ADDRESSING NON-UNIQUENESS IN A WATER DISTRIBUTION CONTAMINANT SOURCE IDENTIFICATION PROBLEM

Emily M. Zechman[1], E. Downey Brill[1], Jr., G. Mahinthakumar[1], S. Ranjithan[1], James Uber[2]

[1]North Carolina State University
Raleigh, North Carolina
{emzechma, brill, kumar, ranji}@ncsu.edu
[2]University of Cincinnati
Cincinnati, OH
jim.uber@uc.edu

Abstract

The source of contamination in a water distribution system may be identified through a simulation-optimization approach. The optimization method searches for the contaminant source characteristics by iteratively estimating the contaminant plume concentrations until they match observations at sensors. The amount of information available for characterizing the source depends on the number and spatial locations of the sensors, as well as on the temporally varying stream of sensed data. The accuracy of the source characterization depends on the amount of observations available. A major factor affecting this accuracy is the degree of non-uniqueness present in the problem, which may cause misidentification of the source characteristics. As more sensors are added to the network, the non-uniqueness may be reduced and a unique solution may be identified. Thus, a key consideration when solving these problems is to assess whether the solution identified is unique, and if not, what other possible solutions are present. A systematic search for a set of alternatives that are maximally different in solution characteristics can be used to address and quantify non-uniqueness. For example, if the most different set of solutions that are identified by a search procedure are very similar, then that solution will be considered as the unique solution with a higher degree of certainty. Alternatively, identification of a set of maximally different solutions that vary widely in solution characteristics will indicate that non-uniqueness is present in the problem, and the range of solutions can be used as a general representation of the amount of non-uniqueness. This paper investigates the use of evolutionary algorithm (EA)-based alternatives generation procedures to quantify and address non-uniqueness present in a contaminant source identification problem for a water distribution network. As additional sensors may decrease the amount of non-uniqueness, several sensor configurations will be tested to investigate and quantify the improvement in uniqueness as more information is used in the source characterization.

Keywords

source identification, non-uniqueness, evolutionary algorithm, water quality sensors

1. INTRODUCTION

Water distribution systems are vulnerable to threat of intentional contamination. For example, a pollutant source introduced into a water distribution network will move through the system and expose the public to a health risk. Detection of the contamination in the distribution system using a sensor network could yield useful observations to manage such contamination threat events. Based on these observations, the location, strength, time and duration of the contaminant source could be determined to direct decision-makers toward containing and mitigating the event.

Given a set of concentration observations at sensors in the network, one approach to identify the contaminant source is to couple a water distribution simulation model with an optimization method to minimize the error between predicted concentrations at the sensor nodes and observations in the network. This representation of the problem is broadly classified as an inverse or system characterization problem. Such an inverse problem holds the potential for non-uniqueness, where a set of different sources with significantly different pollutant release characteristics may be identified to give similar prediction errors. Since the non-uniqueness in a system is related to the amount of data available to identify a source, more data made available through either additional sensors or a longer monitoring time, helps reduce the degree of non-uniqueness in the system. When the available information is not sufficient to identify a unique source, determining only a single solution may mislead a decision-maker to costly mitigation actions that exacerbate the contamination situation. Because the source characterization problem in general is sufficiently ill-posed, it is important to ascertain whether the source characterization identified as a solution to the problem is unique or is one of many potential sources that best fits the concentration profiles observed at the sensors.

One approach to characterize this non-uniqueness is through the determination of a set of solutions that represents the range of possible source characterizations. This requires a search for possible alternative source characterizations that are maximally different in their source characterizations and perform similarly well in predicting concentration profiles at the sensors. The degree of difference among the alternative solutions represents the range of possible source characterizations that fit the observed concentration profiles and helps quantify the uniqueness of an identified source. A method for systematic search for a set of maximally different alternative solutions was mathematically defined by Brill (1979) and was extended to evolutionary algorithm-based search procedures by Zechman and Ranjithan (2004). This paper explores the use of a recently developed implementation of an evolutionary algorithm-based alternatives generation procedure to solve a water distribution network contaminant source identification problem and to characterize the degree of non-uniqueness present in the inverse problem.

2. WATER DISTRIBUTION CONTAMINATION SOURCE DETERMINATION PROBLEM DESCRIPTION

The following mathematical description of the water distribution contamination source determination problem is based upon the model formulation presented by Laird et al. (2005):

$$\underset{\{m_k(t)\}}{Minimize}\ E = \max\left\{\left|\hat{c}_k(t) - \hat{c}_k^*(t)\right|, \forall k \in N_s, \forall t \in \Theta_s\right\} \quad (1)$$

Subject to

$$\frac{\partial \bar{c}_i(t,x)}{\partial t} + u_i(t)\frac{\partial \bar{c}_i(t,x)}{\partial x} = 0, \forall i \in P \quad (2)$$

$$\bar{c}_i[x = I_i(t), t] = \hat{c}_{k_i(t)}(t), \forall i \in P \quad (3)$$

$$\bar{c}_i(x, t = 0) = 0, \forall i \in P \quad (4)$$

$$\hat{c}_k(t) = \frac{\left[\sum_{i\in\Gamma_k(t)} Q_i(t)\bar{c}(x = O_i(t), t)\right] + m_k(t)}{\left[\sum_{i\in\Gamma_k(t)} Q_i(t)\right] + Q_k^{ext}(t) + Q_k^{inj}(t)}, \forall k \in J \quad (5)$$

$$V_k(t)\frac{d\hat{c}_k(t)}{dt} = \left\{\sum_{i\in\Gamma_k(t)} Q_i(t)\bar{c}[x = O_i(t), t]\right\} + m_k(t)$$
$$-\left\{\left[\sum_{i\in\Gamma_k(t)} Q_i(t)\right] + Q_k^{ext}(t) + Q_k^{inj}(t)\right\}\hat{c}_k(t), \forall k \in S \quad (6)$$

$$\hat{c}_k(t=0) = 0, \forall k \in S \quad (7)$$

$$m_k(t) \geq 0, \forall k \in N \quad (8)$$

The model variables are defined in Table 1. The objective function, Eqn. 1, seeks to minimize the maximum difference between $\hat{c}_k^*(t)$ and $\hat{c}_k(t)$, the observed and simulated contaminant concentration values, respectively, at time *t* at sensor node *k*. Eqns. 2-8 specify the hydraulics and contaminant transport in the simulation of the water distribution network. The mass loading profile of the injected contaminant, $m_k(t)$, is the vector of decision variables. The contaminant mass balance is described by Eqns. 2-4 for the pipes, Eqn. 5 for the junctions, and Eqns. 6-7 for the tanks.

Table 1. Definitions of variables used to model the water quality simulation

P, J, S	Complete set of all pipes, junctions, and storage tanks
N	Complete set of all nodes (i.e., $N = J \cup S$)
$t \in [0..t_f]$	Time, t_f is final time step
$x \geq 0$	Displacement along a pipe
$\bar{c}_i(x,t), i \in P$	Contaminant concentration in pipe *i* at displacement x and time t
$\hat{c}_k(t), k \in N$	Contaminant concentration of node *k* and time *t*
$m_k(t), k \in N$	Unknown contamination mass flow rate
$N_s \subseteq N, \Theta_s$	Set of nodes with installed sensors and set of all sample times
$c_k^*(t), k \in N_s$	Measured contaminant concentrations; these values will not be known continuously in time, but rather at discrete sampling points in time, Θ_s
$\Gamma_k(t), k \in N$	Set of all pipes flowing into node k at time t
$I_i(t), O_i(t), i \in P$	Displacement along pipe *i* where fluid is entering and leaving pipe, respectively; these designations are time dependent and change with flow direction
$k_i(t), i \in P$	Index of node connected at inlet of pipe *i*; this designation is time-dependent and changes with flow direction

$u_i(t)$	Known fluid velocity in pipe *i*
$Q_i(t), i \in P$	Known volumetric flow rate in pipe *i* at time *t*
$Q_k^{ext}(t), k \in N$	Volumetric flow rate for known external source (e.g., reservoir flow)
$Q_k^{inj}(t), k \in N$	Volumetric flow rate of unknown contaminant mass injection, $m_k(t)$, in practice this value will not be known, and they are set to a small quantity relative to other network flow rates
$V_k(t), k \in S$	Volume in tank *k* at time *t*

3. SOLUTION APPROACH

The search for the time and location of contaminant injection into the network is a large non-linear programming problem that poses sufficient challenges to optimization techniques. A few search methods to solve the source determination problem have recently been reported, including direct sequential and simultaneous methods (van Bloemen Waanders et al., 2003; Laird et al., 2005).

Another 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. Evolutionary algorithms (EA) (Holland, 1975) are a class of heuristic methods that provide a global search mechanism to identify efficiently near-optimal solutions for large non-linear optimization problems. They have been used in several water distribution network design problems (e.g., Dandy et al., 1996; Savic and Walters, 1997). EAs are effectively used to solve inverse problems such as water distribution network calibration (e.g., Vitkosvsky et al., 2000; Lingireddy and Ormsbee, 2002) and groundwater source contamination identification problems (e.g., Mahinthakumar and Sayeed, 2005). EAs are investigated here as an approach to solve the source determination problem in water distribution networks (Eqns. 1-8). As non-uniqueness is an inherent property of the source determination problem, a method for alternatives generation (Zechman and Ranjithan, 2004) is extended for an EA-based implementation and coupled with the water distribution mathematical model.

3.1 Mathematical Background for Generating Alternatives

The original source identification problem is represented as an error minimization problem in Eqns. 1-8. Let $\boldsymbol{m}^*$ (=$\{m^*_k(t);\ \forall\ k, t\}$) be the best solution identified with E^* being the corresponding minimum objective value (i.e., prediction error). An alternative solution $\boldsymbol{m}$ that is maximally different from $\boldsymbol{m}^*$ can be generated by solving the following model:

$$\underset{\{m_k(t)\}}{Maximize}\ D = d(\boldsymbol{m}, \boldsymbol{m}^*) \tag{9}$$

$$Subject\ to\ \max\left\{\left|\hat{c}_k(t) - \hat{c}_k^*(t)\right|\right\} \le T(E^*) \tag{10}$$

Subject to Eqns. 2-8

where D is a difference function based on $d(\boldsymbol{m},\boldsymbol{m}^*)$, which represents a "distance" measure between two solutions $\boldsymbol{m}$ and $\boldsymbol{m}^*$, and T is a target that is specified in relation to the fitness value E^*. T represents an allowable relaxation, if any, in the objective value and permits a small

degradation in the objective value to provide exploration of the decision space to identify regions where maximally different solutions can be found. The target may be specified to allow no relaxation when exploring for alternate optima for highly non-unique problems.

A set of alternatives can be identified in a sequential approach. Once the best solution $\boldsymbol{m}^*$ is found, the first alternative, $\boldsymbol{m}^1$, is identified by solving the model represented by Eqns. 2-10. To identify the second alternative, the difference function D can be modified to find the solution maximally different from both the best solution ($\boldsymbol{m}^*$) and the first alternative ($\boldsymbol{m}^1$), while the target function remains the same. The difference function can be updated for each new alternative identified, and the search for additional alternatives continues until no significantly different alternatives are found.

A more elegant and computationally efficient search will enable the identification of the set of alternatives simultaneously, rather than iteratively as described above. As an EA uses a population of solutions in its search strategy, it can be structured to evolve to a set of alternatives, enabling a flexible framework for implementing a simultaneous search. EAs are used in this study to determine the source characterization for a water distribution network; accordingly, the search for maximally different alternative solutions is facilitated through evolutionary computation.

3.2 EA-based Approaches for Generating Alternatives

Several EA-based approaches are available for generating a set of alternative solutions for an optimization problem. The most direct method is to execute an EA sequentially, as described and implemented for a genetic algorithm by Harrell (2001). Initial execution of the algorithm identifies the best solution to the original modeled problem, and for each sequential run, the solution identified in the previous run is included in the distance function. This approach may become computationally burdensome due to the repeated executions of the EA.

The niching operator (Mahfoud, 1992) is a well-established EA-based approach for identifying a set of solutions for multimodal problems. Niching is useful for identifying a set of solutions that all perform well with respect to the objective function; however, identification of solutions that are most different from one another involves extensive parameter tweaking or a priori knowledge of the decision space. This was demonstrated by Loughlin et al. (2001). Loughlin et al. (2001) also described a new GA-based procedure (GAMGA – Genetic Algorithms for Modeling to Generate Alternatives) that extends niching to generate maximally different solutions. The GAMGA procedure introduces several additional parameters and algorithmic steps that require careful tuning.

An EA-based approach, the evolutionary algorithm for generating alternatives (EAGA) (Zechman and Ranjithan, 2004), was developed to explicitly identify maximally different alternatives using a set of subpopulations that collectively search in a simultaneous manner. The first subpopulation searches for the optimal solution to the original modeled problem. The remaining subpopulations search for solutions that are maximally distant from all other subpopulations while meeting the specified target on the objective. The target objective becomes more restrictive as the search progresses, tightening as the fitness of the individuals in the first subpopulation improve. Recombination, mutation, and selection operators are applied separately in each subpopulation, and migration is not allowed between subpopulations. EAGA is designed with a minimal number of additional algorithmic parameters. As the structure of EAGA is independent of the search

procedure employed in the subpopulations, it can be used with any type of EA for optimization problems.

4 Niched Co-Evolution Strategies

Evolution strategies (ES) (Schwefel, 1995) is an EA-based search method that is being explored as a solution approach for the source determination problem as described in Eqns. 1-8. Similar to a genetic algorithm, ES searches using a population. At the first generation, the population is initialized with a size of μ individuals. A probabilistic mutation operator is applied to produce λ new solutions each generation. The next set of μ individuals may be selected from the combined array of parent and offspring solutions (denoted as $(\mu+\lambda)$ selection) or from the set of offspring alone ((μ,λ) selection). ES is being explored for its use in overcoming difficulties associated with representation of the water distribution network for the source identification problem.

The construct of EAGA is extended to an ES-based implementation, called the Niched Co-Evolution Strategies (NCES), which uses the basic concept of cooperative co-evolution to evolve a set of subpopulations to identify maximally different alternative solutions by solving the alternatives generation model Eqns. 2-10. The set of subpopulations is used to collectively search for different alternative solutions, where each subpopulation is guided toward a region in the solution space that is distant from other subpopulations. Information about the location of a subpopulation in the solution space (and therefore the set of common solution-characteristics of a subpopulation) is shared such that the subpopulations cooperate in co-evolving toward different regions of the solution space. Selection within each subpopulation depends upon how well the solutions perform with respect to prediction error, as well as upon how far they are from the other with respect to the source characterizations represented by the solutions. NCES is designed to search explicitly for a set of solutions that are as different as possible in the source characteristics and are within a target range of the prediction error (Eqn. 10). Building upon the EAGA procedure described by Zechman and Ranjithan (2004) for a genetic algorithm, it is extended to construct a new algorithm NCES for generating alternatives using evolution strategies. The main steps of the algorithm are described below for a $(\mu+\lambda)$ ES, but could be easily adapted for any alternative ES configuration.

4.1 Algorithmic Steps of NCES

Step 1. Create an initial population with P subpopulations (each with a population size of μ), where P is the number of alternative solutions (i.e., source characteristics) being sought. Let SP_p (p=1,…, P) represent the index for subpopulation p. The first subpopulation (SP_1) is dedicated to the search for the solution with the best objective function (i.e., minimum prediction error) value.

Step 2. In each subpopulation SP_p (p=1,…,P), apply an adaptive mutation operator to generate λ offspring.

Step 3. In SP_1, evaluate the fitness (Eqn. 1) of each solution and identify the solution in the subpopulation with the best fitness (i.e., prediction error). This solution will serve as the benchmark for setting the relaxation constraint Eqn. 10.

Step 4. In SP_p (p=2,…,P), evaluate the fitness of each individual solution (i.e., distance among the different source characteristics as defined in Eqn. 9). Solutions that meet the target constraint Eqn. 10 are assigned a feasible flag, and solutions that fail to meet the target are labeled infeasible.

Step 5. For each solution k in subpopulation SP_p (p=2,…,P), calculate the difference metric (Eqn. 9) in the solution space between that solution and other subpopulations.

Step 6. In each subpopulation SP_p, apply a selection operator. In SP_1, the selection is based on how well a solution maximizes the fitness. The solutions are ranked based on fitness and the top μ solutions survive to the next generation. In SP_p (p=2,…,P), the selection is based on how well the solution meets the constraints Eqns. 2-8, as well as on the value of the difference function. Feasible solutions are ranked first from highest to lowest difference function. Infeasible solutions are then ranked from best to worst fitness values.

Step 7. Check for termination criteria. Stop the algorithm if termination criteria (e.g., a maximum number of iterations) are met. Otherwise, go to Step 2.

5. ILLUSTRATIVE CASE STUDY

A synthetic case study is used to demonstrate the use of NCES for a contaminant event in a water distribution network. The network used is one of the problem instances available as a tutorial within EPANET (Rossman, 2000). This network consists of 97 nodes, including two sources, three tanks, and 117 pipes. EPANET was used to simulate the water distribution system. The network is depicted in Fig. 1, and further details can be found in the EPANET users' manual.

To generate a set of synthetic observations for an illustrative hypothetical contamination event, a non-reactive contaminant source is introduced into the network at node #105 (Fig. 1) and twelve sensors (Fig. 1) are placed in the network to observe the consequent concentration profiles (Fig. 2). The hydraulics in the network is simulated hourly over a 24-hour time period. The hydraulics is assumed to be at steady state within each hour of the simulation. For each hourly hydraulic condition, the contaminant transport is simulated in 5-minute intervals, and the concentration values at the sensors are observed at the end of each 5-minute increment.

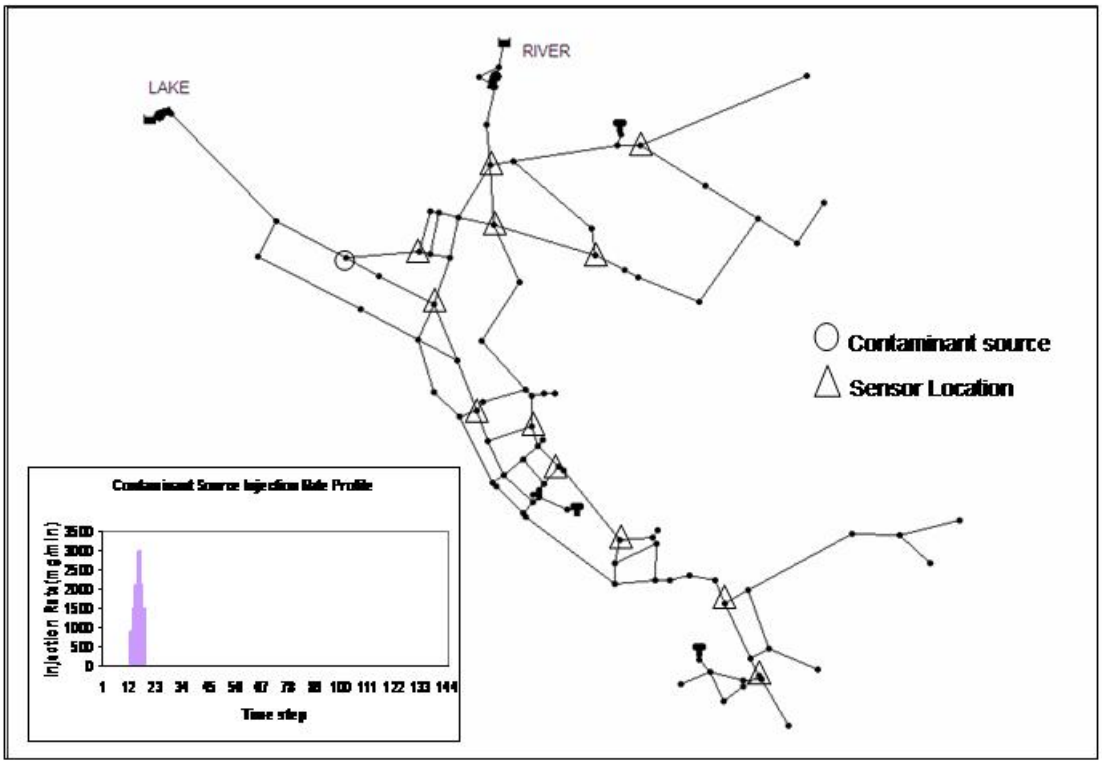

Figure 1. Water distribution network schematic and contaminant source injection rate profile. Contaminant source indicated by the circle, and triangles designate sensor locations.

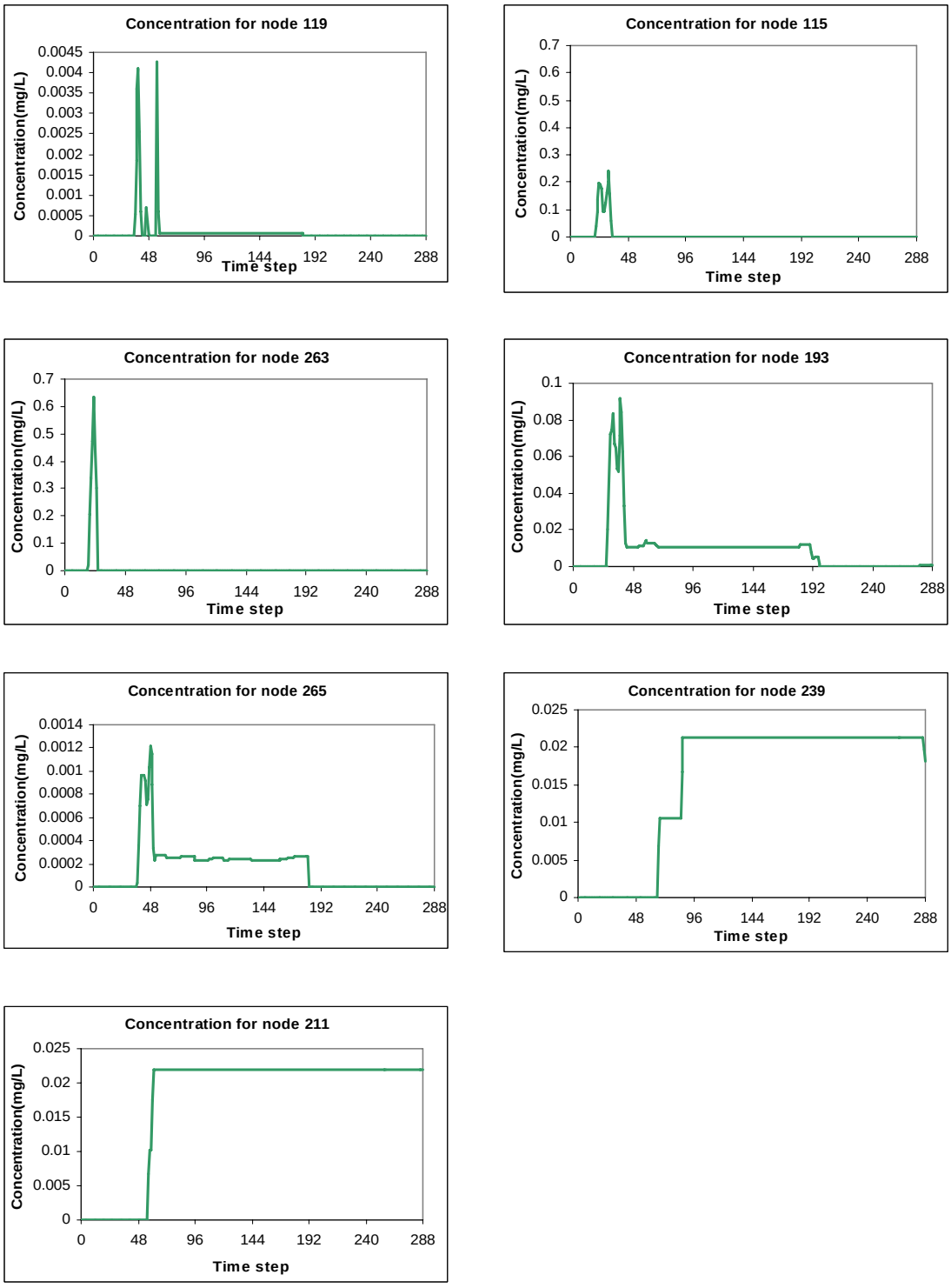

Figure 2. Concentration profiles at seven sensor nodes with non-zero concentration profiles.

The purpose of this case study is to predict alternative source characterizations for the hypothetical contamination event by coupling the EPANET model with the NCES search procedure. Specifically, the search is conducted to determine a set of source characterizations, including the node of the contaminant source, the start time of the contamination, and the

contaminant mass loading profile as it is introduced to the network. This is conducted for varying degrees of observation information, representing different amounts of non-uniqueness in the problem.

6. RESULTS AND OBSERVATIONS

Several on-going investigations address the efficacy of the proposed NCES method to solve the source determination problem in water distribution networks. The investigations are designed to explore many aspects of the problem and solution, including the degree of non-uniqueness present in solving this inverse problem, the settings of the contamination event as well as sensor placements to study the effects of problem complexity, different ways to represent the problem within NCES, and the range of NCES algorithmic parameter settings. A range of algorithmic settings and several problem representations are explored to determine the robustness of the ES-based NCES search method. Additionally, new implementations of ES-based operators are being explored specifically for different solution representations associated with the water distribution network source characterization problem.

Existence of a set of alternatives is used to indicate the uniqueness of the problem through the amount of similarity among maximally different solutions. To demonstrate the use of alternatives to characterize the non-uniqueness, several problem instances are being investigated with varying levels of information available to determine the contaminant source. For example, a set of source characterizations are being identified using information from few sensors. The same source may be characterized using data from a larger set of sensors, and the set of alternatives generated may display more uniqueness in the source characterization than the set of alternatives identified using a small set of sensor data. Similarly, sensor data from a longer monitoring period may result in more unique source characterization. Results from these investigations and associated observations and discussions will be presented at the conference.

References

Brill, E. D., Jr. (1979). “Use of Optimization Models in Public-Sector Planning.” *Management Science,* 25(5), pp. 413-422

Dandy, G. C., Simpson, A. R., and Murphy, L. J. (1996). “An Improved Genetic Algorithm for Pipe Network Optimization” *Water Resources Research,* 32(2), pp. 449-458.

Harrell, L. (2001). “Evolutionary Algorithm-Based Design of a System of Wet Detention Basins under Uncertainty for Watershed Management.” *Proceedings of the 24th Annual Water Resources Planning and Management Conference,* ASCE, Houston, TX, pp. 272-277.

Holland, J. H. (1992). Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press, Cambridge, MA

Laird, Carl L., Biegler, Lorenz T., van Bloemen Waanders, Bart G., and Bartlett, Roscoe A. (2005) Contamination Source Determination for Water Networks. *Journal of Water Resources Planning and Management,* 131(2) 125-134.

Lingireddy, S. and Ormsbee, L. (2002). “Hydraulic Network Calibration Using Genetic Optimization.” *Civil Engineering and Environmental Systems,* 19(1)

Loughlin, D. H., Ranjithan, S., Baugh, J. W., and Brill, E. D., Jr. (2001). "Genetic Algorithm Approaches for Addressing Unmodeled Objectives." *Engineering Optimization*, 33, pp. 549-569.

Mahfoud, S.W. (1992). "Crowding and Preselection Revisited." *Proceedings of the Second Conference on Parallel Problem Solving from Nature*, Elsevier Science Inc., Brussels, Belgium, pp. 27-36.

Mahinthakumar, G. Kumar and Sayeed, M (2005). "Hybrid Genetic Algorithm – Local Search Methods for Solving Groundwater Source Identification Inverse Problems." *Journal of Water Resources Planning and Management*, 131(1); 45-57

Rossman, L. A. (2000) EPANET User's Manual, Risk Reduction Engineering Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH

Savic, D. A. and Walters, G. A. (1997). "Genetic Algorithms for Least-Cost Design of Water Distribution Networks" *Journal of Water Resources Planning and Management*, 123(2), pp. 67-77.

Schwefel, H.-P. (1995) Evolution and Optimum Seeking, Wiley & Sons, New York

van Bloemen Waanders, B. G., Bartlett, R. A., Bigler, L. T., and Laird, C. D. (2003) "Nonlinear Programming Strategies for Source Detection of Municipal Water Networks." *Proceedings of the ASCE World Water and Environmental Congress*, Philadelphia, PA, June 23-26

Vitkosvsky, J. P., Simpson, A. R., and Lambert, M. F. (2000). "Leak Detection and Calibration using Transients and Genetic Algorithms" *Journal of Water Resources Planning and Management*, 126(4), pp. 262-265.

Zechman, E. M., and Ranjithan, S. (2004). "An Evolutionary Algorithm to Generate Alternatives (EAGA) for Engineering Optimization Problems." *Engineering Optimization*, 36(5), pp. 539-553.