RISK ANALYSIS AND ADAPTIVE RESPONSE PLANNING FOR

WATER DISTRIBUTION SYSTEMS CONTAMINATION

EMERGENCY MANAGEMENT

A Dissertation

by

AMIN RASEKH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Civil Engineering

Risk Analysis and Adaptive Response Planning for Water

Distribution Systems Contamination Emergency Management

RISK ANALYSIS AND ADAPTIVE RESPONSE PLANNING FOR

WATER DISTRIBUTION SYSTEMS CONTAMINATION

EMERGENCY MANAGEMENT

A Dissertation

by

AMIN RASEKH

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	James K. Brumbelow
Committee Members,	Emily M. Zechman
	Paolo Gardoni
	Michael K. Lindell
Head of Department,	John Niedzwecki

August 2012

Major Subject: Civil Engineering

ABSTRACT

Risk Analysis and Adaptive Response Planning for Water Distribution Systems

Contamination Emergency Management. (August 2012)

Amin Rasekh, B.S., Civil Aviation Technology College;

M.S., Iran University of Science and Technology

Chair of Advisory Committee: Dr. James K. Brumbelow

Drinking water distribution systems (WDSs) hold a particularly critical and strategic position in preserving public health and industrial growth. Despite the ubiquity of this infrastructure, its importance for public health, and increased risk of terrorism, several aspects of emergency management for WDSs remain at an undeveloped stage. A set of methods is developed to analyze the risk and consequences of WDS contamination events and develop emergency response support tools.

Monte Carlo and optimization schemes are developed to evaluate contamination risk of WDSs for generation of critical contamination scenarios. A multicriteria optimization approach is proposed that treats likelihood and consequences as independent risk measures to find an ensemble of uniformly-distributed critical scenarios. This approach provides insight into system risk and potential mitigation options not available under maximum risk or maximum consequences analyses.

Static multiobjective simulation-optimization schemes are developed for generation of optimal response mechanisms for contamination incidents with two

conflicting objectives of minimization of health consequences and impacts on non-consumptive water uses. Performance of contaminant flushing and containment are investigated. Pressure-driven hydraulic analysis is performed to simulate the complicated system hydraulics under pressure-deficit conditions.

Performance of a novel preventive response action – injection of food-grade dye directly into drinking water – for mitigation of health impacts as a contamination threat unfolds is explored. The emergency response is formulated as a multiobjective optimization problem for the minimization of risks to life with minimum false warning and cost. A multiobjective optimization scheme is used for the management of contamination events for diverse contaminant agents without interruption of firefighting.

A dynamic modeling scheme is developed that accounts for the time-varying behavior of the system during an emergency. Effects of actions taken by the managers and consumers as well as the changing perceived contaminant source attributes are included in the simulation model to provide a realistic picture of the dynamic environment. A dynamic optimization scheme is coupled with the simulation model to identify and update the optimal response recommendations during the emergency.

Machine learning approaches are employed for real-time characterization of contaminant sources and identification of effective response strategies for a timely and effective response to contamination incidents and threats. In contrast to traditional approaches that perform whole analysis after a contamination event occurs, proposed machine learning methods gain system knowledge in advance and use this extracted information to identify contamination attributes after an incident occurs.

To my father and mother

for their endless love, inspiration, and support

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my advisor, Dr. Brumbelow, for his guidance and support during my doctoral studies at Texas A&M University. I especially appreciate him for motivating me to approach challenging problems with innovative solutions and encouraging me to explore my own directions during this research. I sincerely thank him for being supportive, considerate, and patient particularly during the difficult periods of this research. I would also like to greatly appreciate Dr. Emily Zechman, Dr. Paolo Gardoni, and Dr. Michael Lindell for their valuable support and guidance as my research advisory committee throughout the period of my studies.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE   Page

LIST OF TABLES

1. INTRODUCTION

Drinking water distribution systems (WDSs) are critical urban infrastructures that are vulnerable to contamination because of their ubiquity, multiple points of access, and aging infrastructure. Contaminants may be introduced into the system either accidentally during a back-flow or cross-connection incident or intentionally through a malevolent attack. These systems have been recognized as one of several critical infrastructures that are vulnerable to terrorism attacks through the Public Health Security and Bioterrorism Preparedness and Response Act (PL 107-188) (U.S. Government Accountability Office 2004). Accordingly, this research project aims to analyze the risk of WDS contamination incidents and develop a comprehensive response planning framework for the emergency management of contamination hazard intrusions into water distribution systems.

Monte Carlo and risk-based optimization schemes are developed to evaluate contamination risk of WDSs for generation of critical scenarios that are representative of the most vulnerable aspects of the system. Defining attributes of contamination scenarios are identified as contaminant type and amount, contamination location, start time, duration, and time of year it occurs. Documented waterborne outbreaks reported in developed nations are analyzed to empirically estimate statistical characteristics of defining attributes in accidental events. Monte Carlo simulation is conducted to determine the probability distribution of public health consequences, aggregate conditional risk, and significance of different scenario attributes. A multiobjective

_____

This dissertation follows the style of *Journal of Water Resources Planning and Management*.

optimization methodology is proposed to capture the attributes of critical accidental contamination scenarios. The principal risk components of likelihood and health consequences are treated as optimization objectives and are maximized simultaneously to identify an ensemble of non-dominated critical scenarios. The multiobjective approach provides insight into system risk and potential mitigation options not available under maximum risk or maximum consequences analyses.

Decisions on protecting public health against possible water contamination threats should be made with careful consideration of credibility of threat observations and unintended impacts of response implementation on water supply system serviceability. Response optimization frameworks are structured to help water utility operators in making such critical decisions during the intense course of an emergency. Pressure-driven hydraulic analysis is performed to simulate the complicated system hydraulics and propagation of a contaminant through water distribution system under pressure-deficit conditions due to the execution of response actions. Application of this analysis approach relaxes the hard constraint of avoiding negative pressures fundamental to demand-driven models which filters out many potentially effective response plans from the search space. Response actions of contaminant containment and flushing operation rules are optimized for achievement of public health protection with minimal service disruption. Sensitivity analyses are conducted to assess optimal response performance for varying response delay, number of hydrants, and intrusion characteristics. Different methods are explored for quantifying impacts on public health and system serviceability and the sensitivity of optimal response plan to these different

formulations is investigated. It is concluded that the analysts must be cautious of potentially misleading risk reduction recommendations that can be offered due to inherent imperfection in quantitative measures of response criteria that are based on preset exposure thresholds.

This study also explores performance of a novel preventive response action – injection of food-grade dye directly into drinking water – for mitigation of health impacts as a contamination threat unfolds. Dye injection acts as an alerting mechanism that discourages public consumption of potentially contaminated water. Considering the uncertainties in threat observations and the imperfection in system understanding, however, the action has potential for costly false alarms. These could occur when contamination has indeed happened but population segments residing in safe regions are mistakenly alerted or when observations of contamination occurrence turn out to be entirely wrong. The emergency response is thus formulated as a static multiobjective optimization problem for the minimization of risks to life with minimum public warning and execution cost.

A dynamic modeling scheme is developed that accounts for the time-varying behavior of the system during an emergency. Effects of actions taken by the managers and consumers as well as the changing perceived contaminant source attributes are included in the simulation model to provide a more realistic picture of the dynamic environment. A multiobjective-based dynamic optimization scheme is coupled with the adaptive system simulation model to identify and continuously update the optimal response recommendations at every stage of the emergency. A major advantage of this

technique is that it eliminates the need for defining *a priori* the proper diversity preservation parameter.

Contamination source identification involves the characterization of the contamination event attributes using threat observations such as sensor network measurements. The defining attributes of a contamination event may include contaminant type, site(s) of contaminant intrusion, contaminant amount, the time of day the contamination event is initiated, and the intrusion duration. Accurate and prompt determination of these attributes is central to validity of impact assessments conducted and effectiveness of response strategies taken. Focusing on high accuracy, past efforts have successfully applied optimization and back-tracking approaches to deal with this critical task. However, these techniques are typically computationally burdensome and thus not acceptably fast, specifically when they are applied to realistically large water distribution networks. This study accordingly investigates performance of machine learning tools for real-time source identification. In contrast to traditional approaches that perform whole analysis after a contamination event occurs, machine learning methods gain system knowledge in advance and use this extracted information to identify contamination attributes after an incident occurs. Machine learning tools are employed to derive emergency response rules from the large set of response optimization results which may not be decipherable during the extraordinarily critical environment of an emergency. Identification of the trends in most reliable response strategies for each trigger event provides an in-depth understanding of the system that

the emergency managers can use for making rapid and reliable decisions during a contamination events.

Applicability and performance of all proposed approaches is demonstrated on water distribution system of Mesopolis virtual city. With a population of approximately 150,000, Mesopolis resembles the intricacy and interconnectedness of real world water distribution networks that help conducting a more realistic evaluation of the structured risk assessment and response planning frameworks.

2. RISK ASSESSMENT TO CHARACTERIZE CRITICAL CONTAMINATION SCENARIOS

2.1 Introduction

Drinking WDSs are critical urban infrastructures that are expected to deliver safe drinking water to consumers with minimal disruption. Accidental contamination of these systems has historically been recognized as a threat to public health worldwide (Hrudey and Hrudey 2004, 2007). Despite recent technological advances, water contamination outbreaks have persisted in developed nations resulting in sickness and mortality (Reynolds et al. 2008). Documented outbreaks are primarily the result of inadequate knowledge of source water hazards, technological failures, failure to treat water, and human error (Craun et al. 2006; Hrudey and Hrudey 2007). Accompanied by the past decade's concerns over terrorism threats, industry attention and research efforts have focused on development of vulnerability mitigation and emergency response plans. While significant activity has been devoted to terrorism threats based upon hypothetical information with no uncertainty estimated, relatively little effort has focused on addressing accidental contamination threats using evidence documented by the public health community. This chapter presents a set of methods for assessment of these threats that incorporates both likelihood and consequences based on past real contamination events.

The risk assessment process is a set of cogent, well-defined, and systemic activities that provide the decisionmaker with a thorough understanding, quantification,

and evaluation of the risk associated with certain natural hazard or man-made threat (Haimes 2009). It aims to answer the three fundamental questions: 1) What can go wrong?, 2) What is the likelihood that it would go wrong?, and 3) If it does occur, what are the consequences? (Kaplan and Garrik 1981). Risk-based mitigation processes include assessment of event probabilities and impacts and seek measures that consciously avoid unintended consequences. Such strategies are suited to policy decisions where a limited budget must be allocated among complex options to form a defense strategy that minimizes the maximum risk from the actions of an attacker or accidents (National Research Council 2008). The Risk Analysis and Management for Critical Asset Protection framework (American Society of Mechanical Engineers 2006) used by the Department of Homeland Security (Moore et al. 2007) and American Water Works Association (Morley 2010) emphasizes the need to consider the worst reasonable case scenarios in risk management. The U.S. Environmental Protection Agency (USEPA) Response Protocol Toolbox (RPT) (USEPA 2003) requires preparation of a response planning matrix which lists scenarios with different levels of credibility, the potential consequences of a threat, and possible response actions along with their impacts on consumers. However, the RPT provides no detailed instructions on how these scenarios and credibility levels can be identified. Characterization of critical scenarios is thus a necessary initial phase in the overall planning framework for risk mitigation and emergency response. It serves as a guide for a water utility by providing a basis for protection system design and a consistent criterion for evaluating the adequacy of such a design. It can help identify vulnerable system elements and prioritize available resources

to assure cost-effective protective measures while ensuring minimal disruption of reliable supply of water. We define a contamination scenario here as encompassing a defined set of attributes, resulting in a specific level of health impacts, and having a specific probability of occurrence.

Several studies have used Monte Carlo simulation (MCS) to estimate the likely health impacts from contamination intrusion into WDSs (Uber et al. 2004; Nilsson et al. 2005; Khanal et al. 2006; Torres et al. 2009; Pasha and Lansey 2010). Khanal et al. (2006) also performed a generalized sensitivity analysis to determine sensitivity of WDS response to dynamic variables of base demand, storage capacity, injection mass, and injection duration. Exposure levels were found to be most sensitive to variations in base demand and injection mass. Davis and Janke (2011) conducted a sensitivity characterization study that included intrusion duration and contaminant amount for a set of actual systems. Perelman and Ostfeld (2010; in press) proposed a method derived from cross entropy for sampling extreme-impact contamination events for the design of contamination warning systems.

Previous studies have been primarily focused on deliberate intrusions and have either ignored the uncertainties in some scenario attributes or constructed hypothetical probability distributions to quantify these uncertainties. In addition, while application of MCS provides a helpful insight into the variability of exposure levels and significance of different scenario attributes given realistic probability distributions are used, it does not guarantee that the critical scenarios are efficiently identified. Risk-based assessment, which considers both scenario likelihood and consequences (Kaplan and Garrick 1981),

is an attractive tool to identify an ensemble of critical contamination scenarios to include events ranging from rare extreme-impact scenarios to more likely scenarios associated with relatively lower consequences but maximum risk (the product of likelihood and consequences). Additionally, risk mitigation plans can be evaluated and compared by relative risk reduction.

The remainder of this chapter is thus organized in three major sections. First, documented major waterborne outbreaks and real water utility demand information are analyzed to determine generally applicable stochastic properties of contamination scenario attributes. Second, the use of MCS to evaluate aggregate risk and identify relative importance of contamination scenario attributes is discussed and illustrated. Third, a risk-based multiobjective optimization methodology is structured to identify a set of non-dominated contamination scenarios ranging from extreme-impact scenarios to most likely events, and the method is demonstrated. Finally, the chapter concludes with discussion of extensions to these methods.

2.2 Quantification of Uncertainty in Contamination Scenario Attributes

A contamination scenario is defined by a set of attributes including: (1) contaminant type, (2) site(s) of contaminant intrusion, (3) contaminant amount, (4) time of year (which can be represented by the surrogate WDS-wide demand multiplier (DM)), (5) the time of day the contamination event is initiated, and (6) the intrusion duration. We base our quantification of most of the stochastic properties of these attributes on analysis of 70 real accidental water contamination events compiled by Hrudey and

Hrudey (2004). All of these events occurred in community water systems in developed nations (U.S., Canada, Western Europe, Japan, Australia, and New Zealand) during the period of 1974-2004. During the almost identical period 1971-2002, Blackburn et al. (2004) report slightly less than 300 contamination events in U.S. community systems with only about half of those having determined etiology. Accounting for relative population differences and assuming similar rates of contamination occurrence, the Hrudey and Hrudey case study set thus represents a significant fraction of worldwide contamination events in community systems in developed nations. The water systems and event characteristics in the case study set include a wide range of utility sizes (300 to 746,000 users), water sources (roughly split evenly between ground and surface water), and contamination sources; no significant cross-correlations of event attributes are apparent. Review of the relative roles of technological failure and human error found that about 56% of events in the study set included some degree of human error in event causation; in about a quarter of events human error was dominant. Thus, the case study set is sufficiently large and wide-ranging for generalizable analysis, and technological advances have not rendered the cases moot.

2.2.1 Site of Contaminant Intrusion

Epidemiological studies often proceed with great sophistication to identify contaminant introduction to source waters far removed from the WDS itself (e.g., animal sources in raw water reservoirs). However, from the standpoint of WDS modeling and management it is operational failure at a water treatment plant (WTP) that should

remove or inactivate contaminants by design, which allows introduction of these contaminants to the WDS. This reality allows a simplified characterization of contaminant intrusion site as either "WTP" or "distribution system (DS)", the latter consisting of the WDS pipe network and storage tanks. The analysis finds 89% of accidental contamination intrusion sites are at WTPs and 11% are in the DS (9% in DS pipes and 2% in DS storage tanks).

2.2.2 Contaminant Type and Amount

In 65 of the Hrudey and Hrudey (2004) case studies, specific pathogen contaminants were identified; none of the events had specified chemical contaminants. Five pathogens have been comparatively more frequent and were selected for this analysis: Giardia lamblia, E. coli, C. jejuni, Cryptosporidium, and Norwalk-like virus (NLV). Occurrence probabilities for these five contaminants were calculated by frequency of occurrence in the data set (Table 2.1).

Amount of introduced contaminant (measured here as the number of infective doses per capita [IDPC]) was inferred for each event by the following backward procedure. The number of reported disease cases was first multiplied by the ratio of water use per capita to tap water intake to estimate the number of doses for each outbreak. This number was then normalized by population served by the WDS. Expected value of number of doses (*E[IDPC]*) for each specific pathogen is finally calculated via averaging normalized number of doses for all outbreaks associated with that pathogen.

Table 2.1. Statistical data for pathogens

Pathogen type	Occurrence probability (%)	E[IDPC]
Giardia	20.6	195
E. coli	15.9	131
C. jejuni	20.6	197
Norwalk-like virus	15.9	239
Cryptosporidium	27.0	*IDPC* quartiles: min 0.29, 25%: 1.0, med. 2.3, 75%: 87, max 412

Uncertainties in IDPC for other systems can be modeled using the exponential distribution with cumulative distribution function (CDF):

$$\text{CDF}(IDPC) = 1 - \exp(-\lambda \cdot IDPC) \tag{2.1}$$

where λ = the reciprocal of the expected value of *IDPC* (1/*E[IDPC]*). Values of *E[IDPC]* for each pathogen are given in Table 2.1. Testing of the exponential distribution for the five pathogen samples using the Kolmogorov-Smirnov test shows acceptance of the derived exponential distributions at the 20% significance level in 4 cases (see Fig. 2.1 for examples). Cryptosporidium cases were not well modeled by the exponential distribution because they tended toward extremes of apparent *IDPC* values;
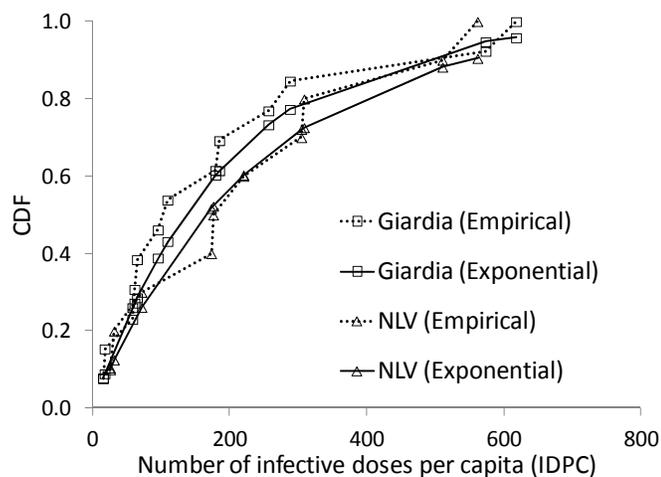
Fig. 2.1. Probability distribution model for contaminant amount

this phenomenon is likely a consequence of the organism's unique properties for morbidity and resilience in the environment (including chlorine-resistance). Table 2.1 includes quartile values of *IDPC* for Cryptosporidium rather than *E[IDPC]*.

This study count a person as sick when the ingested contaminant dose exceeds the infectious does no matter how much. In other words, health impacts for two exposed persons is treated equally when the ingested mass in above infectious dose disregarding the fact that one may have might have significantly ingested more contaminant. To provide a more realistic picture of the problem, on may quantify the amount of contaminant as the number of lethal doses introduced into the system during a contamination event too. This, however, is not feasible to be performed with acceptable level of accuracy since the number of occurred contamination events resulting in deaths is highly scarce.

2.2.3   Time of Year

Aggregate water demand for a WDS typically varies throughout the year with minimum demand in the winter and maximum demand in the summer. These varying demands result in two important effects for contamination events: (1) differing flow velocities and contaminant transport rates, and (2) differing ratios of consumed to non-consumed water at consumer nodes. Data on water use in New York City for 1982 (Protopapas et al. 2000) is used here to identify an appropriate probability distribution function (PDF) for a WDS aggregate demand multiplier, which is used as a surrogate for time of year. The chi-squared goodness-of-fit test was performed on a candidate set of distribution functions (Normal, Log-normal, Gamma, and Beta functions) to model uncertainty in the demand multiplier and the shifted Gamma PDF was found to be the best-fit distribution function. This probability density function (PDF) is expressed as:

$$PDF(DM) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}(DM - \gamma)^{\alpha-1}e^{-\frac{(DM-\gamma)}{\beta}} \qquad (2.2)$$

where α, β, and γ = distribution parameters. Fig. 2.2 illustrates the empirical data and best-fit distribution. Similar assessments for an unpublished, confidential dataset provided by a Texas water utility indicated the suitability of the shifted Gamma distribution. However, the Texas dataset exhibited significantly greater variation than the New York City case. The New York demand multipliers ranged from 0.9 to 1.4, and the Texas multipliers ranged from 0.5 to 2.0, which is to be expected as summer

demands include much higher landscape irrigation in the Texas dataset. Thus, a specific utility would need to analyze its own aggregate demand data to determine appropriate Gamma distribution parameters.
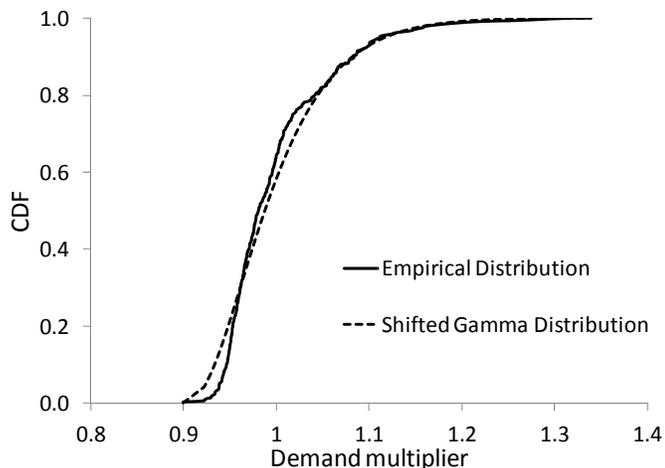
Fig. 2.2. Probability distribution model for WDS-wide demand multiplier

2.2.4 Intrusion Duration and Time of Day

The duration of contaminant intrusion is highly variable in historical events due to the large variation in causes of contamination (e.g., operator error, extreme weather, equipment malfunction) and the possibility that intrusion may not be remedied until high numbers of casualties occur. Bristow and Brumbelow (2006) found in a review of contamination events that the mean time for contamination to be discovered through public health system alert is over 11 days. Analysis of events published by Hrudey and Hrudey (2004) found a highly uncertain and broad range of identified contaminant intrusion of a few days to a month. The hydraulic realities of WDS allow for some simplification of this issue by noting the maximum residence time of water in a WDS

(a.k.a. "water age") as a practicable upper bound for modeling of intrusion duration. Contaminant intrusions lasting longer than this upper bound will lead to linearly increasing exposures of consumers to contaminant, but no further dynamic phenomena remain to be discovered once contaminant has been allowed to reach a maximum spatial extent. As technology to diagnose potential contamination events in real-time improves (e.g., Hart et al. 2009), it is hoped that shorter intrusion durations would be most relevant. Owing to the long contaminant intrusion durations often experienced in actual events, time of contaminant initiation has practically no documentation for these events. However, diurnal flow variations could very well lead to exposure sensitivity to time of initiation, especially for very short intrusion durations (i.e., less than 24 hours). At this time, simple uniform distributions of these parameters are perhaps the best possibilities, but we include them in the framework as possibly discernible and important attributes in the future.

2.3 Propagation of Uncertainties, Aggregate Conditional Risk, and Sensitivity to Attributes

2.3.1 Monte Carlo Simulation

To determine the general effects of uncertainty in the attributes defined above on a WDS's risk circumstances, Monte Carlo simulation (MCS) is an effective tool. MCS is a numerical procedure designed to propagate the uncertainties in system input random variables to determine the uncertainty properties of system outputs by performing a large number of simulations sampling from the appropriate distribution for each input

variable. For accidental water contamination threats, the scenario defining attributes described above are the stochastic input parameters, and system response is consumer exposure. The statistical analysis of scenario attributes presented above thus facilitates the stochastic realizations.

As will be shown in the application example below, MCS for accidental contamination scenarios produces system response (i.e., consumer exposures) with high variance. This numerical product necessitates relatively large numbers of MCS simulations in order to produce reasonable belief intervals on distributions of system outputs (Morgan and Henrion 1990). However, the large number of realizations does provide a usefully large sample by which to assess sensitivity of exposure numbers to individual scenario attributes. Likewise, the MCS analysis allows determination of an empirical distribution function of exposure, which can provide significant understanding of aggregate WDS risk and allow for comparative analysis of potential mitigation options.

2.3.2 Application Example

Adverse health impacts are calculated using the EPANET simulator (Rossman 2000) coupled with an exposure model. Simulation is performed under extended conditions to account for the dynamics of the system and temporal variations of water demand. In our example the contamination transport is simulated as a perfect tracer: density effects, decay, and reaction with wall materials and other dissolved species are

not considered; these effects could be added to the simulation using EPANET's water quality routines if reliable information on these phenomena is available.

The quantity of contaminant ingested by individuals during a contamination event depends on water ingestion patterns and time-varying concentration of contaminant. The timing ingestion model selected for this study assumes that tap water is ingested at the common starting times for the three major meals (7:00, 12:00, and 18:00) and times halfway between these meals (9:30 and 15:00). Daily per capita tap water intake rate was set to 0.93 L/day based upon USEPA (2004). Alterations in water demands after the contamination event unfolds are not considered in this study.

For illustration we will use the virtual city and WDS "Mesopolis" (Johnston and Brumbelow 2008), an open-source virtual city that is developed in both geographic information systems and EPANET (Fig. 2.3) and possesses spatial and temporal features of complex real world WDSs. The WDS is comprised of 2,062 water mains, 876 hydrants, 65 pumps, two treatment plants, one reservoir, and 13 tanks. Demands are exerted at 706 residential, industrial, and commercial/institutional nodes, representing a mid-size community of approximately 147,000 residents. Parameters of the shifted Gamma distribution for demand multiplier are $\alpha = 3.7$, $\beta = 0.18$ and $\gamma = 0.5$. The continuous random variable values are discretized to form a probability mass function to simplify MCS trials. A uniform probability mass function with one-hour intervals from one to four days is assumed here for the duration of contaminant intrusion. The time of day a contamination event is initiated is uniformly distributed throughout the day with one-hour intervals. Since the analysis of past events showed that 89% of accidental
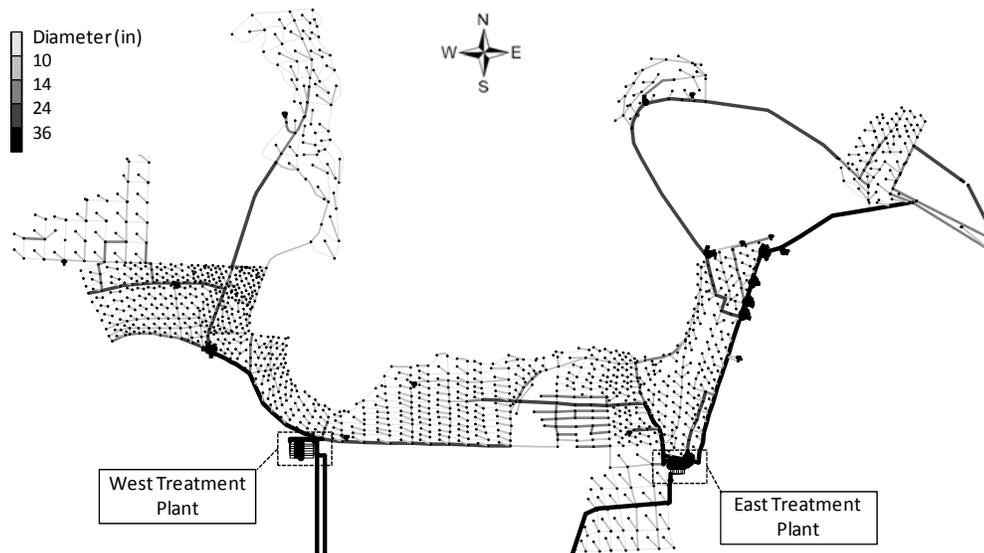
Fig. 2.3. Water distribution system of Mesopolis

events have intrusion at treatment facilities, only the WTPs are considered as possible intrusion locations, as this high probability for WTPs will dominate risk over any particular distribution system location. Contaminant amount follows the exponential distribution defined by the expected values that are estimated through scaling statistical analysis results to the population of Mesopolis (Table 2.1).

To estimate the uncertainty in human exposure and the significance of different scenario attributes 50,000 Monte Carlo simulations were performed. Exceedance probability (i.e., 1-CDF) curves for exposure above infective dose (ID) are depicted in Fig. 2.4. This figure also illustrates the curves for each WTP that are constructed using only the realizations associated with each plant. It is observed that almost 76% of all realizations result in zero exposure, meaning the total ingested contaminant amount for no person exceeds ID. This percentage is smaller for the realizations associated with the

West WTP (65%) while it is noticeably larger when contamination occurs in the East WTP (91%). Although the occurrence probability of non-zero exposure contamination events is much smaller for the East WTP, severity of extreme-impact scenarios is considerably greater for this WTP. Exposure never exceeds 38,000 if the contamination occurs in the West WTP while it may exceed 120,000 for the East WTP. This result is rooted in system hydraulics. While the hydraulics change with time and variations in total municipal demand, the West WTP supplies water only to the consumers in the western region. However, the East WTP supplies water to a much larger area covering almost the whole city. Accordingly, if the contamination occurs in the West WTP, it only affects the hydraulically isolated western region with approximate population of 38,000 while nearly the whole city population is at risk if the East WTP is the source. Dilution of the contaminant plume, however, is much more significant for East WTP because the contaminated area is larger, and this lowers the probability that ingested mass of contaminant exceeds ID and causes sickness.
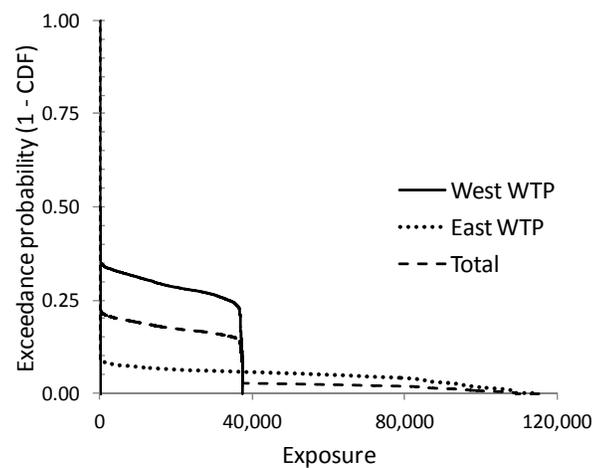
Fig. 2.4. Exceedance probability curves for human exposure

To determine overall risk for the system, the area under an exceedance probability curve may be integrated to produce Aggregate Conditional Risk (ACR) – the expected value of consequences over all foreseeable possible events conditioned on the occurrence of a single event. ACR is a convenient summary statistic of a WDS's circumstances, and comparisons of ACR values for differing assumptions or potential mitigation plans provide insight for relative conditions. For the Mesopolis example, ACR values for the East and West WTPs are, respectively, 5260 and 10800 exposures above ID. Thus, from a risk-management perspective (where risk is the product of likelihood and consequences), the West WTP is the element of higher priority even though the East WTP could experience higher consequence events. Various risk-mitigation plans could then be evaluated by their benefits in ACR reduction versus cost of implementation. It is possible that a decision maker may be more concerned with extreme events (e.g., Perelman and Ostfeld 2010; in press), and the exceedance probability curves also provide useful information for that reasoning where the East WTP would be of greatest concern.

Relative effects of variability in contamination scenario attributes on average exposure number above ID are illustrated in Fig. 2.5 for contaminant type, demand multiplier, intrusion start time, and intrusion duration. NLV and Cryptosporidium are associated with maximum and minimum averaged exposures, respectively, for both WTPs. However, Cryptosporidium has about twice the occurrence probability of NLV (Table 2.1), which underscores the interplay among the attribute uncertainty structures. In general, exposure follows a decreasing pattern with increasing demand multiplier, a
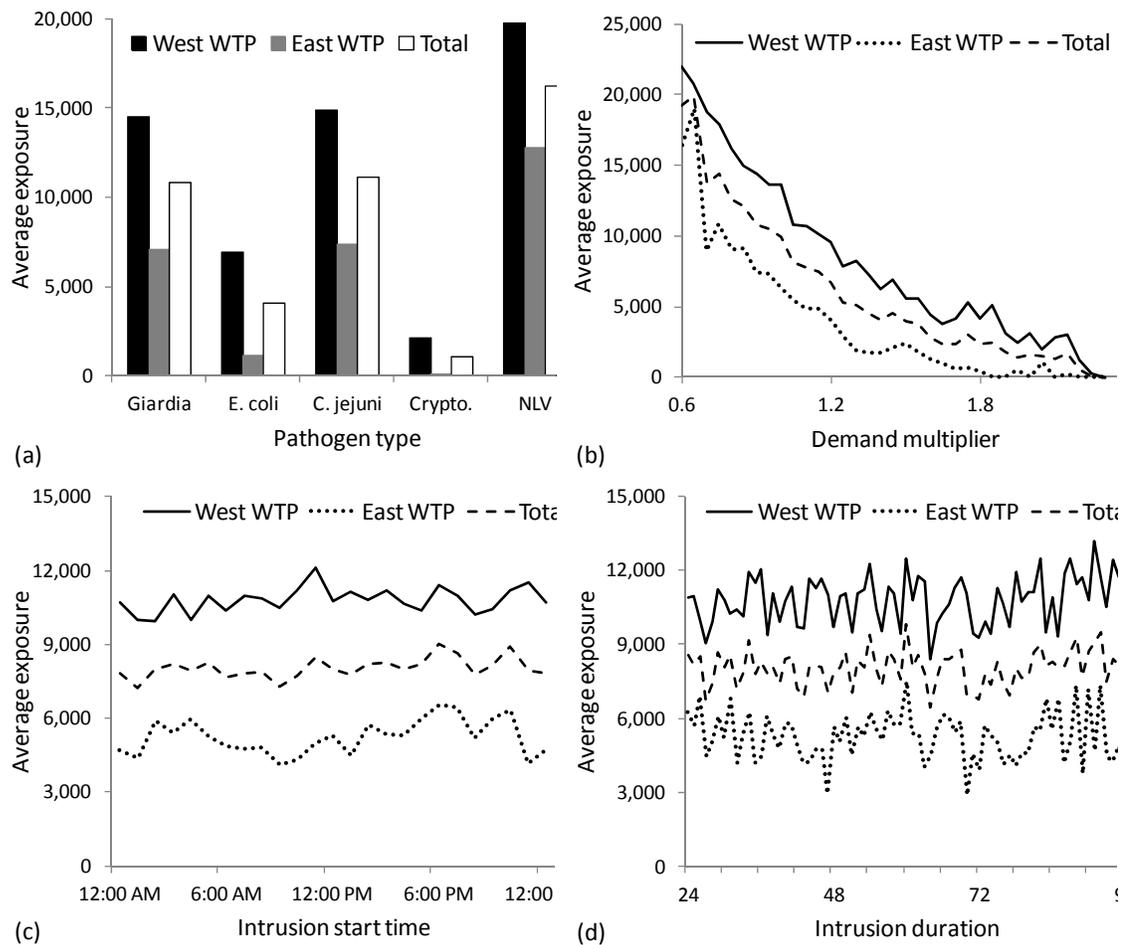
Fig. 2.5. Variability in exposure due to varying scenario attributes

consequence of higher demands forcing contaminant through the WDS faster and decreasing residence time (and exposure opportunities) and higher flows diluting contaminant concentrations. Although the results demonstrate fluctuations in exposure as start time and duration respectively change, no consistent pattern is apparent. Thus, a picture of particular vulnerabilities and their relative levels of importance emerges.

2.4 Identifying Ensembles of Critical Scenarios

2.4.1   Methodology

While the above methods diagnose general risk circumstances for a WDS, vulnerability mitigation planning typically requires identification of some ensemble of critical scenarios (a.k.a. "design basis threats," e.g., Van Leuven 2011). This ensemble constitutes a focused and finite range of scenarios against which to plan, making the process practicable. We apply here mathematical optimization to identify ensembles of critical accidental contamination scenarios using 2 selection criteria – maximum risk and maximum consequences – with full development of the tradeoff relationship between likelihood and consequences. Kanta and Brumbelow (2012) have applied a similar risk-based optimization method to the identification of WDS fire-flow vulnerabilities. In the universe of possible events, managers and decision makers are most concerned about occurrence of the events that are associated with both high consequences and likelihood. A conventional approach is to aggregate probability and adverse effects to estimate the single criterion of risk that provides a measure for ranking events. We will deviate slightly from this single-objective approach for a few reasons. First, as noted above, decision makers are at times very concerned with maximum consequence scenarios and willing to downplay to some extent likelihood. Second, diversity in the critical scenario ensemble is expected to encourage robustness in risk-mitigation planning.   Third, explicit quantification of the tradeoff relationship between likelihood and consequences can generate insight into vulnerability and potential mitigation options.

In contrast to a single-objective approach that searches for a single solution with the maximum product of likelihood and consequence, the multiobjective optimization approach proposed here seeks a set of critical events considering probability and exposure as independent risk components (i.e., 2 independent optimization objectives). The term "maximum-risk frontier" is coined here to describe the set of non-dominated scenarios in these objectives. Based on the general concept of dominance in multiobjective optimization (Deb 2001), scenario x_1 dominates x_2 if two conditions are met: 1) the respective magnitudes of x_1's likelihood and consequences are each greater than or equal to those for x_2, and 2) the value of at least one of these two risk components is greater for x_1. A scenario is defined as non-dominated if there is no scenario in the whole universe of possible scenarios which dominates it. A set of such non-dominated scenarios subsequently construct a maximum-risk frontier.

Occurrence probability of each scenario can be described with the joint probability mass function of the random scenario attributes (Ang and Tang 2007):

$$p = P(C = c, M = m, DM = dm, L = l, T = t, \Delta T = \Delta t) \tag{2.3}$$

where C = pathogen type, M = contaminant amount, DM = water demand multiplier, L = contamination location, T = time of day contamination event is initiated, and ΔT = intrusion duration are the scenario attributes. Statistical independence is assumed among all scenario attributes except contaminant type and amount. The probabilistic properties of the attributes can be taken from the analysis above. While all six attributes could be

defined as decision variables, a reduced set of four continuous decision variables – contaminant amount, demand multiplier, initiation time, and duration of intrusion – will be used here to reduce computational burden and avoid discontinuities associated with the discrete variables – contaminant type and intrusion location. Multiple optimization runs are performed with fixed values of the discrete variables, which also allows for sensitivity analysis to be conducted. This procedure also sidesteps the issue of differing severity of illness associated with different pathogens.

2.4.2   Optimization Algorithm

The event simulation model described above coupled with the multiple independent probability structures of the decision variables suggests that the optimization problem involves both significant nonlinearities and a high potential for multi-modality. Genetic algorithms (GA) have proven to be flexible and effective tools in solving such complex water resources problems (Nicklow et al. 2010). GAs are discussed fully by Deb (2001), but the basic issues in any GA are: (1) representation of decision variable sets as "chromosomes," (2) evaluation of chromosomes through fitness functions, (3) recombination of discrete decisions among sets through a crossover operator, and (4) random perturbation of decision sets through a mutation operator. GAs have been extensively employed for optimal design of WDSs as evidenced by use of GA for sizing of pipes (Krapivka and Ostfeld 2009), placement of early warning sensors (Ostfeld and Salmons 2004), and contamination consequence management (Baranowski and LeBoeuf 2008).

GAs have demonstrated unique ways of handling multiobjective optimization problems. Since they are population-based optimization methods, they offer a means of finding the Pareto optimal front in a single run. Over the past decade, several multiobjective evolutionary algorithms have been proposed (Guliashki et al. 2009). Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al. 2002) is arguably the most popular among existing algorithms available and has been widely employed for multiobjective design and operation of WDSs (Preis and Ostfeld 2008; Alfonso et al. 2010). NSGA-II is an elitist optimization algorithm which uses a fast non-dominated sorting strategy and does not require any user-defined parameter for diversity preservation. The NSGA-II algorithm employed here uses the simulated binary crossover (SBX) operator (Deb and Agrawal 1995) and polynomial mutation (Deb 2001) to create offspring population.

2.4.3   Application Example

The multiobjective optimization scheme is applied to find the maximum-risk frontiers for all possible combinations of pathogens and WTPs. Tuning of the GA optimization parameters was performed through a series of sensitivity analyses, and the final values are reported in Table 2.2. For clarity, we show the frontiers found for two pathogens, Giardia and E. coli, in Fig. 2.6. For exposures below 38,000, scenarios associated with the West WTP are more critical than at the East WTP as they are noticeably more probable for the same exposure level. Contamination scenarios with an exposure above 38,000, however, may only happen if they occur at the East WTP, which

Table 2.2. NSGA-II algorithm parameters

Optimization parameter	Value
Population size	100
Number of generations	150
Tournament size	3
Crossover probability	0.80
Mutation probability	0.05
SBX crossover distribution index	15
Polynomial mutation distribution index	10

is in agreement with the MCS results. In other words, all non-dominated scenarios associated with the East WTP with exposure below 38,000 are dominated by non-dominated scenarios for the West WTP. From a system-wide perspective, thus, the maximum-risk frontier may be split into a high-likelihood and low-severity partition associated with the West WTP and a low-likelihood and high-severity section corresponding to the East WTP. Obtained maximum-risk frontiers could serve as a suitable criterion for assessing efficacy of mitigation policies; a more effective strategy is the one that moves these frontiers more towards the origin.
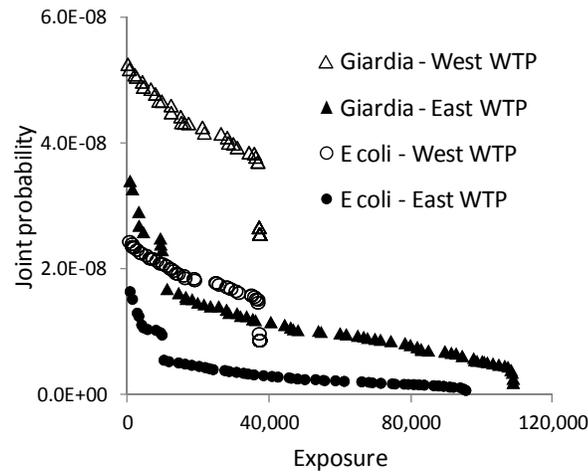
Fig. 2.6. Maximum-risk frontier for different pathogens and treatment plants

As mentioned above, a single-objective approach optimizing risk (the product of likelihood and consequences) would be a more traditional risk-management technique. The multiobjective approach explicitly contains the scenarios that would be found in the single-objective method. This is illustrated in Fig. 2.7 that shows the product measure versus exposure for all obtained non-dominated scenarios for Giardia. The "peak" in each ensemble is the scenario having maximum risk according to the traditional product definition. However, the multi-objective approach allows for greater insight on the relative properties of scenarios and potential mitigation plans when weighing maximum risk versus maximum consequence decision possibilities. This idea is illustrated schematically in Fig. 2.8. A typical maximum-risk frontier is shown by black squares with the maximum risk ensemble bounded by an ellipse and the maximum-consequence ensemble bounded by a rectangle. Risk frontiers for two potential mitigation plans are also shown with their respective maximum risk and maximum consequence ensembles
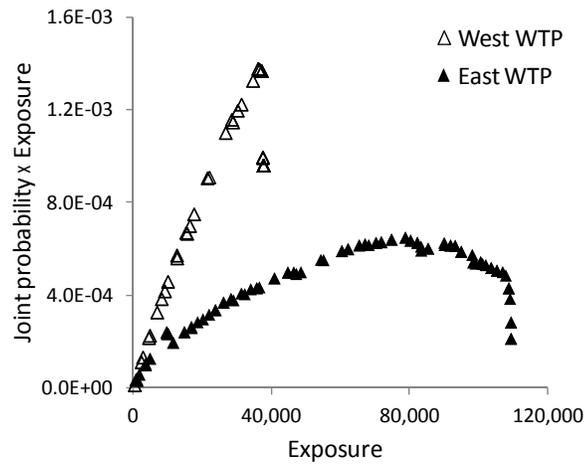
Fig. 2.7. Product risk measure associated with non-dominated contamination

scenarios versus corresponding exposure for the Giardia pathogen

similarly bounded. This drawing demonstrates the concept of two plans improving over

the original system but having differing advantages relative to each other: one minimizes

maximum product, and one minimizes maximum consequences.   A single-objective

approach focused on either criterion would not produce this insight; it is only possible

using this type of multiobjective technique.

Sensitivity of results is shown for the decision space in Fig. 2.9 with contaminant amount and demand multiplier associated with maximum-risk frontier solutions for Giardia. For the most part, the results are not surprising. Greater contaminant amount leads to more exposures for both WTPs, and exposures decrease with demand multiplier for the West WTP as shown in the MCS results (Fig. 2.5(b)). However, the relationship between demand multiplier and exposure is more complicated for the East WTP non-dominated scenarios, which requires an explanation rooted in the WDS hydraulics. The demand multiplier is small for non-dominated scenarios associated with lower exposure because a portion of contaminant amount (which is low for these scenarios as shown in Fig. 2.9(a)) would be transported into the western side of the city and diluted under a
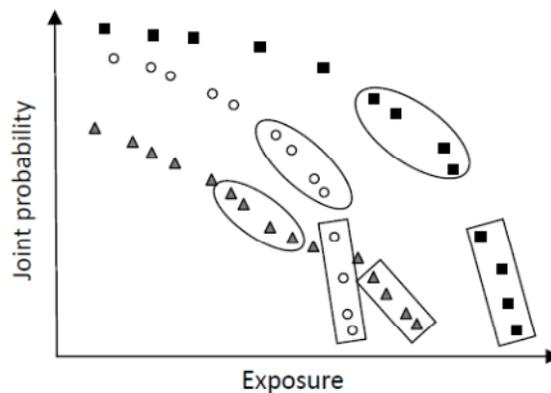
Fig. 2.8. Schematic illustration of comparison of maximum-risk frontiers for baseline system (black squares) versus 2 potential mitigation plans (white circles and gray triangles). The maximum risk product ensemble is outlined in each frontier by an ellipse; the maximum consequences ensemble is outlined in each by a rectangle

high demand multiplier. Comparing Fig. 2.9(b) with the MCS results presented in Fig. 2.5 (b), it is observed that MCS is not capable of capturing this characteristic of critical scenarios associated with the East WTP. Fig. 2.9(b) also shows that the demand multiplier is smaller than one for all non-dominated scenarios for both WTPs, confirming that low demand times of year possess greater vulnerability. With the help of MCS results (Fig. 2.5(b)) and the shape of shifted Gamma distribution (where probability is maximum for a demand multiplier of one and decreases for other values), this can be attributed to the fact that exposure is smaller for larger demand multipliers while they are not necessarily more probable than demand multipliers smaller than one.
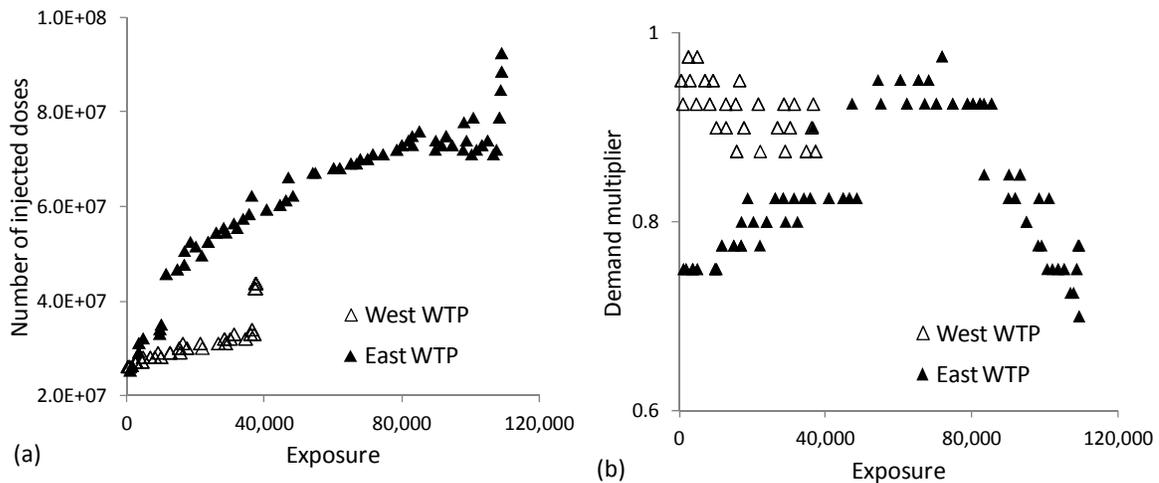
Fig. 2.9. Number of injected infective doses (a) and demand multiplier (b) associated with non-dominated contamination scenarios versus corresponding exposure for Giardia pathogen

2.5 Conclusions and Future Work

The goal of this chapter has been to present a generally applicable set of information and methods for evidence-based risk analysis for accidental contamination events in water distribution systems. The methods presented each have specific utilities: aggregate risk determination using MCS, and critical scenario identification by multiobjective optimization. There also exist interesting complementarities: both can be used for sensitivity analysis to event attributes, and both can be used to evaluate possible risk-mitigation plans but with different measures.

While a large set of contamination case studies was available for determination of event attribute properties, this type of risk-based analysis can only benefit by analysis of more events. It is hoped that the water resources engineering and public health communities can cooperate to improve the quantitative aspects and availability of documentation of contamination events for this purpose.

We have also suggested the possibly conflicting natures of maximum-risk versus maximum-consequences driven decision making. This is likely a fruitful area for further work to determine the magnitude of such conflicts and how far apart decisions reached by these philosophies are. More advanced techniques of risk analysis such as the partitioned multiobjective risk method (Asbeck and Haimes 1984; Haimes 2009) can be extended for this purpose.

Information obtained about the probability of health impacts using Monte Carlo simulations may be used by the future research to construct f-N diagrams. These diagrams basically illustrate the annual contamination probability on the vertical axis vs.

health impacts on the horizontal axis. f-N diagrams are very useful for deciding on acceptable risk levels and consequently evaluating different available risk reductions measures. They have been used by the Unites States Bureau of Reclamation (2003) for dam safety decision making.

Finally, this chapter did not account for human-infrastructure interactions that may occur as a contamination event unfolds – e.g., water demand reductions in response to utility warnings. Chapter 3 will incorporate consumer behavior (such as water demand changes and word of mouth communications) and the stochastic nature of these interactions in the methods presented in this chapter. This incorporation will more accurately mimic system behavior and evaluate contamination risks for the characterization of critical scenarios.

3. STATIC OPTIMIZATION OF CONTAMINANT FLUSHING AND CONTAMINANT MECHANISMS TO MINIMIZE HEALTH IMPACTS AND SYSTEM SERVICEABILITY INTERRUPTION

3.1 Introduction

Decisions on protecting public health against possible water contamination threats should be made with careful consideration of credibility of threat observations and unintended impacts of response implementation on water supply system serviceability. To effectively cope with these threats there is a need to prepare contamination emergency management plans that describe the actions a drinking water utility needs to take in preparation for and in response to a contamination threat or incident. An emergency management plan should be based upon careful risk assessments and cover the four phases of hazard mitigation, emergency preparedness, emergency response, and disaster recovery (Lindell et al. 2006).

A contamination emergency response phase is initiated with an actual (or potential) release of contaminant that is spreading (or will spread) across a WDS, and it extends until the situation is stabilized, when the risk of health impacts has returned to pre-event levels. An emergency response plan explains actions that managers may take in response to the perceived state of the system after the emergency begins, and it considers how best to achieve managers' multiple objectives. These response actions can be classified as "assessment," "corrective," or "protective" actions, depending on whether they collect information about the state of the system, operate on the system to

decrease impacts, or require action by the public to reduce exposure (Perry and Lindell 2007).

Title IV of the Public Health Security and Bioterrorism Preparedness and Response Act of 2002 (US Congress 2002) requires all community water systems serving a population greater than 3,300 to prepare or revise emergency response plans. The Response Protocol Toolbox (RPT) has been prepared by the United States Environmental Protection Agency (USEPA 2003) to help water utilities meet this requirement. It provides general guidelines on how response decisions should be made at the various stages of a contamination event as more information is gained. Because this toolbox is essentially a qualitative document, however, it does not provide specific guidance on how appropriate response strategies should be devised for a particular WDS. This chapter is focused on developing quantitative simulation-optimization models for preparation of emergency response plans that specify functional contaminant containment and flushing operation rules for achievement of conflicting response objectives.

Contaminant containment through isolation valve operations is a corrective response action implemented to prevent contaminant spread to uncontaminated regions of a network and to preclude consumers from withdrawing contaminated water (USEPA 2003). Decisions on combination of valves to be closed and timing must be made carefully and implemented quickly to be effective and minimize accompanying side consequences such as impacts on non-consumptive uses in isolated regions. Genetic algorithms (GA) (Baranowski and LeBoeuf 2008) and heuristic approaches (Poulin et al.

2008) have been employed to find locations of pipe closures necessary to reduce the contaminant concentration during an emergency.

Contaminant flushing is another corrective response action that is executed through opening hydrants to discharge a large volume of contaminated water. System flushing should be planned and implemented carefully so that it is performed at the sections of the system where contaminant concentration is higher. Otherwise, it will worsen the situation by further spread of contaminated water to uncontaminated areas as it can considerably alter flow regime (USEPA 2003). Optimization tools have been used to explore performance of this response action for public health protection (Baranowski and LeBoeuf 2006, 2008; Zechman 2010).

While emergency response plans are primarily implemented to protect public health, achievement of this goal might hinder meeting normal system operation objectives including suppression of urban fire events and service to residential, industrial, and commercial consumers. Water utilities must also operate in resource constrained environments in terms of finance, personnel, etc. An over-emphasis on vigilance against perceived contamination threats could divert needed resources from maintenance and other crucial activities. To date, limited research has addressed this multicriteria nature of the contamination emergency response problem (Preis and Ostfeld 2008; Alfonso et al. 2010). Multiobjective frameworks proposed so far have only considered hydrant and valve locations as decision variables and have not optimized the operation timing. While these studies have considered the number of operational actions as an emergency response criterion, no attempt is made to explicitly address the

important criterion of system serviceability interruption. Most importantly, many previous single and multiobjective studies have used demand-driven analysis (DDA) to simulate WDS behavior, and this assumption inevitably limits the optimization search space to response plans that do not cause excessively low pressure in the WDS. This may unfavorably filter out many possible response plans with high potential to mitigate health impacts.

In the light of these needs, this study develops and integrates pressure-driven analysis (PDA) and multicriteria models for optimization of emergency response plans with explicit consideration of two important response criteria: impacts on public health and system serviceability. The emergency response is treated as both single and multiobjective optimization problems to address utility managers' needs under different situations, provide insight into effective response plans, and assess sensitivity of response to different parameters such as response delay. Operation rules for contaminant containment and flushing locations and timing are explicitly treated as optimization decision variables. Different formulations to quantify impacts on public health and service availability are examined with the help of an exposure model and the PDA. Performance of the proposed schemes is investigated using the Mesopolis virtual city WDS.

3.2 System Simulation Model

3.2.1   Pressure-driven Hydraulic Analysis

Behavior of WDS is generally simulated using standard hydraulic models like EPANET 2 (Rossman 2000) for applications such as design, operation, and rehabilitation. The conventional DDA approach uses nodal demands as an input in the pressure (head) calculations on the premise that these demands shall be satisfied at all conditions. Simulation models based on DDA typically reflect the network satisfactorily under normal conditions where pressures are sufficiently high. However, such models will distort the dynamics of real systems under the abnormal conditions of low pressures such as may be caused during the course of emergency and implementation of some corrective actions (e.g., high flows during hydrant flushing may lead to high head loss and depressed nodal pressures). Considering the limitations of DDA, PDA is employed here to more properly reflect the real behavior of the system.

A comprehensive review of pressure-deficient network predictors is performed by Gupta and Bhave (1996). They conclude that the method using parabolic head-discharge relationship (Bhave 1981; Wagner et al. 1988; Chandapillai 1991) is the best for prediction of such conditions, and this method will be used in this chapter. In this method, a parabolic relationship is assumed between the service head H_i^{ser} for node i and the minimum head $H_i^{\min}$ needed to satisfy the nominal demand Q_i^{nom}:

$$Q_i = \begin{cases} 0 & H_i < H_i^{\min} \\ Q_i^{nom}\left(\dfrac{H_i - H_i^{\min}}{H_i^{ser} - H_i^{\min}}\right)^{1/n} & H_i^{\min} < H_i < H_i^{ser} \\ Q_i^{nom} & H_i > H_i^{ser} \end{cases} \qquad (3.1)$$

where n = exponent corresponding to choice of head loss formula.

The PDA is an iterative process. The nodal pressure heads are first calculated for the nominal demands using the DDA that is done here via EPANET toolkit. The results are then used in Equation (1) to correct the demands and re-estimate the network heads. This procedure is repeated until sufficient convergence is obtained. A convergence criterion thus needs to be devised, such as total change in network heads after each iteration or a preset maximum number of iterations, for every WDS application example.

3.2.2 Exposure Model

Adverse health impact, defined here as either the number of illnesses resulting from a contamination event or the total ingested contaminant mass, is calculated using a PDA-based hydraulic and water quality simulation model coupled with an exposure model. The quantity of contaminant ingested by individuals during a contamination event depends on water ingestion pattern, time-varying concentration of contaminant, and availability of drinking water. An individual is assumed to become ill if the cumulative amount of contaminant ingested during a contamination event exceeds a known infectious dose. The timing ingestion model selected for this study assumes that

tap water is ingested at the common starting times for the three major meals and times halfway between these meals, given that there is water available for drinking at a consumer's demand node. The tap water intake rate used in the exposure model is central to accuracy of evaluated exposure and is obtained from USEPA (2004), provides estimates of per capita ingestion of community water.

3.3 Response Optimization Framework

Emergency response is a progressive, interactive, and adaptive process that includes parallel activities of assessing unusual contamination observations and making appropriate emergency response decisions. As more information is obtained about contamination, emergency management progresses through three threat stages of "possible," "credible," and "confirmatory" accompanied by an increase in seriousness of the threat impacts and magnitude of response decisions. While public health protection is the primary response focus, emergency management should carefully consider other potential consequences on infrastructure serviceability due to response implementation, specifically in the early stages of the process where the attack credibility level is relatively low. At this stage, a multiobjective response plan would be of substantial help for utility operators to identify the balance between actions taken to protect public health and limiting overaction that adversely impacts the ability of the system to meet multiple aspects of its overall mission. Nevertheless, if evaluation of collective threat information that progressively becomes available corroborates the threat warning and indicates that contamination is likely, all available resources must be utilized to minimize the single

objective of health impacts without further consideration of other response criteria. Under these circumstances, the large size of multiobjective optimization results may not be decipherable for making timely emergency decisions and use of a single-objective model becomes preferable.

3.3.1 Mathematical Problem Formulation

Quantification of emergency response criteria is the first step for the preparation of an optimization model. In reality, different measures may be formulated to quantify two important objectives of impacts on public health and system serviceability. This study investigates two distinct formulations for each of these criteria and analyses the sensitivity of optimal response plan to each quantification method.

Some studies have quantified the health impacts in terms of contaminant concentration in system nodes either as total contaminant concentration (Baranowski and LeBoeuf 2006, 2008) or total number of nodes with concentration above a specified threshold (Alfonso et al. 2010). This approach neither accounts for the ingestion timing and rate which can significantly influence estimated impacts (Davis and Janke 2008), nor does it consider the critical fact that the connections serving higher populations are comparatively more vulnerable. In general, health impacts metrics may be differentiated depending upon whether or not they are based on a threshold: i.e., number of people/nodes that experience a concentration or contaminant ingested mass above a threshold (Perelman and Ostfeld 2010; Zechman 2011) or sum of concentration or consumed mass for all nodes/people (Baranowski and LeBoeuf 2008; Preis and Ostfeld

2008). Accordingly, with the help of the developed exposure model, the health impact criterion is quantified here as either the total number of sicknesses due to the contamination event or the total contaminant mass ingested by all people during the whole course of the event:

$$f_{H1} = \sum_{i=1}^{N_p} \alpha_i \qquad , \alpha_i = 1 \quad if \quad m_i > m_d, \quad \alpha_i = 0 \quad otherwise \qquad (3.2)$$

$$f_{H2} = \sum_{i=1}^{N_p} m_i \qquad (3.3)$$

where N_p = total number of people; α_i = a binary index; m_i = total contaminant mass ingested by individual i during the whole course of event that is calculated by the exposure model; m_d = known infectious dose. The function f_{H1} explicitly represents the health impacts in terms of morbidity while f_{H2} projects the health consequences in the more implicit form of ingested mass. However, f_{H2} does not require that infectious dose be known in advance, which greatly simplifies the analysis since this value is often difficult to determine and subject to high variance among individuals.

Interruption to system serviceability is also mathematically expressed using two different measures. The first measure is the total number of hours the volume of water supplied to consumers is below a certain percentage of their demand. Alternatively, the second measure sums up the difference between water demand and supply for all consumers during the whole course of the emergency. Mathematically,

$$f_{S1} = \sum_{t=0}^{t_{end}} \sum_{i=1}^{N_c} \beta_{ti} \qquad , \beta_{ti} = 1 \quad if \quad (S_{ti} < \lambda D_{ti}), \quad \beta_{ti} = 0 \quad otherwise \tag{3.4}$$

$$f_{S2} = \sum_{t=0}^{t_{end}} \sum_{i=1}^{N_c} (D_{ti} - S_{ti}) \tag{3.5}$$

where t_{end} = duration of emergency response phase in hours; N_c = total number of consumers; β_{ti} = a binary index; S_{ti} = volume of water supplied to consumer i at time step t calculated by the PDA model; D_{ti} = water demand of consumer i at time step t; λ = supply deficit threshold that ranges from 0 to 100%. The function f_{H1} treats all consumers equally while f_{H2} places more emphasis on consumers with higher water demand. It is noteworthy that both these metrics are always (unrealistically) calculated as zero if DDA is used.

The objective functions described in Equations (3-2)-(3-5) are minimized through optimization of response actions of contaminant containment and flushing. Implementation of response actions should account for the response delay times between when intrusion starts and when injection of contaminant in the network is identified as likely or confirmed through multiple contamination trigger events. For system flushing, the decisions include the identification of the hydrants that should be opened to flush the contaminated water ($\mathbf{h} = \{h_1, h_2, ..., h_{n_h}\}$) and the time at which hydrants would be opened ($\mathbf{t_h} = \{t_{h,1}, t_{h,2}, ..., t_{h,n_h}\}$) after the response delay. A maximum number of hydrants, n_h, may be opened during the simulation and this depends on personnel and equipment availability. For contaminant containment, the decisions include the time at which

closure valve sets (CVS) would be closed ($\mathbf{t_s} = \{t_{s,1}, t_{s,2}, ..., t_{s,n_s}\}$) after the response delay

and the duration that the CVSs would remain closed ($\Delta\mathbf{t_s} = \{\Delta t_{s,1}, \Delta t_{s,2}, ..., \Delta t_{s,n_s}\}$) where n_s

is the total number of CVSs. While every single pipeline may be theoretically considered

as a possible closure location, this would result in a tremendously large decision space

that may not be handled by the optimizer practically. Therefore, assessments need to be

performed to identify the most reasonable configuration of these CVSs. This step helps

construct a more compact isolation plan, supports the model with engineering

knowledge, and reduces computational burden of the optimization process. Fig. 3.1

shows an arbitrary timeline of contamination emergency period and response execution.

It should be emphasized that the order of actions and time overlaps may be different for

distinct events and emergency management plans.

3.3.2 Optimization Algorithm

Emergency decisions that should be optimized in response to a contamination
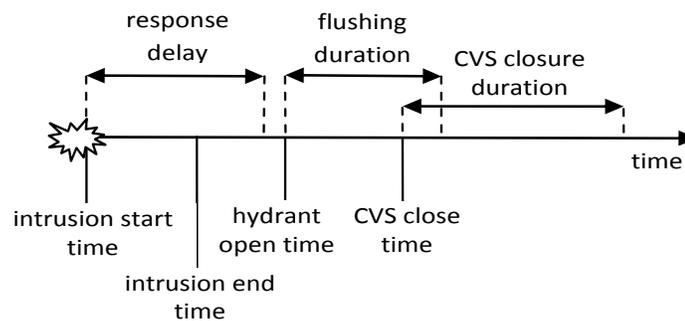
Fig. 3.1. Arbitrary timeline of contamination emergency period

threat are diverse and discrete, and the relationship between decisions and performance is highly complex. Single and multiobjective genetic algorithms are developed here and coupled with the pressure-driven hydraulic simulator and exposure model for identification of emergency response planning.

The single-objective model uses an elitist real-coded genetic algorithm with roulette wheel selection, simulated binary crossover (SBX) (Deb and Agrawal 1995), and polynomial mutation (Deb 2001). In the roulette wheel selection, the probability that a solution will be selected is given by the ratio of its fitness to the total fitness of other members of the current population. For hydrants identification, crossover and mutation operations are performed on longitude and latitude coordinates and the nearest hydrant to the generated coordinates is picked.

In contrast to a single-objective optimization approach that searches for a single solution with the best scalar fitness value, multiobjective optimization seeks a set of trade-off solutions which together define the best multiobjective alternatives surface called the Pareto optimal front (Deb 2001). This study employs Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al. 2002) that is an elitist evolutionary algorithm which benefits from a fast non-dominated sorting strategy and does not need any user-defined parameter for preserving diversity in Pareto optimal surface.

3.4 Application Example

The Mesopolis virtual city is used here to demonstrate the optimization of response actions using the proposed frameworks. The configuration of closure valve sets

is illustrated in Fig. 3.2. Two intentional contamination scenarios are selected for which the response actions are optimized. The settings and characteristics of these scenarios are shown in Table 3.1. The demand multiplier associated with each scenario is representative of aggregate water demand for a WDS that typically varies throughout the year. Contaminant agent is arsenic with an infectious dose of $m_d = 3.5$ mg for a body weight of 70 kg as reported by Office of Environmental Health Assessment Services (1999) and daily water ingestion rate is 0.93 liter/day (USEPA 2004b). Both contamination scenarios occur in the third day of simulation after the system has reached dynamic equilibrium and the total simulation time is 6 days.
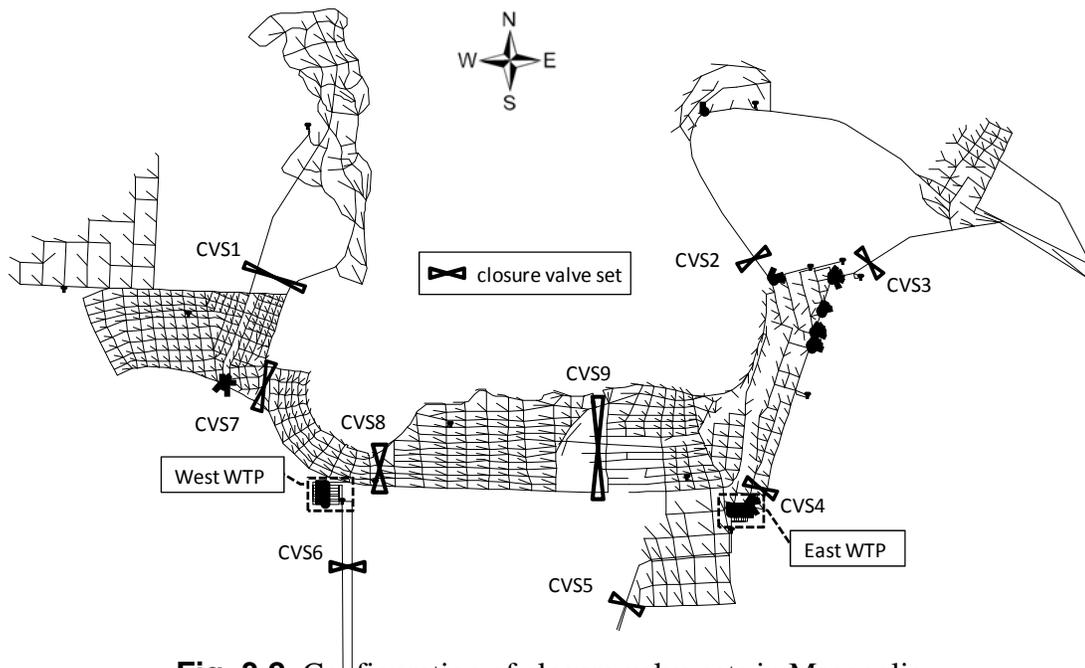
Fig. 3.2. Configuration of closure valve sets in Mesopolis

Table 3.1. Contamination scenarios

Scenario	Location	Load (kilograms)	Demand multiplier	Start time	Duration (hour)	f_{H1} (people)	f_{H2} (grams)
1	West WTP	300	1.00	18:00	6	33,944	304.08
2	East WTP	300	1.00	19:00	5	54,638	397.28

To explore sensitivity of response plan performance, three values for response delay are used: 6, 12, and 24 hours. Similarly, the number of hydrants that may be opened for flushing is set to 3 and 5 in different cases; once opened, all hydrants remain open for 5 hours. Model emitter discharge coefficients for hydrants are set to 166.5 gpm/psi$^{0.5}$ (associated with a 3-inch diameter connection fire hydrant) to calculate pressure-driven outflow in the PDA model. Values of H_i^{ser} and H_i^{min} for the PDA are based on engineering design standard of the cities of Bryan and College Station, Texas (Cities of Bryan and College Station 2005). Under normal conditions, a design head of 35 psi should be maintained throughout the system, and no water is available at a connection if its pressure drops below 20 psi, the minimum allowed during fire flow conditions.

3.4.1 Single-objective Optimization

The single-objective model is applied first to optimize hydrant operation for the minimization of health impacts considering both mathematical representations of f_{H1} and f_{H2}. Appropriate genetic algorithm settings are determined based on sensitivity analyses. The population size is set to 40, and the stopping criterion is set at 150 generations. The crossover rate is set to the fixed value of 0.80, whereas mutation rate decreases linearly from 0.10 to 0.05 as a function of the generation number. SBX crossover distribution and polynomial mutation indices are set to 15 and 5, respectively.

The optimal percentage reduction in health impacts versus the response delay is illustrated in Fig. 3.3 for both contamination scenarios and different numbers of hydrants used for flushing. Expectedly, the effectiveness of impact mitigation practices decreases as the utility operators respond later to a contamination incident. In this sense, f_{H2} indicates a better projection of what we expect than f_{H1} when the West WTP is contaminated (Scenario 1). The results demonstrate the effectiveness of the response strategy of contaminant flushing when the response delay is reasonably short but the performance diminishes if response is implemented after one day. Overall, contamination would be more effectively mitigated if it occurs in West WTP as the affected area is much smaller than that associated with the East WTP contamination and thus more tractable.
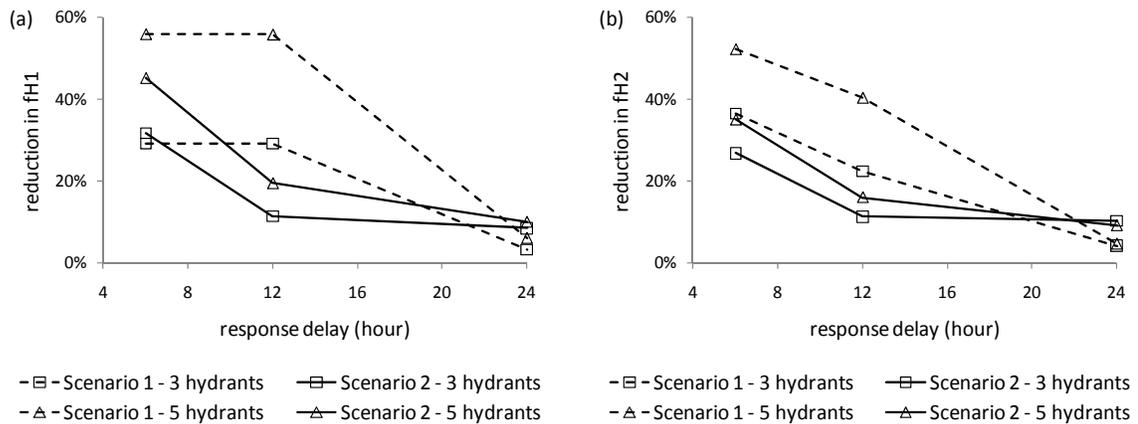
Fig. 3.3. Percentage reduction in health impact using two formulations of (a) total number of sicknesses and (b) total ingested mass of contaminant

Values of f_{H1} and f_{H2} project a scalar measure of the health impacts and do not provide distribution information on how much impacts are mitigated for different population sectors. Cumulative distribution curves of ingested contaminant mass are therefore prepared for optimal plans associated with both formulations as shown in Fig. 3.4. The vertical axis indicates the percentage of total population that has ingested a mass of contaminant below the corresponding value on the horizontal axis. f_{H1} is the population on the vertical axis above the cross point of toxic dose line with each cumulative curve, while f_{H2} is the area between the curve and vertical axis after it is transformed to absolute population values i.e. the percentage values are multiplied by the total population of Mesopolis. For both Scenarios 1 and 2, optimization of response based on f_{H2} evidently outperforms f_{H1} except for the very limited population that ingest a mass of contaminant close to the toxic dose threshold of 3.5 mg. The conclusion
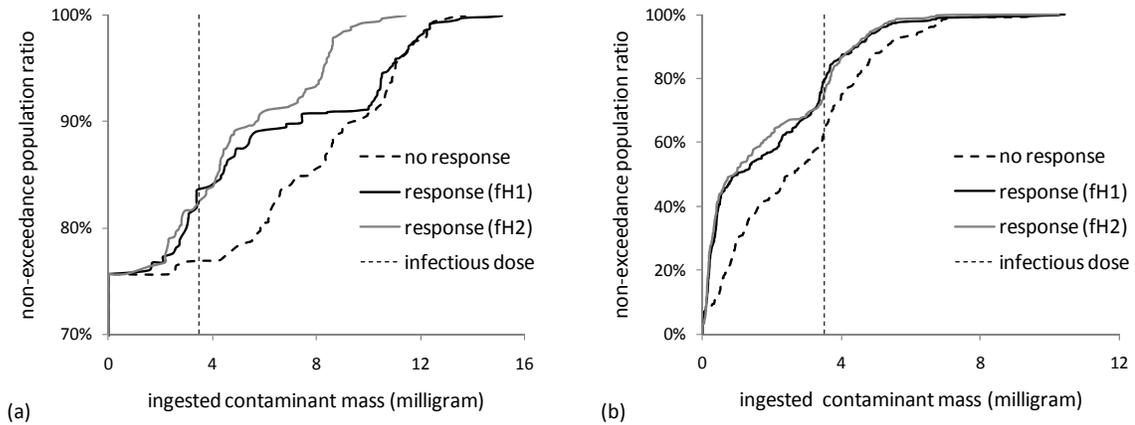
Fig. 3.4. Ratio of total population with ingested mass above different levels for (a) Scenario 1 ($n_h = 3$, delay = 6 hours) and (b) Scenario 2 ($n_h = 5$, delay = 6 hours)

one can draw is that although using f_{H1} would decrease the number of sicknesses more (f_{H1} curve is above f_{H2} curve when they intersect with the toxic dose line), it will not protect the people overall as much as when f_{H2} is used (f_{H1} curve is overall more extended rightward than f_{H2} curve). As observed in Fig. 3.4(a), it may even magnify the original risk for the people residing in highly contaminated areas (ingested mass above 12 mg for this case).

Fig. 3.5 depicts the optimal location and timing for opening hydrants for response delays of 6 and 12 hours when $n_h = 5$ and f_{H2} is considered as the health impacts metric. If West WTP is contaminated, all 5 hydrants should be opened in the western part of the city as this plant supplies water only to the western consumers. If Scenario 2 occurs (East WTP is the contaminant source), the optimal locations are spread around the contaminant source for response delay of 6 hours and move further

toward the east as delay is increased to 12 hours. A similar shift in optimal flushing locations is also observed for Scenario 1 where four out of five hydrants are opened in the western peninsula when delay is 12 hours compared to one out of five for a 6 hour response delay. Provision of these optimal response maps for finer discretization of response delay and other vulnerable nodes would serve as a supportive visual aid for utility managers to respond to the contamination in a more timely and efficient manner.

3.4.2   Multiobjective Optimization

Response actions of contaminant containment and flushing are next optimized simultaneously to minimize impacts on public health and system serviceability. CVSs

Fig. 3.5. Optimal hydrant operation for minimizing health impacts quantified as

total ingested mass of contaminant

are required to be configured before the optimization is performed. While more CVSs may theoretically result in greater reduction in health impacts due to a larger search space, a large set would lead to more elaborate response plans that may be burdensome to execute in practice and would also increase computational burden. CVSs with fewer included valves will minimize the number of operational actions needed to isolate an area. Nine CVSs are accordingly considered as shown in Fig. 4.2 based upon WDS layout, hydraulic simulations, and optimization computation intensity. CVS1, 2, and 3 are located on long mains that transmit water to far sections of WDS. CVS4 allows isolation of a large eastern area through closing only two mains. Closure of CVS 5 and 6 shuts down the East and West WTPs, respectively. CVS7 isolates the highly populated residential area in the western part of the city from both WTPs by closing only two mains. CVS8 allows the western area to be supplied by only West WTP. Finally, CVS9 includes three mains and completely disconnects the eastern network from central and western areas.

Fig. 3.6 shows the obtained Pareto optimal fronts for four possible combinations of objective functions represented by Eqs. (3.2)-(3.5). These fronts are for contamination Scenario 2, with response delay of 6 hours and $n_h = 5$, representing a total number of 28 decision variables (10 for hydrant opening and 18 for CVS operation).

The Pareto-fronts demonstrate that there are significant trade-offs between the response criteria of public health protection and system serviceability. In light of the fact that impacts on public health are considered more crucial, one may pose the question whether we should be concerned about the trade-offs at all. In the trade-off curves found here, each individual corrective action plan is associated with a level of reduction in health impacts and accompanied system disruption. A central point to remember is that, while implementation of such plan would lead to system disruption for certain, reduction in health impacts is conditioned on the credibility of threat observations (i.e., the RPT sequencing of possible, credible, and confirmatory threat stages). For the trade-off analysis to be rational, the horizontal axis should be multiplied by the probability that the contamination has actually occurred in order for it to be consistent with the vertical axis representing unconditioned system interruption. In reality, however, such probability values are not known exactly and must be inferred from an ensemble of uncertain sensor triggers and unusual observations. Thus, the trade-off curves must be understood through this filter of threat uncertainty.
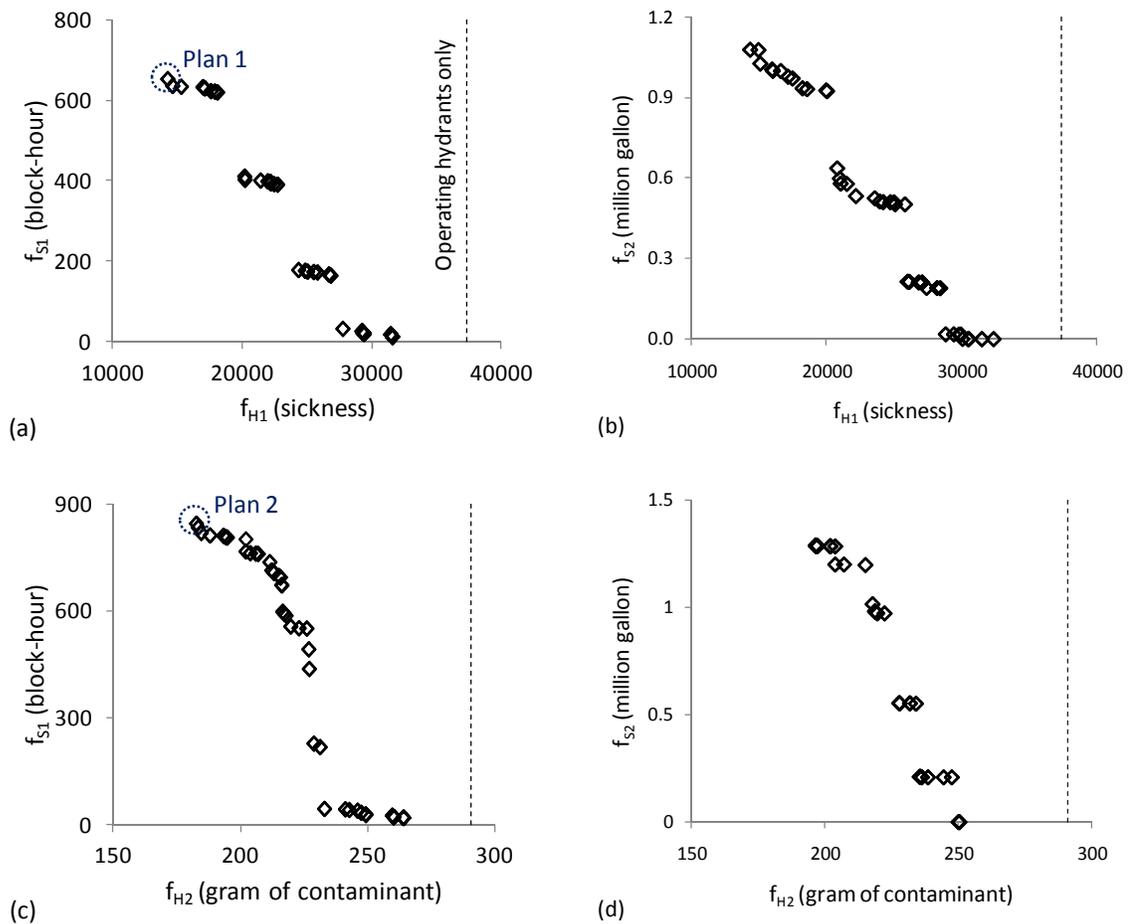
Fig. 3.6. Pareto optimal fronts for impacts on public health (f_{H1} and f_{H2}) and system

serviceability (f_{S1} and f_{S2})

Fig. 3.7 illustrates the two extreme Pareto-optimal plans associated with minimum health impacts using f_{H1} (Plan 1) and f_{H2} (Plan 2) formulations in the decision space. The figure also shows the population frequency curve of contaminant mass ingestion. For Plan 2, closure of CVS8 and CVS4 would block further spread of contaminant to western and eastern regions of the city, respectively, and isolate the

central area. Opening of five hydrants all located in the isolated central region will flush out the contaminant and protects the population in this area from contaminated water that is now even more concentrated due to the isolation. As observed in Fig. 3.7, this will essentially reduce ingested contaminant mass to zero for nearly 40% of the total population. For Plan 1, however, similar mechanics of isolation and discharge are not clearly observed; while closure of CVS4 would protect the eastern region, no action is taken to block further propagation of contamination across the western region as shown in Fig. 3.7. In fact, it is even determined to close the West WTP for 4 hours, increasing the pressure gradient between east and west and further intensifying contamination spread westward. Identification of this particular minimum-health-impact response plan by the optimization model is motivated by the expression of health impacts based upon a preset and fixed threshold. Instead of attempting to protect the whole population, the model tries to only decrease the number of people who have ingested a mass of contaminant above that threshold, which is partially achieved through distributing the injected mass across a larger area to lower the concentration.
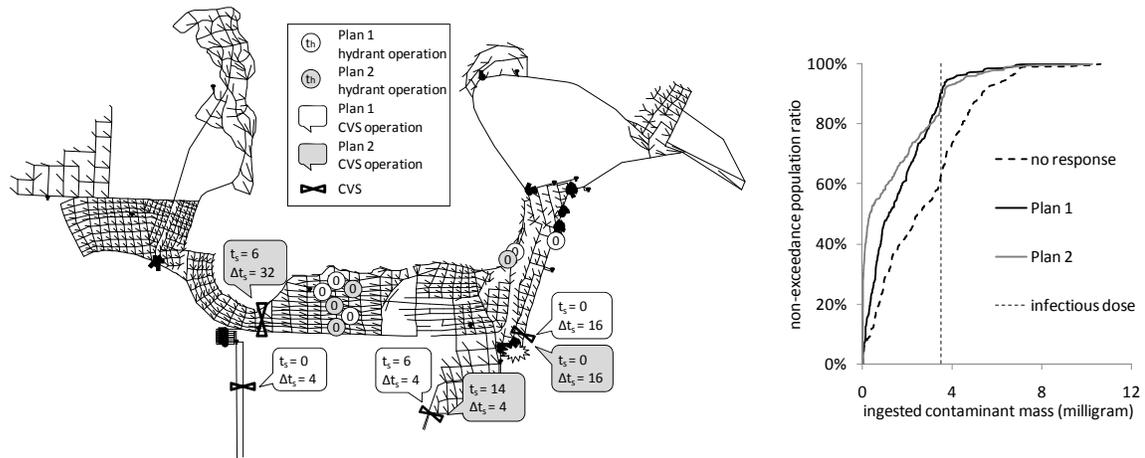
Fig. 3.7. Pareto optimal response plans associated with minimum health impacts using metrics f_{H1} (Plan 1) and f_{H2} (Plan 2) along with ratio of total population with ingested mass above different levels

3.5. Conclusions

Compared to the single-objective approach, the multiobjective optimization of valves and hydrants produces response plan alternatives that reduce health impacts more effectively but at the cost of more service disruption and provide trade-off information between impacts on public health and system serviceability. The single-objective optimization of hydrants operation, on the other hand, is faster to implement, and results in a simpler decision-making process. Introduction of the multiobjective approach, therefore, does not make the use of single-objective approach obsolete but they complement each other.

Several studies performed to address the risk of drinking water contamination have evaluated health impacts based upon a pre-specified exposure threshold. This metric has the advantage of being understandable and explicit as it communicates the

impacts sensibly in terms of number of sicknesses or deaths when thresholds are known accurately. However, findings of this paper indicate that using the alternative metric of total ingested mass of contaminant may lead to more effective optimal impact mitigation recommendations even if exposure thresholds can be estimated accurately. In this sense, quantification of health impacts as total mass of contaminant ingested by whole population may be preferable. This formulation also has the advantage that the optimal response plan found for a specific contamination scenario is still optimal for any other similar attack scenario with different contaminant type and loading. This is because this metric is a linear function of injected mass, whereas number of sicknesses and deaths vary nonlinearly as injected mass changes. Ideally, dose-response curves can be utilized after optimization is performed based upon total ingested contaminant mass metric to provide a more explicit and understandable presentation of adverse health consequences.

The findings show that simultaneous hydrant and valve operation proves to be very effective as it reduces potential health impacts by 74% and 54% for f_{H1} and f_{H2} formulations of health impacts, respectively; up to 45% and 35% reduction may be achieved when hydrants are operated only. However, this high reduction in possible health impacts is accompanied by water demand deficits, and trade-offs must be considered to choose the most appropriate response plan. The observation that reduction in health impacts is generally higher when it is expressed in terms of f_{H1} than f_{H2} should not imply that consideration of f_{H1} leads to more effective response plans but performance associated with each metric must be viewed in the context of its definition independently.

Delay between when contamination starts and when utility operators decide to respond the event substantially reduces effectiveness of risk reduction attempts. Because implementation of response optimization schemes is computationally intensive for large water distribution networks, their direct application after the event begins is not reasonable unless simplifications are made to the network or genetic algorithm search is guided with good solutions obtained from previous runs or expert knowledge. These schemes may be applied to a set of design basis threats during the preparedness phase to infer patterns in optimal response plans and provide insights to be used in the event of contamination. This large set of threat-response data – also called the Response Planning Matrix in RPT (USEPA 2003) – may be also used by decision support systems and data mining models to aid utility operators in rapidly making most effective decisions. Future efforts are required to develop such decision aid tools and learning algorithms.

4. STATIC OPTIMIZATION OF FOOD-GRADE DYE INJECTION ALERTING MECHANISMS TO MINIMIZE HEALTH IMPACTS AND FALSE WARNINGS

4.1 Introduction

As a water distribution system (WDS) contamination threat or incident unfolds, water utility operators may take different assessment, preventive, and protective actions. Assessment actions collect information about the state of the system and may include sensor data analysis, contaminant source identification, and community impacts evaluation (Lindell and Prater 2003; Janke et al. 2006; Davis and Janke 2011; Liu et al. 2011). Preventive actions operate on the system to decrease impacts and may consist of hydrant opening for flushing, valve closure for isolation, and chlorine injection for disinfection (Baranowski et al. 2008; Parks and VanBriesen 2009). Protective actions require action by the public to reduce exposure and might include broadcasting general or targeted protective action recommendations (Zechman 2011).

Food-grade dye injection is a potential method for alerting a local population that its tap water might be contaminated. This strategy is mentioned in Module 5 of the U.S. Environmental Protection Agency Response Protocol Toolbox (USEPA 2004a) as a novel response action. Yet, despite its significant potential for use as an alerting mechanism, no previous analysis is known that has systematically modeled its implementation and investigated its performance for community health protection against municipal water contamination. This challenge is the focus of this chapter.

The suitability of different preventive and protective response actions may be evaluated based upon certain emergency response criteria including magnitude of desired effects, cost, time to implement, and unintended consequences. Performance in reducing public health impacts and accompanying sociopolitical consequences is the most important criterion. A response strategy should be effective, dependable, and robust. Preventive actions of hydrant opening and valve closure, for example, are promising tools, but a lack of understanding of WDS hydraulics could lead to ineffective response or worsened impacts. Chlorine disinfection boosters may not be effective against certain contaminants (e.g., cryptosporidium oocysts). Actions are more desirable if they impose lower direct costs for labor, equipment, and materials. Unintended consequences like indirect costs (system infrastructure recovery expenses, business disruption losses, etc.), firefighting interruption, and public alarm are also important due to utility operators' concerns over budget limitations and public trust. Isolation valve closure could potentially carry high recovery costs (e.g., due to water hammer damages) and significant consequences for industrial and firefighting uses. Execution errors in chorine disinfection practices could cause problems ranging from minor taste and odor changes to significant consumer health risk. The possibility of alerting consumers in uncontaminated regions is a possible adverse side effect of warning broadcasts (commercial radio and TV, reverse 911, sirens, etc.) that may generate antagonism toward officials particularly if contamination reports are not credible enough. A summary of how various preventive and protective response actions perform on the basis of these criteria is presented in Table 4.1. In this table, effectiveness denotes the

Table 4.1. Review of different possible WDS contamination emergency response actions

Option	Effectiveness	Direct Cost	Time to Implement	Unintended Consequences
Hydrant opening	Variable, dependent on event understanding	Low	Hours	Could worsen contaminant spread
Isolation valve closure	High if contaminant spread well understood	Low	Hours	No fire flow or non-consumptive uses; potential damage to WDS
Chlorine disinfection	Variable based on contaminant	High	Days	Potential false sense of security; alarming taste
Preventive action recommendations	Variable, dependent on warning medium and public access to media	Low to medium	Hours	Possible false public alerting; could lead to "over-compliance"
Do nothing	N/A	None	None	Continued exposure of consumers

percentage reduction in direct consequences of the contamination if it has indeed occurred.

This study formulates the decision making problem of food-grade dye injection and structures a multicriteria simulation-optimization framework for determination of best alerting mechanisms. Health impact reduction and false alert prevention are

independently treated as conflicting emergency decision making objectives. The simulation model comprises WDS hydraulics and quality simulators integrated with an exposure model that together evaluate the two criteria of public health consequences and size of the population alerted. A multiobjective genetic algorithm is used to minimize these emergency response objectives through optimizing dye injection practices. Post-processing is performed on Pareto-optimal solutions obtained by the optimization model to address the additional response criterion of implementation costs. The proposed multicriteria emergency management scheme is demonstrated using the WDS of Mesopolis virtual city.

4.2 Simulation Model

Both EPANET (Rossman 2000) and the multispecies extension to EPANET (EPANET-MSX) (Shang et al. 2008) may be used to simulate system hydraulics and propagation of contaminant introduced at a contamination site and dye injected by the utility. EPANET-MSX allows for a single simulation that considers the transport of the contaminant and the dye while EPANET requires separate simulation of the contaminant and the dye. This study uses the standard (single species) version of EPANET for extended simulation because it is less computationally intensive overall and easier to implement by utility operators. The contaminant and dye transport is simulated here as a perfect tracer: decay, density effects, and reaction with wall materials and other dissolved species are not taken into account.

The quantity of contaminant ingested by individuals during a contamination event depends on water ingestion pattern, time-varying concentration of the contaminant and dye, and water consumption choices made by the consumers under an unfolding contamination event. The timing-of-ingestion model considered here presumes that tap water is ingested at the regular starting times for the three main meals (7:00, 12:00, and 18:00) and times halfway between these meals (9:30 and 15:00). The tap water intake rate is here set to 0.93 L/day based upon USEPA (2004b).

Changes in consumers' water consumption choices after they observe dye in the tap water may depend on multiple factors such as age, gender, education, and ethnicity as well as dye color, intensity, and concentration. Different people may react in distinct ways. They may merely ignore color changes and keep drinking tap water as before, cease drinking water only at the times the changes in color are noticeable, or totally stop drinking water for a period of time after they observe the intense dye in the tap water. Moreover, consumers might also suspend contact uses, such as hand washing, dishwashing, and bathing after they observe dye. Other non-consumptive uses, however, may continue, such as toilet flushing, landscape watering, and pipe leaks. Since the hydraulic conditions in the system are dictated by the demands of consumers, these water consumption choices made by the consumers subsequently influence the hydraulic state of the network, and thus the spread of the contaminant plume in the system.

No qualitative or quantitative study or public survey is known that has addressed changes in consumers' ingestion choices after the observation of dye in the tap water. This study is thus performed based upon a certain set of assumptions. It is presumed that

consumers stop drinking water for the rest of the simulation period after they observe dye in the tap water with a concentration above a relatively high threshold. The simulation model checks this observation only at the ingestion times described in the timing ingestion model. To account for the fact that people may not observe the dye in tap water during the night, only the time period between 7:00 A.M. and 10:00 P.M. is considered for the period that people can stop drinking. Moreover, this study does not consider the influence of changes in consumer behavior on the hydraulics of the system. Agent-based modeling framework developed by Zechman (2011) may be employed to incorporate consumers' mobility, reduction of water demand, and word-of-mouth communication in the modeling schemes proposed here.

There exist a variety of food-grade dyes that may be utilized for injection. Allura Red dye (also known as "Red 40" and "E129") is considered in this study because of several advantages. It has an intense red color with high potential to strongly discourage people from ingesting contaminated water. It is approved by the U.S. Food and Drug Administration (USFDA 2012) for food use and does not pose any additional health risk to consumers. Moreover, it is relatively inexpensive and widely available for water utility use. In this study, the concentration threshold that causes people to stop drinking water is assumed equal to the concentration of Allura Red in commercial soft drinks -- roughly 25 mg/L as reported by Lopez-de-Alba et al. (1996; 2001). Public surveys are required for accurate calculation of this threshold and associated variance and uncertainties.

4.3 Optimization Framework

4.3.1   Problem Statement

Public officials with the authority to issue protective action recommendations face a difficult trade-off between health protection and possible false warning. Emergency decisions are made based upon imperfect information, in the form of uncertain threat observations, subjective system understanding, and approximate model predictions. Therefore, there is always a possibility that the contamination trigger events are false, and the managers thus may be falsely alarmed. Even if contamination has truly occurred, there is a possibility that false public alerting may occur; this would happen if people residing in the geographical areas that are not at risk of contamination observe dye in the tap water. Therefore, while injection of dye would discourage public consumption of potentially contaminated water to reduce health impacts, it should be executed such that it only targets regions or consumers that are (or will in the future be) exposed to the contaminant.

Thus, there is a conflict between reduction of potential health consequences (with consideration of threat credibility) and minimizing the magnitude of possible false public alerting. Considering the unquestionable fact that community health consequences are much more significant, one may pose the question whether emergency managers should be concerned about the trade-offs at all. Trade-offs among response objectives (that are the outcome of the optimization framework) are schematically illustrated in Fig. 5.1 to elucidate this critical dilemma. Every single dye injection plan from the trade-off front illustrated here corresponds to a decrease in possible health consequences and an

increase in the extent of public alerting (which could be potentially false). Since the occurrence of health impacts is conditioned on the occurrence of a contamination event, for the trade-off analysis to be realistic, the horizontal axis must be transformed to expected value to be explicitly comparable to the vertical axis, which represents unconditioned public alerting. This can be achieved by multiplying the horizontal axis by the conditional probability that the incident has indeed occurred given that a contaminant sensing system has reported "positive." Such probability values, however, are not known with certainty in reality and must be deduced from a chain of uncertain sensor readings and unusual observations. The optimization model theoretically finds a set of alerting mechanisms that cause minimized false alerting in regions that are not prone to the contamination. Nevertheless, it does not guarantee the alerting in risk areas is not false as it takes occurrence of contamination for granted in this chapter. Future research will incorporate likelihood of contamination (based on sensor network properties) and expected value of health consequences.

Every alerting mechanism has execution expenses, which are another response criterion to consider simultaneously. The problem of dye injection is therefore a multicriteria decision problem, which is addressed here using a multiobjective optimization approach. Since dye injection is a subset of preventive response actions that are more appropriate to be implemented after threat credibility elevates to 'credible' or 'confirmatory' level, the response criterion of cost is considered less important than the two critical criteria of protecting health and avoiding unnecessary public alerting. Accordingly, multiobjective optimization is first performed considering only the two latter criteria, and post-processing is performed later on optimization results to account for the implementation cost criterion.

There exist different possible ways to quantify health impacts due to water contamination and the extent of public alert as a result of dye observation. Public health impacts are expressed here as the total ingested mass (TIM) of contaminant by all
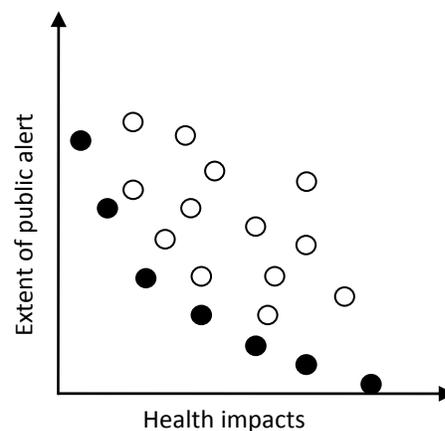
Fig. 4.1. Non-dominated alerting mechanisms (shown as filled circles)

consumers during the course of a contamination incident because it removes the need to specify an infectious or lethal threshold in advance. Public alerting is quantified as the number of people who observe colored water with concentration of dye above the preset threshold and consequently stop drinking water. In mathematical terms, the multiobjective optimization problem for the minimization of public exposure (f_1) and the extension of public alert (f_2) is expressed as:

$$\min \quad f_1 = \sum_{i=1}^{N_C} \sum_{j=1}^{N_I} V_{i,j} \times C_{i,j} \tag{4.1}$$

$$\min \quad f_2 = \sum_{j=1}^{N_C} \alpha_j \tag{4.2}$$

where N_C = number of consumers, N_I = number of water ingestion events for each consumer, $V_{i,j}$ = volume of water ingested by consumer j at ingestion event i, and α_i = a binary variable that is 1 if consumer j is alerted by dye presence and 0 otherwise. The decisions that should be optimized for dye injection include: (1) the location(s) for inserting dye into the WDS, (2) mass of dye injected in each location, and (3) duration of injection. Sensitivity analyses should be performed to assess efficacy of response for different numbers of locations and the response delay between the start of contaminant intrusion and when injection of contaminant in the network is identified as likely or confirmed through multiple contamination trigger events.

4.3.2   Solution Algorithm

The underlying hydraulics of the WDS coupled with the exposure and quality simulation models suggests that the optimization problem involves both significant nonlinearities and a high potential for multi-modality. Genetic algorithms (GAs) have been demonstrated as flexible and powerful tools for solving such challenging optimization problems in the discipline of water resources planning and management (Nicklow et al. 2010). NSGA-II is employed here uses simulated binary crossover (SBX) (Deb and Agrawal 1995), and polynomial mutation (Deb 2001) for reproduction of new solutions (injection location, mass, and duration). To explore new locations, reproduction operators are performed on geographical coordinates of the parent solutions and the nearest intermediate node in the network is selected for offspring solutions.

4.4 Application

Virtual city of Mesopolis is used to demonstrate optimization of dye injection using the proposed framework. Two contamination scenarios are selected for which the dye injecting alerting mechanisms are optimized (Table 4.2). The low rate of contaminant ingestion is consistent with the fact more than 99.8% of WDS inflow goes to non-ingestive uses. The demand multiplier associated with each scenario is representative of aggregate water demand for a WDS that typically varies throughout the year. The contaminant agent is arsenic with a toxic dose of 3.5 mg (milligram) for a body weight of 70 kg as reported by Office of Environmental Health Assessment Services (1999). While optimization is performed here considering only TIM as the

Table 4.2. Contaminant source characteristics and associated health impacts

Scenario	Location	Load (kilograms)	Demand multiplier	Start time	Duration (hour)	# of sicknesses	TIM (grams)
1	West WTP	300	1.00	18:00	6	33,944	323.75
2	East WTP	300	1.00	19:00	5	54,638	402.06

health consequences metric, obtained results may be used to gain approximate information on reduction in number of sicknesses too. Total simulation time is 6 days with both contamination scenarios occurring on the third day of the simulation after the system has reached dynamic equilibrium. Total cost of each dye injection mechanism is considered here to be the sum of cost of all injectors and total mass of dye used. Price of each dye injector unit and unit mass (kg) of dye are set to $10,000 and $180, respectively. A dye injector unit is conceived here as a portable, trailer-mounted apparatus consisting of a pump, motor, dye tank, and hoses that could be connected to a hydrant.

Appropriate NSGA-II parameters settings are determined based on sensitivity analyses. The population size is set to 100 and the stopping criterion is achieved when the total number of generations reaches 150. Crossover and mutation rates are 0.85 and 0.07 and SBX crossover distribution and polynomial mutation indices are set to 15 and 5, respectively. The convergence history during the evolution process for contamination scenario 2, a 6-hour response delay, and 5 dye injection locations is shown in Fig. 4.2.

As the algorithm proceeds, the population evolves from a scattered cluster mostly concentrated in high-health-impact and low-public-alerting zone to a diverse Pareto optimal front.
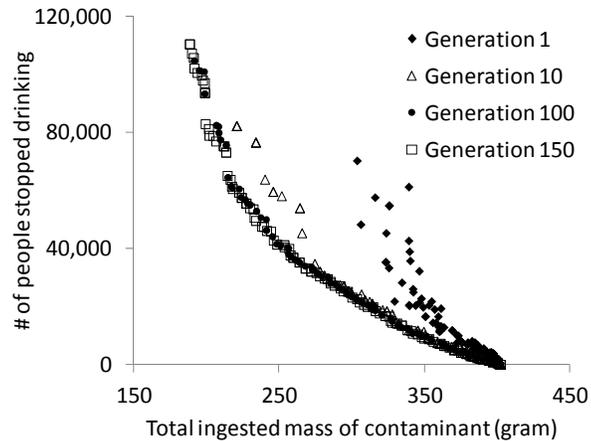
Fig. 4.2. Optimization convergence history for Scenario 2, 6-hour response delay, and 5 dye injectors

Pareto optimal fronts obtained for different configurations of the alerting system (i.e., number of dye injection locations and response delay times) are shown in Fig. 4.3. These results clearly indicate the very significant trade-offs between minimization of health impacts and extent of public alerting, which is not surprising. Comparatively, injection of dye can more effectively reduce health impacts where the West WTP is contaminated (Scenario 1). The reduction in TIM can be as high as 90% which demonstrates the effectiveness of dye injection as a strategy for responding to WDS contamination emergencies. Even when only three dye injectors are utilized after a 12-
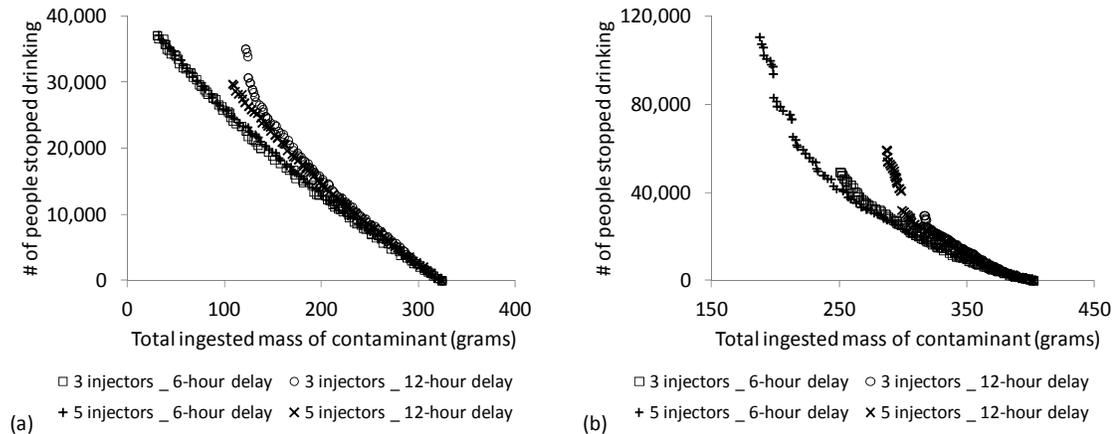
Fig. 4.3. Pareto optimal fronts for contamination Scenarios 1 (a) and 2 (b)

hour response delay, this reduction is quite significant (62%). Difference in effectiveness of dye injection for two scenarios associated with the West and East WTPs is rooted in WDS hydraulics. The East WTP supplies a much larger area than the West WTP so it is more difficult to manage the impacts when the East WTP is contaminated. This also clarifies the noticeable difference between the maximum number of people that are alerted in each scenario. The population of the western region is 37,099 in Scenario 1 whereas there is a much larger population (110,414 people, approximately 75% of the city's population) to be alerted in Scenario 2.

Fig. 4.3(a) indicates that increasing the number of injection locations from 3 to 5 does not result in any noticeable improvement in the solutions for contamination of the West WTP (Scenario 1). If the East WTP is contaminated (Scenario 2), by contrast, this increase in injection locations enhances the effectiveness of non-dominated alerting

strategies as shown in Fig. 4.3(b). This observation may also be explained by the spatial difference in impact area associated with each contamination scenario. Contamination of the East WTP impacts a much larger area and, therefore, an increase in the number of injection locations would allow the water managers to alert more consumers and consequently achieve higher protection.

Fig. 4.4 illustrates the optimal alerting systems associated with minimum health impacts for a 6-hour response delay with 3 and 5 locations for Scenario 1 and 2, respectively. This figure provides insight into the optimal decision space and supports the information presented in Fig. 4.3 on the optimal objective space. Unsurprisingly, the best injection locations are large transmission mains in the network. These locations allow a larger fraction of total population to be alerted to existence of contaminant in the tap water. There are some dye injection locations in the alerting system that are relatively far from contaminant sites. Further analysis inspired by these observations indicates that injection of dye at these points alerts people residing in far regions before contaminant even reaches them. This action, therefore, completely reduces the exposure risk for these population segments.

Results presented in Fig. 4.4 generally show that the alerting system would be more effective if dye is injected during a short time. This result might be due to the assumptions about people's response after they observe dye in their tap water. As mentioned earlier, it is presumed that people stop drinking after dye concentration exceeds a threshold. The optimization model selects shorter injection durations that
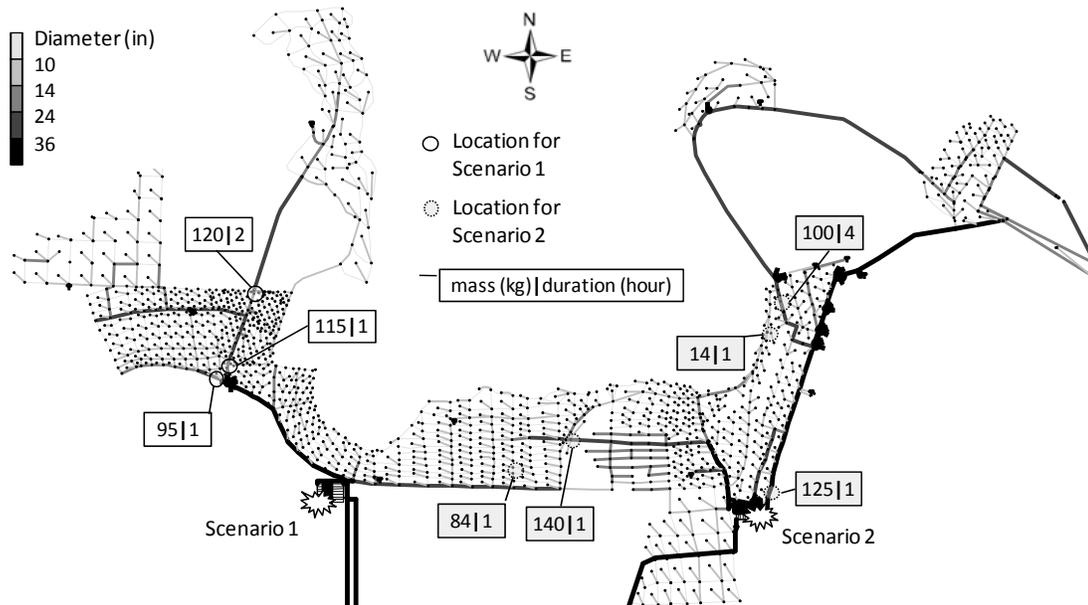
Fig. 4.4. Minimum-TIM non-dominated alerting mechanisms (injection locations, mass, and duration) for a 6-hour response delay with 3 and 5 injectors for Scenarios 1 and 2, respectively

result in higher peaks in the dye concentration time series, which subsequently increases the chance of exceeding the threshold.

Population segments in the city ingest different levels of contaminant mass and observe different concentrations of dye in the water. Fig. 4.5 visualizes ingested contaminant mass per capita and maximum observed dye concentration for different population sectors. These results are associated with the minimum-TIM non-dominated alerting system for 6- and 12-hour response delays with 3 locations for Scenario 1.

Figs. 4.5 (a) and (c) indicate ingested mass of contaminant per capita for the case in which the population takes no protective actions in response to the alerting system whereas Figs. 4.5 (b) and (d) provide the same results for the case in which they do take protective actions. As shown for both response delays, a large fraction of total population (109,298) is located in safe areas (central and eastern region) that are never contaminated. The optimization algorithm configures injection locations, mass, and duration such that this large population segment never observes any dye in the water and, accordingly, is not unnecessarily alerted. It alerts the rest of population that is at the risk of contamination to take protective actions. While people would stop drinking contaminated water if the dye concentration exceeds a fixed threshold, interestingly, the figure indicates that for the 6-hour response delay, maximum observed dye concentration is generally greater for population sectors that are at higher risk. Since these consumers contribute more to the overall health impact, the optimizer tries to locate injection points closer to these regions to more quickly alert those consumers.
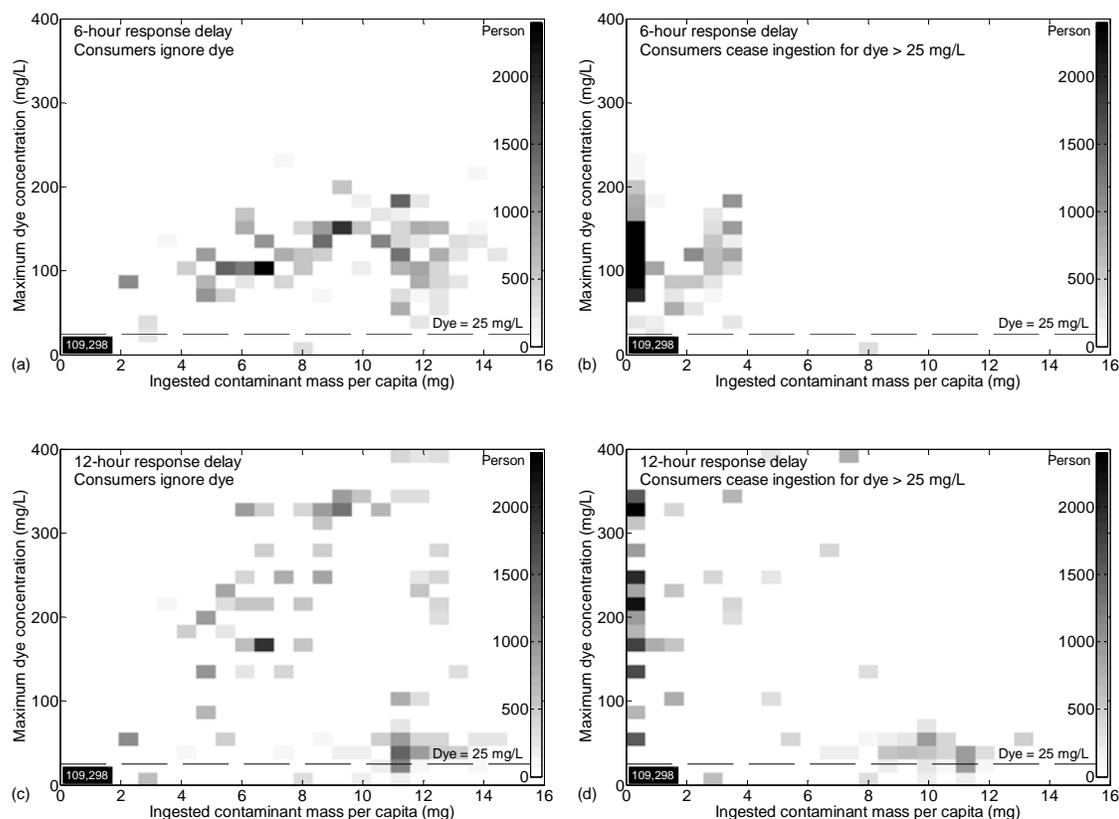
Fig. 4.5. Contaminant-dye targeting performance (a and c) for minimum-TIM alerting protocols with 3 injectors for Scenario 1. The distribution of reduction in health impacts through executing the dye injection protocols is shown by the shifting of cells toward the vertical axis (b and d). The black cell at the origin indicates that 109,298 persons are never exposed to the contaminant or dye.

Figs. 4.5 (b) and (d) indicate how the health consequences are changed for increased delays to the start of dye injection: here, delays of 6 and 12 hours, respectively. A 6-hour delay response both reduces the impacts generally and dampens

the extreme impacts noticeably. If delay reaches 12 hours, however, the response is not as effective in reducing severe impacts. Population segments associated with these high values of ingested contaminant mass are those residing around the large mains near the contamination source (West WTP) where a 12-hour delay is long enough for the consequences to become severe. Nevertheless, people living in more distant areas can be alerted before the contaminant plume reaches them. For these population segments, the ingested contaminant mass becomes zero as shown in Fig. 4.5. In particular, this includes the people residing in the isolated western peninsula, the farthest district from the contamination source. On the contrary, certain exposed population segments never observe a dye concentration above the stop-drinking threshold. The health impacts associated with these consumers are thus never prevented; essentially, the optimizer "gives up" on populations that it cannot help. As expected, Fig 4.5 indicates this population segment increases in number when response is executed later. A fraction of this population comprises the people living in the vicinity of the large water main who are exposed for a short period soon after the contamination starts. Alerting is thus of limited effect when the response delay is close to or longer than this exposure period.

Due to the lower importance of response implementation costs at the stage of the emergency when the presence of contaminant is deemed credible, these expenses were not addressed directly in the optimization process. However, to avoid unnecessary expenses, post-processing can be performed in the optimization results to quantify cost associated with each alert system option. The optimization model is first run for different numbers of injectors ranging from 1 to 9 for Scenario 2 considering public health and

extent of population alerting as objectives. Thereafter, obtained Pareto-fronts are combined and the cost associated with each response is calculated. Non-domination sorting is consequently performed considering the health, alerting, and cost as sorting criteria and the non-dominated mechanisms are determined. Fig. 4.6 indicates the non-dominated solutions in the objective space where color coding is used to illustrate the cost criterion. While the reduction in health impacts becomes negligible after the number of injectors exceeds 6, the response costs consistently rise the number of injectors. Nevertheless, although use of fewer injectors would be just as effective in limiting the impacts at lower costs, this protection may not be achieved unless greater public alerting takes place. This observation is due to the fact that optimization model can take advantage of a greater flexibility in configuring the alerting mechanism when more injectors are used. These results are particularly helpful for estimating the minimum number of injectors that a utility should install in order to guarantee the most effective response in case an emergency occurs. Since these results are for the scenario that the East WTP is contaminated (which results in the largest contamination spread area), the number of injectors selected for this scenario would suffice in case any other location in the city experiences contaminant intrusion.
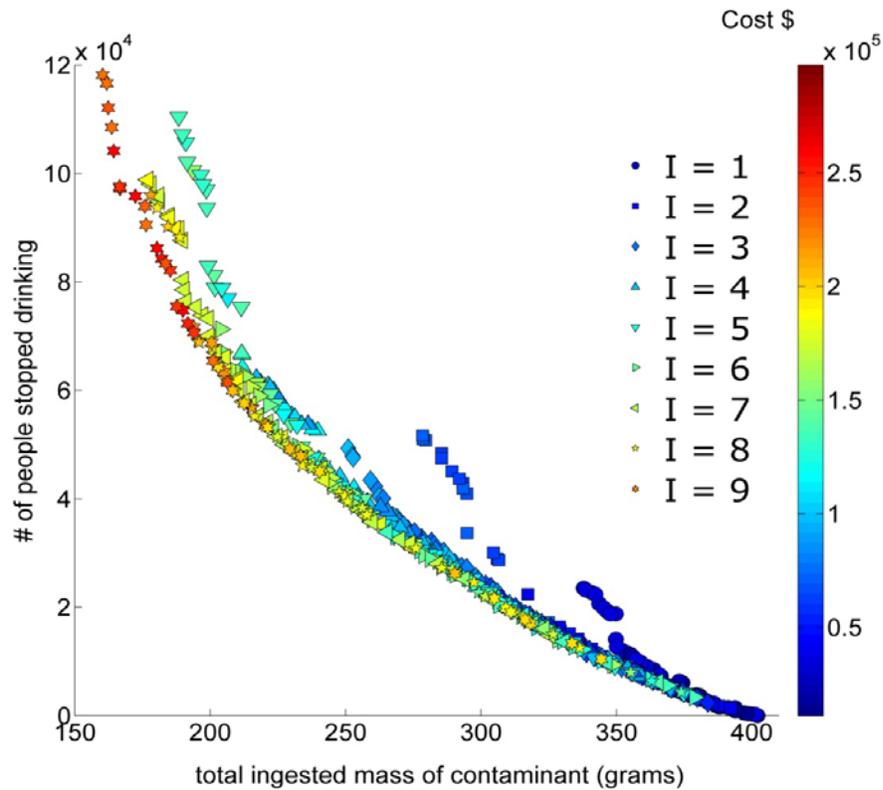
Fig. 4.6. Alerting protocols for Scenario 2 with different trade-offs between public

health, alarm, and execution cost for varying number of dye injectors (I)

4.5 Conclusions and Future Work

This chapter has demonstrated that food-grade dye injection in water distribution

networks holds promise as an effective strategy for reduction of health impacts due to

contamination. It is relatively inexpensive, easy, and quick to implement, causes no

physical damage to system infrastructure, and does not interrupt fire protection and other

non-consumptive uses. However, managers' overaction and lack of the WDS hydraulics

understanding may result in unnecessary public alerting for population segments not

residing in risk areas. The mathematical problem formulation and multiobjective

simulation-optimization framework proposed in this chapter can provide a meaningful guide for utility operators to effectively reduce health impacts without unnecessary public alerting. It serves a population protection function by recommending protective response actions for areas determined or predicted to be at higher risk from contamination impact. The algorithms can be used to guide decisions on equipment and material needs for alerting as well as formulation of response plans.

The optimization results indicate that the dye would be better injected in larger mains at points that are not all located in the vicinity of the contamination source. Injection at larger mains assures more people are notified while spread of injection locations enhances timeliness of warning by notifying people before the contaminant plume reaches them. Assuming that people will stop drinking water after they observe a concentration of dye above a threshold, the entire amount of dye would be better injected during a short time period to maximize peak dye concentration. Injection of dye in more locations would enhance both the level of impact reduction and the number of people alerted for the same level of protection but at the cost of higher direct expenses. This improvement is achieved through a more uniform and guided dye concentration through the system. Nevertheless, no further enhancement is achieved after the number of injectors exceeds a certain level, which can be determined through a sensitivity analysis for the set of contamination scenarios that result in most severe consequences.

Food-grade dye injection modeling and practice is in its infancy and extensive future research is required to address various sociotechnical aspects. Expectations about people's behavior during contamination impacts shapes the way that emergency

managers plan for emergency response. Public surveys should be conducted to collect data on consumers' perception to guide the development of a sociotechnical model that accurately predicts people's water consumption after they observe dye in their tap water. This would help to produce more accurate calculation of health impacts and public alerting for more realistic evaluation of dye injection functions.

Future work should explore the effects of dye injection when it is executed in combination with other response actions such as contaminant containment and warnings broadcast through the media and other mechanisms. Public education would also be of considerable value to prepare people for possible observation of dye in their tap water. This would help to increase compliance and would also alleviate utility operators' concerns that consumers will be unnecessarily distressed and not know what to do when they see intense dye color in their tap water.

5. DYNAMIC OPTIMIZATION OF FLUSHING AND ALERTING

MECHANISMS TO MINIMIZE HEALTH IMPACTS

5.1 Introduction

The models presented in Chapters 3 and 4 applied static optimization approaches to find Pareto-optimal response plans on the implicit assumption that system behavior and contamination source characteristics remain unchanged once a contamination event begins and model computation is started. Mathematically, this assumption implies that the response optimization fitness functions (e.g., minimization of ultimate health impacts) are not time-varying and are not subject to feedback mechanisms driven by dynamically introduced system parameters. In other words, the objective function is not changing during the optimization process. In reality, however, the fitness functions are feedback-influenced by several system parameters that change over time as the emergency proceeds. The decision support model should thus explicitly account for the changing behavior of the system to realistically identify effective contamination risk mitigation decisions in a timely manner.

Dynamic optimization techniques have been successfully applied in different engineering disciplines for solving optimization problems in changing environments. Applications include products pricing (e.g., Besbes and Zeevi 2009), vehicle routing (e.g. Khouadjia et al. 2010), contaminant source characterization (e.g., Liu et al. 2011), chemical batch process scheduling (e.g., Nie et al. *in press*), and mission planning (e.g., Bui et al. *in press*). Dynamic optimization methods methodically transfer useful

knowledge from previous environments and maintain adaptability to guide and speed up the search in changed environments.

This chapter develops a dynamic simulation-optimization model for identification and tracking of time-varying optimal response to provide emergency mangers with realistic, real-time decision support. The adaptive simulation model accounts for multiple sources of uncertainty and variability, including perceived contaminant source attributes, consumers' water use, and emergency management operations. This dynamic optimization scheme uses an evolutionary-computation-based multiobjective approach where the adaptability and diversity in the search process are preserved through defining and maximizing an artificial diversification objective. The proposed decision support scheme is demonstrated and discussed on a WDS that possesses the spatial and temporal complexity of real-world systems.

5.2 Dynamic Environment Simulation

In the context of WDS dynamics during normal operation, a system is expected to exist in a dynamic equilibrium where system behavior follows a repeating consistent pattern that is known with acceptable accuracy. As a WDS is contaminated, the water system exhibits a complex and uncertain behavior that significantly deviates from the normal operation conditions. Knowledge of contaminant source characteristics that dictate emergency response decisions evolves as more information streams in over time. Emergency mangers change system normal operation conditions and alert consumers based upon their current assessment of the state of the system. Warned or sickened

consumers subsequently change their water consumption choices, which consequently affect network hydraulics and contaminant plume spread. Conceptualizing and modeling these different sources of uncertainty and complexity is fundamental to realistic simulation of system behavior and effective reduction of contamination risks.

5.2.1 Contaminant Source Perceived Attributes

The perceived attributes of a contamination event, including the location, strength, time, and duration, are estimated through integrated assessment of different system observations and evidence that streams from physical security alarms, sensor networks, and consumers' complaints. Bayesian and optimization models have been applied to process the streaming data in real-time and update estimated source characteristics based on observations up to the current time (Wesley et al. 2006; Liu et al. 2011). Since the perceived source characteristics dictate the suitability of mitigation strategies taken, the optimization process needs to adapt to these changes to be capable to continuously track the optimum in a time-varying search space. The dynamic model proposed in this chapter adapts optimal response decisions to changing perceived source characteristics in real-time through systematic preservation of diversity in the search procedure.

5.2.2 Water Utility Operations

In the event that a contaminant is introduced to a WDS, water utility operators may take different assessment, preventive, and protective actions to protect public

health. These actions taken by the utility managers will change the normal hydraulic conditions, and thus the propagation of the contaminant plume, in the system. Implementation of response actions thus alters public health consequences and changes the effectiveness of future response decisions. Effectiveness of risk reduction measures also degrades as time passes due to wider spread of contaminant and prolonged exposure of public to the contaminant. The dynamic simulation model developed here adaptively evaluates the effectiveness of different response actions through consideration of the increasing response time delay and the effect of previously executed actions on system behavior. Two response strategies of hydrant operation for contaminant flushing and dye injection for public warning are included in the model. No other warning systems are used here.

5.2.3   Consumer Behavior

Different consumers ingest varying amounts of contaminant depending upon time-varying concentration of the contaminant and dye in their tap water, and the water consumption choices they make under the unfolding contamination incident. The exposure model used in this study assumes consumers ingest tap water at the typical starting times for the three main meals (7:00, 12:00, and 18:00) and times halfway between these meals (9:30 and 15:00). The tap water ingestion rate for every consumer is set to 0.93 L/day in the model.

Consumers cease drinking tap water when they become aware that their tap water is contaminated. This happens either when they are sickened and assume tap water is the

cause or when they observe intense dye color in their tap water. A series of rules needs to be defined for modeling consumers' reactions and their water usage changes. Consumers are sickened and experience symptoms once they ingest a threshold toxic dose and a certain time period passes after the threshold is exceeded. The contaminant agent used here is arsenic with a toxic dose of 3.5 mg for a body weight of 70 kg as reported by Office of Environmental Health Assessment Services (1999). The model assumes that within one hour after this toxic dose is ingested, the consumers experience symptoms. Consumers may also be altered through observation of dye. It is presumed that they become alerted and cease drinking water for the rest of contamination incident once the dye concentration exceeds a relatively high threshold. The simulation model checks this observation of high-intensity dye only at the daily ingestion times described in the time-of-ingestion model. Allura Red dye (also known as "Red 40" and "E129") is chosen in this study among different available food-grade dyes because of several advantages.

These water usage reduction choices made by alerted consumers influence the hydraulic state of the network, and thus the spread of the contaminant plume in the system. Consumers may suspend contact uses, such as hand washing, dishwashing, and bathing after they become alert to the contamination. Other non-consumptive uses, however, may continue, such as toilet flushing, landscape watering, and pipe leaks. Such uses are assumed to comprise on average 60%, 51%, and 43% of the total demand for low, medium, and high density residential demands, respectively, using the information reported by Vickers (2001) for urban water use. It is assumed the residential users reduce

their water usage to these values after they are alerted. Industrial users are assumed to maintain 96% of their total water usage. A more realistic model may be developed through incorporation of consumers' mobility and word-of-mouth using, for instance, the complex agent-based modeling framework developed by Zechman (2011).

5.2.4 Network Hydraulic Simulation

EPANET software is used for hydraulic simulation of the WDS. It is a publicly available hydraulic and water quality modeling program developed by the US Environmental Protection Agency (Rossman 2000). It provides an integrated computer environment for an extended-period hydraulic and quality simulation of WDSs within pressurized pipe networks. The contaminant and dye transport in the network is modeled here as a perfect tracer, meaning the model does not account for decay, density effects, and reaction with wall materials and other dissolved species.

5.3 Dynamic Evolutionary Optimization

Emergency response decisions of hydrant operation and food-grade dye injection should be optimized for effective mitigation of the public health risks. The health impacts are expressed here mathematically as the ultimate total ingested mass (TIM) of contaminant by all consumers during the course of a contamination incident:

$$\min \quad TIM = \sum_{i=1}^{N_C} \sum_{j=1}^{N_I} V_{i,j} \times C_{i,j} \tag{5.1}$$

where N_C = number of consumers, N_I = number of water ingestion events for each consumer, $V_{i,j}$ = volume of water ingested by consumer j at ingestion event i, and $C_{i,j}$ = concentration of contaminant in water volume ingested by consumer j at ingestion event i. Considering multiple sources of uncertainty and variability described in the previous section, Eq. (5.1), which indicates the predicted value of ultimate health impacts at every stage of the emergency, represents a time-varying function. Static optimization algorithms are insufficient for dealing with such changing objective functions. They need to be modified to adapt rapidly to changes in environment for generation of effective response plans at every phase of the emergency. Obviously, the simplest approach to respond to a change in the environment is to consider each change as the emergence of a new optimization problem that needs be solved from scratch. Given sufficient time, this is obviously a feasible approach. However, the time available for re-optimization is normally short during an emergency. Moreover, this approach presumes that a change in the environment can be identified, which is not always true. Dynamic optimization techniques systematically reuse information from previously explored environments to accelerate optimization process in emerging environments.

Evolutionary algorithms (EAs) resemble natural biological evolution, and since evolution is a continuous adaptation process in nature, they are promising candidates for tackling dynamic optimization problems (Jin and Branke 2005). To solve dynamic optimization problems, static EAs should be modified to adapt and recover from the changes during the evolution process. Major modifications in the static EAs are necessary for a timely adaptation to the changing environment to balance between

convergence and exploration. Compared to static EAs, higher emphasis should be placed on exploration after a change occurs, so that the algorithm can react rapidly to the change and track the moving optimum. Different methods have been proposed to deal with this issue, which can be classified into four groups (Jin and Branke 2005; Bui et al. *in press*):

1) *Boost diversity after a change*: the EA is initially run in standard fashion. As soon as a change in the environment is identified, explicit strategies are implemented to generate diversity in the population. A common technique is hypermutation (Cobb 1990), where the mutation rate is significantly increased for a limited number of generations after the change event is detected and then decreased over time. A very high mutation rate essentially results in a re-initialization of the population, whereas a low mutation rate does not boost sufficient diversity of the population. The difficult task of tuning the mutation rate changes is the major drawback of this approach.

2) *Maintain* diversity *throughout the run*: convergence is limited through constant diversification hoping that a diverse population is more promising to adapt to time-varying changes. Random immigrants method (Grefenstette 1992), where new individuals are regularly introduced into the population, thermodynamic genetic algorithm (Mori et al. 1997), where the original objective function is replaced with an entropy-based value, and multiobjective-based method (Bui et al. 2005), where an artificial objective is used to promote diversity, are representatives of this approach.

3) *Memorize good solutions*: the algorithm retains good solutions from past generations. This strategy provides diversity and helps the algorithm retrieve the

optimum in repetitive environments. Diploidy approach (Goldberg and Smith 1987), where redundant representations are used to generate solutions, is a popular instance of memory-based approaches.

4) *Use multiple subpopulations*: the population is clustered into multiple subpopulations that evolve together to explore multiple promising regions of the decision making space. Some representative methods are multinational GA (Ursem 2000), self-organizing scouts (Branke et al. 2000), and the shifting balance approach (Wineberg and Oppacher 2000).

This study employs the multiobjective-based diversity preservation approach, which has been demonstrated as a robust and efficient method by previous research (Toffolo and Benini 2003; Bui et al. 2008). The main advantage of this technique is that it eliminates the need for defining *a priori* the proper diversity preservation parameter. The proper balance between convergence and exploration is systematically preserved during the process through treating diversity as a secondary (artificial) objective, which is optimized simultaneously with the main (true) optimization objective, as schematically illustrated in Fig. 5.1.

Without the loss of generality, this chapter considers only the minimization of health impacts as the single *true* objective function among all objective functions that were considered in the previous chapters. The multiobjective-based diversity preservation approach used here "multiobjectivizes" (Handl et al. 2008) this classic single-objective optimization problem to a bi-objective optimization problem. Health impacts are minimized simultaneously with maximizing the *artificial* objective function
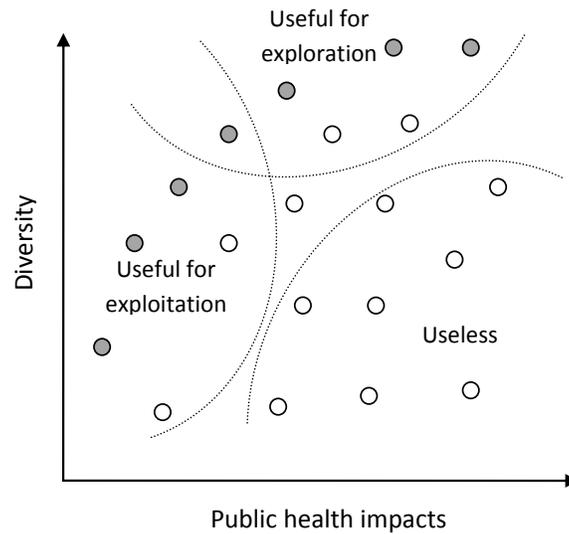
Fig. 5.1. Diversification of GA solutions in multiobjective-based dynamic optimization approach for methodological balance between exploitation and exploration in search process (filled circles represent Pareto-optimal solutions)

of diversity hoping added diversity to the GA population helps tracking the changing optimum in the dynamic environment. Once the multiobjectivization is performed, any traditional multiobjective optimization algorithm may be used to solve the constructed bi-objective optimization problem.

The artificial diversity-preservation metric may be mathematically expressed in different ways. The three following formulations are examined here:

1) Distance from the nearest neighbor (DNN): The artificial objective for a solution x_i is defined as the distance from x_i to its nearest neighbor. Therefore, a pair of very similar

individuals will have a relatively poor artificial objective value, and thus the diversity in

population is encouraged over the search space.

$$DNN(x_i) = \min d(x_i, x_j), j = 1,...,N_p \quad j \neq i \tag{5.2}$$

where N_p is the population size.

2) Distance from the best solution of the population (DBS): The diversity metric is

expressed as the distance from x_i to the current best solution in the population x_{best} (with

respect to the true objective function) to avoid any likely trap caused by local optima.

$$DBS(x_i) = d(x_i, x_{best}) \tag{5.3}$$

3) Average distance from all solutions (ADS): Diversity is quantified as the average

distance of x_i to all other individuals in the population. This formulation prefers

solutions at the edge of population to boost the spread of the population.

$$ADS(x_i) = \frac{1}{N} \sum_{j=1}^{N_p} d(x_i, x_j) \tag{5.4}$$

The bi-objective optimization problem defined by Eq. (5.1) and any of Eqs.

(5.2)-(5.4) may be solved using any classic multiobjective optimization algorithm. Non-

dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al. 2002) is among the most

popular algorithms for solving both classic multiobjective water resources problems and emergent dynamic optimization problems. The algorithm applied in this study uses simulated binary crossover (Deb and Agrawal 1995), and polynomial mutation (Deb 2001) for reproduction of offspring (contaminant flushing and dye injection locations). To explore new locations, reproduction operation is performed on coordinates of the parent solutions and the closest intermediate node in the WDS is chosen as an offspring solution.

5.4 Application

The virtual city of Mesopolis is used to demonstrate optimization of dye injection using the proposed framework. One contamination scenario is selected for optimization of hydrant opening and dye injecting alerting mechanisms. For this scenario, aggregate demand multiplier is 1.00 and 300 kg of arsenic is inserted into intermediate node IN0655 that is located in the vicinity of the East WTP. Injection starts at clock time 00:00 over a period of 3 hours. The simulation duration is 24 hours and the dynamic optimization model run starts after a 6-hour response delay after the injection starts, i.e. 06:00. The ultimate TIM is 146.2 grams if the managers take no action and consumers continue drinking after their ingested mass of arsenic exceeds the toxic dose of 3.5 mg and they observe the symptoms. If consumers change their water use behavior once they have received a toxic dose, the ultimate TIM is reduced to a smaller value of 138.0 grams.

To analyze effect of changes in perceived source attributes, it is assumed that contaminant location and injection duration are first wrongly perceived to be the East WTP and 5 hours, respectively. Perceived ultimate TIM for this wrong estimation of true scenario attributes is 220.0 grams when no action is taken by the managers and consumers. If action is taken by the consumers, this is reduced to 211.8 grams. It is presumed that this wrongly perceived scenario is updated to the true scenario at time 09:00 in the model.

The number of hydrants and dye injectors are each set to 3. Hydrants are opened for 5 hours when used for flushing. The amount of dye injected per each injector is set to 100 kg and the injection duration is 1 hour. The NSGA-II population size is set to 50 and the model is run until clock time 18:00. Crossover and mutation rates are 0.85 and 0.04, and SBX crossover distribution and polynomial mutation indices are set to 15 and 10, respectively. The optimization is run 5 times for every optimization case.

Analysis is first performed to identify the best diversity measure among the three measures of DNN, DBS, and ADS. Both response strategies of hydrant operation and dye injection are used in this analysis. Changes in consumers' water use and perceived scenario attributes are not considered for this analysis. Fig. 5.2 indicates the time series of the mitigated ultimate TIM corresponding to the best solution at every time step. The area confined by the time series curve and the horizontal axis provides a numeric value for the comparison of effectiveness of different measures: a smaller area corresponds to a more efficient measure. This area is 2,182.9, 2,201.2, and 2,143.4 for ADS, DBS, and DNN, respectively. Conclusively, through quantitative and visual comparison, DNN
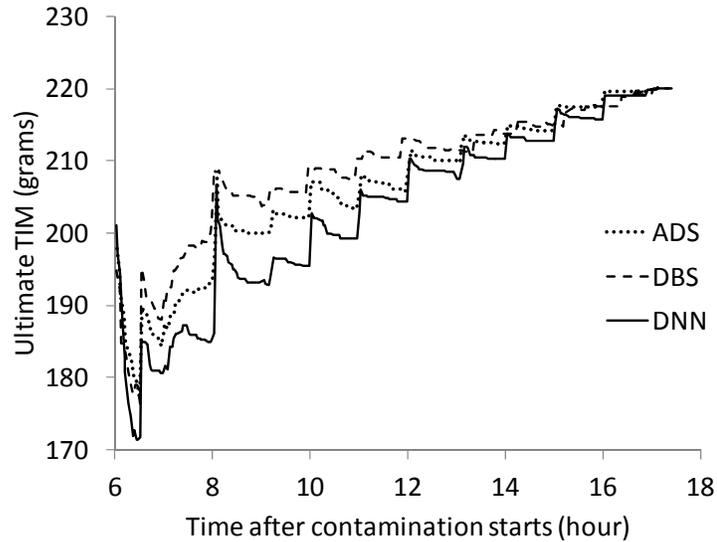
Fig. 5.2. Time series of minimum ultimate TIM obtained using different diversity

measures for the perceived scenario attributes

outperforms ADS and DBS and is selected as the more efficient diversity measure and used for the later analyses in this chapter.

After the dynamic optimization model is run, the quality of optimal solutions suggested by the model changes over time. Longer time allows the model to better explore the search domain and converge to better solutions. This extended delay, however, reduces the effectiveness of solutions because of the prolonged exposure of the public to the contaminant and wider spread of the contamination. Fig. 5.3 shows the time series for minimum-TIM solution offered by the optimization model using hydrant operation, dye injection, or both strategies. The mitigated health impacts reduce during the first hour and generally increases afterwards for all three cases. Considering the simulation-optimization setting described above, visual inspection shows dye injection
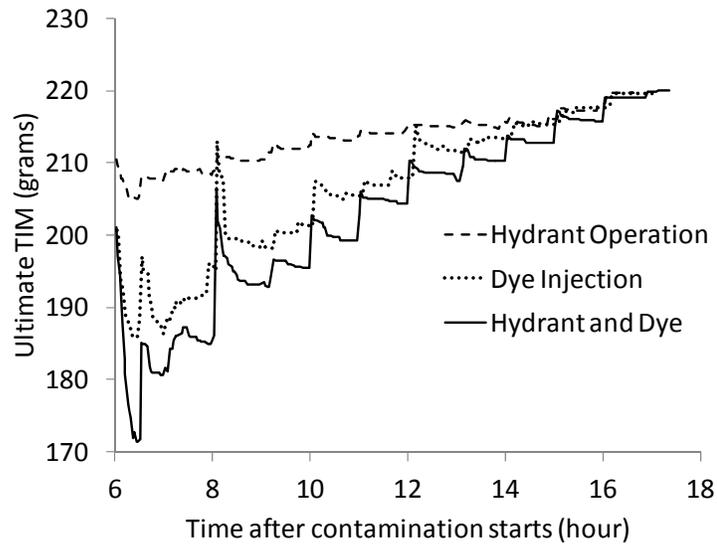
Fig. 5.3. Time series of minimum TIM obtained using strategies of hydrant

operation, dye injection, or both, for the perceived scenario attributes

outperforms hydrant operation, and best performance is achieved when both strategies

are implemented, that is not surprising. The area measure is 2261.2, 2180.6, and 2,143.4

for time series corresponding to hydrant operation, dye injection, and both strategies,

which quantitatively confirms the visual comparison findings. After about 14 hours,

nevertheless, the effectiveness of all three response cases becomes practically similar

due to the significantly long response delay.

Besides the response delay, the effectiveness of emergency response actions

taken at every time step depends on the previous response strategies executed by the

emergency managers and the changes in perceived scenario attributes identified by the

stream of new information that becomes available. Fig. 5.4 shows how the time series

for minimum ultimate TIM changes when these factors are taken into account. Time
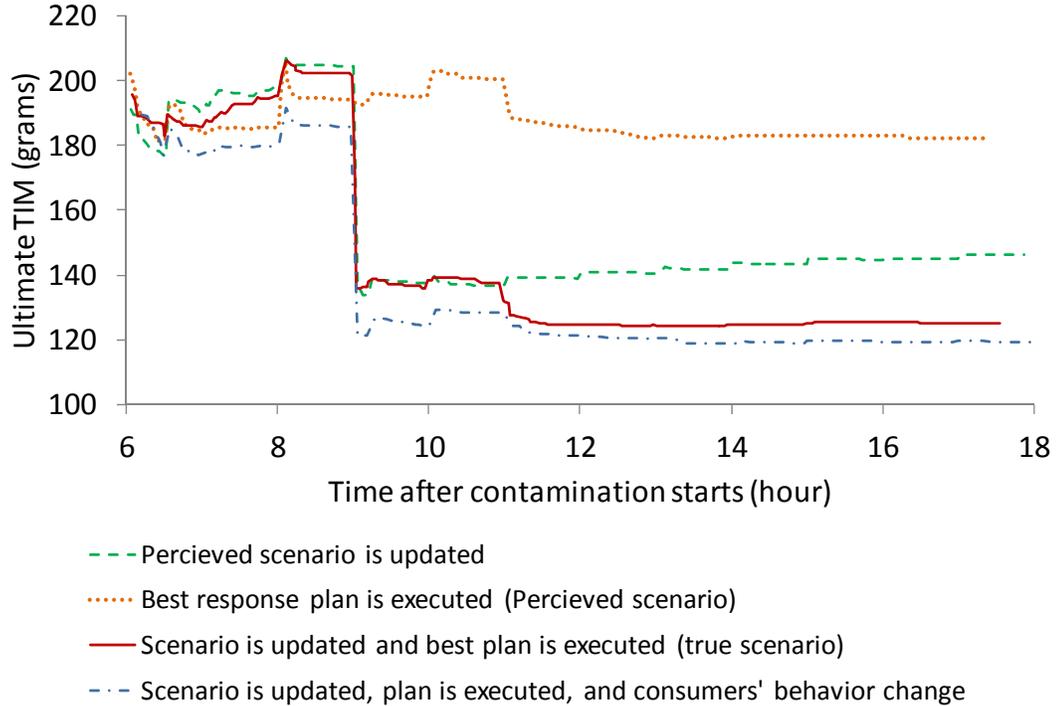
Fig. 5.4. Time series of minimum TIM considering the effects of managers' actions

(best plan is executed at 11:00), perceived scenario changes (scenario is updated at

09:00), and consumers' water use behavior changes

series associated with these changes shown in Fig. 5.4 indicate that the minimum

ultimate TIM drops suddenly after the scenario is updated, which can be attributed to the

smaller exposure zone associated with the intermediate node than the East WTP.

Moreover, it is presumed that the managers execute the minimum-TIM response

plan recommended by the model at clock time 11:00. This change is manifested by a

drop in TIM in Fig. 5.4. The model considers effect of this executed action when

evaluating later actions. It is generally observed that the originally increasing trend of

the time series changes to a decreasing pattern after the response execution time. This implies the increased quality of solutions due to the more convergence surpass the degradation in solutions performance due to the longer response delay.

Fig. 5.4 also illustrates the minimum-TIM time series when the changes in consumers' water use behavior is also taken into account. The general pattern of this time series is noticeably similar to the case in which only the scenario changes and managers' response is considered. However, the ultimate TIM is consistently lower, which can be attributed to the fact that health impacts is smaller when the consumers stop drinking contaminated water after they ingest a mass of contaminant above the toxic dose.

5.5 Conclusions

Dynamic simulation of emergency conditions provides a more realistic picture of the complex process than static models through considering uncertainties and changes that alter system normal operation after contamination starts. Dynamic optimization provides timely and realistic emergency response recommendations through adapting the response to the varying behavior of the system and finding the best balance between exploitation of old search domain information and exploration of emerging search space.

A multi-objective-based dynamic optimization algorithm was used in this study that defines and maximizes an artificial objective function for preserving diversity among solutions. Among three different diversity preservation measures of distance from the nearest neighbor, distance from the best solution of the population, and average

distance from all solutions, distance from the nearest neighbor turned out to be the best metric.

The model employed in this study for the simulation of WDS-consumers interactions and dynamics was very simple. Use of a more advanced sociotechnical model that accounts for the communication of human agents and mobility would better capture real system behavior and is recommended for future research. However, use of more complicated models should not significantly increase the computational burden of every simulation since the dynamic optimization model needs to be run in a real-time manner during the emergency.

6. MACHINE LEARNING FOR REAL-TIME CONTAMINANT SOURCE IDENTIFICATION AND EMERGENCY RESPONSE

6.1 Introduction

Water contamination source identification involves the characterization of the contamination event attributes using threat observations such as sensor network measurements. Effectiveness of emergency response decisions for the mitigation of contamination impacts significantly depends on managers' knowledge of these source characteristics. This knowledge may be enhanced through analyzing measured contaminant concentration time series data using an inverse modeling approach.

Once the contamination source has been identified, a response generation model must be employed to characterize response strategies for the mitigation of impacts. Ideally, the response generation model should be able to generate optimal strategies in real-time. The optimization approach proposed in previous chapters performs well in identifying optimal or near-optimal strategies. This, however, comes at a price, which is computation run time. The optimization model starts from a random set of solutions and increases the quality of solutions through evolution over time. Effectiveness and timeliness of such models may be significantly improved when they are supplemented with a group of good solutions to start the evolution process instead of using completely random solutions. Data mining may be employed to extract knowledge on good solutions during the emergency from the information database developed during the emergency preparedness phase (i.e., well before an actual emergency begins). Such

identified good solutions may be either used to boost the evolution process or be executed independently without performing optimization during an emergency.

The rest of this chapter describes the application of data mining for real-time characterization of contaminant sources and emergency response strategies in two separate sections. The first section is dedicated to real-time contaminant source characterization. The second section describes the schemes developed for real-time emergency response. Each section covers the description of past works and methodologies proposed in this dissertation, as well as demonstration and discussion of proposed models on the Mesopolis virtual city WDS.

6.2 Classification Approach for Source Identification

6.2.1   Literature Review and Statement of the Work

Probabilistic approaches have been explored by several researchers to characterize contaminant sources in WDSs. These approaches are mostly based upon Bayes' theorem to estimate likelihood of possible contamination sources. Dawsey et al. (2006) employed a Bayesian belief networks methodology to integrate sensor data with other validating evidence of contamination events to better characterize sources and reduce false positives. De Sanctis et al. (2008) studied the impact of imperfect sensor measurements on contamination source characterization using a backtracking algorithm. Propato et al. (2010) proposed an entropic-based Bayesian inversion technique, the minimum relative entropy method, to estimate contaminant source probabilities. Wang

and Harrison (*in press*) implemented a Markov Chain Monte Carlo algorithm based on Bayesian analysis for probabilistic source characterization.

Optimization approaches have also been broadly employed to deal with the problem of contaminant source characterization. The source characteristics such as intrusion location and duration are treated as decision variables and the objective is to minimize the difference between observed and simulated concentrations. Laird et al. (2005) employed nonlinear programming to estimate the time and location of contamination source. Guan et al. (2006) demonstrated a simulation-optimization model by coupling a WDS simulation model with a gradient-based local search. Evolutionary computation-based optimization algorithms such as evolution strategies have been investigated to solve the source identification problem and address non-uniqueness of contaminant sources (Zechman and Ranjithan 2009; Liu et al. 2011; Drake and Zechman 2011). Overall, optimization techniques have been demonstrated to accurately determine contaminant sources and can be also modified for adaptive monitoring (Liu et al. 2011). However, these methods are inherently computationally intensive, which is a critical issue considering the very important role of prompt response for mitigation of public health consequences.

The approach presented in this study has the following steps. First, considering the uncertainties in different system parameters, a reasonably large set of realizations is simulated for a bounded set of contamination scenario possibilities, and sensor readings time series are recorded for each realization. Since the contamination scenario is known for each realization-sensor reading dataset, this constitutes a large database of

"classified" and "labeled" sensor readings time series (these terms are further defined in section 6.2.2). In application, this phase of generating the time series dataset is performed during the emergency preparedness phase. Once an actual contamination incident occurs, a time series of contaminant concentration is recorded that is used to estimate the unknown contamination scenario; since the real attributes of the contamination scenario are not definitively known, the sensed time series is referred to as "unlabeled" and "unclassified." The task of classification of unlabeled time series – which can be also interpreted as an inverse problem – will be accomplished here with a data mining technique and pattern matching scheme. The pattern matching framework uses similarity search to compare the unlabeled times series with the labeled ones that exist in the dataset. Two similarity measures of Euclidean distance and correlation metric are used here. The k-nearest neighbors (kNN) classification algorithm (Cover and Hart 1967) is also used for data mining in this study.

6.2.2 Probabilistic Analysis

Uncertainties are unavoidable in design and operation of engineering systems. The randomness in sensor data measurements stems from various uncertainties. These uncertainties are beyond the control of WDS designers and operators. They essentially arise from our inability to predict the accurate consequence of a process due its random nature, lack of complete information, or both. Consideration of these uncertainties is a crucial task in the characterization of contaminant sources.

Ang and Tang (2007) distinguished two broad types of uncertainties: (1) uncertainty associated with natural randomness of the underlying phenomenon (e.g., natural variability of water demands); and (2) uncertainty associated with imprecision in our prediction of reality (e.g., uncertainty in estimation of pipes roughness coefficient in design phase). The former is called *aleatory* uncertainty, while the latter is known as *epistemic* uncertainty.

Khanal et al. (2006) categorized the sources of uncertainty in a WDS contamination event in a different way. They categorized them into static and dynamic parameters. Static parameters are characteristics of the WDS that are not influenced by human behavior, such as pipe diameter. Dynamic variables, on the other hand, are properties of the system that are affected by the behavior of consumers and utility operators, such as demand patterns. Khanal et al. (2006) considered the uncertainties in demand pattern, tank storage, contamination duration, and contaminant mass in probabilistic impact assessment of contamination events. Pasha and Lansey (2010) also included the uncertainties in decay coefficients, pipe diameter and roughness, and nodal demands in such assessments.

For a deterministic model, every contamination event is associated with only one time series of measurements by a given sensor. However, in reality there exist multiple possible time series at this sensor for a specific contamination event due to the probabilistic behavior of the system. Uncertainties in static and dynamic WDS variables may be propagated through Monte Carlo analysis to determine the uncertainties in sensor network measurements. Fig. 6.1 illustrates four different sensor reading time

series associated with two potential contamination scenarios for WDS of Mesopolis. Each of the time series correspond to one realization of the event under uncertainties in demands, pipe diameter and roughness, and tank water level. As observed, while the time series for different realizations for each specific event do not completely match, a common pattern may still be distinguished. The set of realizations for one scenario is called a *class* and the *label* of this class is the corresponding scenario.
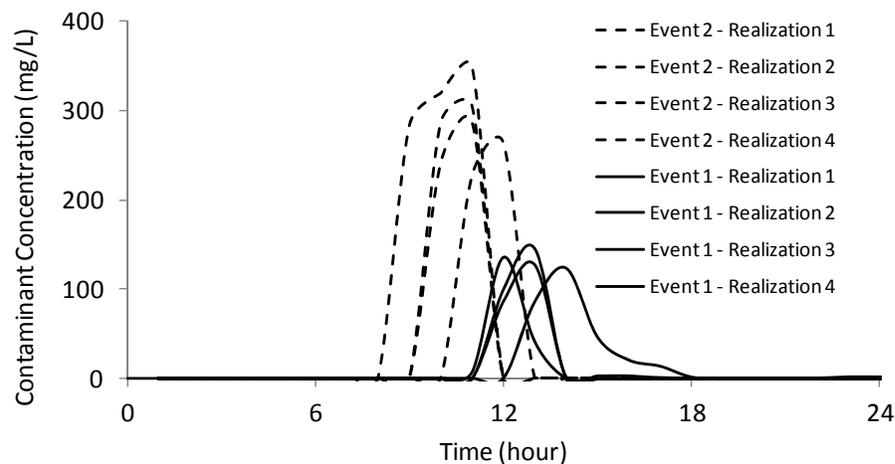
Fig. 6.1. Sensor reading time series for multiple realizations for two contamination scenarios under different parameter uncertainties

6.2.3 Classification of Time Series

6.2.3.1 Similarity Measures

One of the simplest similarity measures for comparing time series is the Euclidean distance metric. Assume that two time sequences **r** and **s** are of the same length n. We can find the Euclidean dissimilarity measure using the familiar formula:

$$d(r,s) = \sqrt{(r-s)(r-s)^T} \qquad (6.1)$$

where superscript T denotes the transpose of a vector. Correlation distance may be also used for determining the level of similarity between two time series. This distance is defined as one minus the sample correlation between points (treated as sequence of values), or mathematically:

$$d(r,s) = 1 - \frac{(r-\bar{r})(s-\bar{s})^T}{\sqrt{(r-\bar{r})(r-\bar{r})^T}\sqrt{(s-\bar{s})(s-\bar{s})^T}} \qquad (6.2)$$

where $\bar{r}$ and $\bar{s}$ are the mean values of corresponding time series.

6.2.3.2 Classification Algorithm

Given an unlabeled time series and a pool of labeled time series, different classification algorithms (classifiers) may be used to label the unlabeled time series. The k-nearest neighbor (kNN) classifier (Cover and Hart 1967) is used here for this purpose. kNN classifies unlabeled instances based on a "voting" of the labels of k closest training samples in the feature space. In the context of this particular study, given a fixed value of k, the k nearest labeled time series to the unlabeled new time series are first identified using any of the similarity measures mentioned above. The label (scenario) that is most frequent in this set of neighbors is selected as the scenario that has generated the

unlabeled time series. Since the size of neighborhood directly influences the final classification decision, analysis should be performed using different values of k for a training and test set to determine the optimal k value.

kNN is a lazy learning algorithm (also known as the memory-based algorithms) since it defers dataset processing until a classification request arises. Because kNN uses local information, it can achieve highly adaptive performance. On the other hand, kNN involves a large storage requirement, and the value of k also needs to be determined properly. Alternative classification algorithms such as quadratic classifier, Artificial Neural Networks (ANN), and Support Vectors Machines (SVM) may be investigated for dealing with these difficulties (Bishop 2006).

6.2.4   Application

A sensor network needs to be first designed to collect the time series for different contamination scenarios. There exist many sensor placement strategies that can be used for this purpose. A comprehensive list of such methods may be found in a review conducted by Hart and Murray (2010). Since this design task is not the focus of the source characterization scheme presented in this chapter, a set of 11 intuitive places are selected to place the sensors (Fig. 6.2.). It is believed, however, that using a methodically designed sensor network would enhance the information content and quality of recorded time series, which accordingly improves performance of the source identification model. Future work will employ a well-demonstrated algorithm to design the sensor network for Mesopolis for this purpose.

In this application example, the region of nodes in which the contaminant is injected is highlighted by a rectangular zone in Fig. 6.2. Without the loss of generality, only the contaminant injection location is considered as the attribute that varies over different scenarios. Since the region includes 341 nodes, 341 different scenarios exist that thus define 341 classes. The injected contaminant mass is 100 kg and the injection start-time and duration are 20:00 and 3 hours, respectively. Global demand multiplier is set to the average value of 1.00. The simulation is performed for 68 hours, and the sensors' readings are recorded at every one-hour time step after the injection starts. Therefore, the length of time series for each single sensor is 48. As the sensor network includes 11 sensors, the total number of sensor reading values is 528 for every scenario (class) of contaminant injection location.

The uncertainty in system parameters is succinctly expressed by its coefficient of variation (COV), which is defined as a parameter's standard deviation divided by its
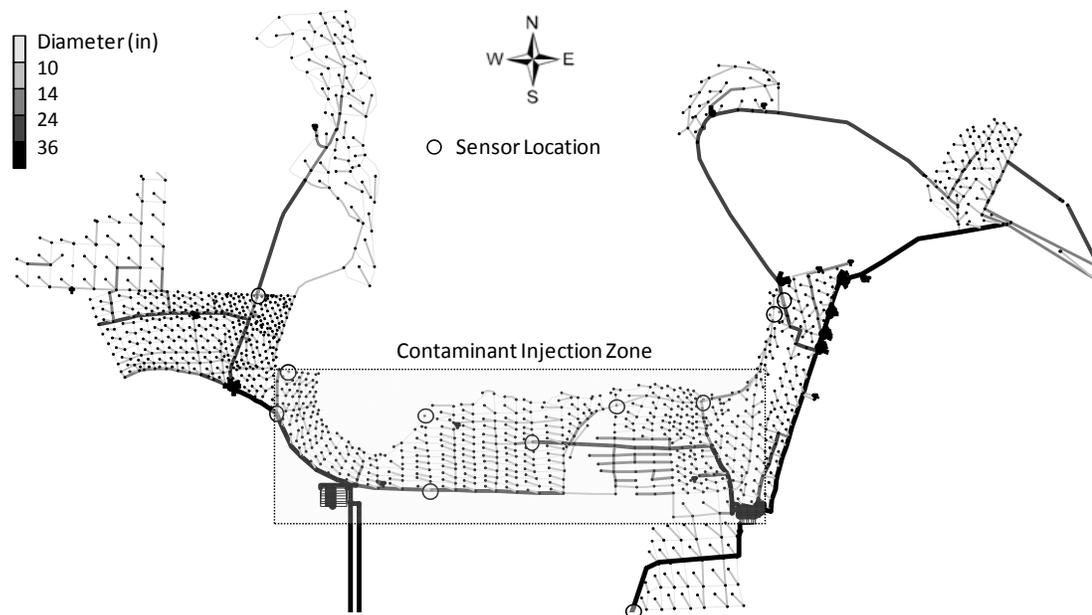
Fig. 6.2. Sensor network and the zone of contaminant injection

mean. In this study, COV is set to a relatively high value of 0.1 for nodal demands, pipe diameter, pipe roughness, and tank water level. A COV of 0.05 is assumed for aggregate system demand. A normal distribution is assumed for all uncertain parameters. Twenty realizations are performed for each injection location. The total size of the contamination event dataset is 6,820, equal to the number of contamination scenarios multiplied by the number of realizations.

Performance of classification algorithms is commonly evaluated based upon the rate of correct classification of samples in a test set using the training set. Classification rate is defined as the number of correctly classified test samples over the total number of samples in the test set. In the context of this example application, correct classification means the injection location is identified correctly by the kNN algorithm. In this study, 80% of the data is randomly selected for training, and the rest is used for testing. A relaxed metric is also defined that assumes classification is correct if the model suggests either the true injection location or its immediate neighbors for a test sample.

Fig. 6.3 shows the classification rates using Euclidian and correlation similarity measures for varying neighborhood size k. Since the randomness in dividing the dataset into training and test sets influences the classification rate, the kNN algorithm is run 5 times for each value of k, and the minimum, mean, maximum rates are reported. The results show that the classification is performed better when the Euclidean similarity measure is used. It is also observed that highest classification rate is achieved for k = 1. This result may be due to the fact that the number of realizations for each scenario

(class) is significantly lower than the total number of classes. Further analyses using more realizations for each scenario are needed to check this hypothesis.
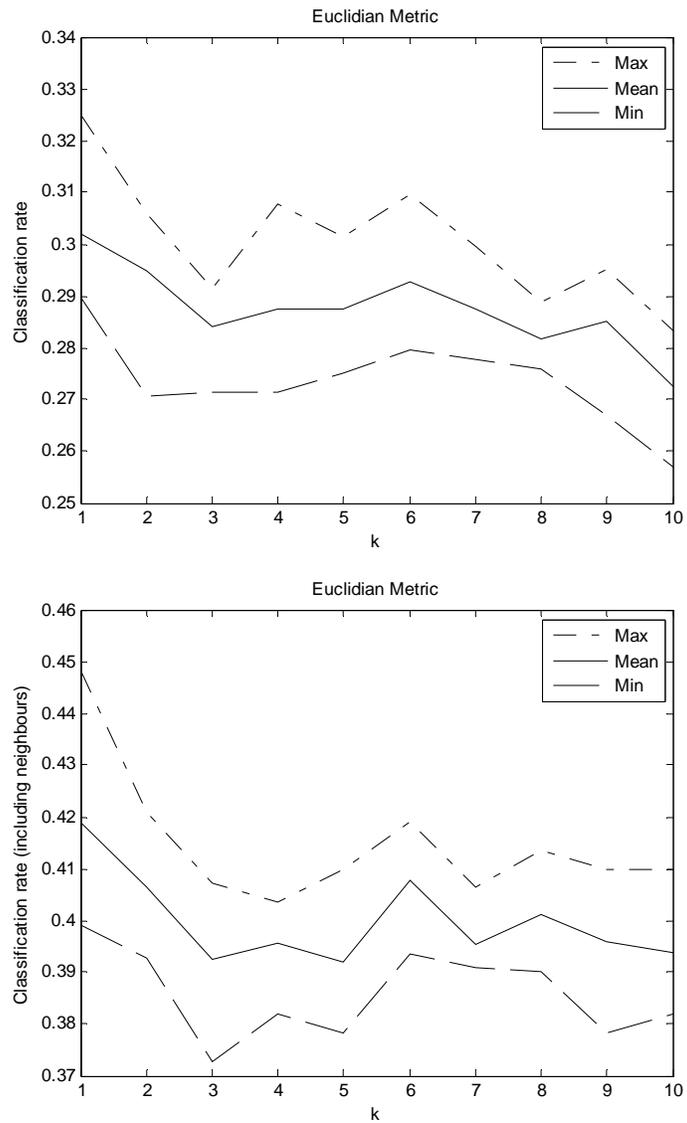
Fig. 6.3. Classification performance for different similarity measures and varying

k values
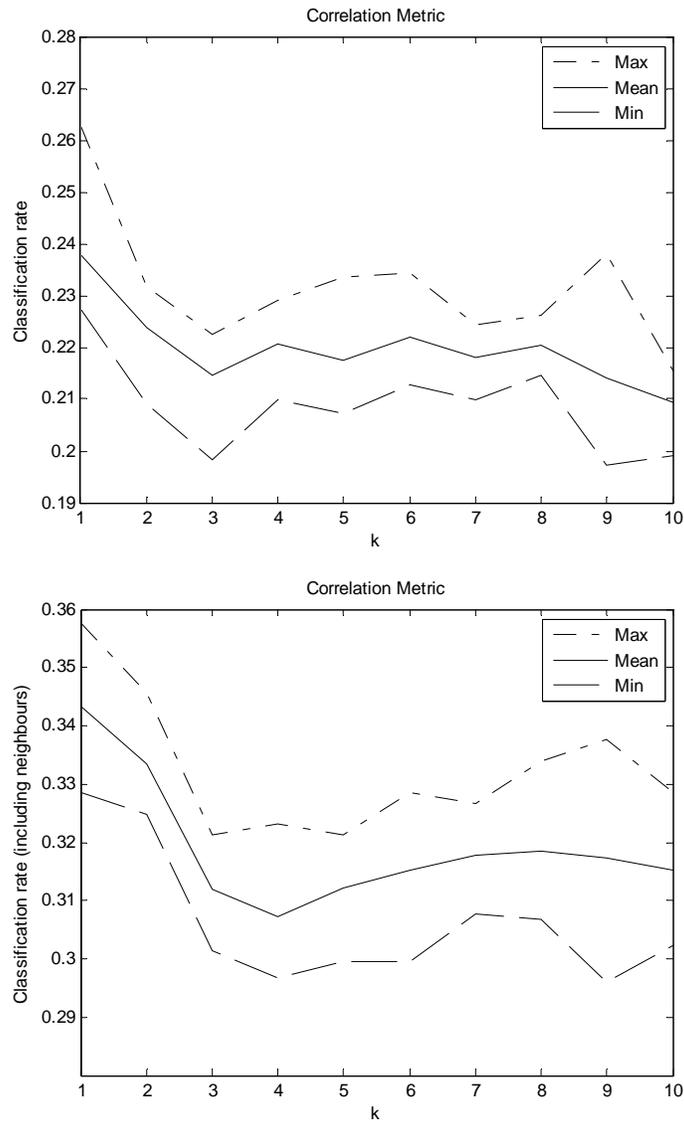
Fig. 6.3. Continued

6.3 Clustering Approach for Emergency Response

Once the real-time contamination source characterization model elaborated above has identified the contamination scenario using the sensor network data, a real-time response identification model should be used to find an effective response plan. Identified

contamination source is used as input information to the real-time response characterization model.

Considering the very large number of possible contamination scenarios and the computational cost of response optimization, preparation of optimal response plans for all possible scenarios during the emergency management preparedness phase is very difficult, if not impossible. A potential approach to limit the number of optimization runs is to discover groups of similar contamination scenarios and find optimal response plans only for representatives of these groups. This process of discovering a number of groups within a dataset examples is called *clustering*. In contrast to classification, clustering is an unsupervised learning process meaning it does not use or require labels of data samples for learning. In the context of contamination scenario clustering, for instance, scenarios do not have any labels attached to them to supervise the learning process. The task of clustering scenarios has three initial requirements: 1) defining a characteristic for comparing different scenarios, 2) defining a similarity measure to determine how similar scenarios are using the defined characteristic, 3) structuring a clustering algorithm.

6.3.1   Scenario Characteristic and Similarity Measures

Comparing two different scenarios may be most simply based upon the sum value of differences between each pair of their corresponding attributes. However, it is not feasible to effectively integrate the differences between each pair of attributes to construct a single comparison metric because the attributes are inherently of different natures (e.g., contamination location vs. injection duration).

An alternative approach is to define a comparison characteristic. In the context of emergency response planning for the mitigation health impacts, the spatial distribution of health impacts associated with every scenario can be used to define this characteristic. The underlying assumption here is that a specific response plan that performs well for Scenario X would also perform reasonably well for Scenario Y if the distributions of health impacts for both scenarios are convincingly similar. Mathematically, this characteristic is defined here as a vector of ultimate total injected mass of contaminant (TIM) for every node in the network with non-zero population. Extended hydraulic and exposure simulation needs to be performed for every scenario to construct its impact vector. Fig. 6.4 shows impact vectors for two potential contamination scenarios in Mesopolis.
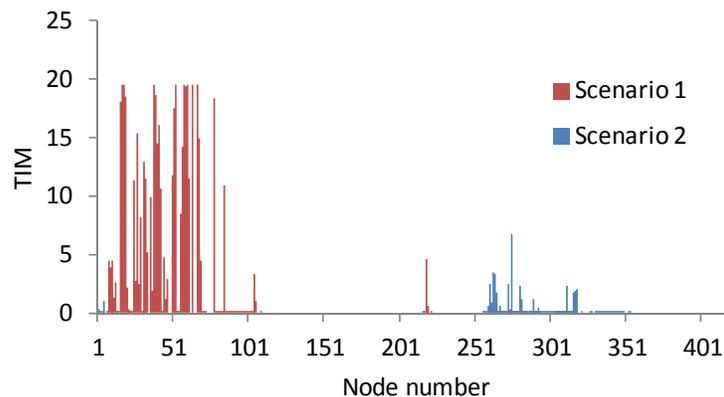
Fig. 6.4. Health impact vector associated with two potential contamination scenarios in Mesopolis

Having defined the scenario characteristic, a similarity measure needs to be defined to determine how much two contamination scenarios are similar based upon this

characteristic. Euclidean and correlation measures are used for this purpose. These measures are calculated using Eqs. (6.1) and (6.2).

6.3.2 Clustering Algorithm

The K-means algorithm (Lioyd 1982) is used in this study for clustering contamination scenarios. Intuitively, we may think of a cluster as comprising a set of data samples whose inter-point distances are small compared with the distances to samples outside of the cluster. In the K-means algorithm, this intuitive notion is formalized through introducing a set of vectors μ_k, where $k = 1, \ldots, K$, in which μ_k is a prototype (representative) corresponding to the kth cluster ω_k. Length of μ_k is equal to the length of the impact vector, which is the number of non-zero population nodes.

K-means is a clustering procedure that attempts to minimize a criterion function J, which is usually called the distortion function, that is defined as

$$J = \sum_{i=1}^{K} \sum_{x \in w_i} d(x - \mu_i) \tag{6.3}$$

where $d(x - \mu_k)$ is the distance (dissimilarity) between data sample x and the mean vector and is given by any of Eqs. (6.1) and (6.2). Distortion function basically sums up the distances from each data sample x to the mean vector μ_k of the cluster that it is assigned to.

The goal in K-means algorithm is assigning data samples to K clusters so that the distortion function is minimized. This is commonly achieved through an iterative process that reassigns each sample to its nearest cluster (represented by its mean vector μ_k) at every iteration. Some degree of controlled randomness may be introduced during the assignment process to reduce the possibility of premature convergence to local minima.

6.3.3 Application

A dataset of contamination scenarios first needs to be prepared. While the presented approach can be generally applied with consideration of all scenario attributes, the illustrative study here is limited to the single attribute of contaminant injection node setting other attributes to fixed values. Contaminant mass and global demand multiplier are set to 100 kg and 1.00, respectively. The simulation duration is 1 day and the contamination starts at 06:00 with a duration of 1 hour. The length of impact vectors is 428 and total number of data samples is 881. After removing injection locations that do not contaminate any non-zero-population node, this dataset size is reduced to 636. To enhance the clustering performance, impact vectors are normalized so that the ultimate TIM for every node in the vector is between 0 and 1.

Since K-means algorithm is an iterative process and starts from randomly assigned samples, multiple model runs are required to assure a more robust performance. The number of iterations and model runs is set here to 1000 and 10, respectively. A controlled random assignment is introduced with a variable probability of 0.1 linearly

decreasing to 0 at the end of each model run. This is principally performed to escape from local minima in the search space. Fig. 6.5 shows the clustering results for K values of 14 and 21 using Euclidean and correlation similarity measures.

Euclidean - K = 14
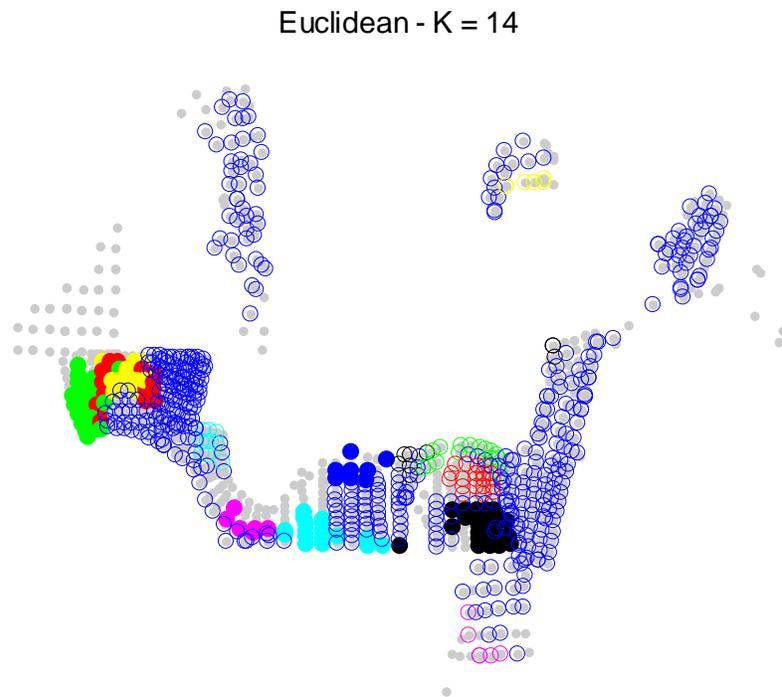
Fig. 6.5. Clustering of scenarios using different K values and similarity measures

Euclidean - K = 21

Correlation - K = 14

Fig. 6.5. Continued
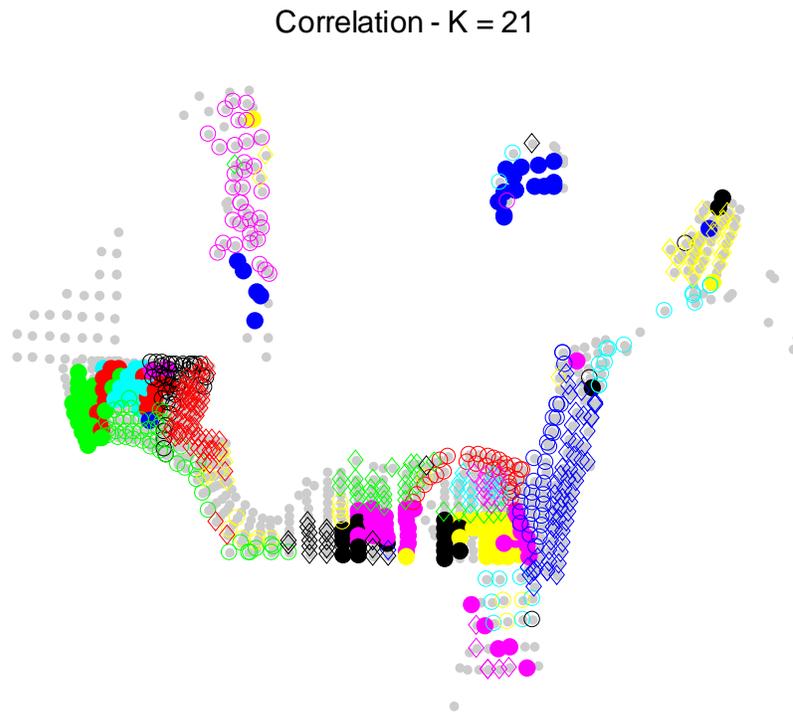
Correlation - K = 21

Fig. 6.5. Continued

Using a previous knowledge on network hydraulics, the results show that the correlation measure is a significantly more informative measure for comparing similarity and clustering the scenarios. Clusters obtained using this measure well capture hydraulics of the system, including the discontinuities, pressure zones, and flow patterns. For instance, for K = 21, this is manifested by generating clusters in isolated regions such as eastern and western peninsulas, and pressure zones such as the clusters illustrated with empty red and blue diamonds.

Fig. 6.5 generally shows that the use of Euclidean measure results in some very large clusters in contrast to the correlation metric. This information is quantitatively illustrated in Fig. 6.6 using pie charts. While it is not necessarily better that all clusters

have the same size, assigning a high percentage of scenarios to a few clusters is not interpreted as an efficient clustering practice for the WDS of Mesopolis since this network is highly complex and includes several pressure zones and isolated regions. To deal with the issue of having too small or too large clusters, more advanced clustering method of Iterative Self-Organizing Data Analysis (ISODATA) (Ball and Hall, 1965) may be also used that automatically merges very small clusters and splits very large clusters.

A next phase of analysis in now performed for methodological demonstration of the effectiveness of scenario clustering for real-time response. The objective of this phase is to compare the accurate but slow optimization method proposed in previous chapters with approximate but real-time machine learning approach presented here. First, optimization is performed for all scenarios in a cluster k and corresponding minimum health impact values are determined. Second, the representative scenario for

Fig. 6.6. Size of clusters obtained using different similarity measures

that cluster is determined, which is the scenario that is most similar to the mean vector μ_k. Third, simulations are performed to calculate health impact values for all scenarios in the cluster when the optimal response plan for the representative scenario is executed. The machine learning approach is most efficient when these impact values are equal to the corresponding minimum values obtained in the first step.

The analysis is performed for three clusters obtained when correlation similarity measure is used and K = 21. These three clusters are called A, B, and C, and are illustrated in Fig. 6.7, which also shows the identified representatives of every cluster with a filled square. The number of scenarios in these clusters is 47, 54, and 30, respectively, and Cluster B is the largest cluster among all 21 clusters. Representatives of clusters A, and B are found to be West and East WTP, respectively. This analytical finding is in agreement with subjective judgment as these two clusters are right downstream of the two WTPs. For the representative for Cluster C, however, no specific comment may be made about where the representative scenario can be since all scenario injection locations are normal intermediate nodes located in central zone of the WDS, where the hydraulics and dynamics of the system is complex and very difficult interpret.
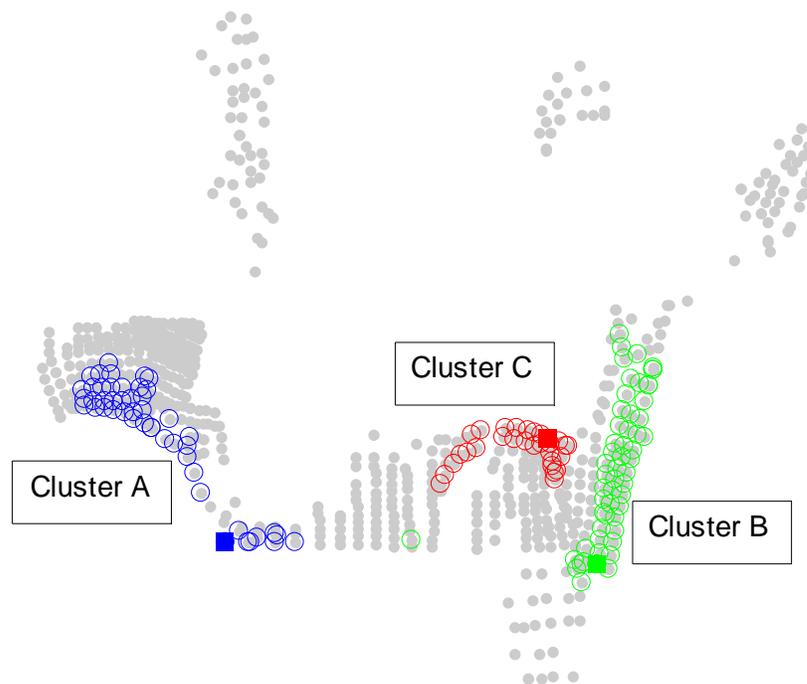
Fig. 6.7. Clusters A, B, and C (circles), and corresponding representative scenarios

(squares)

Two response strategies of hydrant opening and food-grade dye injection are considered. The number of hydrants and dye injectors are 5 and 3, respectively, and the response delay (time period after contaminant insertion is completed) is set to 2 hours. The characteristics of the optimization algorithm are described in Chapters 3 and 4. Genetic algorithm optimization model is run 1 time for every scenario in every cluster. Multiple runs per scenario, however, would provide a better estimation of the global optimum.

Fig. 6.8 shows the values of ultimate total ingested mass for scenarios in Cluster A for three situations: 1) no response is executed, 2) optimal response plan for every

scenario is implemented, and 3) representative response plan is executed for every scenario in the cluster. Scenarios are sorted according to their no-response TIM to facilitate visual comparison. Horizontal axis, thus, conveys no particular information. Results indicate that optimal dye injection outperforms optimal hydrant operation for the mitigation of impacts although the number of dye injection locations is less than contaminant flushing locations. For both response strategies, it is observed that using representative plan can well mitigate the impacts for several scenarios in the cluster. For certain scenarios, it is observed that reduction in TIM is higher when representative plan is used than the determined optimal plan. This shows more than 1 optimization run is required for these scenarios to better estimate the global optimum.

Since dye injection was found to be more effective than hydrant operation, only this strategy is used for Clusters B and C. Results for these two clusters are indicated in Fig. 6.9. These results further demonstrate the effectiveness of the proposed machine learning approach for real-time response to contamination events.
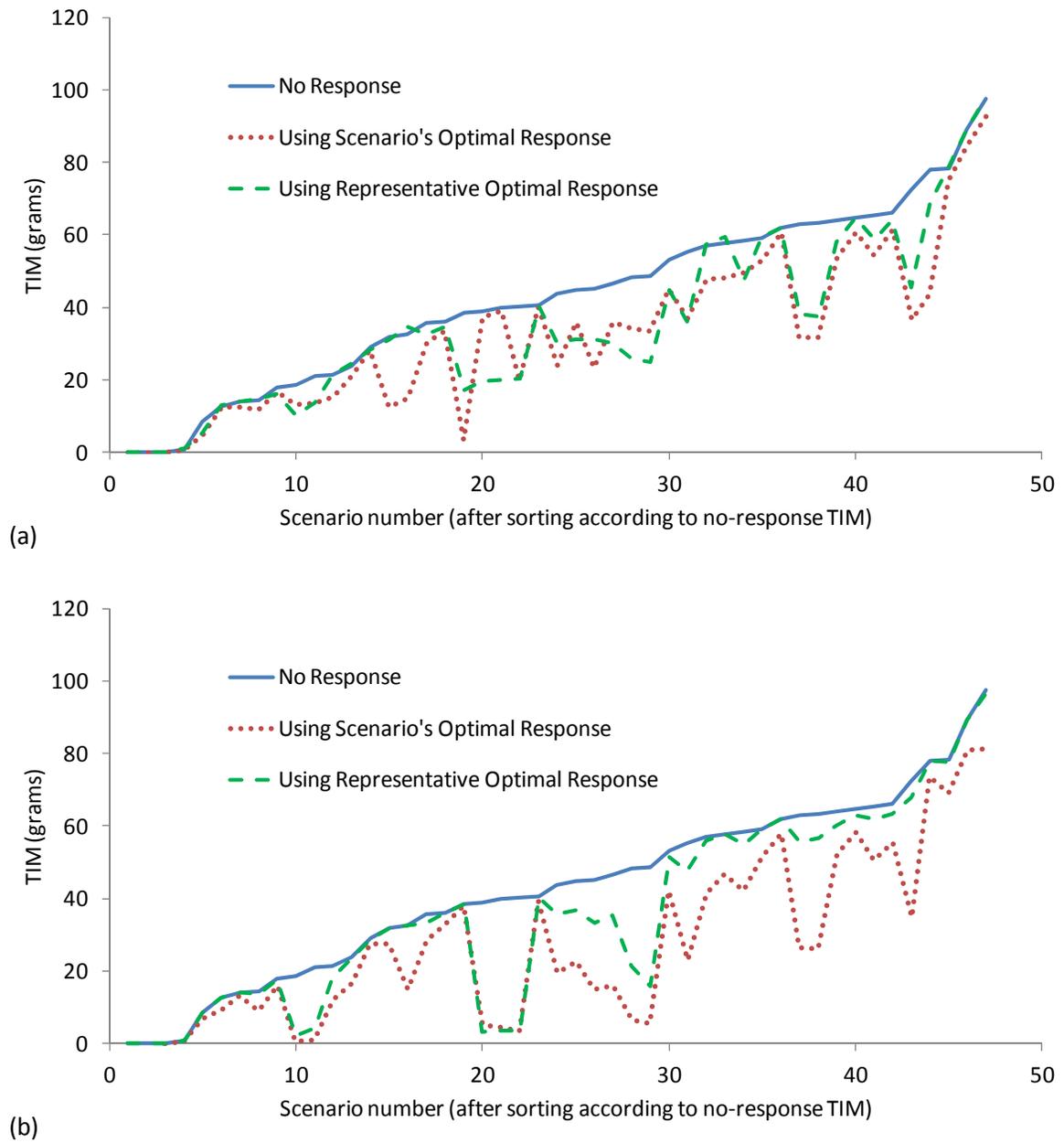
Fig. 6.8. Health impacts for scenarios in Cluster A for different response situations using (a) hydrant opening and (b) food-grade dye injection

(a)

(b)
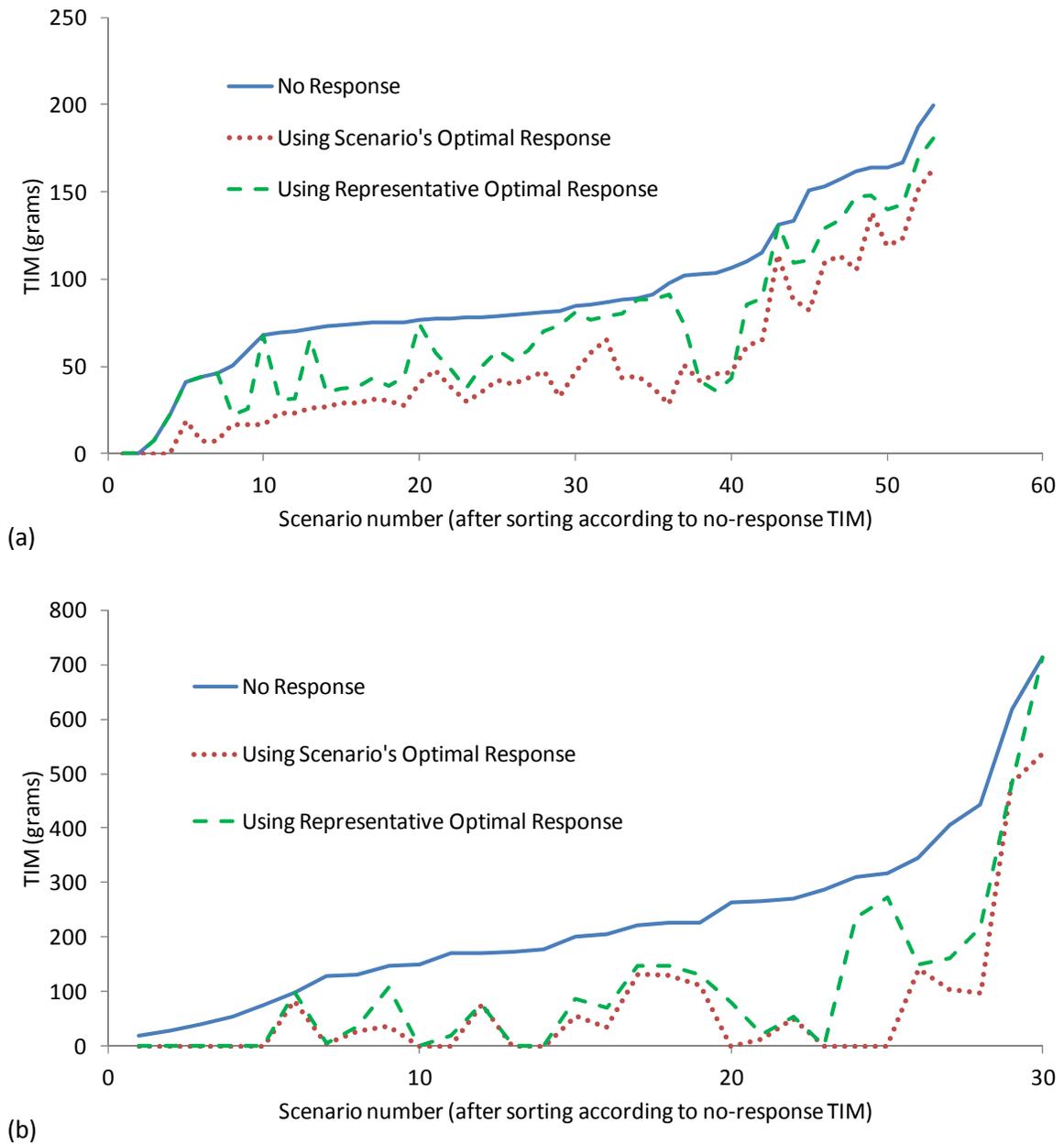
Fig. 6.9. Health impacts for scenarios in (a) Cluster B and (b) Cluster C for different

response situations using food-grade dye injection

6.4 Conclusions

kNN classification algorithm was applied for the characterization contaminant source in WDS under various uncertainties. Application of this approach during an emergency is simple, real-time, and considers the random behavior of the system manifested in sensor network readings. However, it requires a very large number of simulations to be performed during the preparedness phase. Between the two similarity measures of Euclidean and correlation distances, Euclidean measure was shown to have a better performance.

K-means clustering algorithm was used to cluster contamination scenarios for real-time and reasonably effective contamination emergency response. Clustering was based upon similarity of scenarios regarding the corresponding distribution of health impacts. In contrast to the classification study, correlation similarity measure is more effective. More elaborate optimization runs may be performed for all clusters determined using different similarity measures and K values to more accurately evaluate performance of different combinations and tune model settings.

Proposed machine learning schemes have also the potential to be used in conjunction with optimization approaches, which have higher accuracy but are significantly more computationally intensive, during the emergency. Machine learning models can provide the optimization models with a set of good solutions (contaminant source characteristics or response plans) to start the iterations or evolution process and, thus, expediting the convergence to global or near-optimal optimal solutions during an emergency.

7. CONCLUSIONS

This dissertation proposed and developed a comprehensive set of risk assessment, systems analysis, and machine learning methodologies and models to help the emergency managers with risk assessment, emergency preparation, and emergency response for WDS contamination events. This set of decision support schemes provides the mangers with valuable information on vulnerable aspects of the system and effective and timely strategies for achievement of different emergency management objectives. Proposed tools were discussed and demonstrated on a highly complex virtual WDS to assure their usefulness for real-world applications.

The concept of maximum-risk frontier was proposed and demonstrated to be effective in dealing with principal risk measures of event likelihood and consequences for characterization of critical scenarios. A more comprehensive meta-analysis of historical events supplemented with epidemiological studies would enhance the accuracy of probability estimations. Consequences would be also more realistically estimated when sociotechnical modeling that simulates the human-infrastructure dynamics is performed. Research is currently underway to accomplish this more advanced modeling.

Multiobjective optimization was used for multicriteria emergency preparedness and response considering multiple strategies and objectives. Proposed schemes would be of significant help to emergency mangers to reach optimal trade-off between their conflicting objectives during the intense course of an emergency. Dynamic optimization models consistently adapt to the changing environment to provide the managers with up

to date support and information. A comprehensive decision support system may be developed in future that includes all response strategies and objectives in an integrated graphical user interface to facilitate communication between managers and the models.

Machine learning approaches were proposed to provide the managers with timely information on contaminant sources and effective emergency response decisions. Simplicity and real-time performance of these methods, in particular, are believed to be their most valuable characteristics, while their accuracy are not claimed to be as high as the optimization approach. This kind of modeling is in its infancy and extensive future research is required to enhance their performance. This includes application of more advanced and efficient classification and clustering algorithms. Source identification and response recommendation schemes may also be integrated to facilitate their use and enhance their applicability.

All in all, developed static optimization, dynamic optimization, and machine learning models are most useful when are used in an integrated manner. Static optimization is used before the contamination happens to develop a reasonably generalizable database of effective response plans. Real-time source characterization model is used after the emergency occurs to determine the contamination scenario. This scenario is given as an input to the real-time response model to estimate a reasonably effective response plan. This response plan is executed by the managers, either directly, or after some degree of improvisation based upon their subjective and qualitative judgments. Dynamic optimization scheme is informed by these actions and any future decisions to adapt to the new environment and provide up to date response support as the

emergency proceeds and emerging conditions deviate more form those predicted during the preparedness phase. All these different stages of modeling are believed to be fruitful areas for future research.

REFERENCES

Alfonso, L., Jonoski, A., and Solomatine, D. (2010). "Multiobjective optimization of operational responses for contaminant flushing in water distribution networks." *Journal of Water Resources Planning and Management*, 136(1), 48–58.

American Society of Mechanical Engineers Innovative Technologies Institute (2006). *Risk analysis and management for critical asset protection (RAMCAP): the framework*, ASME Innovative Technologies Institute, Washington, DC.

Ang, A., and Tang, W. H. (2007). *Probability concepts in engineering: emphasis on applications to civil and environmental engineering*, John Wiley & Sons, Hoboken, NJ.

Asbeck, E., and Haimes, Y. Y. (1984). "The partitioned multiobjective risk method." *Large Scale Systems*, 6(1), 13–38.

Ball, G. H., and Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification*, Stanford Research Institute, Menlo Park, CA.

Baranowski T. M., and LeBoeuf E. J. (2006). "Consequence management detection optimization for contaminant and isolation." *Journal of Water Resources Planning and Management*, 132(4), 274–282.

Baranowski T. M., and LeBoeuf E. J. (2008). "Consequence management utilizing optimization." *Journal of Water Resources Planning and Management*, 134(4), 386–394.

Besbes, O., and Zeevi, A. (2009). "Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms." *Operations Research*, 57(6), 1407–1420.

Bhave, P. R. (1981). "Node flow analysis for water distribution systems." *Journal of Transportation Engineering*, 107(4), 457–467.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Spring Science + Business Media, New York, NY.

Blackburn, B. G., Craun, G. F., Yoder, J. S., Hill, V., Calderon, R. L., Chen, N., Lee, S. H., Levy, D. A., and Beach, M. J. (2004). "Surveillance for waterborne-disease outbreaks associated with drinking water—United States, 2001–2002." *MMWR Surveillance Summary*, 53, 23–45.

Branke, I., Kaußler, T., Schmidt, C., and Schmeck, H. (2000). "A multipopulation approach to dynamic optimization problems." *Proceedings of 4th International Conference on Adaptive Computing in Design and Manufacture*, Springer, Berlin.

Bristow, E. C., and Brumbelow, K. (2006). "Delay between sensing and response in water contamination events." *Journal of Infrastructure Systems*, 12(2), 87–95.

Bui, L.T., Michalewicz, Z. Parkinson, E., and Abello, E.M. "Adaptation in dynamic environments: A case study in mission planning." *IEEE Transactions on Evolutionary Computation*, in press, DOI:10.1109/TEVC.2010.2104156

Bui, L. T., Nguyen, M., Branke J., and Abbass, H. A. (2008). *Tackling dynamic problems with multiobjective evolutionary algorithms*. In Knowles, J., Corne, D., and

Deb, K. (Eds.), *Multiobjective Problem Solving from Nature* (pp. 77–92), Springer, Berlin.

Chandapillai, J. (1991). "Realistic simulation of water distribution system." *Journal of Transportation Engineering*, 117(2), 258–263.

Cities of Bryan and College Station (2005). *B/CS unified design guideline manual, water, wewer, and streets*. Available: www.bcsunited.net/2005Files/DesignManual2005.pdf

Cobb, H. G. (1990). *An investigation into the use of hypermutation as an adaptive operator in genetic algorithms having continuous, time-dependent nonstationary environments*. Naval Research Librarory, Washington, DC.

Cover, T., and Hart, P. (1967). "Nearest neighbor pattern classification." *IEEE Transactions on Information Theory*, 13(1), 21–27.

Craun, M. F., Craun, G. F., Calderon, R. L., and Beach, M. J. (2006). "Waterborne outbreaks reported in the United States." *Journal of Water and Health*, 4(2), 19–30.

Davis, M. J., and 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(5), 449–456.

Davis, M. J., and Janke, R. (2011). "Patterns in potential impacts associated with contamination events in water distribution systems." *Journal of Water Resources Planning and Management*, 137(1), 1–9.

Dawsey, W. J., Minsker, B. S., and VanBlaricum, V. L. (2006). "Bayesian belief networks to integrate monitoring evidence of water distribution system

contamination." *Journal of Water Resources Planning and Management*, 132(4), 234–241.

De Sanctis, A. E., Boccelli, D. L., Shang, F., and Uber, J. G. (2008). "Probabilistic approach to characterize contamination sources with imperfect sensors." *Proceedings of ASCE/EWRI World Environmental and Water Resources Congress*, Ahupua'a, HI.

Deb, K. (2001). *Multiobjective optimization using evolutionary algorithms*, Wiley, Chichester.

Deb, K., and Agrawal R. B. (1995). "Simulated binary crossover for continuous search space." *Complex Systems*, 9(2), 115–148.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). "A fast and elitist multi-objective genetic algorithm NSGA-II." *IEEE Transactions on Evolutionary Computations*, 6(2), 182–197.

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

Ghosh, A., and Dehuri, S. (2004). "Evolutionary algorithms for multi-criterion optimization: a survey." *International Journal of Computing and Information Sciences*, 2(1), 38–57.

Goldberg, D. E., and Smith, R. E. (1987). "Nonstationary function optimization using genetic algorithms with dominance and diploidy." *Proceedings of International Conference on Genetic Algorithms*, Cambridge, MA.

Grefenstette, J. J. (1992). "Genetic algorithms for changing environments." *Proceedings of 2nd International Conference on Parallel Problem Solving from Nature*, Brussels, Belgium.

Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. M. (2006) "Identification of contaminant sources in water-distribution systems using simulation-optimization method: A case study." *Journal of Water Resources Planning and Management*, 132(4), 252–262.

Guliashki, V., Toshev, H., and Korsemov, C. (2009) "Survey of Evolutionary Algorithms used in multiobjective optimization." *Problems of Engineering Cybernetics and Robotics*, Bulgarian Academy of Sciences, 60, 42–54.

Gupta, R., and Bhave, P. R. (1996). "Comparison of methods for predicting deficient-network performance." *Journal of Water Resources Planning and Management*, 122(3), 214–217.

Haimes, Y. Y. (2009). *Risk modeling, assessment, and management*, John Wiley & Sons, Hoboken, NJ.

Handl, J., Lovell, S.C., Knowles, J. (2008) "Multiobjectivization by decomposition of scalar cost functions." *Proceedings of the 10th international conference on Parallel Problem Solving from Nature*, Dortmund, Germany,

Hart, D. B., McKenna, S. A., Klise, K. A., Wilson, M. P., and Murray, R. (2009). *CANARY user's manual and software upgrades,* Environmental Protection Agency, Washington, D.C.

Hrudey, S., and Hrudey, E. (2004). *Safe drinking water: lessons from recent outbreaks in affluent nations*, IWA Publishing, London.

Hrudey, S., and Hrudey, E. (2007). "Published case studies of waterborne disease outbreaks – evidence of a recurrent threat." *Water Environment Research*, 79(3), 233–245.

Janke, R., Murray, R., Uber, J., and Taxon, T. (2006). "Comparison of physical sampling and real-time monitoring strategies for designing a contamination warning system in a drinking water distribution System." *Journal of Water Resources Planning and Management*, 132(4), 310–313.

Jin, Y. and Branke, J. (2005). "Evolutionary optimization in uncertain environments: A survey." *IEEE Transactions on Evolutionary Computation*, 9(3), 303–317.

Johnston, G., and Brumbelow, K. (2008). "Developing Mesopolis – a "virtual city" for research in water distribution systems and interdependent infrastructures." Available online at:

https://ceprofs.civil.tamu.edu/kbrumbelow/Mesopolis/Mesopolis%20procedural%20documentation.pdf

Kanta, L., and Brumbelow, K. (2012). "Vulnerability, risk, and mitigation assessment of water distribution systems for insufficient fire flows." *Journal of Water Resources Planning and Management*, in press, DOI: 10.1061/(ASCE)WR.1943-5452.0000281

Kaplan, S., and Garrick, B. J. (1981) "On the quantitative definition of risk." *Risk Analysis*, 1(1), 11–27.

Khanal, N., Buchberger, S. G., and McKenna, S. A. (2006). "Distribution system contamination events: exposure, influence, and sensitivity." *Journal of Water Resources Planning and Management*, 132(4), 283–292.

Khouadjia, M. R., Alba, E., Jourdan, L., and Talbi, E. (2010). *Multi-swarm optimization for dynamic combinatorial problems: A case study on dynamic vehicle routing problem.* In Dorigo et al. (Eds.), *Swarm Intelligence* (pp. 227–238), Springer, Berlin.

Krapivka, A., and Ostfeld, A. (2009). "Coupled genetic algorithm – linear programming scheme for least-cost pipe sizing of water-distribution systems." *Journal of Water Resources Planning and Management*, 135(4), 298–302.

Laird, C. D., Biegler, L. T., van Bloemen Waanders, B. G., and Bartlett, R. A. (2005). "Contamination source determination for water networks." *Journal of Water Resource Planning and Management*, 131(2), 125–134.

Lindell, M. K., Prater, C. S., and Perry, R.W. (2006). *Fundamentals of emergency management*, Federal Emergency Management Agency, Washington, DC.

Lindell, M., and Prater, C. S. (2003). "Assessing community impacts of natural disasters." *Natural Hazards Review*, 4(4), 176–185.

Liu, L., Ranjithan, R., and Mahinthakumar, G. (2011). "Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure." *Journal of Water Resources Planning and Management*, 137(2), 183–192.

Lloyd, S. P. (1982). "Least squares quantization in PCM." *IEEE Transactions on Information Theory*, 28 (2), 129–137.

Lopez-de-Alba, P. L., Lopez-Martinez, L., Cerda, V., and De-Leon-Rodriguez, L. M. (2001). "Simultaneous determination of tartrazine, sunset yellow and allura red in commercial soft drinks by multivariate spectral analysis." *Quimica Analitica*, 20(2), 63–72.

Lopez-de-Alba, P. L., Wrobel-Kaczmarczyk, K., Wrobel, K., Lopez-Martinez, L., and Hernandez, J. A. (1996). "Spectrophotometric determination of Allura Red (R40) in soft drink powders using the universal calibration matrix for partial least squares multivariate method." *Analytica Chimica Acta*, 330(1), 19–29.

Moore, D. A., Fuller, B., Hazzan, M., and Jones, J. W. (2007). "Development of a security vulnerability assessment process for the RAMCAP chemical sector." *Journal of Hazardous Materials*, 142(3), 689–694.

Morgan, M. G., and Henrion, M. (1990). *Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis,* Cambridge University Press, New York, NY.

Mori, N., Imanishi, S., Kita, H., and Nishikawa, Y. (1997). "Adaptation to changing environments by means of the memory-based thermodynamical genetic algorithm." *Proceedings of the 7th International Conference on Genetic Algorithms, East Lansing, MI.*

Morley, K. M. (2010). "Advancing the culture of security and preparedness in the water sector." *Journal of American Water Works Association*, 102(6), 34–37.

Murray, R., and Hart, W. E. (2010). "Review of sensor placement strategies for contamination warning systems." *Journal of Water Resources Planning And Management*, 136(6), 611–619.

National Research Council's Committee on Methodological Improvements to the Department of Homeland Security's Biological Agent Risk Analysis. (2008). *Department of Homeland Security bioterrorism risk assessment: A call for change*, The National Academies Press, Washington, D.C.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell L., Hilton A. C, Karamouz, K., Minsker, B., Ostfeld, A., Singh, A., and Zechman, E. (2010). "State of the art for genetic algorithms and beyond in water resources planning and management." *Journal of Water Resources Planning and Management*, 136(4), 412–432.

Nie, Y., Biegler, L. T., and Wassick, J. M. "Integrated scheduling and dynamic optimization of batch processes using state equipment networks." *AIChE Journal*, in press, DOI: 10.1002/aic.13738

Nilsson, K. A., Buchberger, S. G., and Clark, R. M. (2005). "Simulating exposures to deliberate intrusions into water distribution systems." *Journal of Water Resources Planning and Management*, 131(3), 228–236.

Office of Environmental Health Assessment Services (1999). *Hazards of short-term exposure to arsenic-contaminated soil*. Olympia, WA.

Ostfeld, A., and Salomons, E. (2004). "Optimal layout of early warning detection stations for water distribution systems security." *Journal of Water Resources Planning and Management*, 130(5), 377–385.

Parks, S. L. I., and VanBriesen, J. M. (2009). "Booster disinfection for response to contamination in a drinking water distribution system." *Journal of Water Resources Planning and Management*, 135(6), 502–511.

Pasha, M. F., and Lansey, K. (2010). "Effect of parameter uncertainty on water quality predictions in distribution system-case study." *Journal of Hydroinformatics*, 12(1), 1–21.

Perelman, L., and Ostfeld A. (2010). "Extreme impact contamination events sampling for water distribution systems security." *Journal of Water Resources Planning and Management*, 136(1), 80–87.

Perelman, L., and Ostfeld A. "Extreme impact contamination events sampling for real sized water distribution systems." *Journal of Water Resources Planning and Management*, in press, doi: 10.1061/(ASCE)WR.1943-5452.0000206.

Perry, R. W., and Lindell, M. K. (2007). *Emergency planning*. John Wiley, Hoboken, NJ.

Poulin, A., Mailhot, A., Grondin, P., Delorme, A., Periche, A., and Villeneuve, J. P. (2008). "Heuristic approach for operational response to drinking water contamination." *Journal of Water Resources Planning and Management*, 134(5), 457–265.

Preis, A., and Ostfeld, A. (2008). "Multiobjective contaminant response modelling for water distributions systems security." *Journal of Hydroinformatics*, 10(4), 267–274.

Propato, M., Sarrazy, F., and Tryby, M. (2010). "Linear algebra and minimum relative entropy to investigate contamination events in drinking water systems." *Journal of Water Resources Planning and Management*, 136(4), 483–492

Protopapas, A. L., Katchamart, S., and Platonova, A. (2000). "Weather effects on daily water use in New York City." *Journal of Hydrologic Engineering*, 5(3), 332–338.

Reynolds, K. A., Mena, K. D., and Gerba, C. P. (2008). "Risk of waterborne illness via drinking water in the united states." *Reviews of Environmental Contamination and Toxicology*, 192, 117–158

Rossman L. A. (2000). *EPANET 2.0, user's manual*, National Risk Management Research Laboratory, U.S. EPA, Cincinnati, OH.

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

Shang, F., Uber, J., and Rossman, L. A. (2008). *EPANET multi-species extension user's manual*, U.S. Environmental Protection Agency, EPA/600/C-10/002, Washington, DC.

Toffolo, A., and Benini, E. (2003). "Genetic diversity as an objective in multi-Objective evolutionary algorithms." *Evolutionary Computation*, 11(2), 151–167.

Torres, J. M., Brumbelow, K., and Guikema S. (2009) "Risk classification and uncertainty propagation for virtual water distribution systems." *Reliability Engineering and System Safety*, 94, 1259–1273

Uber, J., Regan, M., and Janke, R. (2004). "Use of system analysis to assess and minimize water security risks." *Journal of Contemporary Water Research and Education*, 129(1), 34–40.

United States Bureau of Reclamation (2003). *Guidelines for achieving public protection in dam safety decisionmaking*, United States Department of Interior, Bureau of Reclamations, Denver, CO.

United States Congress (2002). *Public health security and bioterrorism preparedness and response act of 2002*. Public Law No. 107-188, H.R. 3448, 107th Congress. Government Printing Office, Washington, D.C.

United States Environmental Protection Agency (USEPA) (2003). *Response protocol toolbox: Planning for and responding to drinking water contamination threats and incidents*, USEPA, Washington, D.C.

United States Environmental Protection Agency (USEPA) (2004a). *Response protocol toolbox: Planning for and responding to drinking water contamination threats and incidents—Module 5: Public health response guide*, USEPA, Washington, D.C.

United States Environmental Protection Agency (USEPA) (2004b). *Estimated per capita water ingestion in the United States*, EPA-822-R-00-001, Office of Water, Washington, D.C.

United States Environmental Protection Agency (USEPA) (2010). *Revised public notification handbook*, EPA-816-R-09-013, Office of Water, Washington, D.C.

United States Food and Drug Administration (2012). "*Color additive inventories.*" Available at:

<http://www.fda.gov/ForIndustry/ColorAdditives/ColorAdditiveInventories/ucm1156
41.htm> (Last accessed Feb. 2, 2012)

United States Government Accountability Office (2004). *Drinking water: Experts' views
on how federal funding can be spent to improve security*. Report No.: GAO-04-
1098T, Washington, DC.

Ursem, R. K. (2000) "Multinational GA optimization techniques in dynamic
environments." *Proceedings of 2nd Genetic and Evolutionary Computation
Conference*, Las Vegas, NV.

Van Leuven, L. J. (2011). *Water/wastewater infrastructure security: threats and
vulnerabilities*. In Clark, R. M., Hakim, S., and Ostfeld, A. (Eds.), *Handbook of water
and wastewater systems protection* (pp. 27–47), Springer, New York, USA.

Vickers, A. (2001). *Handbook of water use and conservation*, WaterPlow Press,
Amherst, MA.

Wagner, J. M., Shamir, U., and Marks, D. H. (1988). "Water distribution reliability:
simulation methods." *Journal of Water Resources Planning and Management*,
114(3), 276–294.

Wang, H., and Harrison, K. W. "Bayesian update method for contaminant source
characterization in water distribution systems." *Journal of Water Resources Planning
and Management*, in press, doi:10.1061/(ASCE)WR.1943-5452.0000221.

Wesley, J. D., Minsker, B. S., and VanBlaricum, V. L. (2006). "Bayesian belief
networks to integrate monitoring evidence of water distribution system

contamination." *Journal of Water Resources Planning and Management*, 132(4), 234–241.

Wineberg, M., and Oppacher, F. (2000). "Enhancing the GA's ability to cope with dynamic environments." *Proceedings of 2nd Genetic and Evolutionary Computations Conference*, Las Vegas, NV.

Zechman, E. M. (2010). "Integrating complex adaptive system simulation and evolutionary computation to support water infrastructure threat management." *Proceedings of 12th Genetic and Evolutionary Computations Conference*, Portland, Oregon.

Zechman, E. M. (2011). "Agent-based modeling to simulate contamination events and evaluate threat management strategies in water distribution systems." *Risk Analysis*, 31(5), 758–772.

Zechman, E. M., and Ranjithan, S. (2009). "Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system." *Journal of Water Resources Planning and Management*, 135(5), 334–343.

VITA

Amin Rasekh

Texas A&M University

3136 TAMU, College Station, TX 77843-3136

E-mail: aminrasekh@neo.tamu.edu

Tel: (979) 739-9652

Ph.D.	Civil Engineering – Water Resources Engineering	August, 2012
	Texas A&M University, College Station, TX	
M.S.	Civil Engineering – Hydraulic Structures Engineering	August, 2009
	Iran University of Science and Technology, Tehran, Iran	
B.S.	Aircraft Maintenance Engineering	August, 2006
	Civil Aviation Technology College, Tehran, Iran	