Contamination Source Identification in Water Distribution Systems Using an Adaptive Dynamic Optimization Procedure

Li Liu[1]; S. Ranji Ranjithan[2]; and G. Mahinthakumar[3]

Abstract: Contamination source identification involves the characterization of the contaminant source based on observations that stream from a set of sensors in a water distribution system (WDS). The streaming data can be processed adaptively to provide an estimate of the source characteristics at any time once the contamination event is detected. In this paper, an adaptive dynamic optimization technique (ADOPT) is proposed for providing a real-time response to a contamination event. A new multiple population–based search that uses an evolutionary algorithm (EA) is investigated. To address nonuniqueness in the initial stages of the search and prevent premature convergence of the EA to an incorrect solution, the multiple populations are designed to maintain a set of alternative solutions that represent various nonunique solutions. As more observations are added, the EA solutions not only migrate to better solution states but the number of solutions decreases as the degree of nonuniqueness diminishes. This new algorithm adaptively converges to the solutions that best match the available observations. The use of the developed method is demonstrated for two WDS networks. **DOI: 10.1061/(ASCE)WR.1943-5452.0000104.**

CE Database subject headings: Identification; Water distribution systems; Water pollution; Algorithms.

Author keywords: Identification; Water distribution systems; Contamination; Algorithms.

Introduction

Accidental and intentional contamination of a water distribution system (WDS) is becoming an increasingly critical issue. For example, a pollutant source introduced into a WDS will spread through the system rapidly and expose the public to health risks. Detection of the contamination in the distribution system using a sensor network could yield useful observations to identify and manage such contamination threat events. Based on these observations, the location, strength, time, and duration of the contaminant source needs to be determined to direct utility operation decision makers toward containing and mitigating the event. Given a set of concentration observations at sensors in the network, an inverse problem can be constructed to identify the contaminant source characteristics (including location, strength, and release history) by coupling a water distribution simulation model with an optimization method. Possible solutions to this inverse problem are determined by minimizing the error between predicted concentrations and actual observations at the sensor nodes in the network. In the context of a quickly evolving contamination event in a WDS, the correct source characterization must be resolved rapidly as the sensor observations, i.e., contaminant concentrations at the sensors in the network, stream in over time.

Although inverse modeling has been applied to a wide array of system identification problems in engineering, it has the potential for nonuniqueness in that different sources with significantly different pollutant release characteristics but with similar prediction errors may be identified. Because the nonuniqueness in a system is related to the amount of data available for identifying the source(s) of the contamination, more data, made available through either additional sensors or an extended monitoring period, may help reduce the degree of nonuniqueness in the system. If the available information is insufficient to determine whether an identified solution is unique, then it is important to determine whether other possible solutions exist. Knowing that the identified solution is the only possible source characteristic that matches the observations is critical because a nonunique but incorrect solution may yield potentially costly mitigation actions that may inconsequentially exacerbate the spread of the contamination.

The key challenges to solving this problem are the determination of the source characteristics given the available measurement information at any given time, and the assessment of whether the solution that is identified is unique. This determination and assessment require a procedure that is able to (1) adaptively search for the source characteristics as the observation data are dynamically updated over time and (2) assess the degree of nonuniqueness, i.e., whether more than one solution fits the available observations.

Recently, researchers have reported the development of several procedures to identify contaminant source characteristics by using information from sensor networks. A direct sequential technique, reported by van Bloemen Waanders et al. (2003), has been applied to solve a small-scale optimization problem using a standard successive quadratic programming tool. Laird et al. (2005) reported a direct simultaneous approach. Guan et al. (2006) exploited a simulation-optimization approach and demonstrated its applicability

[1]Assistant Professor, School of Civil Engineering, 230009 Hefei, Hefei Univ. of Technology, Anhui, China. E-mail: lliuncsu@gmail.com

[2]Professor, Dept. of Civil, Construction, and Environmental Engineering, North Carolina State Univ., Raleigh, NC 27695. E-mail: ranji@ncsu.edu

[3]Associate Professor, Dept. of Civil, Construction, and Environmental Engineering, North Carolina State Univ., Raleigh, NC 27695. E-mail: kumar@ncsu.edu

Note. This manuscript was submitted on November 19, 2009; approved on June 14, 2010; published online on June 25, 2010. Discussion period open until August 1, 2011; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 137, No. 2, March 1, 2011. ©ASCE, ISSN 0733-9496/2011/2-183–192/$25.00.

to nonlinear contaminant sources and release-history identification by incorporating the reduced gradient method. Preis and Ostfeld (2007) described an approach for determining contaminant source by coupling *EPANET* (Rossman 2000) with a genetic algorithm. Overall, these methods attempt to identify a single solution using a fixed, i.e., not dynamically streaming, set of observations. The contaminant source characterization problem can be formulated as an adaptive problem, as the system (e.g., water consumption and network operations) and the observed data are continually changing while contaminant event management decisions are being made. Therefore, new approaches are needed to solve the problem in an adaptive manner. At any given time, the procedure must be able to identify possible solutions that explain the observations available up to that time. Also, the solution procedure must identify not only the best estimate of the source characteristics that minimize the prediction error, but also identify a set of possible alternative solutions, if any, that similarly predicts the available observations. One method for solving such a problem under a dynamic environment is the multiple hypothesis tracking (MHT) approach developed by Reid (1979), which is able to incorporate dynamic observations and determine the probabilities of alternative data-to-target association hypothesis under the Bayesian sequential estimation framework. A number of researchers (e.g., Bar-Shalom and Fortmann 1988; Kurien 1990; Blackman and Popoli 1999) have demonstrated that the MHT approaches offer the potential to handle multitarget tracking problems by recursively filtering observations coming from sensors. The MHT approach is capable of tracking multiple targets in a dynamic and noisy environment, but it has the disadvantage of its computational complexity. This is particularly severe in most real-world tracking applications where the sensors yield unlabeled measurements of the targets, usually requiring the use of Monte Carlo techniques to estimate the posterior probability for each of the targets (Vermaak et al. 2005). For a WDS contaminant source characterization problem, the MHT approach especially suffers attributable to infinite potential contamination scenarios as well as the uncertainties inherent in the observed data and the state of the system. The resulting significant computational costs from numerous expensive simulation runs would unavoidably affect not only the identification time but also the resultant solution quality.

This paper reports a new adaptive search method that uses a simulation-optimization approach whereby the water distribution network model is coupled directly with a new dynamic optimization method to iteratively evaluate and identify solutions that minimize the prediction error. To assess the nonuniqueness in the solution, the procedure also incorporates a systematic method to identify a set of alternative solutions that are as different as possible in the solution space. Thus, at any stage of the solution procedure, possible solutions that best describe the observations are determined and are used as starting solutions for subsequent searches as more information becomes available. The search method explored in this paper is based on an EA that is coupled with an *EPANET* model of the water distribution network. The applicability of the method is illustrated by using hypothetical examples.

Contaminant Source Identification Problem Description in a WDS

To capture the dynamic nature of the available observation data, the source identification problem is described in terms of the source characteristics based on observations up to the current time. As the number of observations changes over time, the description of the problem is updated at some regular time interval, i.e., the observation frequency. At any instant, the problem can be solved to obtain an estimate of the source characteristics that best explain the currently available data. The following mathematical model is defined to determine, at any given time after the contamination is detected at one or more sensors, the contamination source location, the contamination event start time, and the corresponding contaminant mass loading history. Although the following definition assumes that the contamination is introduced at only one node in the network, the proposed approach can be updated to consider multiple contamination source locations

$$\text{Find } \{L, M_{t_c}, T_0\}$$

$$\text{Minimize } F = \sqrt{\frac{\sum_{t=t_0}^{t_c}\sum_{i=1}^{N_s}[C_{it}^{\text{obs}} - C_{it}(L, M_{t_c}, T_0)]^2}{N_s{}^* t_c}} \quad (1)$$

where F = prediction error; L = contaminant source location, T_0 = contamination event starting time, t_0 = time of first detection of contamination at sensors, t_c = current time step, M_{t_c} = contaminant mass loadings represented as a vector of mass injected at the source from time T_0 to t_c, $M_{t_c} = (m_{T_0}; m^{T_0+1;\ldots;m_{tc}})$, C_{it}^{obs} = observed concentration at sensor i at time step t; $C_{it}(L, M_{t_c}, T_0)$ = model estimated concentration at sensor i at time step t, i = observation (sensor) location; t = time step of observation, and N_s = number of sensors.

Solution Approach

The search for the location and the time history of the contamination injection into the network is a nonlinear programming problem that poses sufficient computational challenges depending on the size of the problem. A few search methods to solve the source determination problem have been reported recently. One approach that can be used to solve the inverse problem is a simulation-optimization, or indirect, approach, in which a search procedure is coupled with a simulation model. EAs (Holland 1975) are a class of heuristic methods that provide a global search mechanism to efficiently identify optimal or near-optimal solutions for large nonlinear optimization problems. EAs have been used in several water distribution network design problems (e.g., Dandy et al. 1996; Savic and Walters 1997). Although EAs can be used effectively to solve inverse problems, such as WDS calibration (e.g., Vitkosvsky et al. 2000; Lingireddy and Ormsbee 2002) and groundwater source contamination identification problems (e.g., Mahinthakumar and Sayeed 2005; Mahar and Datta 1997), the applicability of EAs to dynamic source characterization in water distribution networks has not been fully investigated. Thus, in this paper EAs are investigated as an approach to solve adaptively the source determination problem in water distribution networks that is posed as a dynamic optimization model [Eq. (1)]. This new EA-based approach, called adaptive dynamic optimization technique (ADOPT), is a search procedure that is designed for adaptive optimization, considers dynamically varying streams of sensor observations, and identifies alternative solutions, if any, to assess the degree of nonuniqueness in the system (Liu and Ranjithan 2010).

Evolution Strategy for Contaminant Source Characterization

In this paper, ADOPT is implemented using evolution strategies (ES) (Schwefel 1995) to continually search for optimal solutions while the observations are updated. Similar to other EAs, an ES-based research uses a population of individuals during the optimization process. Beginning with an initial population that typically is randomly generated, the ES explores search the space

J. Water Resour. Plann. Manage. 2011.137:183-192.

through mutation and the selection of individuals with higher fitness values to survive into the next generation. As individuals progress, their average performance is expected to advance gradually. The algorithm terminates when specified stopping criteria, such as the maximum number of generations, no improvement in the optimum, etc., are met.

The ES presents an adaptive capability, particularly in dynamic circumstances, in that it typically adapts its step lengths during the optimization process. The benefit of the ES is its mutative self-adaptation, whereby each individual can be represented as a decision variable along with its mutation step lengths (Yang et al. 2007). During the course of mutation, the step length, once mutated, is used to create a random vector to mutate its corresponding decision variable. Thus, mutation rate changes are based on the quality of individuals instead of using predetermined values for parameters, such as mutation and crossover rates. The ES has been demonstrated to possess a self-learning capability, even in the dynamic context of an optimization problem (Hoffmeister and Bäck 1992).

To identify a contaminant source in a WDS, a search for the key characteristics of contamination, such as location, starting time, duration, and injection rates at different time intervals is required. The injection profile is represented as an array of real variables; and the source location, starting time, and duration are encoded as integer values in this study. The mutation step size is encoded in a similar way to its corresponding decision variable. The fitness representing the prediction error [Eq. (1)] of an individual is updated with increasingly available sensor observations. Because the mutation operator plays a key role in maintaining diversity within the ES, several mutation strategies to enhance performance were investigated in this study.

Conceptual Basis for ADOPT

The two key features of the new method, ADOPT, are (1) optimization in dynamic environments, and (2) identification of alternative solutions. In the context of the water distribution network problem, the number of sensor observations varies with time. That is, the objective function [i.e., the prediction error defined in Eq. (1)] changes with time. The dynamic optimization procedure is structured to continually search for the best solution at each time step, t. Initially (i.e., at t_0 when the contaminant is first detected), the search uses a set of random solutions as the starting point for the search. The solutions that are found to best fit the observations up to the previous time step are used as the starting solutions for the subsequent instance of the problem that, in turn, is updated with the new observation data obtained during the next observation time step. This approach works well for a population-based search procedure, such as EAs, where the population of solutions continually explores the decision space and migrates toward the right solution as the objective function is adjusted dynamically based on updated observation information over time.

To address the issue of nonuniqueness, ADOPT is structured to search simultaneously for a set of alternative solutions. The EA-based search is designed according to the method developed by Zechman and Ranjithan (2004, 2007). It consists of multiple subpopulations of solutions, each converging toward a different solution that best fits the current set of observations. For a systematic and efficient search to identify whether or not different solutions exist, each subpopulation of solutions is designed to migrate to a region in the decision space that is maximally different from that of the other subpopulations. If nonunique solutions exist, then more than one subpopulation will converge to different possible solutions to indicate that the currently available observations are insufficient to precisely and uniquely identify the correct solution.

Algorithmic Steps of ADOPT

The EA-based procedure is structured to search for a set of possible solutions by exploring the decision space via multiple subpopulations. These subpopulations simultaneously search for solutions that are as different as possible from one another. To set a benchmark for the best possible solution, one of the subpopulations searches independently for the solution that best fits the observations. The remaining subpopulations use that benchmark to find other possible solutions that fit the observations equally or nearly as well as the best solution. To identify maximally different solutions, a measure of distance in the decision space between pairs of subpopulations is maximized. This procedure is executed for each observation time step. The number of observations available up to that time step is used to construct the objective function that represents a metric of prediction error. At any point in the search, each subpopulation represents the state of the best solution to fit the available observations. When new observations are added at the next time step, the objective function is appropriately updated, and the search continues from the current state of solutions represented in the subpopulations. By hot-starting the search at any time step based on the previous solutions, the search in the subsequent time step is expected to be conducted more efficiently, thus yielding better convergence. As a solution in one subpopulation becomes similar to one in another subpopulation, one of these similar subpopulations is removed. Eventually, when sufficient observations are available to identify a unique solution, only one subpopulation remains. These steps in the ADOPT procedure collectively identify at any observation time step the solution that best fits the currently available observations. They also reveal other possible solutions, if any, to indicate the uniqueness of the solution. The detailed steps of this procedure are described below.

Step 1. Let time step $t = t_0$, and t_0 represents the time of first detection of contamination at sensors. Create an initial set of random solutions, equally divided among N subpopulations. A solution (i.e., individual) is taken as a pair of vectors, (x, σ), where x represents a set of decision variables, including three integers (representing injection location, starting time, and duration) and an array of real variables (corresponding to the mass loading profile); σ represents a set of mutation strength values, each of which is used to mutate the corresponding decision variable.

Step 2. Increment time step $t \leftarrow t + 1$. Set generation index as $g = 0$. Update the monitoring data with additional measurements and construct the prediction error function.

Step 2.1. Increment generation index $g \leftarrow g + 1$. In the first subpopulation ($p = 1$), evaluate the fitness based on the prediction error. In subpopulation p $(= 2, 3, \ldots N)$, evaluate the fitness based on the prediction error and its distance from all other subpopulations. For each individual k in subpopulation p, its distance to other subpopulations can be formulated as

$$d(k) = \min[\mathrm{Dist}(k, \mathrm{Best}_s) | s = 1, \cdots, N, s \neq p] \tag{2}$$

where Best_s = the best individual of subpopulation s, $\mathrm{Dist}(k, \mathrm{Best}_s)$ = the distance between individual k in subpopulation p and the best individual in subpopulation s, which can be simply calculated as the distance of their node locations.

Step 2.2. In each subpopulation, apply selection and mutation operators, and create a new set of solutions. In the first subpopulation, selection is based on the prediction error function only, whereas in other subpopulations selection is based on both the prediction error and distance functions. A specified degree of relaxation (i.e., a reduction in goodness of fit) from the observed data up to the current time t is used as a basis to evaluate the feasibility of individuals of other subpopulations to be accepted as near-optimal

J. Water Resour. Plann. Manage. 2011.137:183-192.

solutions. The feasibility of individual k can be evaluated in terms of the following mathematical constraint:

$$f(k) \leq \text{tar}^* \sqrt{\frac{\sum_{T=t_0}^{t} \sum_{i=1}^{N_s} (C_{iT}^{\text{obs}})^2}{N_s{}^* t}} \tag{3}$$

where $f(k)$ = the objective value of individual k, tar = the target relaxation of individuals, C_{iT}^{obs} = the observed concentration at sensor i at time T, and N_s = the number of sensors. If an individual k meets the constraint, set as the feasible solution; otherwise, set as the infeasible solution. During the selection process, if the feasible individuals in a subpopulation are dominant, then emphasis is placed on the distance evaluation. If, instead, the unfeasible individuals prevail, then an individual with a higher fitness value is given a higher probability to be chosen. In a self-adaptive ES implementation, the mutation strength, once mutated, is used to mutate the corresponding decision variable and progress along with the individuals. Considering that the individuals evolve in a dynamic environment, the mutation strengths are reinitialized once new data comes in.

Step 2.3. If the stopping criteria (e.g., $g <$ maximum number of generations) are not met, then go to Step 2.1; otherwise, go to Step 3.

Step 3. Eliminate subpopulations that represent duplicate solutions. If two subpopulations converge on the same location, the one that performs relatively worse with respect to the prediction error is eliminated from the optimization process. If only one subpopulation remains, or the current set of solutions is acceptable, then stop.

Step 4. If no more observations are available, then stop; otherwise, go to Step 2.

Illustrative Case Studies

In this section, a number of hypothetical contamination events are studied via two water networks of different degrees of complexity. The purpose of the case studies is to exhibit the ability of ADOPT to predict alternative source characteristics by coupling the *EPANET* model with the ADOPT search procedure. Specifically, the search is conducted to determine, at the end of each observation interval, a set of source characteristics that includes the location, start time, and mass loading profile of the contaminant as the contaminant is introduced into the network. Also, the sensitivity of the proposed approach to a range of parameter settings and contamination scenarios is investigated.

Upon confronting a contamination event, the ES-based ADOPT procedure starts to execute with a prespecified number of subpopulations. Over time these subpopulations are adaptively regulated, according to the sensor observations, to keep track of the best solution and a set of alternatives. The resultant alternative solutions must perform similarly well in terms of the prediction error but are maximally far apart from one another. Given these sensor observations, alternatives can be determined according to the currently available data. For example, the target value, which is used to evaluate the feasibility of the solutions, is set to a relaxation value (120% in this study) of the root mean square of the observed data. Considering that the degree of nonuniqueness diminishes as more measurements are collected, the relaxation value is adjusted accordingly (by 0.7% in this study, based on a set of trials). To avoid unnecessary computational costs, a subpopulation is eliminated from the optimization process when it converges to the same location as another subpopulation. The algorithm terminates when the number of remaining subpopulations reduces to one or when no additional data are monitored.

Table 1. Allowable Range of Source Parameters for the Case Studies

Source parameter	Small example	Micropolis example
Location	Any node (1–97)	Any node (1–1,575)
Starting time (h)	0–4	0–30
Duration (h)	0–5	0–7
Mass injection rate (g/ min)	0–30	0–30

In this study, it is assumed that a conservative contaminant is injected at a single location and that the hydraulic condition is deterministic and steady during each hour. These assumptions are made primarily to allow a convenient and viable investigation and exploration of the proposed approach; however, they are not expected to limit a broader applicability of the approach to problems whose conditions deviate from these assumptions. Table 1 describes the allowable ranges of source parameters for both case studies.

Small Network Example

The first network examined is one of the examples available as a tutorial within *EPANET*. The network is depicted in Fig. 1, and further details can be found in the *EPANET* users' manual (Rossman 2000). This network consists of 97 nodes, including two sources, three tanks, and 117 pipes. Four sensors are distributed randomly in this network (denoted as S1, S2, S3, and S4 in Fig. 1). The base demands and the mean temporal variations of the 24-hour demand pattern assigned to all the nodes are assumed to be those described in the network input file for this network in *EPANET*. The hydraulic conditions in the network are simulated hourly and are assumed to be at steady state within each hour of the simulation. The contaminant transport is simulated in 10-min intervals, and the concentration values at the sensors are assumed to be observed in 10-min increments.

To perform the sensitivity analysis, the investigation starts with a base scenario. A nonreactive contaminant is introduced into the network at Node 205 (see Fig. 1) at 3:00 a.m. for a duration of 2 h; and time-varying mass injection rates of 10, 20, 15, 20, 10, 15, 20, 20, 20, 10, 15, and 10 g/ min, each of which corresponds to a 10-min interval. After several trials under different contamination

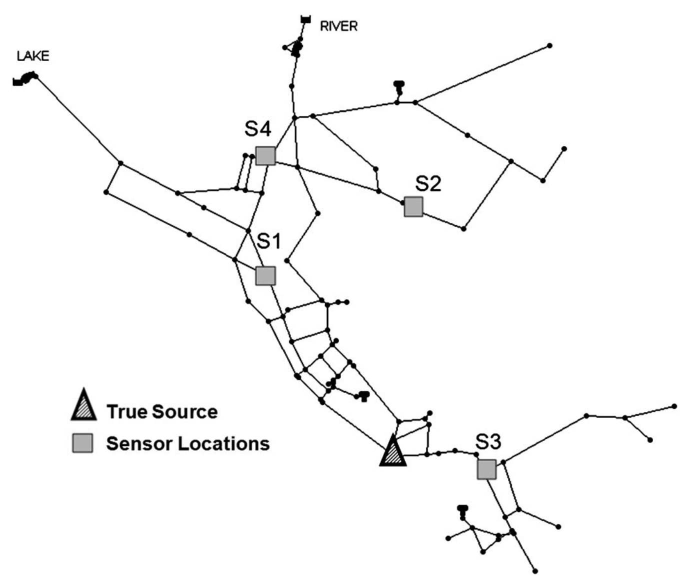

Fig. 1. Layout of the small network

J. Water Resour. Plann. Manage. 2011.137:183-192.

scenarios, this scenario was selected as a base run because the pollutant injected at this location seems to be relatively difficult to identify.

ADOPT, based on a $(\mu + \lambda)$-ES, was applied to the hypothetical event previously described. The parameters used include an initial number of 20 subpopulations, each of which consists of 200 parents (μ) and 300 mutants (λ), and the number of generations is 30 at each 10-min interval. Unless otherwise noted, the same parameter settings were employed in other scenarios as well. To gain statistical significance, 30 random runs were carried out for each case. On average, each random trial using these parameter settings took approximately 5 min on the Neptune system (an Opteron cluster) where *EPANET* simulations were executed in parallel on 10 processors.

Fig. 2 shows the ADOPT results for the base scenario from a typical run. The first detection occurred at 3:40 a.m., and ADOPT identified 14 alternatives for the first set of observations [Fig. 2(a)]. As the sensors collected measurements continually, the ADOPT solutions evolved accordingly. At 6:00 a.m., after the last set of measurements was taken, 10 acceptable solutions were identified, and each of these solutions matched the observation within the specified error limit. The locations of these solutions are shown in Fig. 2(b). In addition to the differences in injection locations, these solutions varied in their starting time, duration, and mass injection profiles, as shown in Fig. 2(c). In spite of such differences, Fig. 2(d) indicates a good agreement between the observed and predicted concentration profiles at sensor S3 for all 10 solutions. No contamination was observed at the other three sensors. The best solution that was identified matched the true source node, starting time, and duration exactly, with slight differences in the mass loading profile.

Parameter Settings

In this section, the sensitivity of the results is evaluated in terms of the differences in the parameter settings for the mutation operators, number of subpopulations, and generation numbers. All experiments were performed for 30 random trials based on the base scenario previously described. All the results are reported in terms of the average and one standard deviation (the error bars in Figs. 3–6) values computed from the solutions obtained for all random trials.

Mutation is the primary key operator in the ES procedure. Thus, determining an appropriate mutation strategy is an important step in the ES-based ADOPT procedure. A self-adaptive strategy was considered, whereby the mutation step size is set through self-adaptation and let to evolve with the decision variables. In a static environment, a mutation step size typically decreases over iterations as the algorithm converges. This decrease in step size is

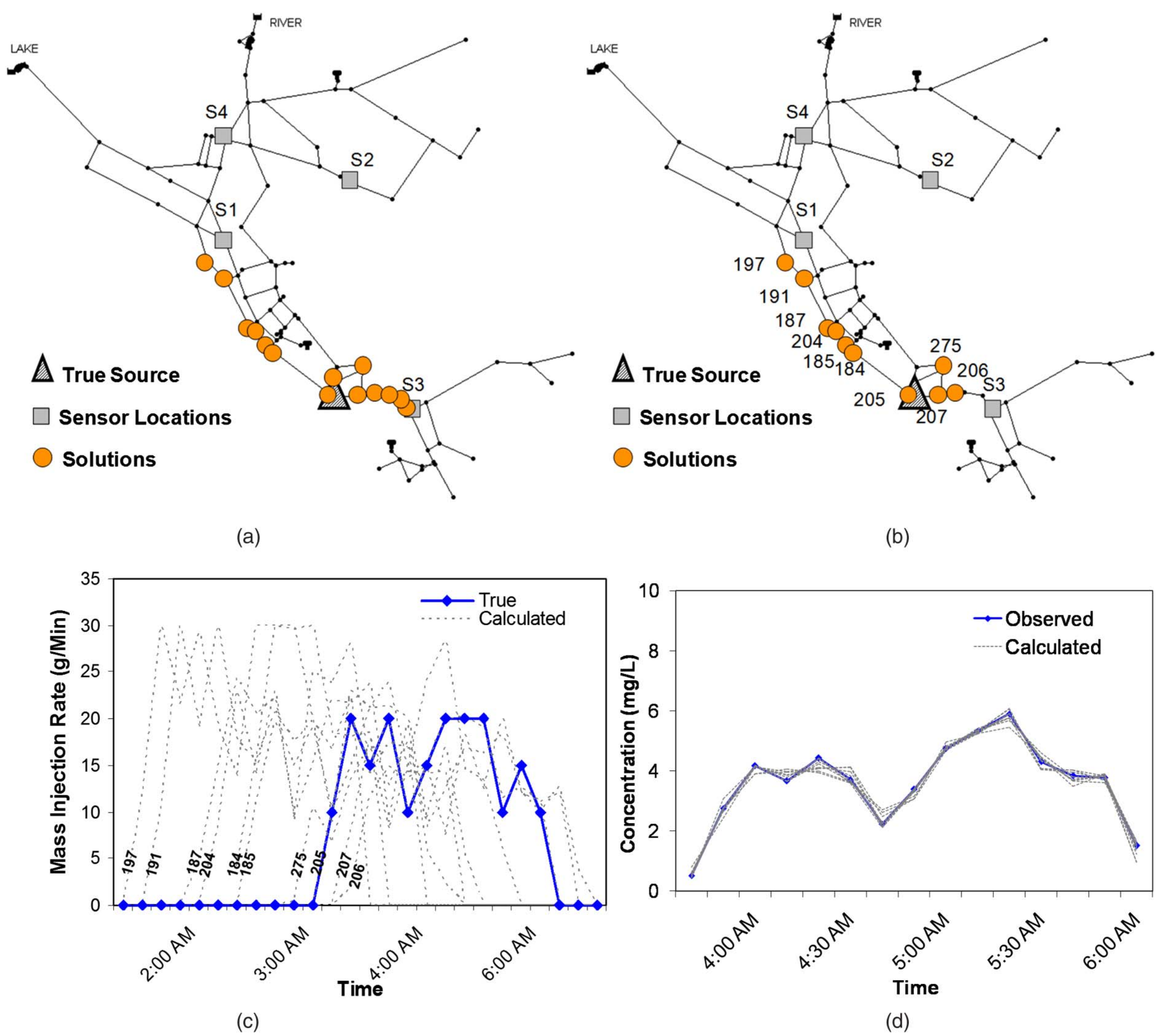

Fig. 2. Results for the base scenario using ES-based ADOPT

J. Water Resour. Plann. Manage. 2011.137:183-192.

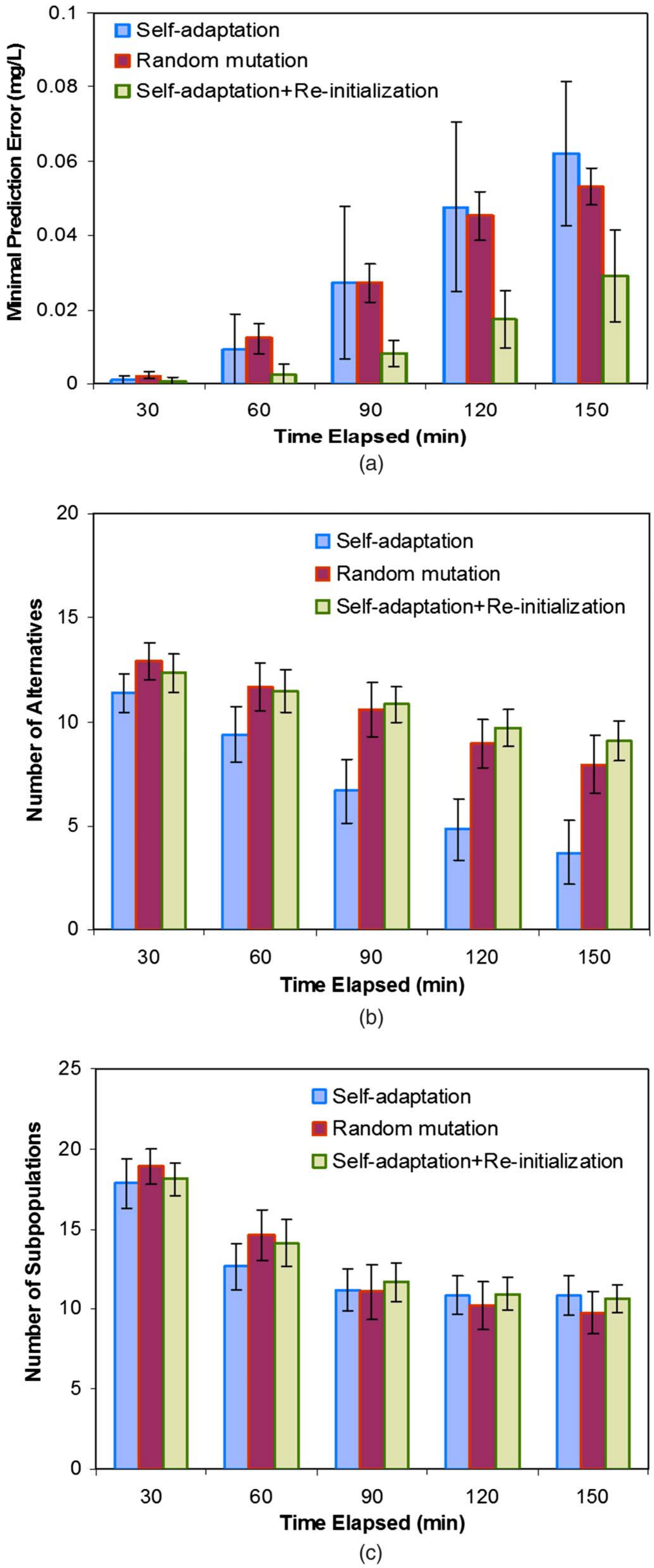

Fig. 3. Results from different mutation strategies based on the base scenario

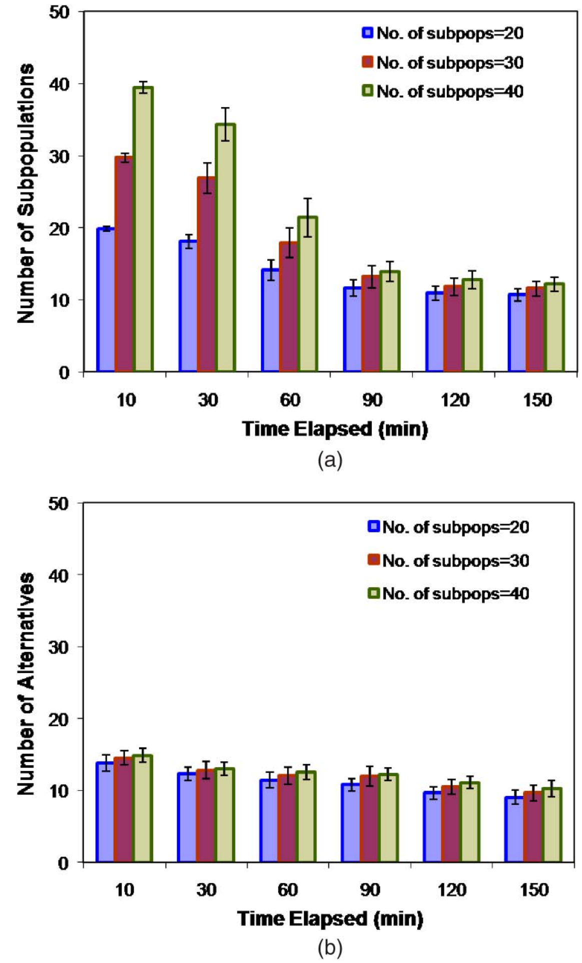

Fig. 4. Effect of initial number of subpopulations on the ADOPT solutions

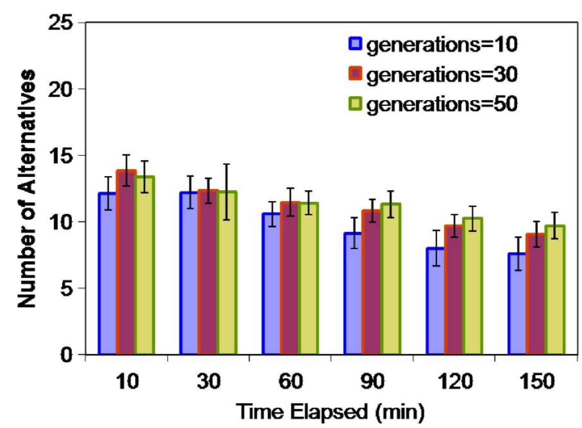

Fig. 5. Effect of number of generations on the ADOPT solutions

especially favorable for fine-tuning at the converged region in the case of a static situation. Although the ES has been shown to possess a self-learning property in a dynamic environment (Hoffmeister and Bäck 1992), more time may be needed to self-adjust step sizes to proper magnitudes than to reinitialize them once an environmental change occurs. As a result, appropriate reconstruction of mutation step sizes may facilitate the adaptability in the dynamic context of the contaminant source characterization.

Several mutation strategies were considered. For each strategy investigated, mutation step sizes were set to the same number as the number of decision variables, one for each dimension. The first set of runs focused on a completely self-adaptive mutation, in which mutation step sizes evolved with the decision variables throughout the entire process, except for the random initialization at the beginning. For the second set of runs, mutation step sizes

J. Water Resour. Plann. Manage. 2011.137:183-192.

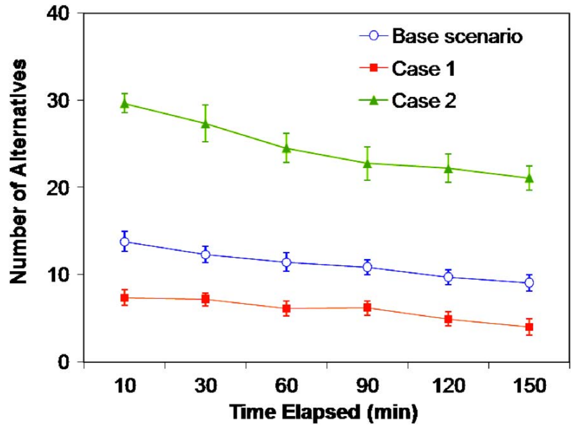

Fig. 6. Effect of the quantity and quality of monitoring data on the number of alternative solutions

were randomized at the start of each generation; that is, the mutation step sizes between generations were unrelated to one another. The last set of experiments was conducted using the self-adaptive mutation strategy between changes, and the step lengths were reinitialized only when new data became available. The results of these three strategies averaged over 30 random trials are illustrated in Fig. 3. Fig. 3(a) plots the average minimal prediction error, which corresponds to the optimal solution among all subpopulations. The third strategy in which reinitialization in addition to self-adaptation was included outperforms the other two schemes. The differences in performance increased with elapsed time. The self-adaptation strategy shows an improvement over the random mutation in the initial stages. However, this advantage gradually disappears over time, reflecting the progressive significance of the appropriate reinitialization as solutions converge to the optimum with the changing environment. To better understand how ADOPT performs under different mutation strategies, the level of nonuniqueness represented by the number of alternatives [Fig. 3(b)], and the number of remaining subpopulations [Fig. 3(c)] which represents the convergence of individuals as well as the computational cost were compared. As shown in Fig. 3(b), the number of alternatives generated corroborates the information presented in Fig. 3(a). Fig. 3(c) reveals a diminishing rate of decline in the number of subpopulations as the number of sensor observations increases. This trend is more significant in the first experiment attributable to the lowest of the degrees of diversity introduced during the search process.

When solving a contamination event using ADOPT, it is desirable for the number of subpopulations to be the same as the number of nonunique solutions present. It is difficult, however, to predetermine the exact number, and therefore an estimate must be specified to execute the procedure. Sensitivity of the performance of ADOPT to this parameter, the initial number of subpopulations, was conducted with 20, 30, and 40 initial subpopulations. Fig. 4(a) shows the number of subpopulations at different elapsed times for the three scenarios. As expected, ADOPT, with more subpopulations converges faster, but at the end of the observation period (150 min), similar numbers of subpopulations were reached. Moreover, the initial number of subpopulations affects only mildly the final set of alternatives identified by ADOPT [see Fig. 4(b)]. As observed in the predicted source characteristics, the proposed approach incorporating more subpopulations yields slightly more accurate predictions.

The sensitivity to the number of generations for each time interval was also investigated. From a practical viewpoint, the solution at any time step should provide the best estimates of the potential source. This requires ADOPT to maintain a balance between convergence and diversity at any time step. More convergence yields better solutions in terms of the objective value at any elapsed time but compromises diversity or becomes mired in local optima. Fig. 5 shows the variation in the number of alternatives over elapsed time for cases with 10, 30, and 50 generations per time step. More alternatives with low prediction error represent a better overall performance. As shown in Fig. 5, the case with 10 generations resulted in fewer alternatives than the other cases. This outcome may be attributable to an insufficient convergence. For the case with 30 generations, more alternatives were identified than the 50 generations case at the initial stages. This phenomenon may be attributable to premature overconvergence in the latter case. As more monitoring data are included, the case with 50 generations achieves a better performance. With regard to the error in the source characteristics (i.e., prediction accuracy), ADOPT results in better estimates at the initial stages for the cases with larger number of generations. At later stages of the search, the differences in the prediction accuracy of the source characteristics among these three cases are very small because the emphasis is placed on the distance evaluation in the selection step when feasible individuals are dominant in the populations, an outcome that is expected at later stages of the evolution process.

Various Contamination Scenarios

Whereas the results from the base scenario indicate the effectiveness of ADOPT in characterizing a WDS contaminant source, a range of contamination scenarios presented here allows the demonstration of the algorithm's robustness given the variations in the amount and quality of the monitoring data as well as the contaminant source characteristics.

To explore the ability of ADOPT to assess nonuniqueness, two scenarios were added to the investigation to characterize the same pollutant as the base scenario. Case 1 incorporates an additional sensor [denoted as 185 in Fig. 2(b)] for comparison to the base scenario, and Case 2 is assumed to have the same sensor distribution as the base scenario. However, these sensors were simulated to recognize the existence of contamination rather than the concentration value. A fixed concentration value specifies the threshold or detection limit to trigger a contamination detection status. A reading that exceeds this threshold (a value of 0.1 mg/L is assumed in this study) represents the presence of contamination (the concentration detection is converted to an output signal of 1); otherwise, it represents the absence of contamination (represented by an output signal of 0). The degree of nonuniqueness is measured in terms of the number of alternatives. Fig. 6 shows comparisons of the number of nonunique solutions identified for the two cases and the base scenario. The degree of nonuniqueness diminishes as the amount of observations increases. In all three scenarios, a decrease in either the number or quality of the sensors leads to a higher degree of nonuniqueness in the solutions.

The behavior of the ADOPT-based approach to different contamination events with various contaminant characteristics was then investigated. Four contamination events were simulated, differing in the injection location, starting time, duration, and/or mass injection rates. A description of the four scenarios is provided in Table 2. Fig. 7(a) shows the injection locations of the four scenarios as well as the base scenario.

The results of the four scenarios compared with the base run are shown in Fig. 7(b). The observations were sufficient to eliminate the nonuniqueness for the scenarios with the contaminant source at Nodes 125 and 105 (Scenarios 1 and 2, respectively). In contrast, a higher level of uncertainty was found for the other scenarios, which

J. Water Resour. Plann. Manage. 2011.137:183-192.

Table 2. True Source Description for Four Contamination Event Scenarios

Source parameter	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Location	Node 125	Node 105	Node 267	Node 171
Starting time	2:30 a.m.	3:00 a.m.	3:00 a.m.	2:30 a.m.
Mass loading (g/ min)	{15, 20, 25}	{10, 20, 25, 25, 15, 10}	{10, 20, 25, 25, 15, 10}	{5, 20, 25, 20, 15, 20, 15, 10, 10}
Duration (min)	30	60	60	90

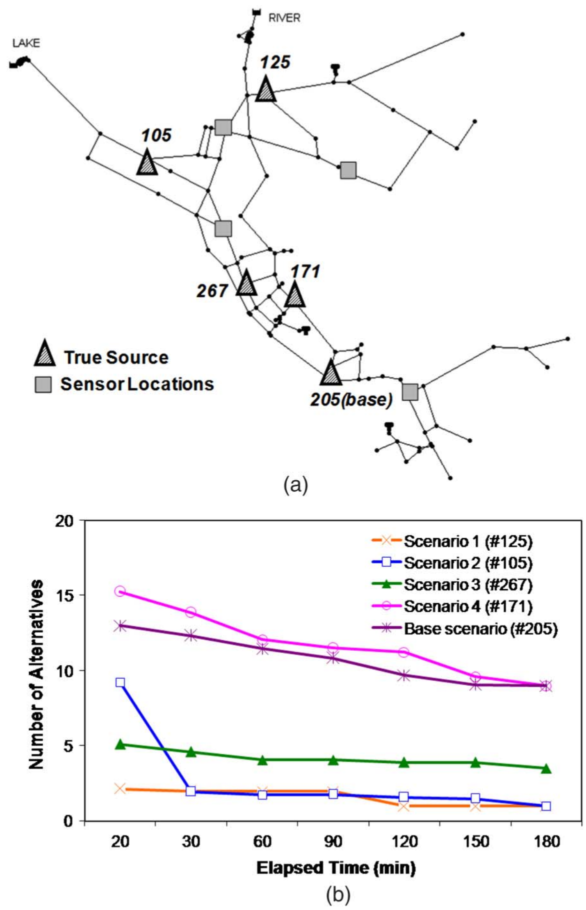

Fig. 7. Four hypothetical contamination scenarios along with the base scenario

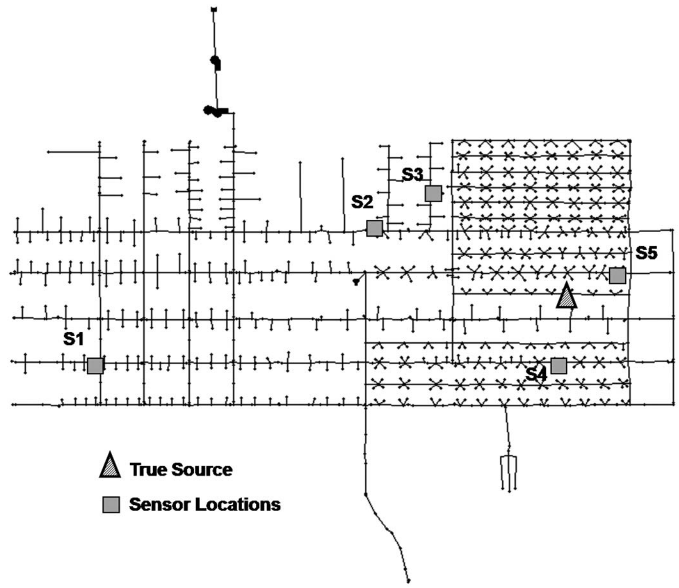

Fig. 8. Micropolis water distribution network schematic

is likely attributable to the location of the true source in relation to the sensor locations in the network. It is also noted that for the contaminant introduced at Nodes 125, 105, and 267 (Scenarios 1, 2, and 3, respectively), the true source location is always identified as one among the alternative set, whereas for the source at Nodes 205 (Scenario 2) and 171 (Scenario 4), the true source location is not always among the set of alternatives. With regard to the distribution of the resulting nonunique solutions for all scenarios, most of the solutions are concentrated in the vicinity of the true source, which may be attributable to the hydraulic similarities among their vicinities. These alternatives gradually converge towards the true source node as more measurements become available.

Micropolis Network Example

To understand the impact of the increasing level of problem complexities on the alternatives identified by ADOPT, a second network was studied. This relatively large network consists of 1,574 junctions, two reservoirs, and one tank, as illustrated in Fig. 8. This example was developed for a virtual city with 5,000 residents, and additional details of this example can be found in Brumbelow et al. (2007). The locations of five ideal sensors are shown in Fig. 8. The contaminant transport is simulated at 10-min intervals, and the concentration values at perfect sensors are assumed to be observed at each 10-min increment.

A conservative contaminant is assumed to be introduced at an intermediate node (labeled as IN 1646) starting at 12:00 p.m. (after a 12-h simulation) for 1 h. The contaminant is injected at the rates of 10, 25, 10, 20, 28, and 17 g/ min, each of which corresponds to a 10-min interval. The allowable ranges for the variables describing the source characteristics are described in Table 1. After several experiments, the algorithm parameter values for ADOPT were specified as follows: The number of subpopulations is 40, and each subpopulation contains 100 parents and 100 mutants. ADOPT was executed for 20 generations at each observation interval. For this contamination event, the first detection occurred at 6:30 p.m., and the observations lasted until 9:00 p.m. ADOPT was executed for 30 random trials in consideration of the randomness of the search process. The computation time for each trial was approximately 75 min on the Neptune system (an Opteron cluster) where *EPANET* simulations were executed in parallel on 20 processors. The results from a typical run are summarized in Figs. 9 and 10.

Fig. 9(a) illustrates the locations of the 39 alternatives generated after considering 20-min interval measurements. It can be observed that the 12 alternative solutions that remain after 2 h of elapsed time are very close to the true source location and one of them is at the true source node, as displayed in Fig. 9(b). This observation reflects that the degree of nonuniqueness diminishes with additional observations. Fig. 10 shows that the simulated profiles of the 12 alternatives correspond well with the observed profile at sensor S3. No abnormal measurements were observed at the other sensors. In addition to determining whether the true source is recovered as a potential solution, the relative rank of alternative solutions is evaluated based on the prediction errors in descending order, in which the solution with the minimal prediction error is ranked first. Overall, a high frequency corresponds to a top rank for the true source node, whereas in a few cases the highest ranking solution is located at a node adjacent to the true source.

Summary and Discussion

In this work, an adaptive dynamic optimization technique for contaminant source identification has been presented and is structured to search for a set of possible solutions by exploring the decision space through multiple subpopulations. Several investigations were conducted to assess the applicability of the proposed ADOPT method. They were designed to discover many aspects of the problem and solutions, including the degree of nonuniqueness, various problem complexities, and the range of the ADOPT algorithmic parameter settings.

This study suggests that ADOPT works adaptively to determine multiple alternatives that match the observations available at any time during the contamination event. Specifically, the level of nonuniqueness assessed by ADOPT diminishes as the number of sensors, measurement quality, or observation period length increases. The use of the time-varying relaxation target helps generate alternatives according to the available sensor observations. The relaxation value is designed to be adaptive and is determined through several experiments. Imposing a high relaxation value unavoidably slows down the convergence of the search, and conversely, a low relaxation value may result in premature convergence, potentially skipping over the true source during the search. Additional work is necessary to investigate the relaxation setting to eliminate the impact on solutions irrespective of contaminant characteristics.

In the context of a contamination event, contaminant source recovery involves identifying the injection location as well as the release history. The characterization of the release history can be represented as a large array of decision variables, which may make the search inefficient. Effective fine-tuning would be helpful for the algorithm to exploit localized structures. Although the self-adaptive mutation strategy in the ES is useful in adaptively refining the search process, a changing environment necessitates the reconstruction of the mutation step sizes. The results suggest that the dynamic reinitialization strategy is increasingly advantageous to the search as time goes on.

Whereas the ADOPT procedure showed an ability to adaptively determine solutions that match the observations available at any time, the identified results are not unique in most cases. This is attributable to the lack of unique solutions to the problem itself as the contaminant source characterization is complicated by the limited observation data and uncertainty in the system state, as well as the arbitrary nature of the contamination scenario in which the contaminant potentially can be injected at any node with varying strength in the network. Nevertheless, solutions obtained using ADOPT incorporate the true source as one of the alternatives, and the number of the solutions identified diminishes as more observations become available. Compared to the MHT method, a significant advantage of the ADOPT approach is that ES-based multipopulation strategy allows quick identification of alternative solutions (i.e., a set of best hypotheses) using its ability to identify good solutions rapidly for real-world complex problems, thus avoiding enumerating many unnecessary hypotheses and reducing computational costs. On average, the computation time of the micropolis example at each 10-min observation interval was around 5 min on the Neptune system (an Opteron cluster) where *EPANET*

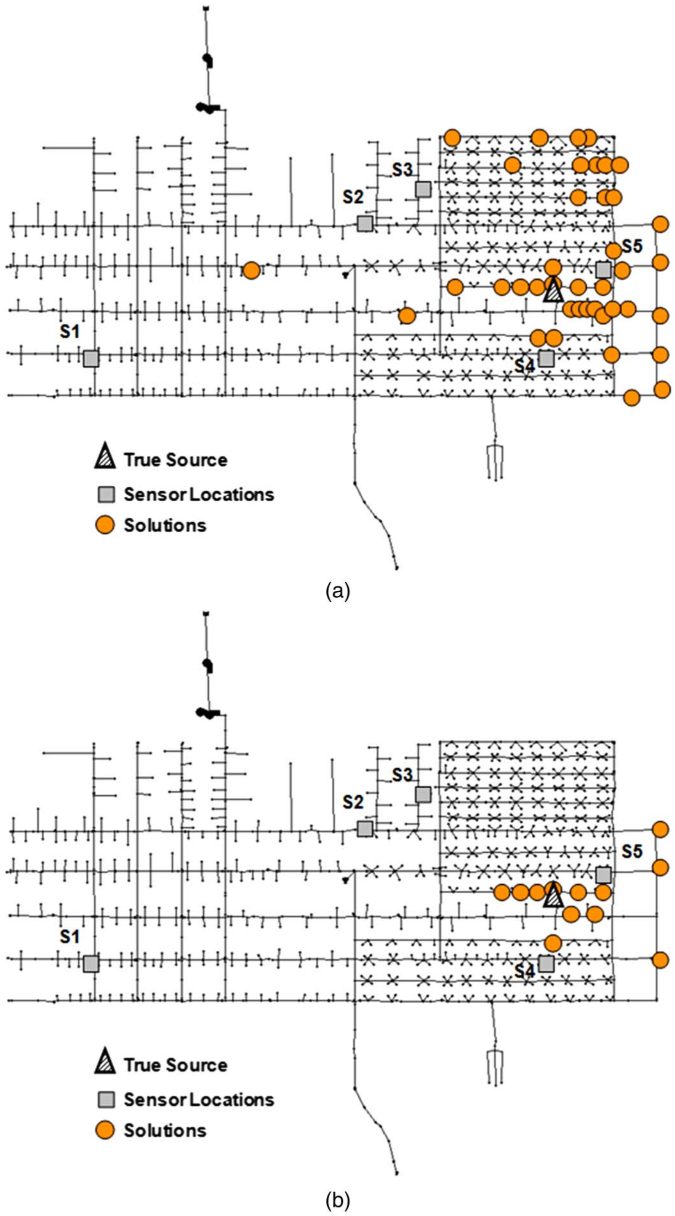

Fig. 9. Identified solutions for the hypothetical contamination event in the micropolis network

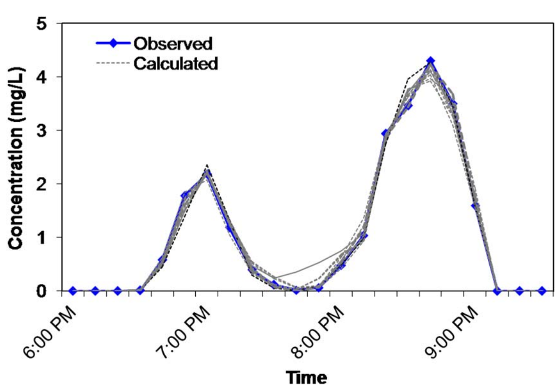

Fig. 10. Comparison of concentration profiles between the observed and calculated

simulations were executed in parallel on 10 2.2-GHz processors, suggesting that real-time application can be achieved.

Although ADOPT has been demonstrated successfully to solve a limited set of scenarios, the example problems solved are based on several assumptions, such as the source is present only at a single injection location, the contaminant is nonreactive, known demands, the demands at all nodes throughout the simulation period are known, the simulation model is defined perfectly, and the network conditions are fully and correctly specified, which may seem impractical in a realistic scenario. Future work will extend this algorithm to handle problems of increasing complexity and facilitate the application to actual WDS contamination events. Another issue associated with ADOPT is its degree of efficiency, especially for application to a large network where simulation models are relatively more time-consuming. The research team intends to investigate enhancements, such as pruning out infeasible nodes using a prescreening procedure for reducing the search space.

Acknowledgments

This work is supported by National Science Foundation (NSF) under Grant No. CMS-0540316 under the DDDAS program.

References

Bar-Shalom, Y., and Fortmann, T. E. (1988). *Tracking and data association*, Academic Press, San Diego.

Blackman, S. S., and Popoli, R. (1999). *Design and analysis of modern tracking systems*, Artech House, Norwood, MA.

Brumbelow, K., Torres, J., Guikema, S., Bristow, E., and Kanta, L. (2007). "Virtual cities for water distribution and infrastructure system research." *Proc. of World Environmental and Water Resources Congress*, Tampa, FL, 15–19.

Dandy, G. C., Simpson, A. R., and Murphy, L. J. (1996). "An improved genetic algorithm for pipe network optimization." *Water Resour. Res.*, 32(2), 449–458.

Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. M. (2006). "Identification of contaminant sources in water distribution systems using simulation-optimization method: Case study." *J. Water Resour. Plann. Manage.*, 132(4), 252–262.

Hoffmeister, F., and Bäck, T. (1992). "Genetic self-learning." *Proc. of the 1st European Conf. on Artificial Life*, MIT Press, Cambridge, MA, 227–235.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*, Univ. of Michigan Press, Ann Arbor, MI.

Kurien, T. (1990). "Issues in the design of practical multitarget tracking algorithms." *Multitarget-multisensor tracking: Advanced applications*, Y. Bar-Shalom, ed., Artech House, Norwood, MA.

Laird, C. L., Biegler, L. T., Van Bloemen Waanders, B. G., and Bartlett, R. A. (2005). "Contamination source determination for water networks." *J. Water Resour. Plann. Manage.*, 131(2), 125–134.

Lingireddy, S., and Ormsbee, L. (2002). "Hydraulic network calibration using genetic optimization." *Civ. Eng. Environ. Syst.*, 19(1), 13–39.

Lui, L., and Ranjithan, S. (2010) "An adaptive optimization technique for dynamic environments." *Eng. Appl. Artif. Intell*, 23(5), 772–779.

Mahar, P. S., and Datta, B. (1997). "Optimal monitoring network and ground-water-pollution source identification." *J. Water Resour. Plann. Manage.*, 123(4), 199–207.

Mahinthakumar, G. K., and Sayeed, M. (2005). "Hybrid genetic algorithm —Local search methods for solving groundwater source identification inverse problems." *J. Water Resour. Plann. Manage.*, 131(1), 45–57.

Preis, A., and Ostfeld, A. (2007). "A contamination source identification model for water distribution system security." *Eng. Optim.*, 39(8), 941–951.

Reid, D. B. (1979). "An algorithm for tracking multiple targets." *IEEE Trans. Autom. Control*, 24(6), 843–854.

Rossman, L. A. (2000). *EPANET user's manual. risk reduction engineering laboratory*, U.S. EPA, Cincinnati.

Savic, D. A., and Walters, G. A. (1997). "Genetic algorithms for least-cost design of water distribution networks." *J. Water Resour. Plann. Manage.*, 123(2), 67–77.

Schwefel, H. P. (1995). *Evolution and optimum seeking*, Wiley, New York.

Van Bloemen Waanders, B. G., Bartlett, R. A., Biegler, L. T., and Laird, C. D. (2003). "Nonlinear programming strategies for source detection of municipal water networks." *Proc. of the ASCE World Water and Environmental Congress*, Philadelphia, 23–26.

Vermaak, J., Godsill, S. J., and Perez, P. (2005). "Monte Carlo filtering for multi-target tracking and data association." *IEEE Trans. Aerosp. Electron. Syst.*, 41(1), 309–332.

Vitkosvsky, J. P., Simpson, A. R., and Lambert, M. F. (2000). "Leak detection and calibration using transients and genetic algorithms." *J. Water Resour. Plann. Manage.*, 126(4), 262–265.

Yang, S., Ong, Y., and Jin, Y. (2007). *Evolutionary computation in dynamic and uncertain environments*, Vol. 51, *Studies in computational intelligence*, Springer, New York, 3–28.

Zechman, E. M., and Ranjithan, S. (2004). "An evolutionary algorithm to generate alternatives (EAGA) for engineering optimization problems." *Eng. Optim.*, 36(5), 539–553.

Zechman, E. M., and Ranjithan, S. (2007). "Evolutionary computation-based approach for model error correction and calibration." *Adv. Water Resour.*, 30(5), 1360–1370.

J. Water Resour. Plann. Manage. 2011.137:183-192.