# Integrating evolution strategies and genetic algorithms with agent-based modeling for flushing a contaminated water distribution system

Emily M. Zechman

## ABSTRACT

Water utilities can prepare for water distribution hazards, such as the presence of contaminants in the pipe network and failure of physical components. In contamination events, the complex interactions among managers' operational decisions, consumers' water consumption choices, and the hydraulics and contaminant transport in the water distribution system may influence the contaminant plume so that a typical engineering model may not properly predict public health consequences. A complex adaptive system (CAS) approach couples engineering models of a water distribution system with agent-based models of consumers and public officials. Development of threat management strategies, which prescribe a set of actions to mitigate public health consequences, is enabled through a simulation–optimization framework that couples evolutionary algorithms with the CAS model. Evolution strategies and genetic algorithm-based approaches are developed and compared for an illustrative case study to identify a flushing strategy for opening hydrants to minimize the number of exposed consumers and maintain acceptable levels of service in the network.

**Key words** | agent-based model, evolutionary algorithms, sociotechnical systems, water distribution system security

**Emily M. Zechman**
Department of Civil Engineering,
North Carolina State University,
CB 7908, Raleigh,
NC 27695,
USA
E-mail: emzechma@ncsu.edu

## INTRODUCTION

Water distribution infrastructure systems are vulnerable to the threat of both intentional and accidental hazards that may introduce a foreign contaminant to the system and expose segments of the population to severe health threats (Geldreich 1991). Through a loss of pressure in the network, for example, bacteriological pathogens may migrate into pipes (Hrudey & Hrudey 2004); alternatively, chemical toxins may be introduced intentionally at treatment plants and accessible pipes (Kroll 2006). Identification and management of threats to water distribution security involve many critical steps to effectively protect public health. For example, vulnerability analysis identifies critical locations in the system where the introduction of contaminants would have harmful consequences for public health, and this information can be used to develop plans for infrastructure hardening and protection (Perelman & Ostfeld 2010). Early warning systems place sensors within the network for quick detection of unusual water quality conditions, and these systems can be effectively designed through modeling and optimization methodologies (see Hart & Murray (2010) for a review). Event detection methods are developed to analyze sensor data to determine if fluctuations in water quality should be attributed to normal system operations or a breach in security. Source characterization methodologies use sensor data in a simulation – optimization framework to identify the contaminant source and release profile (e.g., Guan *et al.* 2006; Di Cristo & Leopardi 2008; Zechman & Ranjithan 2009).

Once a contamination event has been detected and characterized through a threat management system, water system authorities may attempt to remove the contaminant

through flushing. By opening hydrants, water can be expelled from the system until the contaminant is no longer detected in the system. If the area that is impacted is extensive, multiple hydrants may be opened in succession, moving from the source of the contamination in the downstream direction (U.S. Environmental Protection Agency 2001; Irby 2005). This approach was implemented in 1986 for a case in Lacey's Chapel, Alabama, to flush sodium hydroxide (Watts 1998), and the U.S. Environmental Protection Agency (2001) has described 12 backflow incidents in which contamination was introduced to the drinking water system, and hydrants were opened for flushing. To prepare a utility for these incidents in the future, a number of researchers have developed approaches to plan response actions that would utilize source characterization information and identify hydrants and valves that should be manipulated (Jeong & Abraham 2006; Baranowski & LeBoeuf 2008; Poulin *et al.* 2008, 2010; Preis & Ostfeld 2008; Guidorzi *et al.* 2009; Alfonso *et al.* 2010). While these approaches are not currently applied in a real-time manner to flush contaminants, algorithms and computational frameworks can be applied to aid the realistic operation of a water utility. For example, a library of response actions can be generated for a set of diverse contamination events to shed insight about the hydrants that would most effectively flush out likely events. It is expected that new computational frameworks may become available that utilize high performance computing architectures to execute search algorithms in a real-time manner.

Identifying strategies for opening hydrants relies largely on water distribution system simulation, and typical simulation and computing analysis regards consumer demands as exogenous variables. Contamination of a water distribution network, however, is a dynamic event during which consumer behaviors and demands may fluctuate. Feedback occurs between the pipe network and a set of agents, including consumers, the media, public health providers, and water utility managers. Each actor receives information from other actors, takes actions that may directly or indirectly affect the hydraulics of the water distribution network, and passes information to other parties. The emergent state of public health is the result of a set of complex adaptive interactions, and the predictability of response strategy performance may be complicated by the adaptive reduction

of residential demands. The changes in hydraulic conditions due to the reduction of residential demands alters the prediction of public health consequences and can therefore effect the performance of strategies that are designed to reduce consumer exposure. Simulation studies that consider infrastructure in isolation and do not represent dynamic interactions may not predict the emergent condition of the system.

To fully and realistically assess water distribution system contamination, a complex adaptive system (CAS) (Holland 1995; Miller & Page 2007) approach provides an adaptive, dynamic simulation to model the actions and reactions of the actors during a hazardous event. A CAS approach was developed to simulate water distribution system contamination events and was demonstrated for a small virtual city (Zechman 2011) and a mid-sized virtual city (Shafiee & Zechman 2010). These cases demonstrated that considering the changes in consumer demands alters the predicted location and strength of the contaminant plume, compared with hydraulic calculations alone. Simulations showed that as a result of consumers who change their water use after becoming sick, higher amounts of contaminant remain in the system and expose more consumers to a critical level. The word-of-mouth and communication mechanisms prevent unexposed consumers from becoming ill, and consumers change their demands earlier than if they had not received any warning. The total reduction in demands influence hydraulics, where flow volumes and directions were changed significantly from expected values for a set of contamination events.

Because consumer behaviors can have an influence on the propagation of the contaminant plume, this research presents a simulation–optimization approach that couples a CAS model with evolutionary computation (Holland 1975) optimization methodologies to develop a methodology for identifying optimal threat management plans while allowing for the adaptations in consumer behaviors. A set of hydrants is scheduled to be opened in the network once the contaminant is released to best protect public health and to maintain a minimum pressure in the network using a specialized evolution strategies (ES)-based approach. The ES-based approach was developed for applications in water distribution systems analysis (Zechman & Ranjithan 2009), and in this work it is further extended for searching

for hydrants locations and formally compared with existing genetic algorithm (GA) approaches. The use of this framework is demonstrated for a small virtual city, which provides a realistic case study.

## AGENT-BASED MODELING OF A WATER DISTRIBUTION CONTAMINATION EVENT

When a contaminant is introduced to a water distribution pipe network, the location of the contaminant plume and the consequences to public health in the network are influenced by the actions and interactions among many key players. The contaminant and the state of public health recursively affect the reactions and responses of the actors in the system, creating dynamic feedback loops among the network and the actors. For example, utility managers may first be alerted to the presence of a real or perceived threat by water quality sensors in the pipe network, public health services, the media, or consumer complaints. For plausible threats, a utility manager may alter network hydraulics by isolating portions of the system and controlling flows in the network, influencing the movement of the plume. Consumers are made aware of contaminated water through media, word-of-mouth communication from other consumers, and as they experience symptoms due to exposure. Consumers may choose to comply with boil water orders, continue limited water use, or use bottled water, and as a result, the volume of water use may fluctuate significantly from normal demand patterns. Hydraulic conditions are driven by consumer demands, and the spread of the contaminant plume dynamically changes with varying demands. The shift in the contaminant plume may warrant reconsideration of management decisions that were developed using expected demands and flows in the network. The dynamic nature of the event may be heightened as the event progresses; public health notifications may become available which influence consumer demands, and additional water quality sampling data may be acquired to update and improve management decisions.

The CAS approach is a modeling paradigm that simulates interacting actors as a dynamic network of individual actors, and agent-based modeling (ABM) (Axelrod 1997; Miller & Page 2007) is a tool for simulating a CAS. ABM represents individuals as agents that both act autonomously to achieve their own goals and react to the actions of other agents and to environmental conditions through a set of rules. ABM, or individual-based modeling, pieces together the lowest level of subsystems to give rise to a broad view of the emergent system, in contrast to top-down modeling, which is a more traditional approach for engineering systems and uses a stepwise approach to analyze a system by separating sub-systems from the highest level. Lower-level heterogeneous activities of agents are needed in ABM simulation, and for the water contamination event, behaviors that are simulated include the timing of drinking water, communication about the event, and mobility of persons to travel to different locations in a network. The premise of ABM is that the interactions among the agents can drive emergent system properties; in the case of the water contamination event, the properties of interest are public health and changes in network hydraulics that result from the decentralized, heterogeneous decisions and communication among the consumers.

An ABM framework was developed to simulate consumers as agents and capture their interactions with the water distribution system and with response actions taken by utility operators as a contamination event unfolds (Shafiee & Zechman 2010; Zechman 2011). The agent-based model and water distribution model are tightly coupled to allow simulation of the hydraulic impact of the behaviors of the agents within the water distribution system. The water distribution system model calculates nodal water quality at each time step, and the concentration of contaminant at each node is passed to agents that are located at the node during the current time step. Actions of the consumers to change their water usage affect demands and change the hydraulics of the system. The change in water usage is aggregated at each node for a set of consumers, and reduced demands change the inputs for executing the water distribution model. In addition, actions of the utility operator to open hydrants in the network alter the pressures and flow directions and volumes in the pipe network. The commercially available ABM software AnyLogic (XJ Technologies 2010) is coupled with the EPANET water distribution system model (Rossman 2000) to enable simulation of a contamination event. The number of consumer agents exposed above the critical dose for a contaminant is evaluated at

the end of a simulated time period to represent the public health consequences. Each consumer is represented as an agent and behaves according to a set of attributes and rules, described briefly below.

## Mobility

Agents are made mobile as they visit different nodes during a day to represent a population that commutes to work and visits places of business. Modeling mobility is necessary to determine if an individual agent will drink water at a node that receives contaminated water. Each agent is assigned to one residential node and one non-residential node. The maximum number of individual consumers at each node is allocated using population data. For each simulated time step, the number of agents at commercial and restaurant nodes is calculated using information from the water distribution system data. Water demands fluctuate during the day at these non-residential nodes and the population at a node increases and decreases with demands. The number of agents at each node and time step is calculated as the normalized water demand factor multiplied by the maximum population. At industrial nodes, the number of consumers remains constant, based on a 24-hour operation with three equal shifts, and at the beginning of each 8-hour shift, a random set of consumer agents are assigned to an industrial node.

At each time step, agents are chosen with uniform probability to leave their residence node and occupy their non-residential node. The number of agents at non-residential nodes increases through the morning, reaches a peak in the early afternoon, and decreases to zero into the evening. As the number of agents at a non-residential node decreases, agents are released and return to their assigned residential node.

## Water consumption

Models for the timing and volume of consumed water are needed to calculate the exposure of an individual agent. Agents drink water at different times of a day, and similarly, the level of contaminant in the water at a node varies with the movement of the contaminant plume. Davis & Janke (2008) developed an ingestion timing model, which specifies the times at which consumers ingest tap water. The timing model specifies that a consumer ingests tap water at three times per day, corresponding to meals, and each consumer also ingests tap water midway between meals. For each agent, the timings of the five daily ingestions are generated probabilistically using a set of distributions, which are based on U.S. census data that describe statistical information about time use (Davis & Janke 2008). The total volume of water that each consumer ingests per day is also generated using a probabilistic distribution, as developed by the U.S. Environmental Protection Agency (2000), with an expected value of 0.92 L. The total volume of water that is consumed by each agent is divided evenly among the five daily ingestions.

## Exposure and demand reduction

The exposure of each agent is calculated based on his ingestion of contaminated water. For any contaminant, a critical dose describes the amount of contaminant that an individual should ingest to experience symptoms. Once an agent ingests a cumulative critical dose of the contaminant the agent is flagged as 'exposed'. Exposed agents reduce their consumption for contact uses, but are modeled to maintain a moderate level of demand, representing toilets, pipe leakage and outdoor water use. A consumer's residential node is specified as multi-family housing or single-family housing, and consumers that reside in multi-family housing will reduce their demand to 43% of normal uses, and consumers that reside in single-family housing, to 60% (Vickers 2001).

## Word-of-mouth communication

A consumer who becomes exposed to a contaminant through drinking tap water will warn other consumers, who are expected to reduce their water usage and notify other consumers. An agent who changes his water use due to exposure or a warning message from peers responds by alerting a set of connected consumers 1 hour after changing his water usage. If the agent resides at a single-family residence, he will contact all agents located at his residence, representing family members. If the agent resides at a multi-family residence, he will contact one neighbor located at the same node. Consumers that are located at

non-residential nodes (representing commercial or industrial nodes) at the time they change their water use, will contact one consumer at the same node, along with consumers at their residential nodes.

Consumers that receive an alert from other consumers will react probabilistically. If a consumer receives a message from a family member, the notified agent certainly stops using water for contact purposes. If the consumer receives the message from an associate at a non-residential node (work colleague) or a neighbor at a multi-family housing unit, the notified consumer will change his water usage with a probability of 75%.

## PROBLEM STATEMENT FOR IDENTIFYING HYDRANT STRATEGIES

While a dynamic modeling framework provides simulation of contamination scenarios to evaluate the probabilistic outcome of actions taken by both consumers and decision makers, optimization can be used to identify the most effective responses from a large set of possible options. Studies have developed strategies for opening hydrants and manipulating valves to minimize the contaminant in the network (Baranowski & LeBoeuf 2008), while others minimized both the number of hydraulic operations required and the amount of contaminant consumed through multi-objective approaches (Preis & Ostfeld 2008; Guidorzi et al. 2009; Alfonso et al. 2010). Because this framework uses an ABM approach, the number of exposed consumers can be minimized directly. A hydrant strategy, which specifies the location of hydrants and the timing for opening and closing hydrants, is identified to best protect consumers from exposure. The model formulation is as follows:

$$\underset{h, t_s, t_d}{\text{Minimize}} \; C_e + V \tag{1}$$

$$\text{subject to } V = \left\{ \begin{array}{ll} -\text{Penalty} \times P_{\min} & \text{if } P_{\min} \leq 0 \\ 0 & \text{otherwise} \end{array} \right\} \tag{2}$$

The objective of the optimization model is to minimize $C_e$, which is the number of consumers exposed above the critical dose. The number of exposed consumers is

calculated using the ABM simulation framework and is a function of the decision vectors, including the identification of a set of hydrants that should be opened to flush the system ($h = h_1, h_2, \dots h_n$), the time step at which each hydrant is opened ($t_s = t_{s,1}, t_{s,2}, \dots t_{s,n}$), and the duration, or number of time steps, that each hydrant would remain open ($t_d = t_{d,1}, t_{d,2}, \dots t_{d,n}$). A maximum number of hydrants, $n$, may be opened at any time during the simulation. All decision variables are represented as integer values. The parameter $P_{\min}$ represents the minimum pressure at any time and at any node in the network during the simulation period. A linear penalty $V$ (Equation (2)) is added to the objective function to enforce that the minimum pressure remains above an allowable pressure. *Penalty* is a coefficient for the violation, and is set equal to five for this study.

This research develops a simulation–optimization approach to identify effective response plans which are selected from a large set of potential solutions. Highly fit solutions are selected through a framework that couples agent-based and water distribution models with a heuristic optimization methodology (Figure 1). Sets of solutions are generated by the optimization methodology, and each solution specifies a hydrant strategy, or a vector of locations, times, and durations for opening hydrants (e.g., $h$, $t_d$, and $t_s$). To evaluate a solution, these values are encoded in the input data for the water distribution model as a set of additional demands at hydrant nodes, with beginning and start times for the additional demand as specified by that solution. The ABM is executed for each hydrant strategy to
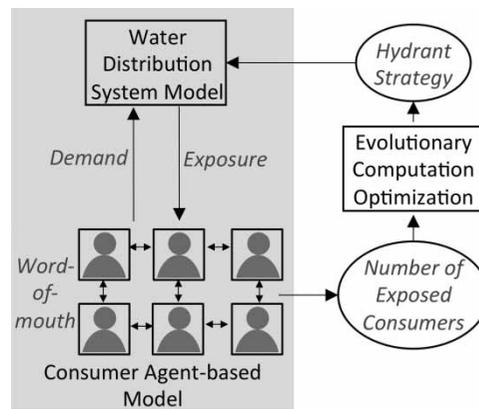


**Figure 1** | Flowchart representing the simulation–optimization approach. The grey box encloses the ABM framework that couples agent simulation with a water distribution system model.

simultaneously simulate the interactions of the consumer agents, the propagation of the contaminant plume in the network, and flushing the contaminant through opening hydrants. At the completion of one simulation, the pressures at all the nodes are evaluated to calculate any violation of the constraint (Equation (2)), and the number of exposed consumers is tallied to represent the objective function. These data are passed to the optimization algorithm, which assigns a fitness value to the solution. Each solution that is generated by the optimization algorithm is evaluated using this approach. The optimization algorithm evaluates a large set of solutions (a population), and over many iterations, improves the solution vectors to converge the population of solutions to a nearly-optimal set of values.

Optimization approaches developed in previous work use gradient-based methods (Baranowski & LeBoeuf 2006), and GAs (Jeong & Abraham 2006; Poulin *et al.* 2008; Guidorzi *et al.* 2009; Alfonso *et al.* 2010) to optimize response plans. The formulation for this problem was developed and solved using the OptQuest algorithm (XJ Technologies 2010) to identify hydrant strategies in a preliminary study (Zechman 2010). The OptQuest algorithm uses a small population of solutions and a local search; it is able to identify good solutions quickly, but does not routinely return highly fit solutions. The research presented here explores a more robust solution approach to identify hydrant strategies through the use of evolutionary computation-based algorithms. An ES-based approach that was developed for directing a search among nodes in a water distribution network is described, further developed, and compared with existing GA approaches here.

## Genetic algorithm

The GA (Goldberg 1989) has gained prominence in solving water-resource related problems (Nicklow *et al.* 2010). A GA uses a population of solutions that converge to a near-optimal solution after iterating through several generations. The archetypical GA combines solutions that may be quite diverse in their solution characteristics to produce two solutions that represent a combination of the parent solutions. Especially during early generations, the GA takes broad steps through the decision space using the crossover operator. Mutation is applied at a very small rate to

change decision variables of a few solutions to new random values and inject diversity in the population. Binary tournament selection is applied to select good solutions with some stochasticity; less fit solutions have a non-zero probability of surviving to the next generation. Two crossover approaches, a uniform crossover and an arithmetic crossover, are implemented and compared for solving the hydrant strategy problem. An individual represents the decision vector as an array of discrete variables ($h$, $t_s$, and $t_d$) using integer values.

## Evolution strategies

ES (Rechenberg 1973; Schwefel 1981) conducts a search similar to a GA by using a population of solutions, which converges over many generations to a near-global optimal solution. ES begins with an initial population of $\mu$ solutions and generates a set of $\lambda$ new solutions by applying a mutation operator that makes incremental changes to decision variables and creates a search that moves in a step-wise approach through the decision space.

### Mutation

To mutate the value of a gene, a normal distribution is sampled to generate a new random value. The current value of the gene serves as the mean of the distribution, and the standard deviation of the distribution is determined adaptively. Using an adaptive mutation operator (Schwefel 1995), the standard deviation is specified as an additional gene for each decision variable and is mutated with the other genes in the solution. Therefore, two genes are required to represent any decision variable: the first gene represents the value of the variable, $x_i$, and the second gene represents the standard deviation, $\sigma_{xi}$, that is used to mutate $x_i$. At each generation, when a solution is selected for mutation, $\sigma_{xi}$ is mutated, and then used as the standard deviation to mutate $x_i$. As the fitness of an individual improves, it will be reproduced more, and many of its offspring will be generated with small standard deviation values, which ensure that smaller steps will be taken in the decision space as the population converges. The array of genes for the hydrant optimization problem includes the vectors of genes $h$, $t_s$, and $t_d$. These decision variables are

represented as integers, and any real-valued number that is generated through mutation is rounded to the nearest integer. To enable adaptive mutation, the second array of $3n$ genes are represented as $\sigma_h$, $\sigma_{ts}$ and $\sigma_{td}$. The vector $\sigma_h$, is represented as integer values, and $\sigma_{ts}$, and $\sigma_{td}$ are real numbers.

Zechman & Ranjithan (2009) developed an ES-based mutation operator for searching among nodes in a water distribution system. This operator uses a graph of nodes that are connected by one link, or pipe, in a water distribution system. Decision variables ($h$) represent nodes, and the associated mutation variables ($\sigma_h$) represent the number of links that can be explored to generate a mutated decision variable. A mutation parameter for $\sigma_{h,i}$ of two, for example, indicates that a new node can be located within two steps away from the current node, $h_i$. For the hydrant optimization problem, a graph of connectivity among hydrants is created that specifies which hydrants are hydraulically connected. Two hydrants are considered hydraulically connected where there is no other hydrant in the shortest flow path between the pair of hydrants. This mutation operator is based on the mutation operator that was developed for identifying the source of contamination in a water distribution system (Zechman & Ranjithan 2009; Drake & Zechman 2011), and it has been applied to resize pipes in a water distribution system to meet fire flow requirements (Kanta et al. 2012) and route siren vehicles during a water contamination event (Shafiee & Zechman 2011b).

### Selection

In a typical ES, the selection operator is deterministic, ensuring that one copy of each of the best $\mu$ solutions in a new population survive to the next generation. The elitist graduated overselection strategy (Fernandez & Evett 1997) allows quicker convergence than a deterministic selection. To execute elitist overgraduate selection, the parent ($\mu$) and children ($\lambda$) solutions are ranked based on their fitness values. A pool of candidate solutions is created, where the size of the pool is specified by the user as the parameter, *upper quantile*. For example, the size of the upper quantile may be 10–20% of the array of $\mu + \lambda$ solutions. The top ranked solutions are placed in the candidate pool, and one solution is selected from the candidate pool with uniform probability to survive to the next generation. The solution

that is selected is replaced in the pool. To select the next individual, the size of the candidate pool is increased by one, and the next best solution is added to the pool. The effect of this selection is that several duplicates of highly fit solutions may appear in the next generation. This encourages more exploitation around highly fit individuals, as the top ranked solutions have a higher probability of being selected. Overselection also allows a small probability for less fit solutions to survive, which can increase the diversity in a population that is becoming dominated by a few solutions. For example, under deterministic selection, the top $\mu$ solutions are allowed to survive, while in overselection, the top ($\mu$ + upper quantile) solutions are given a non-zero chance of surviving.

## MICROPOLIS: A SMALL VIRTUAL CITY

Micropolis is a virtual city that is used to demonstrate the optimization of response actions using the CAS framework (Brumbelow et al. 2007). The water distribution system is comprised of 575 water mains, 52 hydrants, eight pumps, two reservoirs, and one tank (Figure 2). Demands are exerted at 1,236 residential, industrial, and commercial/institutional nodes, representing a small community of 5,232 residents. Each resident is simulated as an agent. Agents ingest water, become exposed, change drinking patterns, communicate with other agents, and move between residential and non-residential nodes as described above during the simulation. Consumers are assigned characteristics as shown in Table 1.

### Contaminant event simulation

A contaminant is introduced at the water tank, which is located in the central section of the city (Figure 2). The contaminant is simulated as conservative, and a dose of 9,000 kg is injected over a 6-hour period, beginning at 12 a.m. The event is simulated for a period of 48 hours after the contaminant is injected.

The contamination event is first simulated without any mitigation actions. To account for stochasticity in the simulation, the event is simulated 30 times, and the range of exposure for three cases is shown in Figure 3. The event is
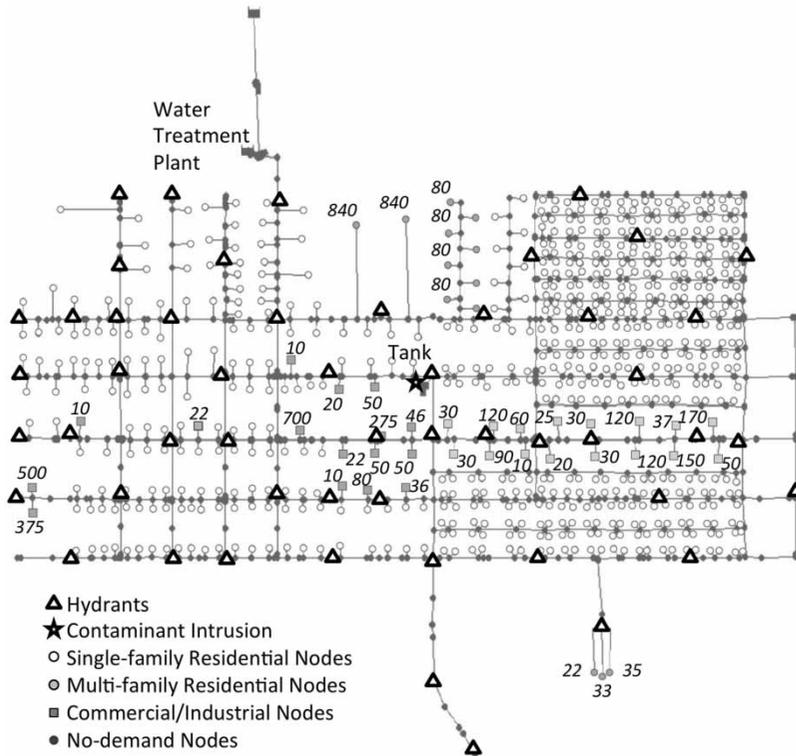
**Figure 2** | Micropolis network and location of consumers. Labels indicate maximum population over a 24-hour time period at commercial/industrial nodes and multi-family residential nodes. Single-family residential nodes are initialized with 1–6 consumers at each node.

**Table 1** | Consumer characteristics. Time to respond indicates the time elapsed before a consumer changes water demands after exposure or word-of-mouth (WOM)

| Parameter | Value |
|---|---|
| Critical dose | 0.25 mg |
| Time to respond after exposed | $U(1,6)$ hours |
| Time to respond to WOM | $U(1,6)$ hours |
| Positive response to family WOM | 100% |
| Positive response to other WOM | 75% |



**Figure 3** | Number of exposed consumers, as the event progresses, as modeled using the ABM framework. Bounds represent the upper and lower limit of predicted exposure for thirty simulations of the event.

simulated without protective actions (reduction in demands) and without word-of-mouth, and the average number of exposed consumers over 30 simulations is 2,484. The same event is modeled with protective actions of the consumers, but without word-of-mouth. The number of exposed consumers increases to an average of 3,128, as those consumers who are exposed stop drinking, leaving more contaminant in the system. In this case, exposed consumers do not warn others, so that un-exposed consumers continue to drink and gradually consume enough contaminant to
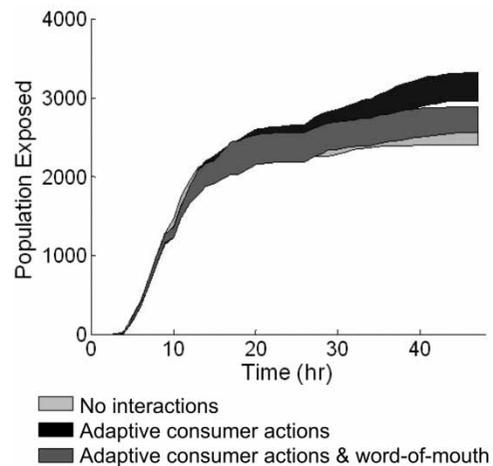
become exposed. Finally, by including all behaviors of the consumers, exposure is reduced due to the word-of-mouth mechanism, and the average number of exposed consumers over 30 simulations of the event is 2,694. The propagation of

the contaminant plume is altered by these interactions; as the information about the event spreads through the network of consumers, demands are reduced further. More contaminant remains in the network, and flow directions and volumes are altered, compared with what would be predicted by an engineering model alone. Further sensitivity analysis of the ABM framework for the Micropolis system is detailed by Zechman (2010, 2011).

# RESULTS

Both the GA-based and ES-based approaches are applied to solve the hydrant optimization problem. Up to five hydrants can be opened at any time after 1:00 a.m. (1 hour after the contaminant is introduced to the system) and the optimization model identifies the hydrants that should be opened, the time at which each one should be opened, and the duration for which each one remains open, up to 10 hours. A limit on the flushing duration was set through preliminary trials, which demonstrated that opening most hydrants for more than 10 consecutive hours produces negative pressures in the pipe network. To simulate an open hydrant, a new demand of 500 gpm is simulated at a hydrant node. A constraint (Equation (2)) included in the model formulation ensures that the pressure remains in the allowable range of pressures.

Four algorithms are tested, including the GA using arithmetic crossover; GA using uniform crossover; ES using deterministic selection; and ES using elitist graduated overselection. The algorithmic settings are listed in Tables 2 and 3 for the GA and ES, respectively. The generation number and population size were consistent throughout the trials to allow direct comparison between the algorithms. Each execution of the ABM simulation takes 100 s on a desktop personal computer, and each execution of an optimization algorithm requires 2–3 days. Due to the computational

**Table 2** │ GA settings

| Algorithm parameter | Value |
| --- | --- |
| Population size | 50 |
| Generations | 50 |
| Crossover rate | 80% |
| Mutation rate | 1% |

**Table 3** │ ES settings

| Algorithm parameter | Value |
| --- | --- |
| Population size ($\mu$) | 50 |
| Generations | 50 |
| Number of children ($\lambda$) | 50 |
| Upper quantile | 10 |

burden of the simulation–optimization approach, each algorithm was executed for ten random trials. Using the same population size for each of the four algorithms allows direct comparison of the results.

Due to the probabilistic nature of the ABM of consumer behavior, there is stochasticity in the performance of any one solution. In preliminary studies (Zechman 2010), a noisy optimization approach was tested, where each solution is evaluated for ten realizations, and the fitness is calculated as the average of the realizations. This approach increases the computational burden and does not result in solutions with more robust performance. Therefore, each solution is evaluated within the optimization process in the present formulation for only one realization. Once the algorithm has terminated, the best solution in the population is post-processed and is evaluated for thirty realizations. The average of the thirty realizations is defined as the fitness of the best solution, and the average fitness across the 10 best solutions generated through 10 trials is reported for each algorithm (Figure 4). The solutions identified by ES are
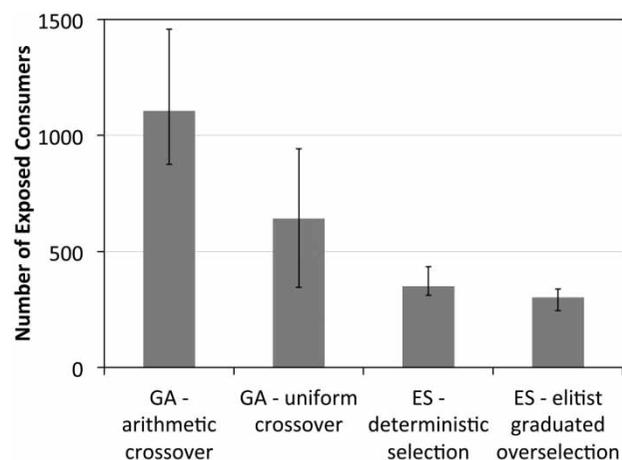


**Figure 4** │ Average fitness value for ten trials for each algorithm. The fitness of one trial is the average of thirty realizations of the best solution. Error bars indicate range of performance over ten trials.

able to reduce the number of exposed consumers lower than those solutions identified by the GA, using either uniform or arithmetic crossover. Using the elitist graduated overselection reduces the average fitness further. Based on hypothesis tests (Milton & Arnold 1995), uniform crossover finds a better-performing solution than arithmetic crossover with a 99% confidence interval; ES using deterministic selection performs better than GA using uniform crossover with a 99% confidence interval; and elitist graduate overselection performs better than deterministic selection with a 98% confidence interval.

Convergence plots for a representative run of each of the four algorithms provide insight about the performance of each algorithm (Figure 5). Using an arithmetic crossover (Figure 5 top left) does not lead the GA towards improved solutions, as the nodes are represented as indices, and finding a weighted average of two indices does not necessarily lead to a new index that is topographically or hydraulically similar to either of the parent indices. The arithmetic operator generates an excessive amount of undirected exploration of the decision space, as can be seen in the

early generations where the population of solutions has a wide variation in objective function values (Figure 5 top left). In later generations, two different solutions have emerged in the population with similarly good fitness values. Due, however, to the inefficiency of the crossover operator, the two solutions are not combined to generate a new, fitter solution.

The uniform crossover operator limits exploration of the decision space better than the arithmetic operator (Figure 5 top right). The limitation of the uniform crossover is that while existing values for genes (i.e., node numbers) are swapped among solutions, no new values are created for decision variables beyond the values that are present in the initial population at the first generation. Only the mutation operator can introduce new genetic material to the population of solutions. As shown by the convergence plot, the search is susceptible to premature convergence to local optima.

The ES search with deterministic selection shows a smoother convergence (Figure 5 bottom left). At each generation, 50 solutions are mutated, and the objective values
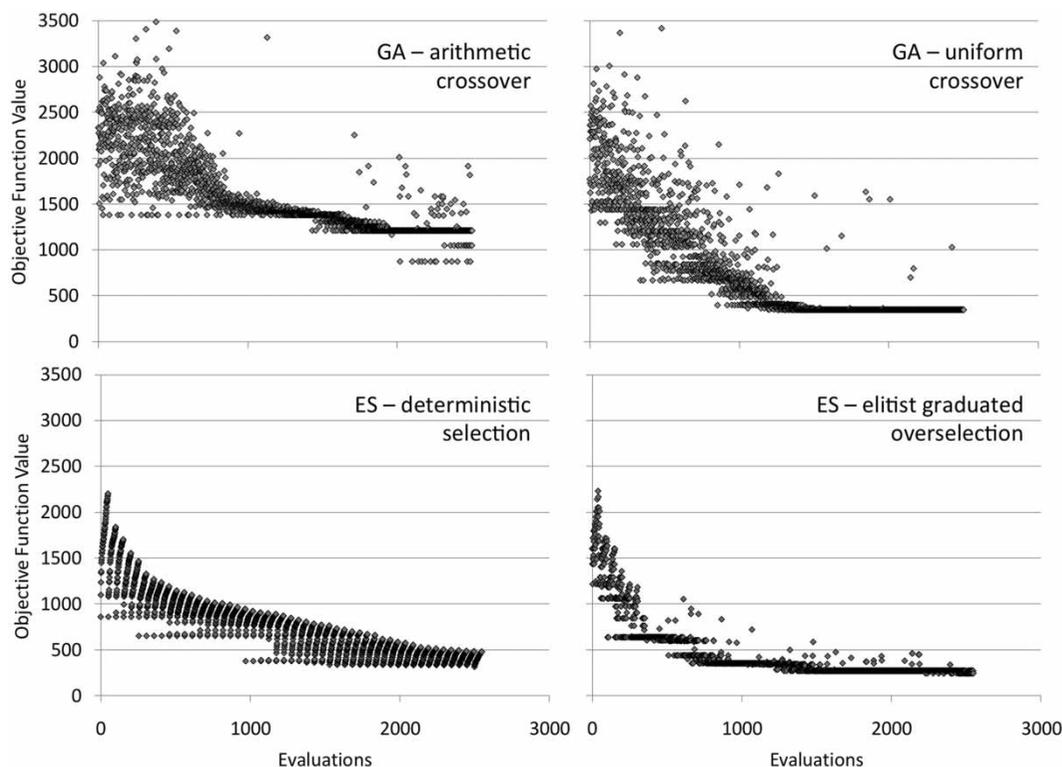


**Figure 5** | Convergence of four algorithms, shown for the best trial for each algorithm.

show the incremental nature of the search as it progresses through generations. The substitution of the elitist graduated overselection strategy (Figure 5 bottom right) significantly changes the convergence behavior. As can be seen by repeated objective values, several copies of one solution survive each generation, and the search focuses on highly fit solutions early in the search. Although this extra selection pressure runs the risk of premature convergence, it is able to focus the search in areas of good solutions.

## Delays in response planning

Ideally, a utility operator should respond as quickly as possible to flush contaminants through opening hydrants. It is expected, however, that there would be a delay in the response, and a utility operator may not have enough information to respond immediately. Additional response plans are identified here for utility managers who respond after some delay. Four settings are simulated, where hydrants are opened after 3:00, 5:00, 7:00, and 9:00 a.m. For each setting, the ES with overselection was executed for five random trials.

As the contaminant is not controlled as early, the number of exposed consumers increases with the increased delays in response (Figure 6). The best solution found for each setting is shown in Figure 7, with details for the time and location for opening and closing each hydrant. For the 1:00, 3:00, and 5:00 a.m. scenarios, each solution in the set (10 solutions for 1:00 a.m. and
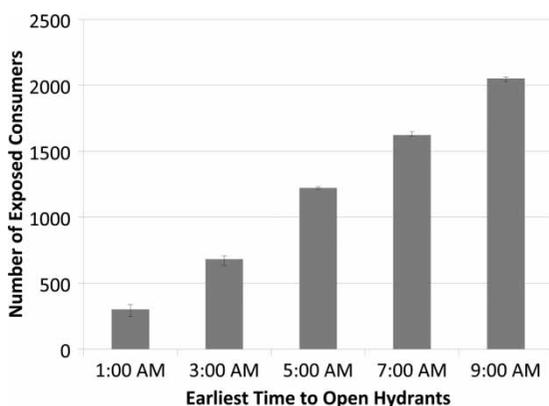
five solutions for 3:00 and 5:00 a.m.) had one hydrant in common (shown in bold in Figure 7). The common hydrant is, however, not the same for the different delay scenarios. In general, hydrant strategies specify that centrally located hydrants should be opened for early responses; as the delay in response increases and the contaminant moves through the system, the hydrants that should be opened are located further from the source. As shown in Figure 6, the variability in the performance of the set of solutions decreases with increasing delays. For hydrant strategies that can be executed quickly, there may be many potential combinations of hydrants that could be opened, while for more delayed hydrant strategies, the number of consumers that can be protected is limited, and only a few strategies exist.

Sensitivity analysis was conducted to determine the sensitivity of a solution to the number of hydrants that should be opened by ranking the hydrants based on their contribution to the number of consumers protected (Figure 8). For the best solution for a 1:00 a.m. response time, each of the five hydrants was simulated alone to determine the most effective hydrant. Each hydrant was simulated for 30 realizations, and the average performance is reported as the fitness. To determine the second most effective hydrant, each of the remaining four hydrants was tested in combination with the first most effective hydrant for 30 realizations, and so forth. This analysis was repeated for the best solutions corresponding to 3:00, 5:00, 7:00, and 9:00 a.m. strategies. For the 1:00, 3:00, and 5:00 a.m. strategies, Hydrants 1–4 have added value in protecting more consumers. For the 7:00 a.m. strategy, Hydrants 1–3 add value to the solution, and for the 9:00 a.m. strategy, only Hydrants 1 and 2 are effective. These results are compared with the average number of consumers exposed when no hydrants are opened (2,694 exposed consumers). As shown by the 9:00 a.m. strategy, as a utility responds later to an event, actions are less effective, and only two hydrants are utilized in protecting the greatest number of consumers. In addition, it does not appear that more than five hydrants can be utilized for the early response strategies, regardless of the speed of response. In this problem formulation, a hydrant can remain open for 10 hours at most; this setting could be explored to allow a hydrant to remain open for a longer time to increase the number of protected consumers.
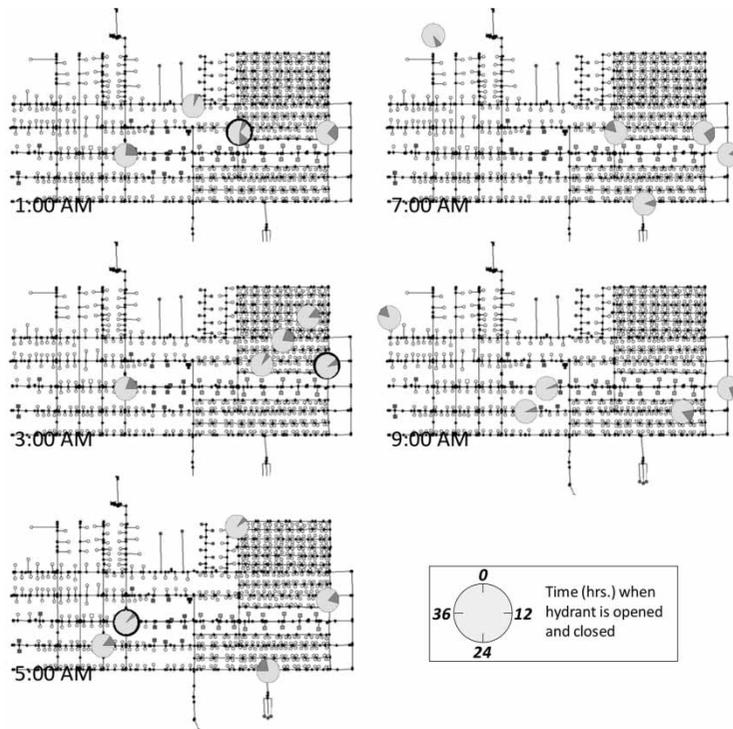


**Figure 6** | Average fitness for solutions for delays in response times. Error bars indicate the range in the number of exposed consumers across five random trials.

**Figure 7** | Hydrant opening strategy corresponding to the best solution for different delays. Bold hydrant indicates the decision that appears in all solutions for response times of 1:00, 3:00, and 5:00 a.m.
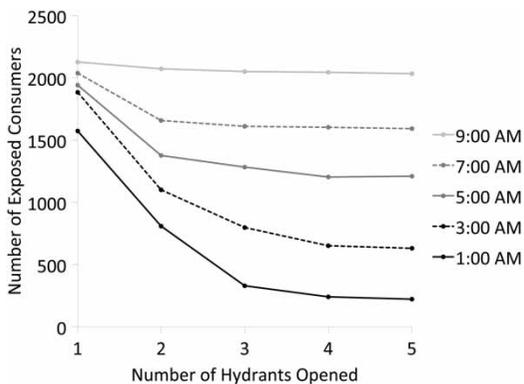


**Figure 8** | Sensitivity analysis of performance of opening sub-sets of hydrants. Results are shown for the best solution for each response time (solutions shown in Figure 7). Number of exposed consumers (y-axis) is the average of 30 realizations of each strategy.

## CONCLUSIONS AND DISCUSSION

This research demonstrates the utilization of a newly developed CAS model of water distribution systems in a simulation–optimization framework to provide a methodology for identifying the most effective response to contamination of a water distribution system. The CAS model generates new analysis for water contamination threat events by simulating the adaptive and dynamic behaviors of consumers and bypasses the conventional assumption that consumer demands remain constant during an emergency. While calibration exercises may not be possible for most water distribution contamination events due to the lack of data that would describe such an event, the modeling framework can be used to compare 'what-if' scenarios, explore the potential impacts of consumer behaviors and adaptations, and simulate a range of probable health impacts given variability in consumer choices. The CAS model provides a count of the number of exposed consumers and is coupled with evolutionary algorithm-based approaches to minimize public health consequences by specifying hydrants that should be opened to flush contaminants from the system. The purpose of this framework is to assist utility managers in developing response plans for use in the event of contamination. As expected, longer delays in response decrease protection of public health, and when actions are taken later in the event, a

smaller number of hydrants can be effectively utilized to reduce the number of exposed consumers.

Two evolutionary computation-based approaches were tested here for identifying hydrant strategies. An ES-based approach with a specialized mutation operator designed for network-based problem domains was demonstrated to perform significantly better than two conventional GA-based approaches. While the network mutation operator has been demonstrated and applied in previous works, here it is further developed for identifying hydrants, rather than intermediate nodes, and is formally compared with a GA-based approach. The ES-based search has many other applications in solving water distribution design and management problems, along with potential for application in other network-based domains.

For the case study examined here, there are several combinations of hydrants that could be opened to result in the same number of protected consumers. At the same time, many hydrant combinations could increase the number of exposed consumers beyond the number of consumers who would be exposed without any response strategies. For example, the GA with arithmetic crossover identifies strategies that expose more than 3,000 consumers (Figure 5, upper left), compared with an average of 2,694 consumers who are exposed when the utility manager does not take any response action at all. It is critical, therefore, to design a strategy for opening hydrants that will not expose additional consumers, but maximize public health protection. Due to the complexities of hydraulic flows in a water distribution system and the changing hydraulics caused by consumer actions, a formal optimization algorithm is needed and is demonstrated here to identify the most effective hydrant strategies.

On-going research is investigating development and application of the coupled ABM-evolutionary computation framework to manage threats for a larger, more realistic municipality (Shafiee & Zechman 2011a, b, 2012; Zechman et al. 2011). The framework described here demonstrates a new approach that can be taken to assist utility operators in responding to hazardous events. Further methodological development and faster computational implementations can be developed to enable an ABM-optimization framework for application in a real-time approach to find optimal response plans once a contaminant source has

been identified. A more accurate approach for modeling hydrants uses a pressure-driven approach, and on-going work is modifying the simulation to include pressure-driven simulation for new case studies, building on the flexible methodology and framework that is developed here. The results identified here are specific to this contamination event; events of larger doses or longer durations may require more hydrants for mitigation. To develop an understanding of how to proceed in a hazardous event without immediate identification of a contaminant source, a library of response actions can be developed by optimizing response plans *a priori* for a large set of contaminant event scenarios. Future work will couple the framework described here with additional methodologies to develop a more comprehensive threat management planning toolkit.

## ACKNOWLEDGEMENTS

## REFERENCES

Alfonso, L., Jonoski, A. & Solomatine, D. 2010 Multiobjective optimization of operational responses for contamination flushing in water distribution networks. *Journal of Water Resources Planning and Management* **136**, 48–58.

Axelrod, R. 1997 *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton University Press, Princeton, USA.

Baranowski, T. M. & LeBoeuf, E. J. 2006 Consequence management detection optimization for contaminant and isolation. *Journal of Water Resources Planning and Management* **132**, 274–282.

Baranowski, T. M. & LeBoeuf, E. J. 2008 Consequence management utilizing optimization. *Journal of Water Resources Planning and Management* **134**, 386–394.

Brumbelow, K., Torres, J., Guikema, S. D., Bristow, E. & Kanta, L. 2007 Virtual cities for water distribution and infrastructure

system research. *Proceedings of the ASCE/EWRI World Environmental and Water Resources Congress*, Tampa, FL, May 2007.

Davis, M. & Janke, R. 2008 Importance of exposure model in estimating impacts when a water distribution system is contaminated. *Journal of Water Resources Planning and Management* **134**, 449–456.

Di Cristo, C. & Leopardi, A. 2008 Pollution source identification of accidental contamination in water distribution networks. *Journal of Water Resources Planning and Management* **134**, 197–202.

Drake, K. & Zechman, E. M. 2011 Using niched co-evolution strategies to address non-uniqueness in characterizing the source of contamination in a water distribution system. *Proceedings of the ASCE/EWRI World Environmental and Water Resources Congress*, Palm Springs, CA, May 2011.

Fernandez, T. & Evett, M. 1997 The impact of training period size on the evolution of financial trading systems. *Proceedings of the Genetic Programming Second Annual Conference*, AAAI, Stanford, CA, July, 1997.

Geldreich, E. 1991 Microbial water quality concerns for water supply use. *Environmental Toxicology Water Quality* **6**, 209–223.

Goldberg, D. 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA.

Guan, J., Aral, M., Maslia, M. & Grayman, W. 2006 Identification of contaminant sources in water distribution systems using simulation-optimization method: case study. *Journal of Water Resources Planning and Management* **132**, 252–262.

Guidorzi, M., Franchini, M. & Alvisi, S. 2009 A multi-objective approach for detecting and responding to accidental and intentional contamination events in water distribution systems. *Urban Water Journal* **6**, 115–135.

Hart, W. E. & Murray, R. 2010 Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. *Journal of Water Resources Planning and Management* **136**, 611–619.

Holland, J. 1975 *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, Ann Arbour, USA.

Holland, J. 1995 *Hidden Order: How Adaptation Builds Complexity*. Addison-Wesley, Reading, MA, USA.

Hrudey, S. E. & Hrudey, E. J. 2004 *Safe Drinking Water: Lessons from Recent Outbreaks in Affluent Nations*. IWA Publishing, London.

Irby, K. 2005 Managing water quality through uniform directional flushing. In: *Water Quality in the Distribution System* (W. Lauer, eds.). American Water Works Association, Denver, Colorado, pp. 885–897.

Jeong, H. S. & Abraham, D. M. 2006 Operational response model for physically attacked water networks using NSGA-II. *Journal of Computing in Civil Engineering* **20**, 328–388.

Kanta, L., Zechman, E. M. & Brumbelow, K. 2012 A multi-objective evolutionary computation approach for redesigning water distribution systems to provide fire flows. *Journal of Water Resources Planning and Management* **138**, 144–152.

Kroll, D. 2006 *Securing our Water Supply: Protecting a Vulnerable Resource*. PenWell Corporation, Tulsa, Oklahoma.

Miller, J. & Page, S. 2007 *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press, Princeton, USA.

Milton, J. S. & Arnold, J. C. 1995 *Introduction to Probability and Statistics*. McGraw-Hill, Inc., New York.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A. & Zechman, E. M. 2010 State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management* **136**, 412–432.

Perelman, L. & Ostfeld, A. 2010 Extreme impact contamination events sampling for water distribution systems security. *Journal of Water Resources Planning and Management* **136**, 80–87.

Poulin, A., Mailhot, A., Grondin, P., Delorme, L., Periche, N. & Villeneuve, J.-P. 2008 Heuristic approach for operational response to drinking water contamination. *Journal of Water Resources Planning and Management* **134**, 457–465.

Poulin, A., Mailhot, A., Periche, N., Delorme, L., Periche, N. & Villeneuve, J.-P. 2010 Planning unidirectional flushing operations as a response to drinking water distribution system contamination. *Journal of Water Resources Planning and Management* **136**, 647–657.

Preis, A. & Ostfeld, A. 2008 Multiobjective contaminant response modeling for water distribution systems security. *Journal of Hydroinformatics* **10**, 267–274.

Rossman, L. 2000 EPANET user's manual. Technical report, U.S. Environmental Protection Agency Risk Reduction Engineering Lab, Cincinnati, Ohio.

Rechenberg, I. 1973 *Evolutionsstrategie: Optimierung technischer Systeme und Prinzipien derbiologischen Evolution*. Frommann-Holzboog, Stuttgart.

Schwefel, H.-P. 1981 *Numerical Optimization of Computer Models*. Wiley, Chichester, UK.

Schwefel, H.-P. 1995 *Evolution and Optimum Seeking*. Wiley, New York.

Shafiee, M. & Zechman, E. M. 2010 An agent-based modeling approach for simulating contamination events applied to the mesopolis water distribution system. *Proceedings of the ASCE/EWRI World Environmental and Water Resources Congress*, Providence, RI, May 2010.

Shafiee, M. & Zechman, E. M. 2011a An agent-based modeling approach to evaluate protective action strategies in a water distribution contamination event. *Proceedings of the ASCE/EWRI World Environmental and Water Resources Congress*, Palm Springs, CA, May 2011.

Shafiee, M. & Zechman, E. M. 2011b Sociotechnical simulation and evolutionary algorithm optimization for routing siren vehicles in a water distribution contamination event. *Proceedings of the Genetic and Evolutionary Computation Conference 2011 (GECCO), Evolutionary Computation and*

*Multi-agent Systems and Simulation (ECoMASS) Workshop*, Dublin, Ireland, July 2011.

Shafiee, E. & Zechman, E. M. 2012 Integrating evolutionary computation and sociotechnical simulation for flushing contaminated water distribution systems. *Proceedings of the Genetic and Evolutionary Computation Conference 2012, Evolutionary Computation and Multi-agent Systems and Simulation (ECoMASS) Workshop*, Philadelphia, PA, July 2012.

U.S. Environmental Protection Agency 2000 Estimated per capita water ingestion in the United States. Technical Report EPA-8222 00-008, U.S. Environmental Protection Agency Office of Water, Washington, DC.

U.S. Environmental Protection Agency 2001 Potential contamination due to cross-connections and backflow and the associate health risks. Technical Report, Office of Water (4601M), Office of Ground Water and Drinking Water, Distribution System Issue Paper, Washington, DC.

Vickers, A. 2001 *Handbook of Water Use and Conservation*. Waterplow Press, Amherst, MA.

Watts 1998 *Stop Backflow News: Case Histories and Solutions*. Watts Backflow Prevention Products, Andover, MA.

XJ Technologies 2010 Anylogic 6 User's Manual, http://www.anylogic.com/

Zechman, E. M. 2010 Integrating complex adaptive system simulation and evolutionary computation to support water infrastructure threat management. *Proceedings of the Genetic and Evolutionary Computation Conference 2010 (GECCO), Evolutionary Computation and Multi-agent Systems and Simulation (ECoMASS) Workshop*, Portland, OR, July 2010.

Zechman, E. M. 2011 Agent-based modeling to simulate contamination events and evaluate threat management strategies in water distribution systems. *Risk Analysis* **31**, 758–772.

Zechman, E. M. & Ranjithan, S. 2009 Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system. *Journal of Water Resources Planning and Management* **135**, 334–343.

Zechman, E. M., Brumbelow, K., Lindell, M., Mumpower, J., Rasekh, A. & Shafiee, M. 2011 Agent-based modeling for planning emergency response to contamination emergencies in water utilities. *NSF CMMI Engineering Research and Innovation Conference*, Atlanta, GA, January 2011.