Adaptive Contamination Source Identification in Water Distribution Systems Using an Evolutionary Algorithm-based Dynamic Optimization Procedure

Li Liu[1], 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
{lliu2, emzechma, brill, kumar, ranji}@ncsu.edu
[2]University of Cincinnati
Cincinnati, OH
jim.uber@uc.edu

Abstract

Accidental drinking water contamination has long been and remains a major threat to water security throughout the world*. Consequently, contamination source identification is an important and difficult problem in the managing safety in water distribution systems. This problem involves the characterization of the contaminant source based on observations that are streaming from a set of sensors in the distribution network. Since contamination spread in a water distribution network is relatively quick and unpredictable, rapid identification of the source location and related characteristics is important to take contaminant control and containment actions. As the contaminant event unfolds, the streaming data could be processed over time to adaptively estimate the source characteristics. This provides an estimate of the source characteristics at any time after a contamination event is detected, and this estimate is continually updated as new observations become available. We pose and solve this problem using a dynamic optimization procedure that could potentially provide a real-time response. As time progresses, additional data is observed at a set of sensors, changing the vector of observations that should be predicted. Thus, the prediction error function is updated dynamically, changing the objective function in the optimization model. We investigate a new multi population-based search using an evolutionary algorithm (EA) that at any time represents the solution state that best matches the available observations. The set of populations migrates to represent updated solution states as new observations are added over time. At the initial detection period, non-uniqueness is inherent in the source-identification due to inadequate information, and, consequently, several solutions may predict similarly well. To address non-uniqueness at the initial stages of the search and prevent premature convergence of the EA to an incorrect solution, the multiple populations in the proposed methodology are designed to maintain a set of alternative solutions representing different non-unique solutions. As more observations are added, the EA solutions not only migrate to better solution states, but also reduce the number of solutions as the degree of non-uniqueness diminishes. This new dynamic optimization algorithm adaptively converges to the best solution(s) to match the observations available at any time. The new method will be demonstrated for a contamination source identification problem in an illustrative water distribution network.*

Keywords

source identification, water distribution systems, sensors, non-uniqueness, dynamic optimization, evolutionary algorithm

1. INTRODUCTION

Accidental and intentional contamination of water distribution networks is becoming an increasingly critical issue. For example, a pollutant source introduced into a water distribution network 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 could be determined to direct decision-makers toward containing and mitigating the event. Given a set of concentration observations at sensors in the network, an inverse problem could 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 model predicted concentrations and real observations at the sensor nodes in the network. In the context of a fast-evolving threat event in water distribution networks, 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.

While inverse modeling is applied to a wide array of system identification problems in engineering, it possesses the potential for non-uniqueness, in that 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 the source of contamination, more data, made available through either additional sensors or a longer monitoring time, may help reduce the degree of non-uniqueness in the system. If the available information is insufficient to determine a solution as unique, then it is important to determine whether a solution identified is unique or what other potential solutions exist. Knowing that the identified solution is the only possible source characteristics that matches the observations is critical since a non-unique but incorrect solution may yield potentially costly mitigation actions that may inconsequentially exacerbate the contamination situation.

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

Recently, researchers have reported the development of several procedures to identify the contaminant source characteristics by using information from sensor networks. A direct sequential technique, which is reported by van Bloemen Waanders et al. (2003), was applied to solve a small scale optimization problem with a standard successive quadratic programming tool. Laird et al. (2005) reported a direct simultaneous approach. These methods attempt to identify a single solution using a fixed, i.e., not dynamically streaming, set of observations. New approaches are needed to solve the problem in an adaptive manner. At any time, the procedure must identify possible solutions that explain the observations available up to that time point. Also, the solution procedure must identify not only the best estimate of the source characteristics that minimizes the prediction error, but also a set of possible alternative solutions, if any, that similarly predict the available observations.

This paper reports a new adaptive search method that uses a simulation-optimization approach where the water distribution network model is directly coupled with a new dynamic optimization method to iteratively evaluate and identify solutions that minimizes the prediction error. To assess the non-uniqueness 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 search as more information is made available. The search method explored in this paper is based on evolutionary algorithms (EAs), which are coupled with an EPANET model of the water distribution network. The applicability of the method is illustrated using a hypothetical example.

2. CONTAMINANT SOURCE IDENTIFICATION PROBLEM DESCRIPTION IN WATER DISTRIBUTION SYSTEMS

To capture the dynamic nature of the amount of available observation data, the source identification problem is described in terms of determining the source characteristics based on observations up to the current time. As the amount of observations changes over time, the description of the problem is updated at some regular time interval, e.g., observation frequency. At any instant, the problem should be solved to get the estimate of the source characteristics that best explains the currently available data. The following mathematical model is defined to determine, at any time after the contamination is detected at one or more sensors, the contamination source location, contamination event starting time and the corresponding contaminant mass loading history. While the definition provided below assumes that the contamination is introduced at only one node in the network, the same model can be updated to consider multiple contamination source locations.

.

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

$$\textit{Minimize} \quad F = \sum_{t=t_0}^{t_c}\sum_{i=1}^{N_s}\left|C_{it}^{obs} - C_{it}(L, M_{t_c}, T_0)\right| \qquad (1)$$

where F – prediction error

L – contamination 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; $M_{t_c} = \{m_{T_0}, m_{T_0+1}, \cdots, m_{t_c}\}$

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

N_s – number of sensors.

3. SOLUTION APPROACH

The search for the location and time history of contaminant injection into the network is a 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 methods (van Bloemen Wannders et al., 2003) and particle-tracking methods (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). While 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; Mahar and Datta, 1997), applicability of EAs to source characterization in water distribution networks is not fully investigated. Thus, EAs are investigated here as an approach to solve the source determination problem in water distribution networks. An EA procedure is developed to solve the dynamic optimization model described above. The EA-based approach used in this paper is a new search procedure (ADOPT—ADaptive OPtimization Technique) that is designed for adaptive optimization that considers dynamically varying stream of sensor observations and for identification of alternative solutions, if any, to assess the degree of non-uniqueness.

3.1 Conceptual Basis for ADOPT

The two key features of the new method are: 1) optimization under dynamic environments, and 2) identification of alternative solutions. In the context of the water distribution network problem, the amount of sensor observations varies with time. Thus, the objective function (i.e., the prediction error defined in Eqn. 1) that is being minimized is changing 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 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 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 is continually exploring the decision space and migrating towards the right solution as the objective function is dynamically adjusted based on updated observation information over time.

To address the issue of non-uniqueness, the new procedure is structured to search simultaneously for a set of alternative solutions. The EA-based search is designed to consist of multiple sub-populations of solutions, each converging towards a different solution that best fits the current set of observations based on the method developed by Zechman and Ranjithan (2004). To systematically and efficiently search to identify whether multiple solutions exist, the search in each sub-population of solutions is designed to migrate to a region in the decision space that is maximally different from the other sub-populations. If non-unique solutions exist, then more than one sub-population will converge to a possible solution to indicate that the currently available observations are insufficient to resolve the non-uniqueness.

3.2 Algorithmic Steps of ADOPT

The EA-based procedure is structured to search for a set of possible solutions by exploring the decision space through multiple subpopulations. These subpopulations simultaneously search for solutions that are as different as possible from each other. 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 well as the best solution. To identify maximally different solutions, some measure of distance in decision space between pairs of subpopulations is maximized.

This procedure is executed for each observation time step. The amount 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 next time step is expected to be conducted more efficiently, 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 would remain. These steps in the ADOPT procedure collectively identify at any observation time step the solution that best fits the currently available observations, and reveal other possible solutions, if any, to indicate the uniqueness of the solution.

Step 1. Create an initial set of random solutions, equally divided among N subpopulations.

Step 2. Increment time step $t \leftarrow t + 1$. Set generation index $g = 0$. Update 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 prediction error. In subpopulation p (= 2,3,…N), evaluate the fitness based on prediction error and its distance to all other subpopulations.

Step 2.2. In each subpopulation, apply selection, recombination and mutation operators, and create a new set of solutions.

Step 2.3. If stopping criteria (e.g., $g <$ max no. of generations) is not met, then go to Step 2.1; else go to Step 3.

Step 3. Eliminate subpopulations that represent duplicate solutions. 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; else go to Step 2.

4. ILLUSTRATIVE CASE STUDY

A synthetic case study is used to demonstrate the use of ADOPT 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). 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.

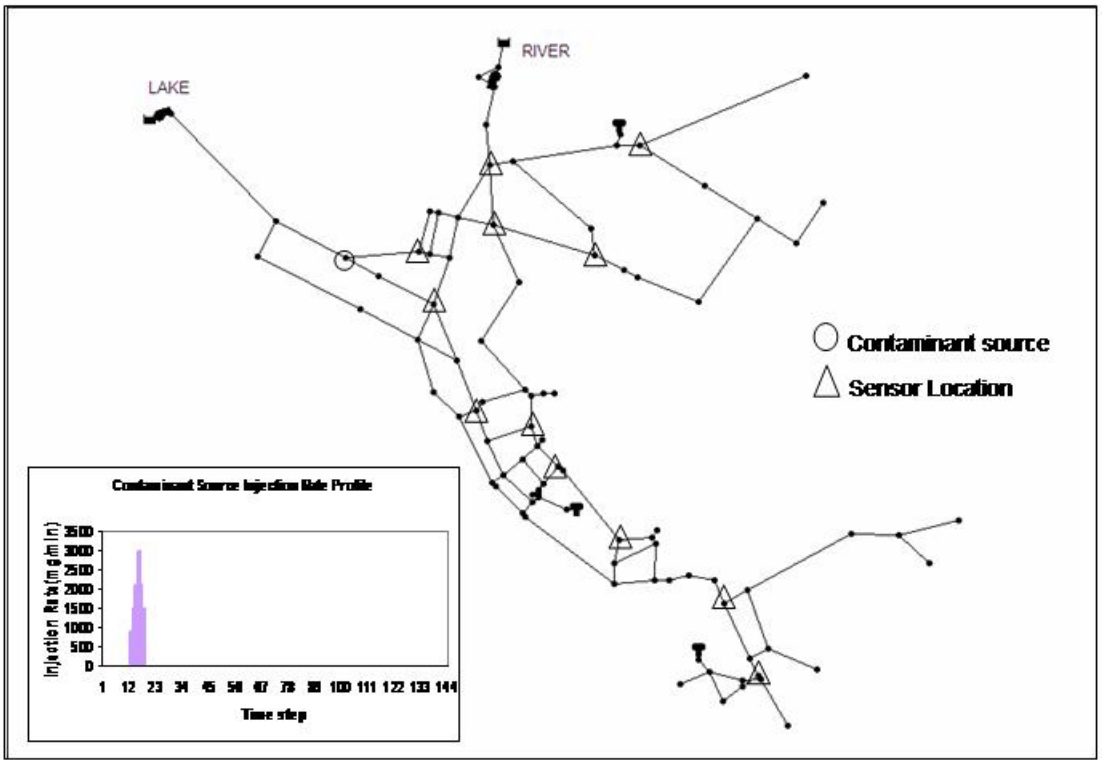

Figure 1. Water distribution network schematic and contaminant source injection rate profile for the hypothetical case study. Contaminant source is indicated by the circle, and triangles designate sensor locations.

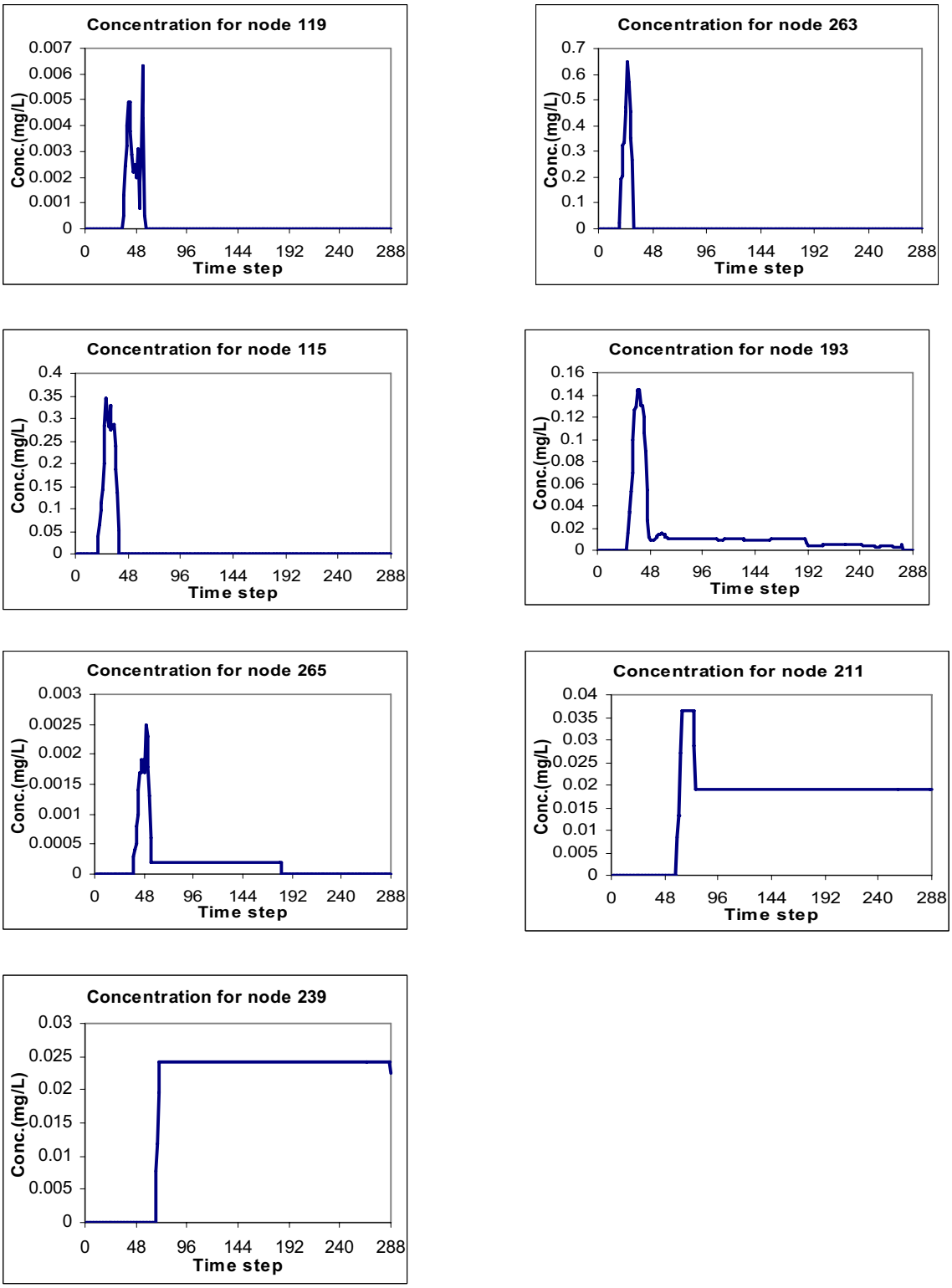

Figure 2. Concentration profiles at seven sensor nodes with non-zero concentration profiles. The other five of the twelve sensors did not record any contamination.

The purpose of this case study is to predict alternative source characterizations for the hypothetical contamination event 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 characterizations, including the node of the contaminant source, the start time of the contamination, and the contaminant mass loading profile as it is introduced into the network. This is conducted for varying degrees of observation information, representing different amounts of non-uniqueness in the problem. The investigation also tested the method for different mass loading profiles at the contaminant source.

5. RESULTS AND OBSERVATIONS

Several on-going investigations address the efficacy of the proposed ADOPT method to solve the source determination problem under dynamic environments 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 ADOPT, and the range of ADOPT algorithmic parameter settings. A range of algorithmic settings and several problem representations are explored to determine the robustness of the GA-based ADOPT search method. Additionally, new implementations of GA-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 under dynamic environments, 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 smaller 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

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.

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).

Mahar, P. S. and Datta, B. (1997). "Optimal monitoring network and ground-water-pollution source identification." *Journal of Water Resources Planning and Management*, 123(4), pp. 199-207.

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.

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.