the chromosome with 1,000 bits contains only five bits which has a value "one" and all other bits will be "zero." With such a chromosome, most of the offspring children generated by the crossover operation may tend to be the same as their parents. In this case, the solution will improve very slowly increasing the computational cost. In this study, the PGA

(Guan and Aral 1999; Aral et al. 2001; Guan and Aral 2005; Guan et al. 2006a,b) is used for the solution of the optimization model described above to determine the optimal number of sensors and sensor placements. The PGA works on the subdomain of junctions that is a subset of the complete set of junctions in the water distribution system. The solution of the problem in the subdomain is based on the conventional genetic algorithm procedures and the subdomain selection and subdomain evolution is based on earlier concepts developed by Guan and Aral (1999) and Aral et al. (2001) which will not be repeated here.

For a specified number of sensors, this procedure can also be used to identify the optimal locations of sensors. In this case, the algorithm may be linked to another loop to check if the solution satisfies a predefined reliability constraint [Eq. (13)]. The reliability is estimated by Eq. (11). If the reliability is less than the specified reliability, this would indicate that more sensors are required. If the reliability is larger than the specified reliability, this would imply that the number of sensors may be reduced to save on cost. This process of increasing or decreasing the number of sensors is systematically determined in the proposed algorithm according to the following criteria:

$$\begin{cases} |r - r_c| \leq \varepsilon & \text{stop calculation} \\ r < r_c & \text{increase number of sensors} \\ r > r_c & \text{decrease number of sensors} \end{cases} \quad (14)$$

where ε =preselected allowable error. The change in the number of sensors can be approximated by

$$M^{(l+1)} = M^{(l)} + \Delta M \quad (15)$$

where l =index of iterations; ΔM =incremental number of sensors to add or subtract. In the first iteration, $l=0$, ΔM may be assigned a value

$$\Delta M = \begin{cases} 1 & r^{(l)} < r_c \\ -1 & r^{(l)} > r_c \end{cases} \quad (16)$$

In general, the reliability of the system is a monotonically increasing function of the number of sensors and linear interpolation can be used to determine ΔM . Therefore, in the subsequent iterations ΔM is determined using the equation below

$$\Delta M = \left\lceil \frac{r_c - r^{(l)}}{r^{(l)} - r^{(l-1)}} [M^{(l)} - M^{(l-1)}] \right\rceil \quad (17)$$

where $\lceil \cdot \rceil$ yields a ceiling integer value. In Eq. (17), the sign of ΔM depends on the term $[r_c - r^{(l)}]$ if r is monotonically increasing as a function of the number of sensors. If $r_c > r^{(l)}$, the sign is positive, the system will increase the number of sensors required. If $r_c < r^{(l)}$, the sign is negative, the algorithm will decrease the number of sensors required. However, as described above, the objective function synthetically trades-off the four criteria and may yield a case with $r^{(l)} \leq r^{(l-1)}$ under $M^{(l)} \geq M^{(l-1)}$ or $r^{(l)} \geq r^{(l-1)}$ under $M^{(l)} \leq M^{(l-1)}$. For such cases, ΔM should be set to the value given by Eq. (16) because it is well known that increasing number of water sensors will increase the reliability of the system.

It must be pointed out that the PGA works on a subset of junctions rather than all junctions. This will improve the efficiency of the genetic algorithm in solving the $\{0,1\}$ integer programming problem. However, this may result in a local optimal solution. This situation can be progressively overcome through the updating of the subdomain during the iterative solution. The final solution may converge to the global optimal solution as long as all junctions are included into the subdomain at least once.

Table 3. Parameters Used in the PGA

Parameters	Value
Population size	50
Crossover ratio	0.8
New member generation ratio	0.2
Elitism ratio	Best member
Mutation ratio	0.2
Maximum generation for each subdomain	30

In the proposed algorithm, the EPANET water distribution system modeling platform is an important component that is used to perform extended-period simulation of hydraulic and water quality behavior within a pressurized pipe network (Rossman 2000). Because these data will be repeatedly used in the evaluation of the objective function, the EPANET simulation is conducted before optimization and the data that are used in the subsequent optimization steps are recovered from the stored data in the computer memory. This approach reduces the computational burden.

Numerical Applications

The model and the methodology proposed above is used in the solution of the two water distribution networks provided in the BWSN application (BWSN 2006). The performance of the water sensor network designed is synthetically evaluated using the four criteria previously described.

Water Distribution System 1. Network 1 is a relatively simple water distribution system consisting of 169 pipes, 129 junctions, a reservoir, two elevated storage tanks, and two pumping stations (BWSN 2006). For the design of the water sensor network, 20 contamination scenarios for each junction are randomly generated within the first 24 h, resulting in a total of 2,580 scenarios. During the scenario generation stage, the writers did not consider the multiple injection point case since it is assumed that the critical scenario would be the one source case (Case A in BWSN exercise) in view of the objective function selected here. During optimization, the number of candidate junctions in the subdomain was selected as 50 and the number of subdomains as 20.

The reliability of the system is specified to be 85%. The candidate junctions are selected using the roulette wheel method based on the detection time. Parameters used in the PGA are given in Table 3.

The case for a specified number of sensors is addressed first. The optimal results obtained for the number of sensors ranging from 5 to 20 are given in Table 4, where Z_1 , Z_2 , Z_3 , and Z_4 are the measures identified earlier. In Table 4, the network junction identification (junction ID) is identified with a letter "J" and a number which indicates the junction identifier.

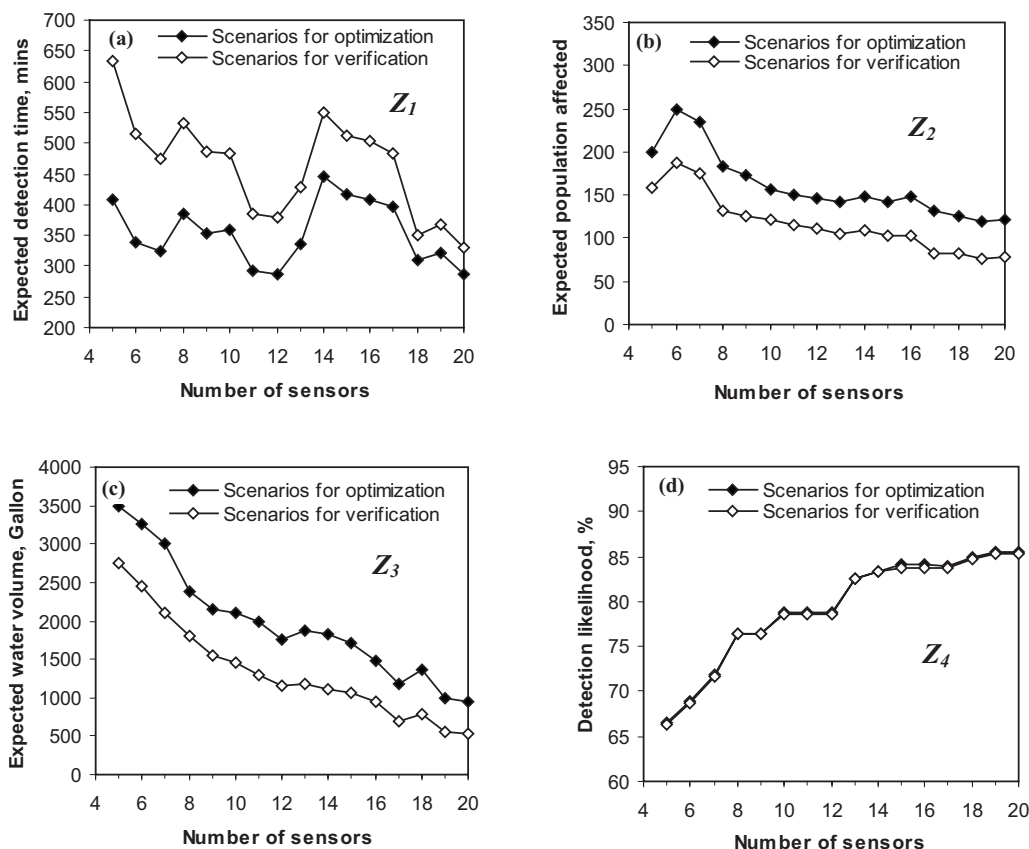
The relation between the number of sensors used and the value of the four measures are shown in Fig. 1. Based on these results, it can be clearly seen that the expected contaminated water volume [Fig. 1(c)] and population affected [Fig. 1(b)] decrease as the number of sensors increase. The reliability [Fig. 1(d)] increases as the number of sensors increase, although there are small fluctuations in this outcome. The expected time of detection [Fig. 1(a)] has a decreasing pattern with larger fluctuations as the number of sensors increase. This fluctuation is expected considering the multiobjective nature of the problem, as discussed below, for the

Table 4. Optimal Solutions and Performance Obtained during Design Phase for Network 1

Number of sensors	Junction ID	Z_1 (min)	Z_2 (ind)	Z_3 (gal.)	Z_4 (%)
5	J17, J31, J81, J98, J102	409.05	200.56	3482.09	66.51
6	J20, J68, J84, J98, J102, J118	339.23	248.97	3254.67	68.84
7	J20, J68, J82, J84, J98, J103, J118	324.95	235.25	2996.49	71.94
8	J17, J23, J46, J68, J83, J101, J103, J118	384.10	183.05	2377.35	76.43
9	J17, J23, J39, J46, J68, J83, J101, J103, J118	353.32	172.99	2147.19	76.43
10	J17, J23, J39, J45, J68, J83, J101, J102, J118, J122	360.18	156.86	2115.58	78.76
11	J17, J23, J37, J45, J68, J76, J83, J89, J101, J118, J122	292.09	149.44	1996.28	78.76
12	J17, J23, J27, J37, J45, J68, J76, J83, J89, J101, J118, J122	287.31	146.70	1752.63	78.76
13	J11, J17, J20, J23, J37, J45, J68, J76, J82, J83, J101, J118, J122	335.14	141.57	1883.78	82.64
14	J5, J10, J17, J23, J37, J45, J68, J76, J82, J83, J98, J118, J122, J126	444.96	148.24	1826.67	83.41
15	J3, J10, J17, J23, J35, J45, J68, J74, J82, J84, J94, J101, J118, J122, J126	417.34	142.92	1708.94	84.03
16	J3, J10, J23, J35, J45, J68, J74, J82, J84, J94, J101, J102, J116, J118, J122, J126	407.81	147.72	1475.21	84.03
17	J5, J10, J17, J20, J23, J34, J35, J45, J68, J74, J82, J84, J94, J101, J102, J118, J122	395.22	132.19	1178.33	83.88
18	J4, J11, J17, J21, J27, J31, J35, J46, J68, J75, J79, J82, J83, J96, J100, J118, J122, J126	310.91	125.58	1372.86	84.85
19	J4, J11, J17, J21, J27, J31, J34, J46, J68, J75, J82, J83, J95, J100, J102, J118, J122, J126	320.07	119.98	1004.37	85.50
20	J4, J11, J17, J21, J27, J31, J34, J35, J46, J68, J75, J79, J82, J83, J98, J100, J102, J118, J122, J126	287.22	122.23	956.08	85.50

fluctuations on the reliability measure. For the scenario with 17 sensors (Table 4), there is a trend reversal for the reliability. The value of the measure, Z_4 , decreases from a value of 84.03% for 16 sensors to 83.88% for 17 sensors. Although this decrease is not significant, it is a consequence of the objective function used. This outcome should be evaluated in light of the decrease in contaminated water volume from 1,475.21 gal. for 16 sensors to

1,178.33 gal. for the 17-sensor case (Table 4). This outcome reflects the multiobjective nature of the objective function used. The four measures of the designed water sensor network (Z_1 - Z_4) yield a reasonable trade-off among the measures when the proposed objective function is used. If a single measure is chosen as the objective function, that measure is expected to improve as the optimization outcome. However, for such an analysis, the other

**Fig. 1.** Relationship between design measures Z_1 - Z_4 and the number of sensors

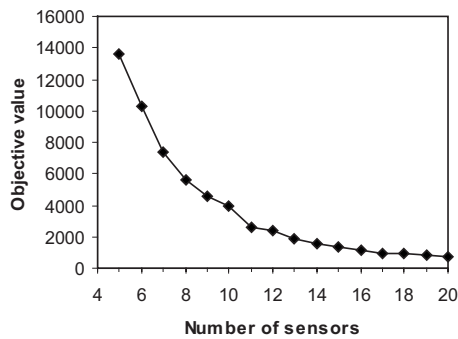


Fig. 2. Relationship between the objective function value and the number of sensors

three measures may become worse progressively, for example, if we select the maximization of the reliability as a single objective. In the five-sensor case, the optimal sensor placement would be at junctions J11, J45, J83, J100, and J117 and the reliability reaches a level of 81.74%. However, the remaining three measures Z_1 , Z_2 , and Z_3 become 927.64 min, 341.37 individuals (ind), and 13,117.22 gal., respectively (compare these results with the results shown in Table 4 for the five-sensor design). This comparison would indicate a worsening condition for these measures. This case can be tested because any objective function can be used in the algorithm the writers have developed and in this verification example we have only used an objective function that maximizes the reliability. We will not go into the details of the use of alternative objective functions since it is not the theme of the study presented here. Given the objective function identified in Eq. (8b), the value of the objective function monotonically decreases as the number of sensors increase however the values of individual measures may fluctuate and that is expected (Fig. 2).

In order to measure the effectiveness of the water sensor network designed, the software provided in BWSN is used to generate the intrusion scenarios. Twenty contamination scenarios for each junction within the first 24 h are randomly generated. The four measures (Z_1 – Z_4) are calculated and the outcome is given in

Table 5. Performance of the Optimal Solutions for Verification Scenarios for Network 1

Number of sensors	Z_1 (min)	Z_2 (ind)	Z_3 (gal.)	Z_4 (%)
5	632.77	158.00	2758.23	66.32
6	515.86	188.29	2461.44	68.64
7	474.92	175.05	2112.90	71.74
8	531.77	132.11	1814.91	76.36
9	487.15	125.37	1556.04	76.36
10	482.30	122.50	1465.37	78.68
11	385.75	115.61	1286.16	78.68
12	378.14	110.52	1161.94	78.68
13	427.71	105.61	1187.25	82.56
14	549.90	109.42	1121.36	83.26
15	513.15	102.15	1071.49	83.80
16	504.10	103.88	947.58	83.80
17	484.63	82.94	687.29	83.68
18	350.30	82.86	792.27	84.65
19	366.49	76.67	550.11	85.31
20	330.42	78.00	525.00	85.31

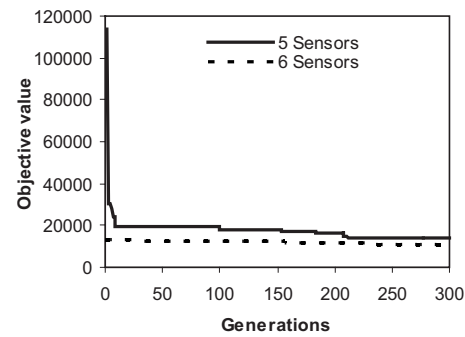


Fig. 3. Iteration sequence of the PGA

Table 5. These results are also shown in Fig. 1 as the “verification” analysis. The reliability is almost identical for both optimization and verification scenarios. The other three measures have similar trends although there are differences between “design” stage results and the verification stage results. During optimization, the demands and concentrations are intensively used for the evaluation of the objective function. For this purpose, in the design stage, the hourly average data were stored and used in the optimization algorithm. In the verification phase, the data specific to each time step is used (i.e., 30 min for demands and 5 min for concentration). This is the main reason that the measures for the expected contaminated water volume (Z_3) and population affected (Z_2) in the verification simulations are smaller than those in the design simulations. This also indicates of the robustness of the proposed algorithm since a more crude data are used in the design phase as opposed to the verification phase.

For the five-sensor case, the objective function value is 13,656.12 and decreases to 3,916.70 for the 10-sensor case (Fig. 2). The rate of decrease tends to reduce as the number of sensors increase. The objective function value for the 20-sensor case becomes 723.96. Fig. 3 shows that the PGA is convergent since the objective function value for the best solution in each subdomain generation decreases monotonically. At the start, for the first several generations, the objective function value shows a significant improvement. Then the convergence rate tends to decrease. Although the results presented cannot be claimed to be the best solutions, they are indeed very good solutions based on the performance measure values obtained and the comparisons made with the other solutions submitted to the BWSN session, as discussed below.

When the number of sensors used in a design is treated as a variable, the reliability measure can be used to determine the number of sensors required for a preset reliability level. For a preset reliability level of $r_c=85\%$ and an allowable error of 0.5%, the iteration process described for constraint Eq. (13) works as the outcome shown in Table 6 when we begin with the five-sensor case. When the number of sensors is initially set as 5, ΔM is set as 1 using Eq. (16). Thus, the number of sensors in the next solution is set as 6. In the first iteration, ΔM increases to 9 using Eq. (17). This leads to an M value of 15 for the second iteration. In the third iteration, the reliability is about the same as the one in the second iteration, thus ΔM takes on the value 1. The reliability in the fourth iteration is slightly less than the third iteration but this value is still less than the preselected reliability level to be achieved, thus ΔM is determined by Eq. (16) and is set as 1. In this process, as demonstrated for this case, five iterations are required to achieve the reliability level of $\sim 85\%$. The final results indicate that at least 18 sensors are needed in the system to satisfy

Table 6. Iteration Process for the Reliability Constraint

Iteration	Sensors	Reliability (r)	$\Delta r = r_c - r$	$\frac{\Delta r \times (M^{(l)} - M^{(l-1)})}{r^{(l)} - r^{(l-1)}}$	ΔM
0	5	66.51	18.49		1
1	6	68.33	16.67	8.74	9
2	15	84.03	0.97	0.56	1
3	16	84.03	0.97	NA	1
4	17	83.88	1.12	-7.25	1
5	18	84.85	0.15		0

Note: NA=not available.

this preselected reliability level. Obviously, the computation demonstrated above is done algorithmically within the model.

Water Distribution System 2. For Network 2, the water distribution system is complicated as it is characterized by 12,523 junctions, 14,822 pipes, two reservoirs, two tanks, four pumps, and eight valves (Table 1). The simulation time is 2 days and system operations are characterized by five distinct demand patterns. In this case the time interval used for hydraulic and water quality simulations are selected as 60 and 5 min, respectively. The optimization calculation for this case requires a large amount of computer memory and computational time. For Network 2, we have used a total of 300 randomly generated scenarios for the junctions with largest demands during the design phase as opposed to 2,580 scenarios used in Network 1. This may, of course, limit the representative nature of the scenarios selected for the complete domain but the sensor design obtained from these limited scenario cases compete rather well with the other designs submitted to the BWSN problem and this is again an indication of the robustness of the method suggested here. The two cases with five and 20 sensors each will be explored for this application. In order to test the optimal solution, a total of 1,000 verification scenarios are randomly generated. The performance measures obtained through the application of the proposed algorithm for Network 2 are given in Table 7. For the case with five sensors, the optimal sensor placement is given in Table 10 and the corresponding objective function value is 553,200. For the case with 20 sensors, the optimal sensor placement is given in Table 11 and the corresponding objective function value is 73,026. Obviously, an increase in the number of sensors improves the performance of the water sensor network. The comparison of results for the design and verification cases indicates that the reliability is almost identical for both cases but there are some differences in the results for the other three measures, again, the verification results indicating better performance as was observed in Network 1. Based on the objective function used in this study, a significant change and improvement is observed in the value of Z_4 . This may imply that this parameter is a sensitive parameter in the design. For large-scale water distribution systems such as the one reported here as Network 2, the use of five or 20 water sensors are

not enough to guarantee the reliability of the system at high percentages. It is expected that the performance of the network will improve significantly as the number of water sensors used further increases.

Comparison of Results Submitted to the BWSN Exercise

The solutions submitted to the BWSN (2006) included the application of various models and algorithms for the optimal sensor placement problem discussed above. At the conference, most writers identified the best sensor locations using their preferred algorithms. Some of these solutions focused on optimizing all four measures, others selected some or, in some cases, only one of the four measures as the criterion for optimization. It is possible to compare these solutions based on the four criteria identified in this paper which was the primary criteria to satisfy for the BWSN challenge. The comparison provided below should not be interpreted as an effort to identify which solution is the best or optimal because nobody knows what the true optimal solution is for this problem and no solution is perfect for all measures. Furthermore, as stated above, some of the solutions presented to the BWSN session studied restricted cases of the general problem investigated in this paper, thus evaluating them, considering all four measures, may not yield a fair comparison. However, the challenge was not to consider only a few of the measures but instead the challenge was to consider all four measures in the analysis in this exercise. On the other hand, evaluating these submissions using a methodology such as Pareto analysis will not be also appropriate since in Pareto method the domination analysis is done by giving importance to one measure against the others. This evaluation method would favorably evaluate the designs that are based on one or two measures as would be selected in the Pareto analysis evaluation.

Keeping these issues in mind, in order to evaluate the performance of the water sensor networks submitted to BWSN, we attempted to identify a unique criterion to test the performance of each design. The available data at hand was the sensor locations

Table 7. Comparison for Performance of the Optimal Solutions for Network 2

Case		Z_1 (min)	Z_2 (ind)	Z_3 (gal.)	Z_4 (%)
Sensors	Scenarios				
5	Design	163.83	1,631.86	22,361.66	22.33
	Verification	787.10	1,753.00	119,641.00	23.43
20	Design	126.88	1,364.18	7,089.26	32.00
	Verification	635.10	967.30	43,226.00	31.53

Table 8. Performance of Optimal for Network 1 with Five Sensors

Reference	Junction ID of sensors	Z_1 (min)	Z_2 (ind)	Z_3 (gal.)	Z_4 (%)
Guan et al. (2006a,b)	17, 31, 81, 98, 102	632.8	158.0	2,578.0	66.32
Dorini et al. (2006)	10, 31, 45, 83, 118	1,067.1	259.8	8,292.9	80.17
Berry et al. (2006)	17, 21, 68, 79, 122	533.9	140.4	2,577.1	61.07
Huang et al. (2006)	68, 81, 82, 97, 118	535.3	281.5	4,593.4	67.67
Krause et al. (2006)	18, 32, 46, 84, 121	836.4	257.7	5,535.3	62.48
Eliades and Polycarpou (2006)	17, 83, 101, 123, 126	762.0	309.6	6,933.0	77.35
Ghimire and Barkdoll (2006a)	126, 30, 118, 102, 34	429.9	359.1	4,368.3	37.04
Ghimire and Barkdoll (2006b)	126, 30, 102, 118, 58	424.6	333.8	4,103.9	40.12
Wu and Walski (2006)	45, 68, 83, 100, 118	699.6	304.9	8,708.1	78.76
Ostfeld and Salomons (2006)	100, 117, 68, 83, 45	705.1	263.8	8,502.8	77.98
Propato and Piller (2006)	17, 22, 68, 83, 123	704.0	163.8	3,278.9	72.70
Gueli (2006)	122, 118, 109, 100, 84	768.0	398.8	9,266.7	72.70
Preis and Ostfeld (2006)	12, 81, 85, 101, 116	739.3	325.1	10,162.8	66.12
Trachtman (2006)	NA				

Note: NA=not available.

submitted to the BWSN session by different writers, that is, the five- and 20-sensor locations for Network 1 and Network 2 and we chose to study the Case-A scenario of the BWSN challenge. Using these sensor locations and using random attack scenarios, it is possible to calculate the value of the four measures for each of the cases submitted. In our evaluation, the number of random attack scenarios we selected is 6,000 for both networks. In our evaluation we have used the same random attack scenarios for all cases and performed our analysis for all results submitted. The results for the performance values for the four measures are given in Tables 8–11 for all cases submitted to the BWSN session.

During the BWSN session (BWSN 2006) it was mentioned that, as an evaluation criteria, the scenarios that were not detected by a solution were excluded from the calculation of the expected time of detection, the population affected, and the contaminated water volume calculations. This concept was not well received

during the session. In our opinion, if one ignores the nondetect scenarios, this would lower the values of measures Z_1 , Z_2 , and Z_3 , artificially—the desired outcome. It also reduces the value of the measure Z_4 , which is not desired. This may lead to an overall evaluation criterion which would be very difficult to resolve. A design which has artificially low measures of the first three measures and also a relatively low reliability measure may be ranked as a good solution. This is not the correct way to interpret the performance (Aral et al. 2008). Again, in our opinion, to resolve this issue, one can choose to do the following. For the scenarios that are not detected, the time of detection can be set to be the end of the simulation period. The population and contaminated water volume can then be calculated from the initiation of injection time to the end of the simulation duration. This would penalize the nondetect scenarios equally in all competitive entries for all measures. This approach does not imply that at the end of simulation time the detection has occurred, thus for these cases, Z_4 measure

Table 9. Performance of Optimal for Network 1 with 20 Sensors

Reference	Junction ID of sensors	Z_1 (min)	Z_2 (ind)	Z_3 (gal.)	Z_4 (%)
Guan et al. (2006a,b)	4, 11, 17, 21, 27, 31, 34, 35, 46, 68, 75, 79, 82, 83, 98, 100, 102, 118, 122, 126	330.42	78	525	85.31
Dorini et al. (2006)	0, 10, 14, 17, 31, 34, 39, 45, 49, 68, 74, 82, 83, 90, 100, 102, 114, 122, 124, 128	404.5	71.8	656.3	85.69
Berry et al. (2006)	3, 4, 17, 21, 25, 31, 34, 37, 46, 64, 68, 81, 82, 90, 98, 102, 116, 118, 122, 126	281.1	68.2	426.1	77.14
Huang et al. (2006)	8, 11, 42, 46, 52, 68, 70, 75, 76, 82, 83, 95, 97, 99, 100, 109, 111, 117, 123, 126	369.4	147.2	1,801.6	84.96
Krause et al. (2006)	18, 84, 121, 32, 46, 101, 12, 125, 69, 91, 22, 36, 35, 117, 122, 113, 123, 77, 11, 20	413.7	120.4	1,828.8	78.62
Eliades and Polycarpou (2006)	10, 11, 14, 17, 19, 21, 31, 34, 35, 45, 68, 74, 83, 90, 100, 102, 114, 123, 124, 126	368.5	106.4	1,016.7	90.12
Ghimire and Barkdoll (2006a)	126, 30, 118, 102, 34, 17, 58, 68, 93, 27, 42, 82, 45, 35, 83, 89, 99, 70, 18, 32	374.4	105.3	773.9	79.33
Ghimire and Barkdoll (2006b)	126, 30, 102, 118, 58, 68, 17, 93, 82, 34, 99, 98, 89, 83, 100, 96, 70, 27, 32, 35	367.4	106.7	822.1	77.13
Wu and Walski (2006)	10, 12, 19, 21, 34, 35, 40, 45, 68, 75, 80, 83, 98, 100, 102, 114, 118, 123, 124, 126	63.7	142.0	1,167.2	90.12
Ostfeld and Salomons (2006)	11, 100, 118, 99, 17, 46, 81, 37, 124, 82, 26, 122, 5, 31, 68, 83, 19, 86, 109	434.5	93.2	1,127.5	84.90
Propato and Piller (2006)	11, 17, 34, 37, 38, 45, 49, 68, 76, 83, 90, 100, 102, 106, 114, 118, 123, 124, 125, 126	427.0	106.0	947.4	88.03
Gueli (2006)	112, 1, 103, 24, 21, 102, 35, 19, 116, 85, 61, 73, 114, 31, 7, 8, 64, 28, 93, 124	222.7	87.7	1,195.8	57.80
Preis and Ostfeld (2006)	NA				
Trachtman (2006)	NA				

Note: NA=not available.

Table 10. Performance of Optimal for Network 2 with Five Sensors

Reference	Junction ID of sensors	Z_1 (min)	Z_2 (ind)	Z_3 (gal.)	Z_4 (%)
Guan et al. (2006a,b)	321, 3770, 4084, 4939, 7762	787.1	1753.0	119,641	23.43
Dorini et al. (2006)	636, 3585, 4684, 9364, 10387	1,275.5	2,440.0	227,262	30.60
Berry et al. (2006)	3357, 4684, 10874, 11184, 11304	758.1	1,515.6	95,255	26.60
Huang et al. (2006)	3355, 5088, 5430, 9005, 9550	944.2	2,481.6	220,293	23.33
Krause et al. (2006)	10875, 4685, 11305, 3358, 11185	910.4	2,045.6	177,301	24.60
Eliades and Polycarpou (2006)	532, 1486, 3231, 4359, 4609	1,552.5	2,936.6	334,004	32.10
Ghimire and Barkdoll (2006a)	9271, 1486, 4482, 5585, 4609	1,243.0	2,892.4	333,001	10.70
Ghimire and Barkdoll (2006b)	9271, 1486, 4482, 5585, 4609	1,243.0	2,892.4	333,001	10.70
Wu and Walski (2006)	3709, 4957, 6583, 8357, 9364	1,191.3	2,663.0	262,817	32.33
Ostfeld and Salomons (2006)	1958, 4845, 3713, 2447, 7906	1,491.7	3,238.4	349,889	29.13
Propato and Piller (2006)	NA				
Gueli (2006)	NA				
Preis and Ostfeld (2006)	NA				
Trachtman (2006)	NA				

Note: NA=not available.

is still recorded as a nondetected contamination event lowering the reliability values as it should. This idea stems from the fact that the first three measures need to reflect the nondetect cases appropriately as well, in some form, rather than artificially lowering their values by not including these scenarios in the calculation of Z_1 , Z_2 , and Z_3 . That is why the values of the four measures we are reporting here are different than the values reported during the BWSN session in Cincinnati in 2006. During the design phase analysis of this study, this criterion was used in the calculation of the four measures and the comparative analysis given below is also based on the same criteria.

The next phase of the comparative analysis is the normalization of the results given in Tables 8–11. For this purpose two approaches can be used to normalize the outcome to a range [0, 1]. The first option is to identify the maximum values for Z_1 , Z_2 , Z_3 , and Z_4 from solutions submitted ($Z_{1\max}$, $Z_{2\max}$, $Z_{3\max}$, $Z_{4\max}$) and calculate the normalized values W_i using Eq. (18)

$$W_i = 1 - Z_i/Z_{i\max} \quad i = 1, 2, 3$$

Table 11. Performance of Optimal for Network 2 with 20 Sensors

Reference	Junction ID of sensors	Z_1 (min)	Z_2 (ind)	Z_3 (gal.)	Z_4 (%)
Guan et al. (2006a,b)	174, 311, 1486, 1905, 2589, 2991, 3548, 3757, 3864, 4184, 4238, 5091, 6995, 7145, 7689, 8826, 9308, 9787, 10614, 12086	635.1	967.3	43,226	31.53
Dorini et al. (2006)	647, 928, 1478, 1872, 2223, 2848, 3573, 4650, 5076, 5366, 6835, 7422, 8336, 8402, 9204, 9364, 10874, 11271, 11528, 12377	911.4	1,331.2	90,748	41.33
Berry et al. (2006)	636, 1917, 3357, 3573, 3770, 4132, 4240, 4594, 5114, 6583, 6700, 7652, 8999, 9142, 9722, 10614, 10874, 11177, 11271, 12258	515.2	539.7	16,459	37.17
Huang et al. (2006)	73, 108, 1028, 1112, 1437, 2526, 3180, 4036, 4648, 5363, 5826, 5879, 6581, 8439, 8580, 8841, 9363, 9616, 10216, 10385	793.8	1,247.1	76,305	35.07
Krause et al. (2006)	10875, 4685, 11305, 3358, 11185, 1479, 9143, 1905, 4033, 9365, 4241, 4133, 3636, 2580, 3837, 6701, 9000, 3748, 8835, 3230	734.6	959.4	45,939	34.73
Eliades and Polycarpou (2006)	375, 532, 579, 1426, 1486, 3231, 3679, 3780, 4234, 4359, 4511, 4609, 5087, 5585, 6922, 7670, 7858, 8629, 9360, 9787	1,152.7	1,800.3	153,082	42.20
Ghimire and Barkdoll (2006a)	9271, 1486, 4482, 5585, 4609, 4359, 9787, 532, 5953, 12341, 4808, 4662, 4638, 3864, 1667, 3806, 1590, 7858, 9303, 12220	1,029.1	1,974.4	197,376	30.97
Ghimire and Barkdoll (2006b)	9271, 1486, 4482, 5585, 4609, 4359, 9787, 532, 5953, 12341, 4808, 4662, 4638, 3864, 1667, 3806, 1590, 7858, 9303, 12220	1,029.1	1,974.4	197,376	30.97
Wu and Walski (2006)	871, 1334, 2589, 3115, 3640, 3719, 4247, 4990, 5630, 6733, 7442, 7714, 8387, 8394, 9778, 10290, 10522, 10680, 11151, 11519	837.4	1,374.1	78,829	42.80
Ostfeld and Salomons (2006)	8256, 9426, 8875, 8804, 7547, 8534, 134, 11246, 8243, 11904, 8217, 8102, 5053, 3494, 6993, 5222, 5905, 3077, 8376, 3995	556.5	690.4	26,254	11.80
Propato and Piller (2006)	NA				
Gueli (2006)	NA				
Preis and Ostfeld (2006)	NA				
Trachtman (2006)	NA				

Note: NA=not available.

Table 12. Performance Calculated Using Eqs. (18) and (20)

Reference	N1A5	N1A20	N2A5	N2A20
Guan et al. (2006a,b)	0.65	0.60	0.61	0.60
Dorini et al. (2006)	0.38	0.54	0.47	0.49
Berry et al. (2006)	0.66	0.63	0.67	0.67
Huang et al. (2006)	0.55	0.28	0.43	0.40
Krause et al. (2006)	0.45	0.28	0.46	0.46
Eliades and Polycarpou (2006)	0.45	0.47	0.39	0.37
Ghimire and Barkdoll (2006a)	0.43	0.47	0.33	0.31
Ghimire and Barkdoll (2006b)	0.47	0.46	0.34	0.31
Wu and Walski (2006)	0.43	0.39	0.45	0.46
Ostfeld and Salomons (2006)	0.45	0.42	0.42	0.41
Propato and Piller (2006)	0.63	0.44	NA	NA
Gueli (2006)	0.32	0.47	NA	NA
Preis and Ostfeld (2006)	0.33	NA	NA	NA
Trachtman (2006)	NA	NA	NA	NA

Note: N1A5=Network 1, BWSN Case A, five sensors; N1A20=Network 1, BWSN Case A, 20 sensors; NA5=Network 1, BWSN Case A, five sensors; N2A20=Network 2, BWSN Case A, 20 sensors; and NA=not available.

$$W_4 = Z_4 / Z_{4 \max} \quad (18)$$

or one can also identify the maximum and minimum values for Z_1 , Z_2 , Z_3 , and Z_4 from solutions submitted: ($Z_{i \max}$, and $Z_{i \min}$; for $i=1,2,3,4$) and calculate W_i using Eq. (19)

$$W_i = \frac{(Z_{i \max} - Z_i)}{(Z_{i \max} - Z_{i \min})} \quad i = 1, 2, 3$$

$$W_4 = \frac{(Z_4 - Z_{4 \min})}{(Z_{4 \max} - Z_{4 \min})} \quad (19)$$

In Eqs. (18) and (19), the larger the values of Z_1 , Z_2 , and Z_3 are, the smaller W_1 , W_2 , and W_3 will be, and the larger the value of Z_4 is, the larger W_4 will be. This trend follows the interpretation of these four measures as they are used in the optimization criteria.

Estimation of the overall performance of each case for each solution submitted maybe calculated by

$$\text{Score} = (W_1 + W_2 + W_3 + W_4) / 4 \quad (20)$$

or

$$\text{Score} = [(W_1 + W_2 + W_3) / 3 + W_4] / 2 \quad (21)$$

The score obtained from Eq. (20) weighs the four measures equally, and the score obtained from Eq. (21) gives a larger weight to the reliability measure. This score is also included here since there was a discussion at the BWSN session (BWSN 2006) on the point of view of reliability being the more important measure. The overall performance of the results submitted to the BWSN session with the use of two different normalization methods and two different averaging methods are given in Tables 12–15. Combining two different normalization procedures with the two averaging processes yields four different scores in the estimation of the overall performance of the sensor placement solutions submitted. More important, these evaluation scores are linear and not biased based on a domination ranking process such as the Pareto front analysis, as was presented during the BWSN session in Cincinnati in 2006.

Table 13. Performance Calculated Using Eqs. (18) and (21)

Reference	N1A5	N1A20	N2A5	N2A20
Guan et al. (2006a,b)	0.71	0.71	0.62	0.60
Dorini et al. (2006)	0.59	0.68	0.50	0.49
Berry et al. (2006)	0.70	0.70	0.71	0.67
Huang et al. (2006)	0.65	0.50	0.46	0.41
Krause et al. (2006)	0.56	0.47	0.53	0.46
Eliades and Polycarpou (2006)	0.62	0.65	0.38	0.38
Ghimire and Barkdoll (2006a)	0.44	0.61	0.31	0.32
Ghimire and Barkdoll (2006b)	0.48	0.59	0.31	0.32
Wu and Walski (2006)	0.61	0.59	0.50	0.46
Ostfeld and Salomons (2006)	0.63	0.59	0.48	0.42
Propato and Piller (2006)	0.72	0.62	NA	NA
Gueli (2006)	0.51	0.53	NA	NA
Preis and Ostfeld (2006)	0.49	NA	NA	NA
Trachtman (2006)	NA	NA	NA	NA

Note: N1A5=Network 1, BWSN Case A, five sensors; N1A20=Network 1, BWSN Case A, 20 sensors; N2A5=Network 1, BWSN Case A, five sensors; N2A20=Network 2, BWSN Case A, 20 sensors; and NA=not available.

Discussion of Results

In this study, we have proposed an optimization model and a solution process for the design of water sensor networks which provide an effective procedure for the analysis of large-scale water distribution systems. Although we have used a single-objective function in this model, the proposed objective function reflects the requirements of the four criteria identified in the BWSN challenge and may achieve the purpose of a multiobjective function with a reduced computational effort. This objective function is not linked with the specific solution methodology suggested in this study, thus it can be embedded in other solution methods to achieve the same effect or even can be used in Pareto analysis. The proposed model also introduces the reliability constraint concept to determine the optimal number of sensors necessary to satisfy a prespecified reliability level of detection for the system. The proposed optimization model is based on a $\{0,1\}$

Table 14. Performance Calculated Using Eqs. (19) and (20)

Reference	N1A5	N1A20	N2A5	N2A20
Guan et al. (2006a,b)	0.82	0.79	0.80	0.79
Dorini et al. (2006)	0.45	0.70	0.57	0.60
Berry et al. (2006)	0.85	0.83	0.89	0.90
Huang et al. (2006)	0.68	0.29	0.51	0.47
Krause et al. (2006)	0.53	0.27	0.56	0.56
Eliades and Polycarpou (2006)	0.55	0.60	0.44	0.41
Ghimire and Barkdoll (2006a)	0.48	0.56	0.35	0.32
Ghimire and Barkdoll (2006b)	0.53	0.54	0.36	0.32
Wu and Walski (2006)	0.52	0.47	0.55	0.56
Ostfeld and Salomons (2006)	0.56	0.51	0.49	0.47
Propato and Piller (2006)	0.80	0.53	NA	NA
Gueli (2006)	0.35	0.55	NA	NA
Preis and Ostfeld (2006)	0.37	NA	NA	NA
Trachtman (2006)	NA	NA	NA	NA

Note: N1A5=Network 1, BWSN Case A, five sensors; N1A20=Network 1, BWSN Case A, 20 sensors; N2A5=Network 1, BWSN Case A, five sensors; N2A20=Network 2, BWSN Case A, 20 sensors; and NA=not available.

Table 15. Performance Calculated Using Eqs. (19) and (21)

Reference	N1A5	N1A20	N2A5	N2A20
Guan et al. (2006a,b)	0.77	0.81	0.79	0.79
Dorini et al. (2006)	0.63	0.75	0.58	0.59
Berry et al. (2006)	0.75	0.75	0.92	0.90
Huang et al. (2006)	0.69	0.47	0.53	0.48
Krause et al. (2006)	0.55	0.39	0.63	0.56
Eliades and Polycarpou (2006)	0.67	0.73	0.40	0.42
Ghimire and Barkdoll (2006a)	0.32	0.59	0.30	0.33
Ghimire and Barkdoll (2006b)	0.38	0.57	0.31	0.33
Wu and Walski (2006)	0.67	0.65	0.59	0.55
Ostfeld and Salomons (2006)	0.69	0.62	0.56	0.48
Propato and Piller (2006)	0.81	0.67	NA	NA
Gueli (2006)	0.51	0.38	NA	NA
Preis and Ostfeld (2006)	0.47	NA	NA	NA
Trachtman (2006)	NA	NA	NA	NA

Note: N1A5=Network 1, BWSN Case A, five sensors; N1A20=Network 1, BWSN Case A, 20 sensors; N2A5=Network 2, BWSN Case A, five sensors; N2A20=Network 2, BWSN Case A, 20 sensors; and NA=not available.

integer programming approach for which the optimal solutions are difficult to achieve for problems with large number of variables. To overcome this difficulty, PGA is used to solve the optimization problem (Guan and Aral 1999; Park and Aral 2004; Aral et al. 2001). Based on a conventional genetic algorithm, the PGA works on a subdomain of junctions and the final optimal solution is obtained through progressive evolution of these subdomains. This approach improves the efficiency of the genetic algorithm solution and reduces the computational burden. During the optimization process, the EPANET algorithm is used as the simulator to calculate the water supply and concentration distribution in the network for different contaminant intrusion events. How one uses the EPANET software in the application also affects the efficiency of the solution of the optimization problem. In the application discussed here, there are two approaches that can be used during this stage. In the first approach, EPANET can be used before optimization to obtain all the data necessary for the evaluation of the objective function. These data can be stored in the computer memory and directly used by the optimization algorithm. In this manner, one can reduce the computation time. The disadvantage of this approach is that it requires a large amount of computer memory. This memory, however, seems to be readily available within the current computational platforms. In the second approach, EPANET simulations can be conducted during the optimization procedure. This approach requires the repeated application of the EPANET model for all scenarios for each solution during optimization, resulting in lengthy computational times but less computer memory storage requirement. In this study, the first approach is used. Whichever method is used, reducing the required computational time and computer memory is one of important issues for the optimal design of a water sensor networks for large water distribution systems. The subdomain concept is used effectively to solve this problem in this study.

The design of water sensor placement in a water distribution system may be formulated as a multiobjective optimization model that includes the four measures described above (Z_1 - Z_4). In this case the model can then be solved by some trade-off method or by Pareto analysis. In either case, either the trade-off coefficients should be determined a priori or in the case of Pareto analysis,

since it would yield a number of feasible optimal solutions, the selection of the final design would involve the trade-off preference of the decision maker. These are important and arguable issues one has to answer when a single solution is required for a multiobjective problem. In the model proposed here, the four design measures are incorporated into a single-objective function. The trade-off coefficients inherently imbedded in the objective function are defined in terms of the four measures. Thus, they are optimized during the solution process. The solutions obtained by the model indeed reflect the interactive effect of the four measures without emphasizing one over the other.

Evaluation of the performance of water sensor placement design using the four design measures is also another important issue. In this paper, a simple normalization process is chosen to directly compare the outcome of the solutions submitted to the BWSN session (BWSN 2006) for all four measures. This comparison is based on the premise that all four measures are equally important when Eq. (20) is used or nearly equally important when Eq. (21) is used. As can be seen from the comparative results presented in the tables above, there is a clear trend in the higher ranked solutions for each network for different number of sensor placement cases. The solutions presented by Berry et al. (2006) and Guan et al. (2006a,b) (the solution discussed in this paper) is consistently ranking as the higher ranked solutions. The solution submitted by Berry et al. (2006) is consistently ranking high (12 first place rankings, 2 second place rankings and 2 third place rankings out of sixteen scores). The solution by Guan et al. (2006a,b) always ranks second or first for all evaluation cases (14 second-place ranking and two first-place ranking out of 16 scores), with the solution submitted by Berry et al. (2006) ranking higher more frequently. For certain cases and this occurs for Network 1 the Berry et al. (2006) solution drops to third rank. The solution submitted by Propato and Piller (2006) and Dorini et al. (2006) is also showing good performance for some cases although Propato and Piller solution was not provided for the second network and was evaluated only for the first network. The solutions submitted by Huang et al. (2006), Krause et al. (2006), Eliades and Polycarpou (2006), and Ostfeld and Salomons (2006) mostly ranked third or fourth. This ranking was achieved only for some cases while for other cases their ranking is lower. The other solutions are mostly not close to the top rankings for all cases as they are evaluated here. As stated earlier, probably the reason behind this is that some of these solutions did not consider all four measures in the optimization model equally or ignored the nondetect cases and this affected the overall score of these solutions since these solutions miss near-optimal values for the remaining measures, as was discussed earlier. This emphasizes the importance of using a solution method that balances the optimization of all measures considered in the problem rather than one or a few of the measures, since all four measures are equally important in the design.

Conclusions

The two water distribution networks of the BWSN are used to demonstrate the performance of the model and the solution algorithm proposed. Based on the computational results, the following conclusions can be drawn:

- The proposed optimization model can effectively solve the design of water sensor network in the water distribution system. The objective function used in the model considers the effect of the four measures in one objective function. An important

characteristic of the model is that the imbedded trade-off coefficients in the objective function are defined in terms of the four measures of the problem and they are optimized as the solution progresses.

- The proposed model also introduces the reliability constraint solution to determine the minimum number of sensors necessary to achieve a certain reliability constraint. Using this model, one can obtain the minimum number of sensors needed and their placement while satisfying a specified reliability for the system.
- For a large-scale water distribution network, the computer memory and computational time requirements remain to be the main obstacle. The algorithm proposed in this study tends to overcome these obstacles through the application of the PGA based subdomain approach.
- The water sensor networks designed by the model yielded good performance in comparison with the other solutions provided to the BWSN session.

Disclaimer

The findings and conclusions in this article have not been formally disseminated by the Agency for Toxic Substances and Disease Registry and should not be construed to represent any agency determination or policy.

References

- Aral, M. M., Guan, J., and Maslia, M. L. (2008). "A multi-objective optimization algorithm for the solution of water sensor placement problem in water distribution systems." *Proc., World Water and Environmental Resources Congress*, ASCE/EWRI, Cincinnati.
- Aral, M. M., Guan, J. B., and Maslia, M. L. (2001). "Identification of contaminant source location and release history in aquifers." *J. Hydrol. Eng.*, 6(3), 225–234.
- Bahadur, R., Pickus, J., Amstutz, D., and Samuels, W. (2001). "Locating optimum water quality monitoring stations in water distribution system." *Proc., World Water and Environmental Resources Congress* (CD-ROM), ASCE, Reston, Va.
- Bahadur, R., Samuels, W. B., Grayman, W., Amstutz, D., and Pickus, J. (2003). "PipelineNet: A model for monitoring introduced contaminants in a distribution system." *Proc., World Water & Environmental Resources Congress 2003 and Related Symposia* (CD-ROM), ASCE, Reston, Va.
- Berry, J., Hart, W., Phillips, C., and Watson, J.-P. (2006). "A facility location approach to sensor placement optimization." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Berry, J. W., Fleischer, L., Hart, W. E., Phillips, C. A., and Watson, J. P. (2005). "Sensor placement in municipal water networks." *J. Water Resour. Plann. Manage.*, 131(3), 237–243.
- BWSN. (2006). "Battle of the water sensor networks." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Congressional Research Service. (2002). "Terrorism and security issues facing the water infrastructure sector." *The Library of Congress*, Congressional Publications, Washington, D.C.
- Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S., and Savic, D. (2006). "An efficient algorithm for sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Eliades, D., and Polycarpou, M. (2006). "Iterative deepening of pareto solutions in water sensor networks." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Ghimire, S., and Barkdoll, B. (2006a). "A heuristic method for water quality sensor location in a municipal water distribution system: Mass-release based approach." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Ghimire, S., and Barkdoll, B. (2006b). "Heuristic method for the battle of the water network sensors: Demand-based approach." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Guan, J., and Aral, M. M. (2005). "Remediation system design with multiple uncertain parameters using fuzzy sets and genetic algorithm." *J. Hydrol. Eng.*, 10(5), 386–394.
- Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. (2006a). "Optimization model and algorithms for design of water sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. M. (2006b). "Identification of contaminant sources in water-distribution systems using simulation-optimization method: A case study." *J. Water Resour. Plann. Manage.*, 132(4), 252–262.
- Guan, J. B., and Aral, M. M. (1999). "Progressive genetic algorithm for solution of optimization problems with nonlinear equality and inequality constraints." *Appl. Math. Model.*, 23(4), 329–343.
- Gueli, R. (2006). "Predator-prey model for discrete sensor placement." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- House Science Committee. (2001). "Protecting water supply systems from terrorists; testimony of J. J. Danneels." November 15, 2001.
- Huang, J. E., McBean, E., and James, W. (2006). "Multi-objective optimization for monitoring sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting accidental contaminations in municipal water networks." *J. Water Resour. Plann. Manage.*, 124(4), 192–198.
- Krause, A., et al. (2006). "Optimizing sensor placements in water distribution systems using submodular function maximization." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Kumar, A., Kansal, M. L., and Arora, G. (1999). "Discussion of 'Detecting accidental contaminations in municipal water networks.'" *J. Water Resour. Plann. Manage.*, 125(5), 308–309.
- Mays, L. W. (2004). *Water supply system security*, McGraw-Hill, New York.
- Ostfeld, A., and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *J. Water Resour. Plann. Manage.*, 130(5), 377–385.
- Ostfeld, A., and Salomons, E. (2006). "Sensor network design proposal for the battle of the water sensor networks." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Park, C. H., and Aral, M. M. (2004). "Multi-objective optimization of pumping rates and well placement in coastal aquifers." *J. Hydrol.*, 290(1–2), 80–99.
- Preis, A., and Ostfeld, A. (2006). "Multi-objective sensor design for water distribution systems security." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3* (CD-ROM), ASCE, Reston, Va.
- Propato, M. (2006). "Contamination warning in water networks: General mixed-integer linear models for sensor location design." *J. Water Resour. Plann. Manage.*, 132(5), 377–385.
- Propato, M., and Piller, O. (2006). "Battle of the water sensor networks."

- Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3 (CD-ROM), ASCE, Reston, Va.*
- Rossman, L. A. (2000). *EPANET 2 user's manual, National Risk Management Research Laboratory, U.S. EPA, Cincinnati.*
- Trachtman, G. (2006). "A 'Strawman' common sense approach for water quality sensor site selection." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3 (CD-ROM), ASCE, Reston, Va.*
- Watson, J.-P., Greenberg, H. J., and Hart, W. E. (2004). "A multiple-objective analysis of sensor placement optimization in water networks." *Proc., Water Distribution Systems Analysis Symp., ASCE/ EWRI, Cincinnati.*
- WDSA. (2006). "Water distribution systems analysis symposium." *Proc., 8th Annual Water Distribution Systems Analysis Symp. (CD-ROM), ASCE, Reston, Va.*
- Wu, Z., and Walski, T. (2006). "Multi-objective optimization of sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution Systems Analysis Symp., Session D-3 (CD-ROM), ASCE, Reston, Va.*