ABSTRACT

VANKAYALA, PRAVEEN. Contaminant Source Identification in Water Distribution Networks under the conditions of Uncertainty. (Under the direction of Dr. G. Mahinthakumar.)

Recent attacks on the national water infrastructure have led to growing need for improved management and control of municipal water networks. Water distributions systems are susceptible to accidental and intentional chemical or biological contamination that could result in adverse health impact to the consumers. This study focuses on the contaminant source identification problem for a water distribution system under uncertainty in demand data. Due to inherent variability in water consumption levels, demands at consumer nodes in a water distribution system remain one of the major sources of uncertainty in source identification problem. In this research, the nodal demands are considered to be stochastic in nature and are varied using Gaussian and Auto Regressive models. The source identification problem is solved using the simulation-optimization method where EPANET water distribution system model acts as a simulator. Contaminant concentration observations are synthetically generated at arbitrarily selected sensor locations by specifying the location and mass loading of a potential contaminant source that is arbitrarily located at one of nodes in the system. Genetic Algorithm (GA) is used as the optimization model with the goal of finding the source location and concentration, by minimizing the difference between the simulated and observed concentrations at the sensor nodes. Two variations of GA, stochastic GA and noisy GA are applied to the same problem for comparison. Results show that noisy GA is robust and is less computationally expensive than stochastic GA in identifying the contaminant source location and concentration for both of stochastic demand models.

CONTAMINANT SOURCE IDENTIFICATION IN WATER DISTRIBUTION NETWORKS UNDER THE CONDITIONS OF UNCERTAINTY

by

PRAVEEN VANKAYALA

A thesis submitted to the Graduate Faculty of

North Carolina State University

In partial fulfillment of the

Requirements for the degree of

Master of Science

in

CIVIL ENGINEERING

Raleigh, NC

JULY 2007

Approved by:

_____	_____
Dr. Sankar Arumugam	Dr. S. Ranji Ranjithan

_____

Dr. G. Mahinthakumar

(Chair of Advisory Committee)

DEDICATION

To my late Grandmother

To my beloved Sri Sri Sri Thridandi Ranga Ramanuja Jeeyar Swami

To my family

BIOGRAPHY

Praveen Vankayala was born in the city of Kakinada, Andhra Pradesh, India and grew up in Amalapuram, India. He completed his schooling from Vijnana Vihara Residential School Gudilova, Visakhapatnam, India. He was the school house leader and was also athletics and cricket team member that won many competitions. He also represented for the school in many extra-curricular activities. He topped the school in CBSE tenth grade examinations and secured highest percentage in Mathematics and English. He then joined the Undergraduate program in Civil Engineering at Indian Institute of Technology (IIT), Madras, India. He received scholarship from the government of Andhra Pradesh during the four years of undergraduate curriculum. During his stay at IIT Madras, he was the hostel volley ball captain, and team member in athletics and cycling events. After graduating from IIT Madras, he then joined the Master's program in Civil Engineering at North Carolina State University, Raleigh in the fall of 2005. He has been working as a Research assistant on this project under the able guidance of Dr. G.Mahinthakumar. After completing his masters, he plans to pursue a professional career in water resources engineering.

ACKNOWLEDGEMENTS

This thesis is successfully completed due to wonderful support, guidance and encouragement I have had at various stages. I would like to express my sincere thanks to everyone who helped me and made my graduate school days a wonderful experience.

First of all, I would like to extend my whole-hearted appreciation to my advisor, Dr. G. Mahinthakumar who has given me the opportunity to work with him and be a part of his research group. He has always provided me with invaluable guidance and constructive criticism to support my work I would also thank Dr. Shankar Arumugam who has been a great resource and has provided me with an abundance of valuable inputs in Time Series modeling. My special thanks to Dr. Ranji Ranjithan for his valuable comments and suggestions to my project. Finally, I would like to thank Dr. Steve Buchberger for providing me with the residential demand data without which this thesis may not be possible. Many other faculty and staff members have been very helpful, and I thank them all for their efforts.

My special thanks to Li and Baha for their useful suggestions and feedback. Friends and fellow lab mates also have lent a hand in one way or another. I am thankful to all my friends in Raleigh and all others for making my stay in NCSU more enjoyable. My parents, grandfather and all my relatives for their constant encouragement and unending moral support throughout my life.

Lastly, I thank his holiness Sri Sri Thridandi Ranga Ramanuja Jeeyar Swami, for all his grace and blessings throughout my life.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

1. INTRODUCTION

1.1 Overview

The primary aim of the urban water distribution networks is to efficiently deliver the water to the consumers .They are designed to supply water for basic domestic purposes and for commercial and industrial uses. A water distribution network consists of complex networks of pumps, pipes, and storage tanks which provide high-quality treated and safe water to millions of consumers around the globe (Nilsson *et al.,* 2005). Water distribution networks are closed systems, and hence, they tend to be more secure than raw water sources. Nonetheless, water distribution networks are susceptible to various types of threats. The recent attacks against targets within United States forced many government planners to consider the possibility that water infrastructure may be prone to terrorist attacks (Clark and Deininger 2000). The water distribution systems are vulnerable to full range of threats that include a direct physical attack on the water supply infrastructure, sabotage of supervisory control and data acquisition instruments, and a deliberate chemical or biological contaminant injection (Murray *et al.*, 2004). Any water outlet, such as a fire hydrant or even a household water tap, can be an access point for backflow contamination in to the network (Laird *et al.*, 2005). There are number of factors that make water supply systems vulnerable to the above attacks: they are spatially diverse, susceptible to intrusion directly from a large number of locations in the network, and contain many components. These attacks can be prevented by

providing physical security to the water distribution plants but it can be enforced only to a limited extent, primarily confining to territories of the water distribution system, storage tanks and treatment systems while leaving other access points such as fire hydrants, household taps unprotected. If a contaminant is injected by breaching the physical security, the only way to detect the source will be associated with chemical concentrations and hydraulics. Early warning devices like sensors can be used to detect contaminants at critical points of the network, but due to the spatial diversity of the drinking water network and high procurement costs of sensors, it is almost impossible to install sensors all around the network. A significant amount of research had been carried out to identify the optimal location of sensors that can provide improved measurements for identifying the location of source of contamination.

1.2 Motivation

Water distribution systems are particularly more sensitive to deliberate microbiological contamination in to the network. Upon injection, a contaminant would move rapidly in to the system and would be difficult to track the contaminant path. This could result in adverse health impacts among the residents of the area served by the water distribution system. An outbreak of water borne disease in a small community in Western Missouri illustrates this point. The Gideon outbreak (1993) which caused in hospitalization of 50 people from a small town Gideon was a result of consumption of Gideon municipal water. Traces of *Salmonella* serovar *typhimurium* were found in the water samples and nearly 10 people exhibited diarrhea and eventually died. The Missouri department of health was

informed that a water supply link as the cause for the outbreak (Clark and Deininger 2000). Therefore, there is a need to identify the contaminant source location and its concentration correctly and quickly, to prevent a widespread outbreak. This type of problem is termed as a Source identification problem which involves finding the contaminant source location (typically a node in the network) and its release history from observed concentrations at several sensor locations in the network.

1.3 Objective

The main objective of this study is to study the effects of uncertainty in demand data on the source identification problem and solving this problem using Stochastic and Noisy genetic algorithms. Demands at consumer nodes in a water distribution system are one of the major sources of uncertainty in source identification problem. The source identification is an inverse problem of a standard water quality simulation which can be solved in both direct and indirect ways. In this project, it is formulated as an optimization problem with the goal of finding the source location and concentration by minimizing the difference between the simulated and observed concentrations at the sensor nodes. This approach is commonly termed “simulation-optimization” as the optimization algorithm drives a simulation model to solve the problem. EPANET [26] water distribution system model is used as a simulator for this simulation-optimization problem and optimization problem is coded using genetic algorithm tool box in MATLAB [30].

1.4 Organization of the Thesis

The structure of the thesis is as follows:

This chapter described the need for the study on source identification problem and the objectives of the study. Chapter 2 reviews related literature and highlights significant findings in existing papers in relation to the above objectives of the study. Chapter 3 presents the methodology of the model development. Chapter 4 discusses the modeling results with an illustrative case study and the final chapter summarizes the important outcomes from the study and recommendations for further research.

CHAPTER 2

2. LITERATURE REVIEW

2.1 Background

Over the past few years, a significant amount of research work had been done in the areas of identifying optimal monitoring sensor locations [*Ostfeld and Salomons (2004), Propato (2006) etc.*], modeling of residential water demands [*Buchberger and Wu (1995), Buchberger et al. (2003) etc.*], optimization model for water distribution system design etc [*Lansey and Mays (1989) etc.*].Relatively, less attention is paid towards the source identification problem. But, the focus on water infrastructure security and consumer health impacts has increased since the Bioterrorism Act was introduced in 2002 (Bioterrorism Act, 2002). *Laird et al. (2005)* presented an origin tracking algorithm for solving the inverse problem of contaminant source identification based on a nonlinear programming framework. An origin tracking algorithm is used to reformulate the pipe expressions and characterize the time delays associated with network pipes which removes the need to discretize along the length of the pipes. *Preis and Ostfeld (2006)* developed a new methodology for contaminant source identification in water distribution systems using a hybrid model trees-linear programming algorithm. The model trees which are developed through a learning process substitute the EPANET simulations. The linear rule classification structure of the model trees is then coupled with a linear programming formulation to solve the inverse problem.

Another approach to solve the source identification problem is a simulation-

optimization approach or indirect approach, in which a search procedure is coupled to a simulation model. *Guan et al. (2006)* proposed simulation-optimization method to solve a non-linear contaminant source and release-history identification problem for a complex water distribution system. This approach is based on optimization analysis, using the EPANET water distribution model as the simulator. In this methodology, a reduced gradient algorithm with an optimal step length is selected to search for release histories from among potential contaminant sources. *Liu et al. (2006)* developed an evolutionary algorithm based adaptive dynamic procedure (ADOPT) to adaptively solve the contaminant source identification problem in water distribution systems. This ADOPT procedure adaptively searches for the source characteristics such as location and concentration as the observation data is dynamically updated over time. But, there might be a time lag associated with these demands to be dynamically updated. Therefore, the demands which must to be updated at time *t*, might be updated at time step greater than *t+1*. This method also assesses the degree of non-uniqueness, i.e., whether more than one solution fits the available observations.

The aforementioned source identification approaches assume that the hydraulic inputs such as demands for example, are known deterministically. The inputs are assumed to be known with 100% certainty and that no noise or error exists in the measurements. Little research has been done so far on source identification problem coupled with uncertainty. An aspect of water distribution systems that has not yet been considered as part of source identification problem is that the demands with in the system are unknown and typically of a much shorter time scale than the length of the hydraulic time step used in distribution system models. Given these observations, this study considers the use of Genetic Algorithm to locate

the origin of contaminant source attack and help mitigate the effects of such attack.

2.2 Demands and Uncertainty

2.2.1 Demands

The consumption or use of water, also known as *water demand*, is the driving force behind the hydraulic dynamics occurring in water distribution systems. Anywhere that water can leave the system represents a point of consumption, including a customer's faucet, a leaky main, or an open fire hydrant. There are three types of basic demand types:

Customer demand – is the metered portion of the total water consumption required to meet the non-emergency needs of users in the system,

Unaccounted-for water – is the portion lost due to leakages in the system and

Fire flow demand – is a computed system capacity requirement for ensuring adequate protection is provided during fire emergencies (*Haestad et al. 2003*).

Water demand is assumed to be constant (i.e. they don't change over time) for steady state simulation calculations. However, in reality, Water demand varies continuously over time according to several time scales (hourly, daily, weekly, or seasonally). The demand may also vary due to changing factors such as climate, population, and degree of industrialization. Therefore, these demand fluctuations have to be incorporated in the base demands to accurately reflect the dynamics of the system, while performing model simulations.

The temporal variations in water usage for a municipal water system typically follow a 24-hour cycle called a *diurnal* demand pattern. Figure 2.1 illustrates a typical diurnal curve for a residential area. There is relatively low usage at night when most people sleep,

increased usage during the early morning hours as people wake up and prepare for the day, decreased usage during the middle of the day, and finally, increased usage again in the early evening as people return home.

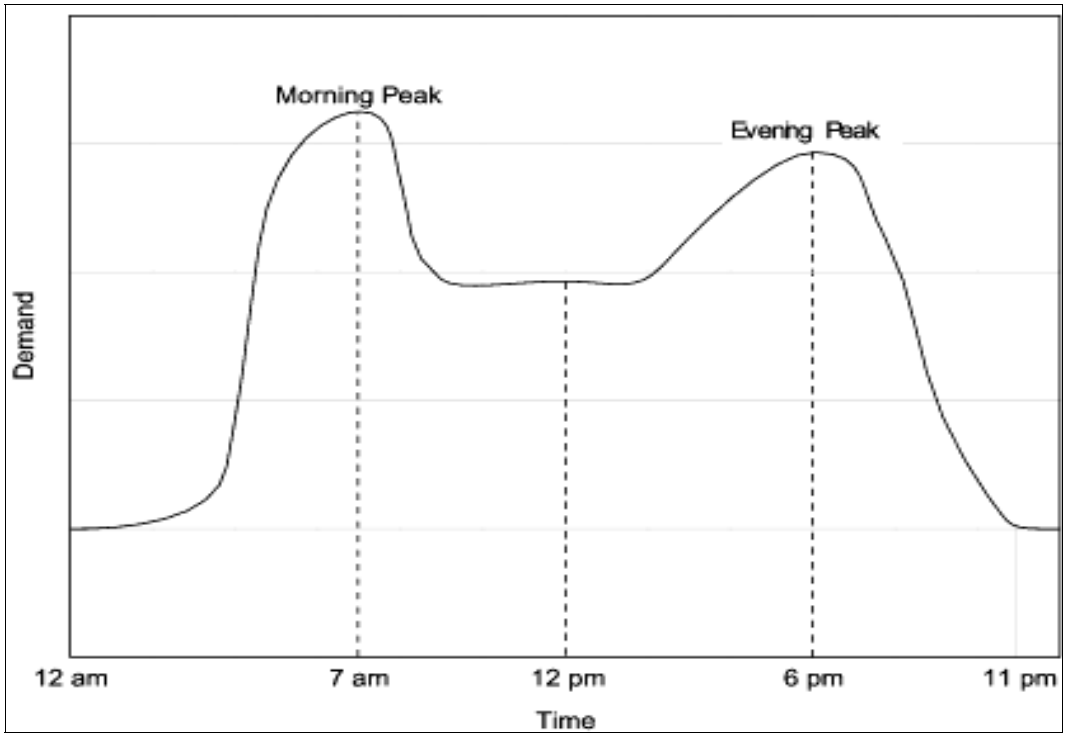


Figure 2. 1 A typical diurnal curve (source: Haestad *et al*. 2003).

However, the system usage changes not only on a daily basis, but also weekly and annually. Weekend usage patterns are different from weekday patterns and seasonal variations might also have an impact on the usage as discussed before.

Customer demand is the major portion of total water consumption which can categorized in several ways. Residential and Industrial demands are one such type of

classification. However, the usage patterns are as diverse as the customers themselves. Figure 2.2 illustrates the different diurnal demand curves for various classifications.

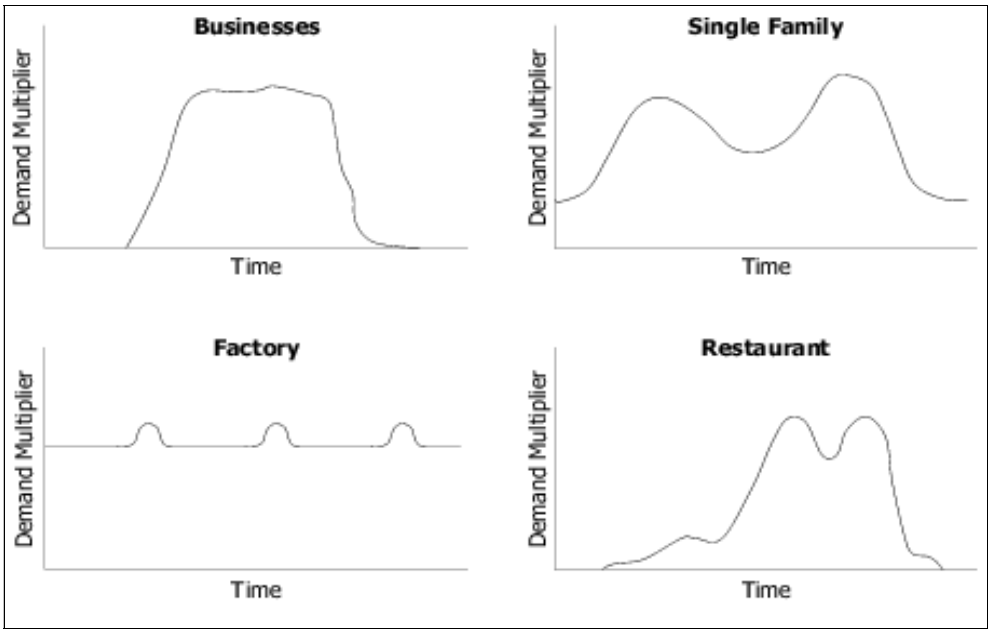


Figure 2. 2 Diurnal curves for different categories (source: Haestad *et al*. 2003)

2.2.2 Uncertainty

The vast majority of mathematical models in engineering use deterministic approaches to describe various processes and systems. However, most of the real life problems incorporate uncertainty one way or another. Uncertainty refers to degree of variability at the observations and lack of precise knowledge. For a contaminant source identification problem in a water distribution system, uncertainty may be present in the magnitude of observed concentration values at water quality sensors due to sampling bias or

errors, fluctuations in instantaneous demands. Uncertainty may also be introduced in water distribution system analysis due to simplifying assumptions used to model the hydraulics or water quality in the system. This work is limited to only variations in demand at consumer nodes.

The driving force behind the hydraulic dynamics occurring in water distribution systems can be attributed to the consumer water demand (*Walski et al. 2003*). It is unrealistic to study a water distribution system with a deterministic approach, which assumes that the system is functioning properly with purely conventional demand conditions only. A probabilistic approach that allows distribution systems to be studied under a wide variety of loading conditions provides a more complete description of the system's behavior and constitutes the necessary approach for an objective evaluation of the system's hydraulic reliability (*Gargano and Pianese, 2000*). Due to inherent variability in instantaneous water consumption levels, values of demands at nodes in a water distribution system remain one of the major sources of uncertainty in the source identification problem. Uncertainty in demand leads to uncertainty in head at the nodes, which in turn might affect the over all system performance. Various models of stochastic demands have been developed to incorporate the random component of the demands.

Uber *et.al* proposed a time series model for the variation in total demand. They developed AR (Autoregressive) model which employs a simple time series model to represent random demand statistics. In this model, the temporally varying demand is defined as a random variable and the basic demand model is expressed as the sum of the mean demand and a random variation. The junction demands are obtained from the total demand

by a simple deterministic disaggregation process. *Buchberger and Wu* (1995) developed a model where residential water demands occur as a non-homogeneous Poisson-Rectangular-Pulse (PRP) process, thus providing a stochastic model for indoor residential water demands as a starting point for quantifying the instantaneous temporal and spatial variability of flow through municipal water distribution systems. The PRP model gives instantaneous water demands at nodes throughout the entire network. *Cristina and Leopardi (2006)* performed uncertainty analysis based on Monte Carlo procedure to test robustness of the source identification selection methodology. In this project, time series model is applied to model the temporal variability of residential water demands for the source identification problem. The general methodology of the model is discussed in the next chapter [3.4].

2.3 Genetic Algorithms

Genetic Algorithms (GA) are a class of combinatorial methods used to solve complex optimization problems. Genetic Algorithms search for optimal solutions or designs using procedures analogous to natural selection or evolution. *Holland* (1975) performed the initial work on GA at the University of Michigan and another text by

Goldberg (1989) provides a comprehensive introduction of GA theory, operation, and research. In this project, two variations of GA are used which are explained in the following chapter.

2.3.1 Noisy Genetic Algorithms

Noisy Genetic Algorithms (Noisy GA) are simply ordinary genetic algorithms that operate in a noisy environment. Basic GA acts as a foundation for Noisy GA. The ‘noise’

that exists in certain environments can be defined as any factor that hinders the accurate evaluation of the fitness of a given trial design or chromosome. A noisy environment is commonly encountered while solving real-world problems, where knowledge about the domain is scarce and uncertainty is present. This type of "noise" prevents accurate evaluation of the fitness of individual members of the population. To illustrate a case of inaccurate fitness evaluation, consider a trial design that has a high fitness value under certain conditions, but when evaluated under a different set of conditions, has a poor fitness value. It is unknown which evaluation represents the "true" fitness of this trial design. As a result of this inaccurate evaluation of the individual's fitness, the user-defined objective function (also known as the fitness function) that operates in a noisy environment is termed a "noisy" fitness function (Miller 1997).

Miller (1997) investigated the effects of noise on GA convergence rates and population sizing in order to determine the optimal sample size (i.e., the number of noisy fitness function evaluations) for GAs that employ sampling fitness functions. GA performance is a function of both the convergence rate of the GA and the number of generations the GA can run within a computational time constraint. The optimal sample size is the size at which the performance gains from faster convergence rates and decreased population sizes balances the performance loss due to larger sample sizes.

Noisy Genetic algorithms haven't been applied for the source identification problem in water distribution networks. *Smalley et al. (2000)* and *Gopalakrishnan et al. (2001)* applied Noisy GA to risk-based ground water remediation problems. Similar multiple realization GAs for optimization under parameter uncertainty was applied by *Aly and Peralta*

(1999) to groundwater remediation design problem and by *Harrell (2001)* to watershed management problem. The next chapter explains the random demand models, optimization methods and the methodology for solving the problem.

CHAPTER 3

3. METHODOLOGY

3.1 Introduction

This chapter presents a discussion of the methodology chosen for solving the source identification problem based on the literature review and available data. As discussed in section [1.3], the contaminant source characterization problem is solved using simulation-optimization method for a water distribution system. This approach is based on optimization analysis, using EPANET water distribution system model as a simulator. The EPANET water distribution system simulator *(Rossman 2000)* performs extended-period simulation of hydraulic and water quality behavior within pressurized pipe networks. This well-documented and tested public domain simulator provides a convenient platform for implementing the optimization-simulation approach.

In this problem, EPANET is used to simulate concentrations at randomly selected monitoring locations by specifying the release history of potential contaminant source that is arbitrarily located with in the water distribution system. This information is used by the optimization model which acts as a predictor to estimate the release history and location of the contaminant source based on the similarity between simulation results and assumed values at selected monitoring locations.

The contaminant source characterization problem is solved using the numerical computing software – MATLAB [30]. The EPANET software is linked to MATLAB

with the help of EPANET toolkit. EPANET toolkit is a dynamic link library (DLL) of functions and an extension of EPANET simulation package that allows developers to customize EPANET software accordingly. The genetic algorithm optimization procedure is employed for solving this problem via genetic algorithm optimization toolbox in MATLAB. In this chapter, the following are discussed: Contaminant Source Characterization problem, demand variability models and Optimization models.

3. 2 Contaminant Source Characterization Problem

The source characterization problem includes finding the contaminant source location and its temporal mass loading history ("release history"). The release history includes start time of the contaminant release in the WDS, duration of release, and the contaminant mass loading during this time. Coupling simulation and optimization has provided a powerful tool for solving problems in the groundwater management area such as identification of contaminant sources in aquifers (Aral et al. 2001). This approach may also be an effective way to solve the contaminant source identification problem.

The architecture of the source identification problem is explained in figure 3.1. In this diagram, $x(t)$ = concentration of the contaminant source; $y_i(t)$ = concentration at a specified monitoring location; m = number of monitoring locations and t is the simulation time.

In the contaminant source identification problem, the first and foremost important step is varying the demands at the nodes. This is made possible by varying the total hourly demands of the system using two random models, Gaussian and Autoregressive models

respectively, which are discussed in detail in the following sections. These total hourly demands can be converted to nodal demands by simple deterministic disaggregation

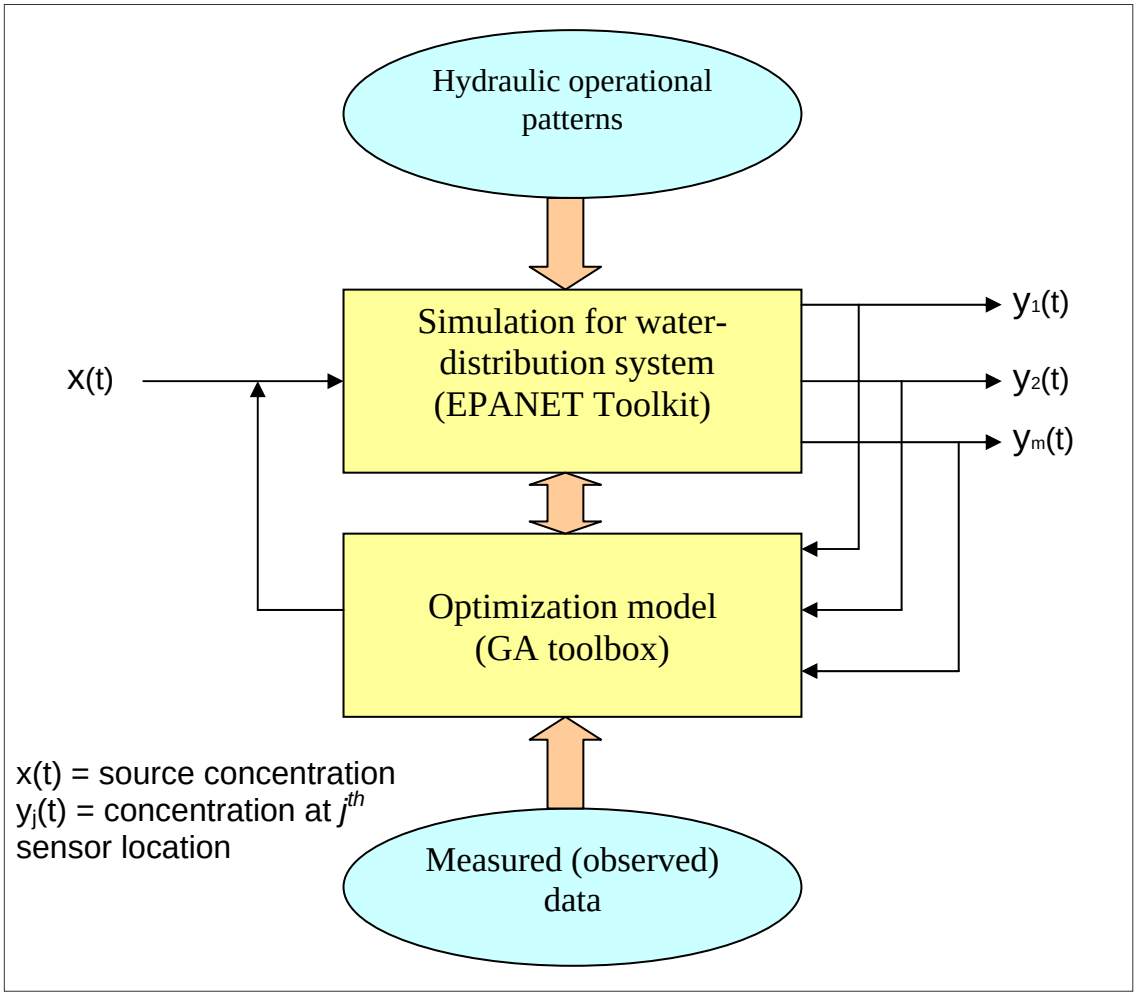


Figure 3. 1 Architecture of the source identification problem.

process. The demand variability is incorporated in the problem using demand patterns where, the demand patterns are multiplied by fixed base nodal demands to give the variable nodal demand.

The simulated concentrations are generated at monitoring locations using water quality simulation models i.e. EPANET. It is assumed that the water distribution system

operates under known hydraulic operational patterns. The predictor step uses genetic algorithm optimization model to estimate the concentrations at the selected sensor locations and optimize the difference between the assumed and observed concentrations at the monitoring locations.

Figure 3.2 illustrates a typical water distribution network. Some of the nodes in the water distribution system are identified as sensor locations. A single contaminant source is assumed to be present at an arbitrary location (one of the nodes) in the water distribution system. It is also assumed that the contaminant is non-reactive, i.e. it doesn't undergo any chemical reactions.

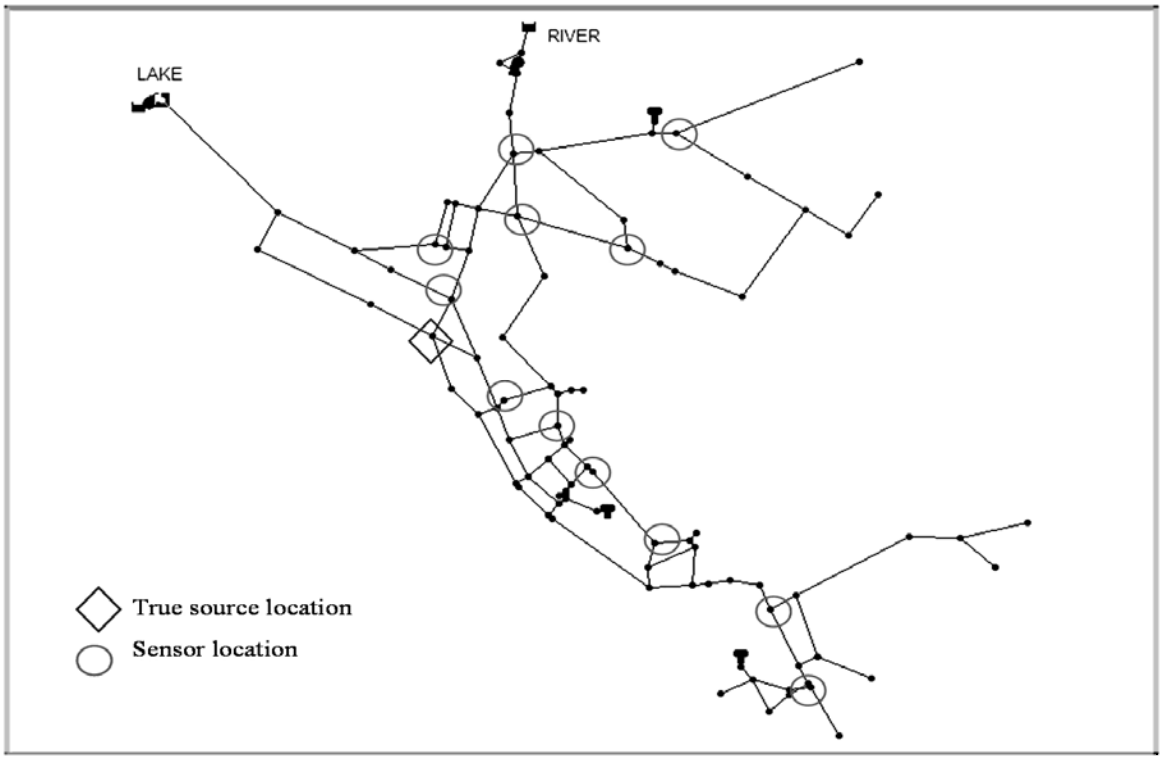


Figure 3. 2 A typical water distribution system with true contaminant source location and sensors location. (source: EPANET input file)

As the time progresses, the contaminant spreads across the system and the nodes which are identified as sensor locations report the concentration readings at specified time intervals. Significant amount of research work has already been done in optimal placement of sensors in water distribution systems and it is not considered in this project. In the current scenario, the sensor locations are arbitrarily selected. It is assumed that a contaminant source is present at a specified node in the system for a predetermined period of time and a water quality simulation is run. The concentration measurements which are obtained from the simulation are then recorded for each sensor during every time step and constitute a set of true source concentration profile readings.

In order to identify the true source, it is started out with a guess for the source location and its release history (i.e. start time of the contaminant, duration and the concentration of the contaminant). In this problem the start time of the contaminant and the duration are assumed to be constant. Thus, it's a two variable problem where location and mass loading (concentration) are the variables. For every guess, the water quality simulation is run for every time step to obtain the concentration profiles for all the sensor locations. The difference in the concentration reading values between the true source location and the predicted source location are obtained for each sensor location and for every time step. The greatest value of the difference among the sensor locations is selected for each time step. The objective function of the optimization method is to minimize the maximum error value. The goal is to come up with a predicted source that has exact source location and similar concentration value that falls in the range of the true source concentration.

Mathematically, the source characterization problem can be expressed as

Find: {contaminant source location (L), contaminant concentration (C_t) }

Objective Function: **Minimize the prediction error** i.e.,

$$F = \sum_{t=t_0}^{T} \sum_{k=1}^{N_s} \left\| C_k^{obs} - C_k^*(L, M_t) \right\|$$

Where,

L – Node ID of the contaminant source location.

$M(t_c)$ – Contaminant concentration

t_o - Contaminant start time

t – Current time step

C_k^{obs} - Observed concentration

$C_k^*(L, M_t)$ - Simulated concentration

N_s - Number of sensors

K - observation (sensor) location

3.3 Models of Random Demands

Estimates of future demands are often based on past water usage. Records for the total system demand may be used to estimate the average and peak demands at every

junction in the system. It is difficult to predict future demands far in advance with any mechanistic models. The demands in a water distribution system vary spatially. Spatial variation implies that the demand varies from node to node within a system. In addition to spatial variation, the demand at a certain node may also vary temporally i.e. hourly, daily, weekly etc. It may also vary due to other factors such as climate, population and industrialization. In this problem, the demands are considered to be stochastic in nature and are varied using a Gaussian model and a time series model to represent the random demand statistics.

3.3.1 Gaussian model

In the Gaussian model, the demand is varied by perturbing the demand values at every node for each time step. Mathematically, it is expressed as:

$$d_i^j = d_i^j + \text{var} * d_i^j * randn \qquad \textbf{(3.1)}$$

Where, d_i^j is the demand at network junction *i*, and time step *j*

Var is the percentage of variation in the demand at each time step *j*

randn is normally distributed random number within 95% Confidence Interval.

In this problem, the percentage variation value (var) is assumed to be 0.25. Demand at each node is varied using equation 3.1 and thus random demands are generated using Gaussian model. The next section explains another method for varying the demands using a time series model.

3.3.2 Auto Regressive (AR) model

Autoregressive (AR) model is a stochastic, stationary, linear, non-seasonal model used for representation of certain practically occurring series such as weather conditions, hydrologic data etc. This model is used for representing the stochastic nature of the nodal demands in a water distribution system. In this model, the current value of the demand is expressed as a finite, linear aggregate of previous values of the demands plus a white noise term. This form of time series model is intuitively appealing and has been widely applied to data sets in many different fields. This model has been used for forecasting the demand values for each hydraulic time step in this problem.

The simplest form of AR model AR(1), can be expressed mathematically as follows,

$$d_t - \mu = \phi_1 (d_{t-1} - \mu) + \varepsilon_t \qquad \textbf{(3.2)}$$

Where μ is the mean level of the process,

ϕ_1 is the non-seasonal AR parameter

d_t, d_{t-1} are the observations at time t and t-1 respectively and,

ε_t is the white noise term that is uncorrelated and normally distributed with a mean of zero and variance of $\sigma_{\varepsilon_t}^2$ i.e. N(0, $\sigma_{\varepsilon_t}^2$) . The variance is calculated from the available demand data.

The observation d_t measured at time t depends only upon the time series value at time t-1 plus a white noise term $\sigma_{\varepsilon_t}^2$ and is termed as AR process of order 1, shortly AR (1).Here, the order determines the number of prior time series observations, the current time

series observation depends on.

The two primary stages in building an Autoregressive model are model estimation and model validation. A data set is needed in order to estimate the parameters of the time series model. Residential water demands measured at 21 homes in Milford, Ohio for a period of 31 days are considered for estimating the parameters of the AR model (Buchberger *et al.* 2003). This data is part of 365,000 residential water demands collected from 21 single family homes in Milford, Ohio during a seven month period. This collection of field data represents one of the most extensive high resolution records (one-second interval) of continuous residential water use ever assembled. The data file has demands stored as rectangular pulses where each pulse is identified by three variables, pulse start time, duration and intensity. Therefore, these rectangular pulses have to be converted to average hourly demands. The following pulse aggregation procedure explains the conversion:

1. Consider the demand data for one particular day. The total demand at any time is obtained by summing up the intensities of all active pulses.

2. Hourly averaged demand is obtained by dividing the total demand with its time interval for each time step. This can be mathematically expressed as:

$$D_T(i) = \frac{1}{T} \int_{(i-1)T}^{iT} D(t)dt \qquad \textbf{(3.3)}$$

Herein, $D_T(i)$ is the hourly averaged total demand at time step i ; $D(t)$ is the intensity of aggregated pulse at time t; T is the time scale.

Average hourly demands are thus calculated by this procedure for all the days and for

each house. Once these are obtained, the total hourly demands averaged over number of days are obtained for each house which is then used in model estimation. These hourly demands are for the residential network in Milford, Ohio. In order to obtain the hourly demands for the example water network, the hourly demands are multiplied by weighting factors for obtaining new hourly demands. In order to study the source identification problem an example water distribution network is considered and it is explained in detail in Chapter 4.

AR model estimation is done for each time step (total time steps are 24) to get an accurate estimation of the forecasted hourly demands. In the model estimation process, the main aim is to identify the order of the autoregressive model for each time step. The primary tools for reaching the goal are autocorrelation plot and the partial auto correlation plot. The sample autocorrelation and the sample partial auto correlation plot are compared to the theoretical behavior of these plots when the order is known. Figure 3.3 and figure 3.4 show the sample auto correlation and sample partial auto correlation plots for a particular time step respectively. The approximate 95% Confidence interval is determined utilizing equation [3.4] and by assuming that the sample ACF is not significantly different from zero after lag zero.

$$95\%\ \text{C.I.} = \left[-\frac{1.96}{\sqrt{N}}, +\frac{1.96}{\sqrt{N}} \right] \qquad \textbf{(3.4)}$$

Where, N is the number of observations

As can be seen in the Figure 3.3, the estimated ACF has significant value - non-zero values at lower lags and tends to follow a damped exponential curve. Because, the theoretical ACF of an AR process behaves in this fashion, this indicates that the AR model with order

greater than one may be appropriate to fit the residential demand data. Partial auto correlation plots are helpful in identifying the order of an autoregressive model. The partial autocorrelation of an AR(*p*) process is zero at lag *p+1* and greater. If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to help identifying the order.

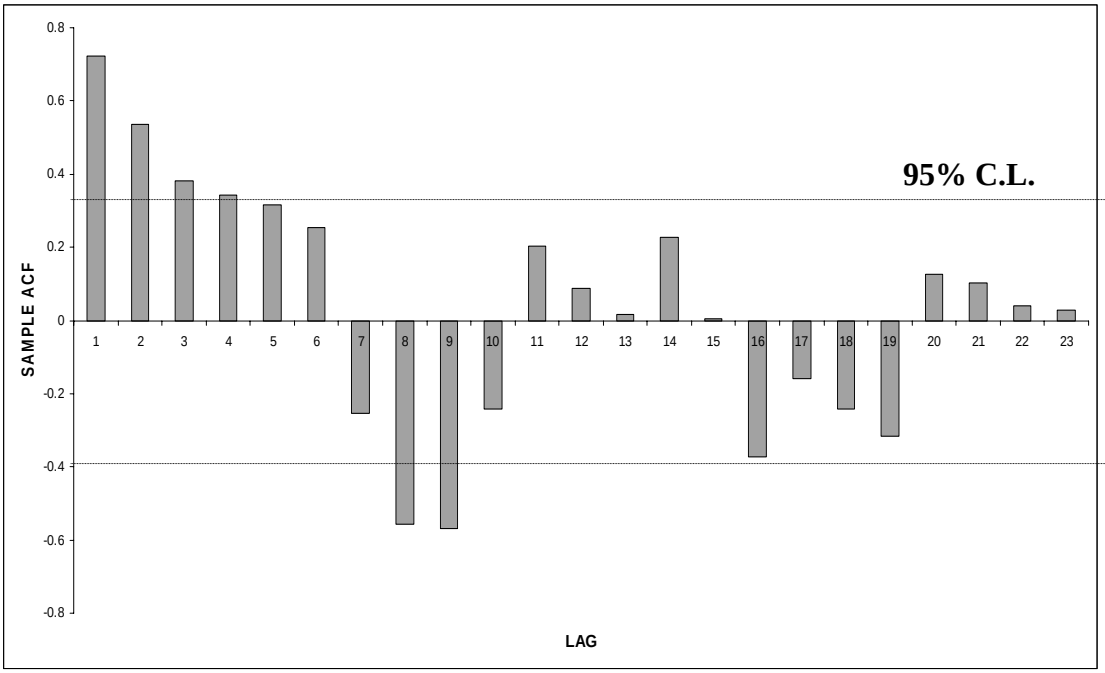


Figure 3. 3 Sample ACF and 95% confidence limits for the average hourly residential demands of Milford, Ohio.

The graph of the estimated PACF for the average hourly residential demands is shown in the figure 3.4. The approximate 95% confidence limit for the partial autocorrelation is calculated by substituting N=30 in to equation [3.4] and plotting it above and below the horizontal axis. It can be seen that there are rather larger values (exceeding the

confidence limit) for the estimated PACF at lags 1,3,14 and 17. The unexpected big values at lags 14 and 17 could be due to chance alone or else the limited size of the sample which was used to estimate the PACF at lags 14 and 17.

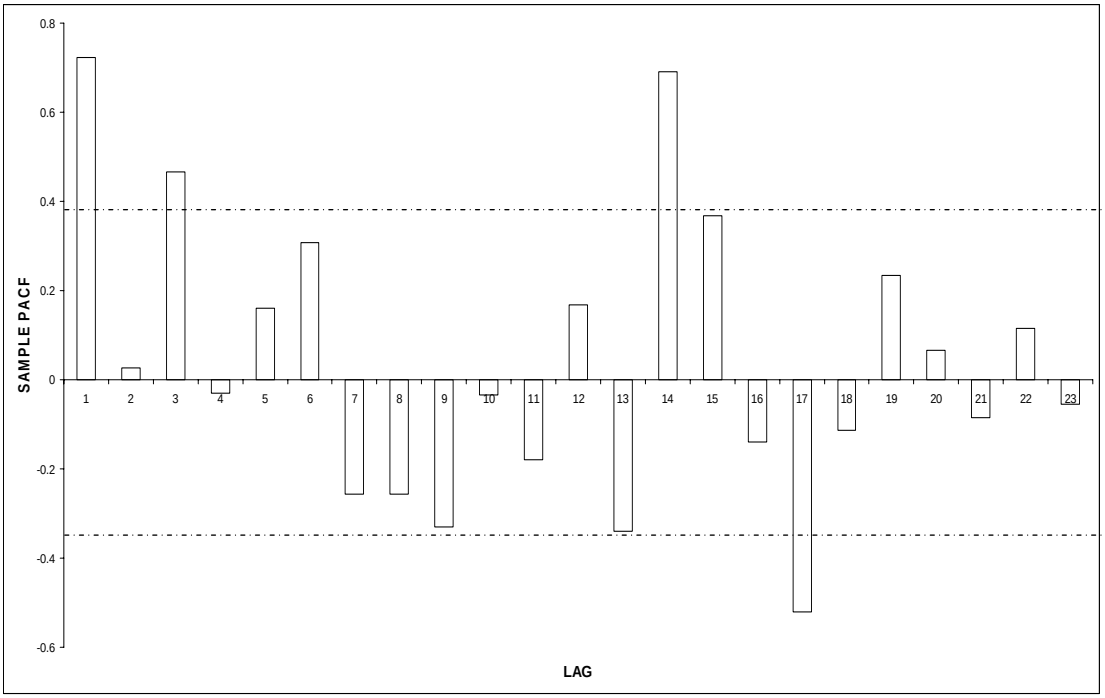


Figure 3. 4 Sample PACF and 95% Confidence limits for the average hourly residential demands of Milford, Ohio.

As the estimated PACF cuts off after lag 3, this implies that an AR(3) model should perhaps be fitted to the data. In addition, as the sample PACF at lag 2 is not very large, the ϕ_2 parameter should be constrained to zero in the AR (3) model in order to reduce the number of model parameters. The model parameters are obtained using the Yule Walker equations. In this manner, the sample auto correlation functions and sample partial auto correlation functions are obtained for all the 24 time steps and thus determine the order of the

AR model for each time step.

The parameters of an AR model are solved using set of linear equations which are known as Yule-Walker equations. The Yule-Walker equations for an AR model of order *p* can be mathematically expressed as:

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots\cdots + \phi_p\rho_{k-p} \qquad \textbf{(3.5)}$$

where, k = 1,2,…*p*

$\phi_1, \phi_{2,} \cdots, \phi_p$ are model parameters and $\rho_1, \rho_2, \cdots, \rho_p$ are theoretical auto correlations.

The Yule-Walker equations can be written in matrix form as:

$$\phi = D_p^{-1}\rho_p \qquad \textbf{(3.6)}$$

where,

$$\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ . \\ . \\ . \\ \phi_p \end{pmatrix}, \rho_p = \begin{pmatrix} \rho_1 \\ \rho_2 \\ . \\ . \\ . \\ \rho_p \end{pmatrix}, D_p = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \text{L} & \rho_{p-1} \\ \rho_1 & 1 & \rho_1 & \text{L} & \rho_{p-2} \\ . & . & . & \text{L} & . \\ . & . & . & \text{L} & . \\ . & . & . & \text{L} & . \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \text{L} & 1 \end{pmatrix}$$

In order to obtain the Yule-Walker estimates for the AR parameters, the theoretical autocorrelations ρ_k 's in equation [3.6] are replaced by their estimates r_k , k=1,2,…*p* (i.e.) sample autocorrelations.

For the above example dataset, the AR(3) parameters are calculated using the equation [3.6]

and the estimated model is written as:

$$d_t - \mu_t = 0.7057(d_{t-1} - \mu_{t-1}) + 0.05(d_{t-2} - \mu_{t-2}) + 0.466(d_{t-3} - \mu_{t-3}) + \varepsilon_t \quad \textbf{(3.7)}$$

where, μ_k is the mean of the hourly demand at time step k = t , t-1, t-2, t-3.

d_{t-k} is the hourly demand at time *t-k* and

ε_t is the white noise term that is uncorrelated and normally distributed with a mean of zero and variance of $\sigma_{\varepsilon_t}^2$ i.e. N(0, $\sigma_{\varepsilon_t}^2$)

The AR models are similarly estimated for all the time steps by the above procedure, thus varying the hourly residential demand using a time series model. The next section discusses the optimization methods for solving the source identification problem.

3.3 Optimization Algorithms

Identifying mass loading history and location of contaminant source is now formulated as an optimization problem. The optimization problem is solved using Genetic Algorithm and Direct Search Toolbox available in the computing software MATLAB [30]. The extensive details about the GA toolbox are provided in the matlab user guide [30] and are not discussed here. The following sections describe the optimization algorithms that are used for solving the contaminant source characterization problem.

3.3.1 Stochastic Genetic Algorithm approach

Stochastic GA is similar to Standard GA that is applied individually to each demand realization. In this study, a set of demand realizations are generated using both the random models; Gaussian model and Autoregressive model separately. The source identification problem is solved by varying the nodal demands (using one demand realization at a time) and standard GA. This is repeated for all the demand realizations separately. Figure 3.5 illustrates the stochastic GA approach.

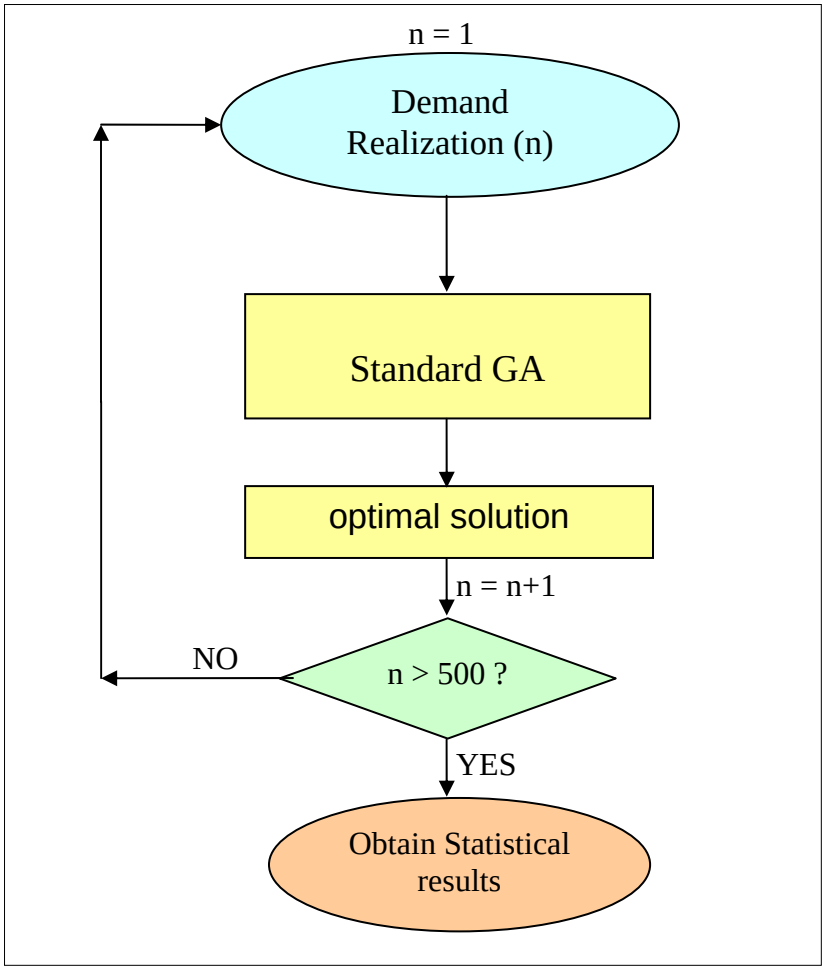


Figure 3. 5 Stochastic GA procedure

The standard GA procedure for one such evaluation is explained as follows:

The first step involves creating initial population of individuals using a real encoding. The population size depends on the size of the problem. The population of individuals is generated randomly, covering the entire range of possible solutions (search space) and the random seeds of the genetic algorithm are fixed for each trial. The fitness of each individual in the population is evaluated based on the objective function and constraints. The individuals then are mapped to a ranking fitness value in which the individuals are sorted linearly in order of decreasing evaluation value. Selection of the individuals which continue onto the mating pool is done by the tournament selection method. Tournament selection randomly chooses two individuals from the population and selects the individual with higher fitness value. This selection process is repeated until the desired number of replacement individuals has been chosen. The selected individuals known as parent individuals are then randomly paired up with a mating partner to perform crossover operation. Crossover is the exchange of a portion of bits between parent individuals. Uniform crossover is used in this problem in which the individual bits in an individual are compared between two parents.
The bits are then swapped with a fixed probability. The population size remains constant through the generations, or iterations by replacing the parent individuals with the new children individuals. If no crossover takes place between parent individuals, then these strings are copied to the new population. The final operator is mutation, where randomly selected bits within new population are changed with a given probability of mutation. Gaussian mutation is used for this problem. This process of selection, crossover and mutation is repeated over many generations until a stopping criterion is met. In this problem, the

demand values at the nodes are unchanged for all the generations in a single trial. This process is repeated for a number of trials to obtain a better estimate of the above approach. The overall procedure is illustrated in the figure 3.6

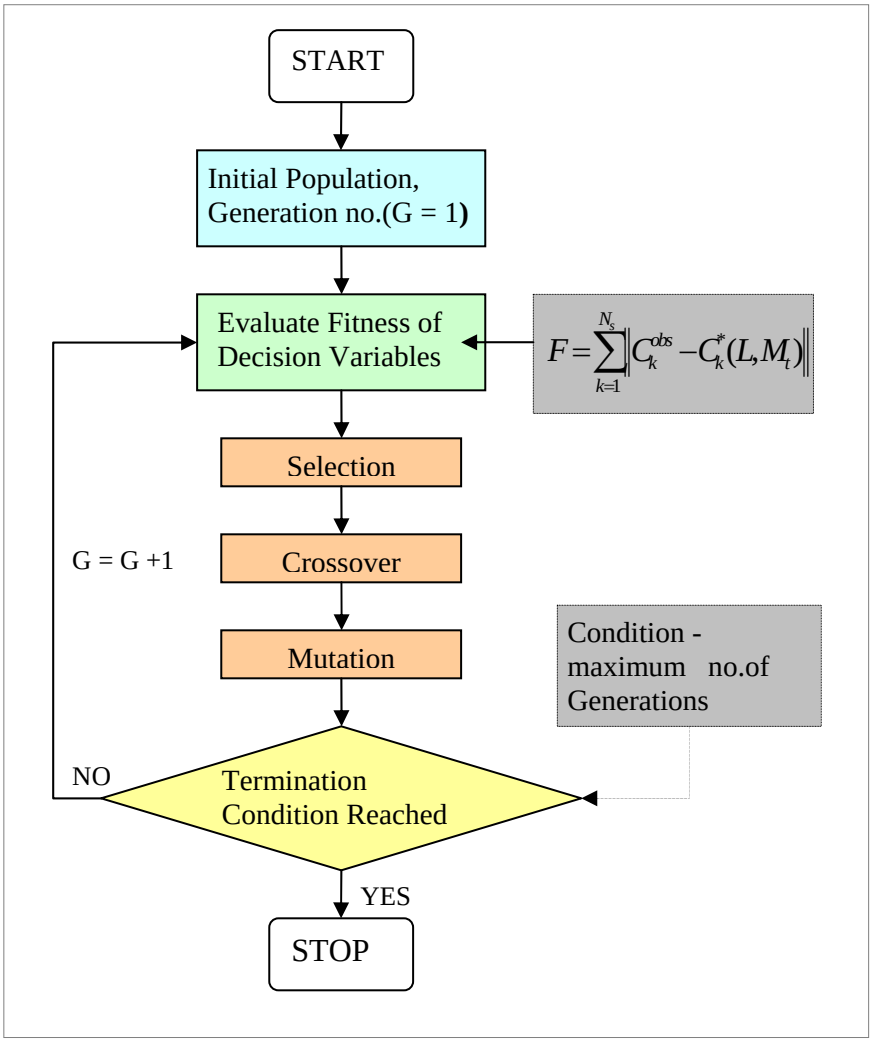


Figure 3. 6 Standard GA procedure

3.3.2 Noisy Genetic Algorithm approach

The noisy genetic algorithm approach is similar to standard genetic algorithm

approach, except that it differs in the fitness evaluation step. It was first developed by Miller and Goldberg (1997) and later applied to groundwater remediation problems by Smalley et al. (2000) and Gopalakrishnan et al. (2001). Aly and Peralta (1999) and Harrell(2001) used multiple realization genetic algorithms under parameter uncertainty to

ground water remediation and water shed management problems respectively. Hilton and Culvert (2005) developed a robust genetic algorithm under uncertainty of hydraulic conductivity values for Ground water remediation design problem.

In this noisy genetic algorithm, multiple realizations of the uncertain variable i.e. demand are created for each generation, and the same set of realizations are used for all the individuals in a given generation. The overall fitness of each individual is based on the average of fitness values from each of the multiple realizations. In other words, the differences of the assumed and observed concentration values are calculated for each realization and averaged over the multiple realizations to determine its overall fitness.

Mathematically, it can be expressed as follows:

$$F_{avg} = \frac{1}{K}\sum_{r=1}^{K} F_r$$

R = Total number of demand realizations in a generation.

F_{avg} = Overall fitness of the individual.

F_r = prediction error (fitness value) for r^{th} demand realization given by -

$$F_r = \sum_{k=1}^{N_s} \left\| C_k^{obs} - C_k^*(L, M_t) \right\|$$

L – Node ID of the contaminant source location.

$M(t_c)$ – Contaminant mass loading

t – Current time step

C_k^{obs} - Observed concentration

$C_k^*(L, M_t)$ - Simulated concentration

N_s - Number of sensors

K - observation (sensor) location

The Noisy genetic algorithm procedure is described in detail below.

Step 1. An initial population of individuals is created using a real encoding

Step 2. A set of multiple realizations of the uncertain variable, demand are created using AR model or Gaussian model at the beginning of each generation. Every individual in a generation uses the same set of demand realizations but different set of realizations are used in different generations. Where as, in the standard GA procedure the demand value remains the same for all the generations. The fitness of each individual is evaluated for all the demand realizations separately. The overall fitness of each individual in a generation is based on the average of fitness values from each of the multiple realizations. Figure 3.7 illustrates the steps involved in calculating the average fitness for an individual (*p*) in a given generation (J).The average fitness calculation process is repeated for all the individuals in a generation with the same set of demand realizations. This is analogous to performing a Monte-Carlo simulation using a small sample size in a fitness evaluation step of the genetic algorithm. Thus, the noisy genetic algorithm alters the optimization formulation such that the

objective function minimized, subjected to constraints on the concentration and the demand values. Eight realizations of the demand values were used in each GA generation in this problem.

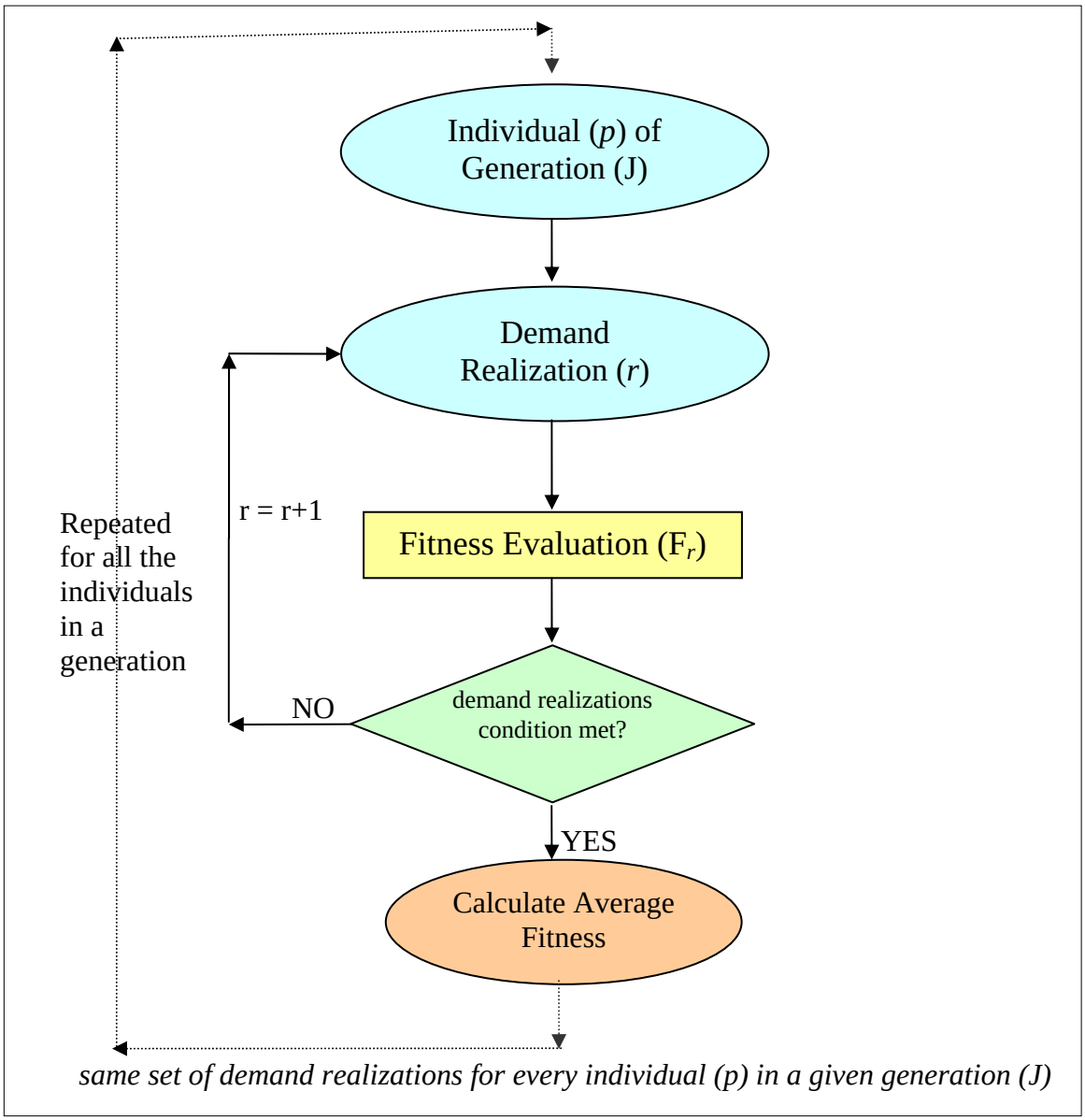


Figure 3. 7 Average fitness calculation flowchart for an individual in a generation of Noisy GA

Step 3. Selection of the mating pool is done through tournament selection, in which two individuals from the entire population are randomly paired off and compared, based on their fitness function values. Tournament selection is repeated until the number of parent individuals is selected.

Step 4. The next step is crossover, where children strings are produced through crossover between two parent individuals. Uniform crossover is used in this problem in which the individual bits in the string are compared between two parent individuals. The bits are then swapped with a fixed probability of 0.8.

Step 5. The final operator is mutation, where randomly selected bits within new population are changed with a given probability of mutation. Gaussian mutation is used for this problem.

Step 6. This generational process is repeated until a termination condition has been reached, such as fixed number of generations reached or a solution is found that satisfies minimum criteria.

The following chapter implements the methodology discussed with the help of a case study and the results are analyzed.

CHAPTER 4

4. CASE STUDY

4.1 Introduction

The methodology discussed in the previous chapter is applied to a water distribution network adapted from EPANET (Rossman, 2000) and is illustrated in figure 4.1. This example network is one of the problem instances available as a tutorial within EPANET. The nodes in the input file are re-numbered such that the nodes adjacent to each other are numbered consecutively. The network consists of one hundred and seventeen pipes, ninety two network nodes, three elevated storage reservoirs, two sources of water supply, and two constant speed pumps. The network data are reported in the EPANET users' manual. All the nodes in the network are assumed to supply water to the residential areas only.

A single non-reactive contaminant source is introduced at node 15 (figure 4.1) in order to generate a set of synthetic observations for an illustrative hypothetical contaminant event. The concentration of the contaminant source is assumed to be 3000 mg/L. A total of twelve sensors are arbitrarily placed in the entire water distribution network. Epanet is used for solving the hydraulic and transport equations. The hydraulics in the above network is simulated for every one hour over a 24 hour time period and it is assumed to be at steady state within each hour of the simulation.

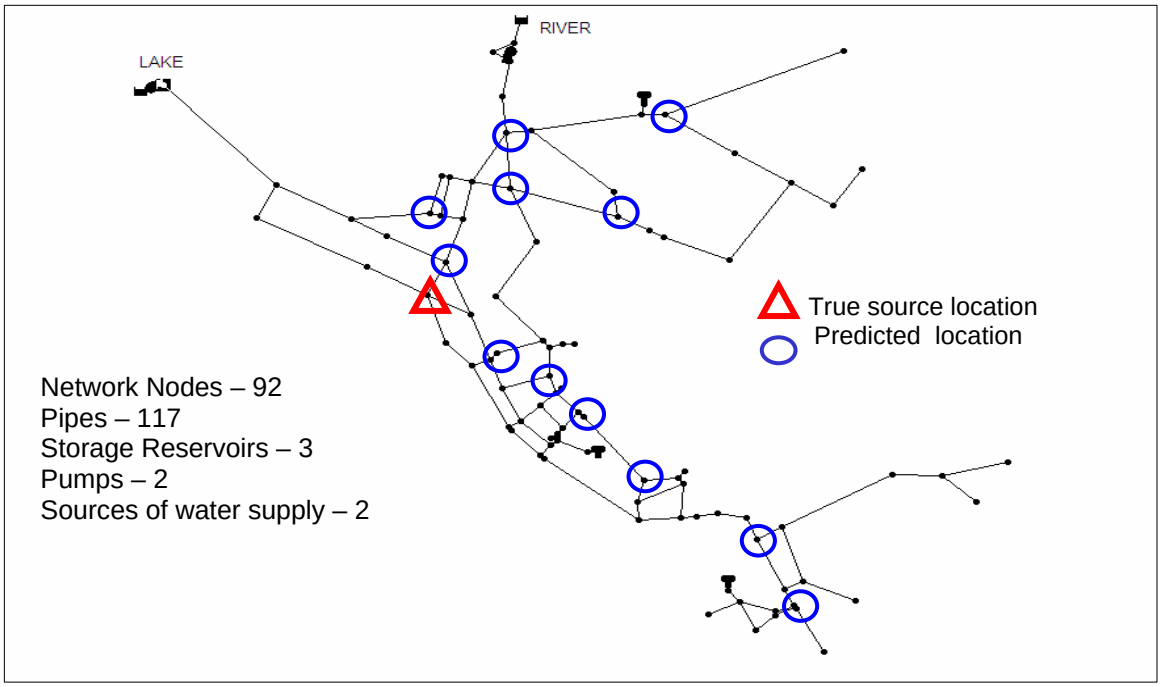


Figure 4. 1 Application – Water Distribution network example.

The daily mean total hourly demand pattern of the water distribution system is shown in the Figure 4.2. The hourly demands are obtained by multiplying the fixed base demand of the nodes with corresponding hourly demand patterns. The stochastic demand modeling methods, discussed in the previous chapter are used to generate hourly demand realizations of the entire network. Figures 4.3 and 4.4 illustrate the demand realizations generated from Gaussian model and AR model respectively. On comparison with Figure 2.1 in Chapter 2, the demand realizations resemble the typical residential diurnal demand curve. For each hourly hydraulic condition, the contaminant transport is simulated in 5-minute time intervals, and the concentration values at the sensors are observed at the end of each 5-minute increment.

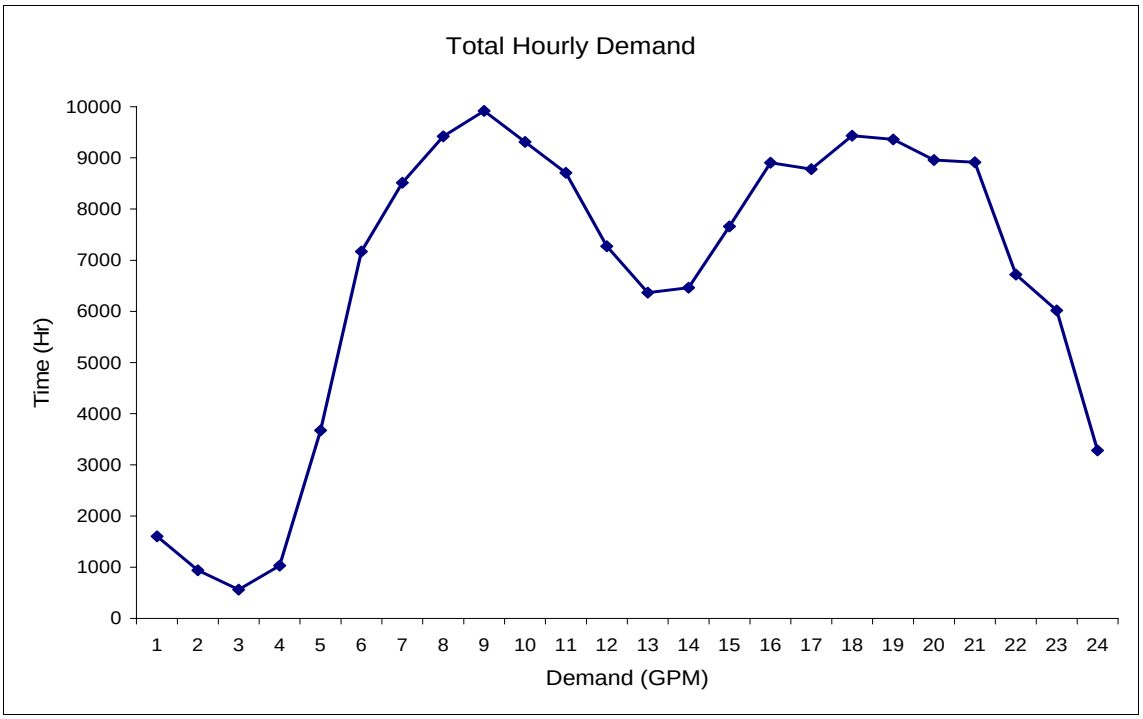


Figure 4. 2 Hourly Demand Pattern of the water distribution network

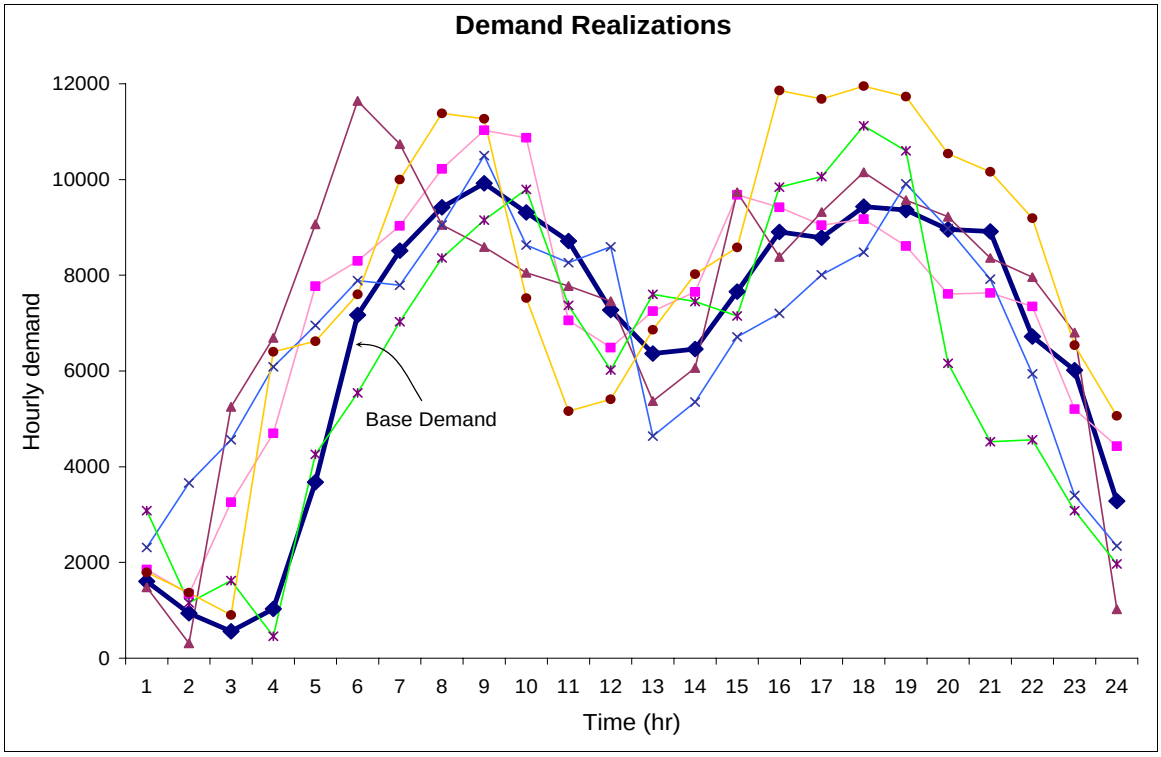


Figure 4. 3 Example Demand Realizations – Gaussian model

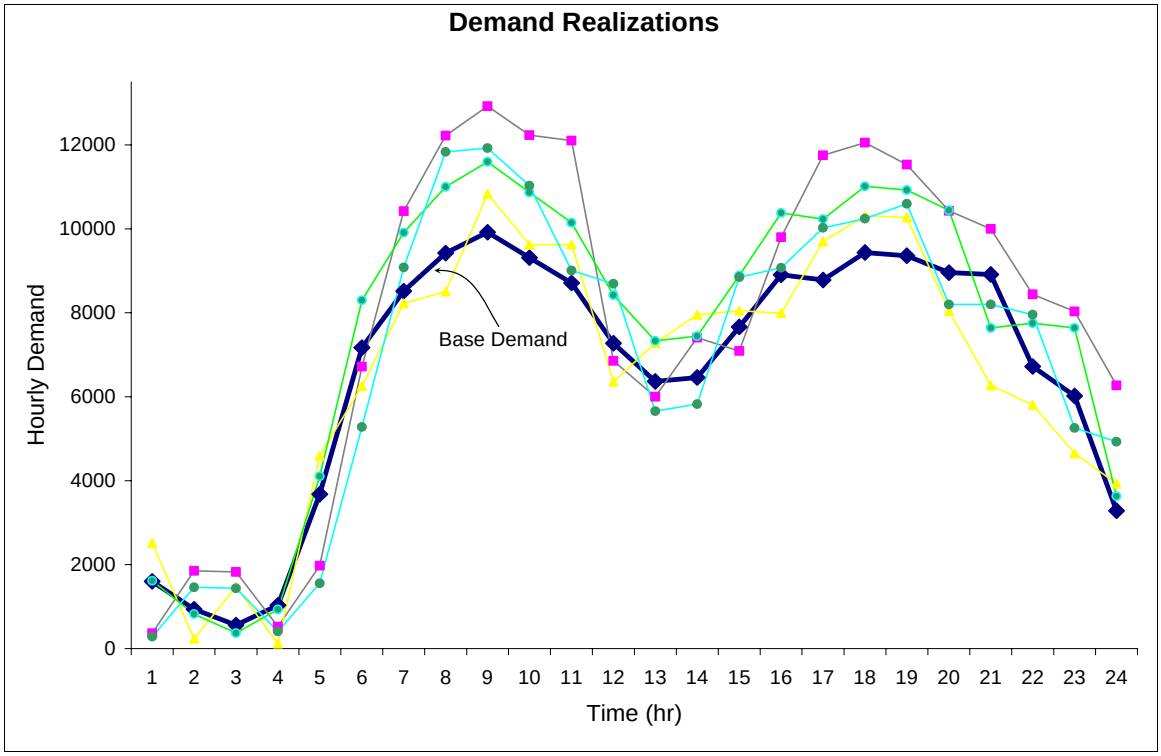


Figure 4. 4 Example Demand Realizations – AR model

The goal of the contaminant source identification problem is to predict the exact source location and concentration. The problem is solved using the Genetic Algorithm and Direct Search tool box provided in MATLAB. The stochastic GA parameters and Noisy GA parameters are listed in Tables 4.1 and 4.2 respectively.

Two random demand models and two optimization methodologies are discussed in the previous chapter. Therefore, the example network is analyzed using the four cases listed below:

Case 1: Gaussian model with Stochastic GA optimization method

Case 2: Gaussian model with Noisy GA optimization method.

Case 3: Autoregressive model with Stochastic GA optimization method.

Case 4: Autoregressive model with Noisy GA optimization method.

The following section discusses the results obtained by simulating the four cases to the example water network system.

Table 4.1 Summary of stochastic GA parameters

GA Parameter	**Value**
Population size	80
Generations	40
Tournament selection	2
Crossover	0.8
Mutation	Gaussian
No.of realizations	500
Contaminant source	
Location (node index)	15
Concentration (GPM)	3000

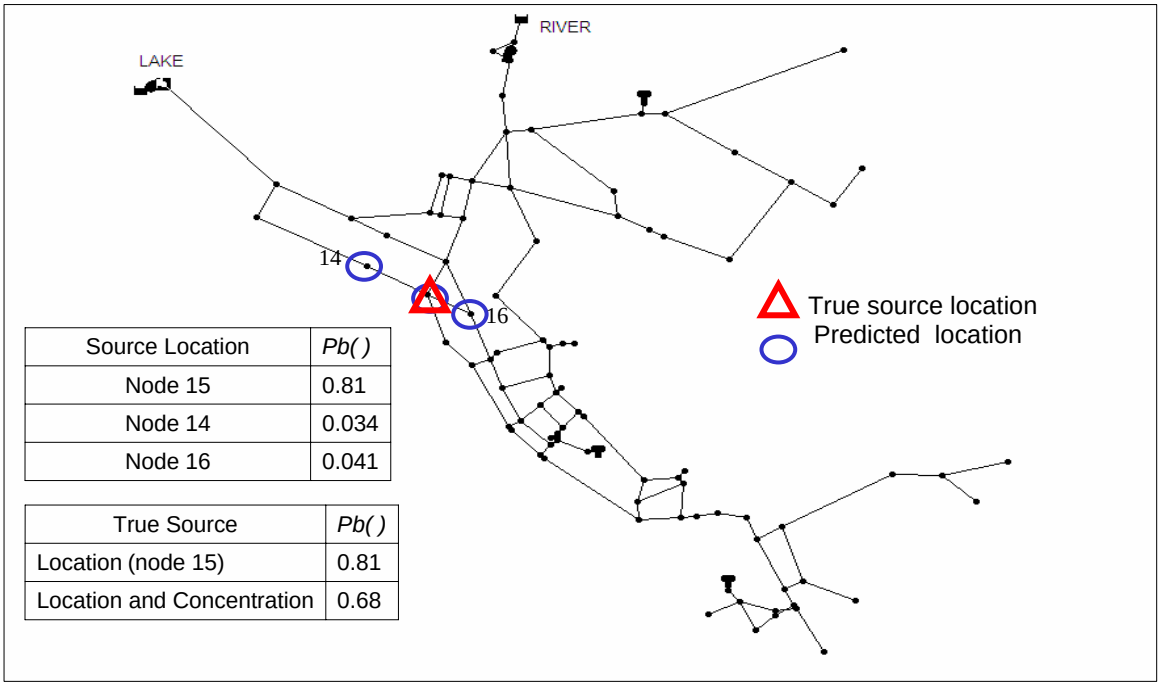
Table 4.2 Summary of Noisy GA parameters

GA Parameter	**Value**
Population size	80
Generations	40
Tournament selection	2
Crossover	0.8
Mutation	Gaussian
No.of realizations per Generation	8
Contaminant source	
Location (node index)	15
Concentration (GPM)	3000
Total No.of runs	100

4.2 Gaussian Model: results

4.2.1 Stochastic GA approach

The hourly demands are varied using Gaussian model and stochastic GA optimization approach is used to solve for the source identification problem. The problem is solved separately for a set of 500 different demand realizations and the results are shown in Figure 4.5. The concentrations are rounded to nearest thousand. For example if the predicted concentration is 2674.21, it is rounded to 3000. The top three probabilities of the nodes being the true source location are listed in figure 4.7. The highest probability of 0.81 corresponds to the true source location. This indicates that this approach identifies the true source location as node index 15 for 81% of the total 500 evaluations. The probability of finding the true source location (node index 15) and the concentration (i.e. 3000) is 0.68.



Source Location	*Pb( )*
Node 15	0.81
Node 14	0.034
Node 16	0.041

True Source	*Pb( )*
Location (node 15)	0.81
Location and Concentration	0.68

Figure 4. 5 Stochastic GA – Gaussian model results

4.2.2 Noisy GA approach

The demands for each hour are varied using the Gaussian model and Noisy GA optimization approach is employed to solve for the source identification problem. Multiple realizations of the uncertain variable i.e. demand, are created for each generation, and the same set of realizations are used for all the individuals in a given generation. A total of eight demand realizations are generated for each GA generation in this problem. The problem is solved for 100 such evaluations and the results are shown in Figure 4.6. The concentrations are rounded to the nearest thousand. The top three probabilities of nodes being the true source location are listed in figure 4.6. The node index with highest probability i.e. 0.87 is identified as the source location. The probability of predicting the true source location and the concentration (i.e. 3000) is 0.84.

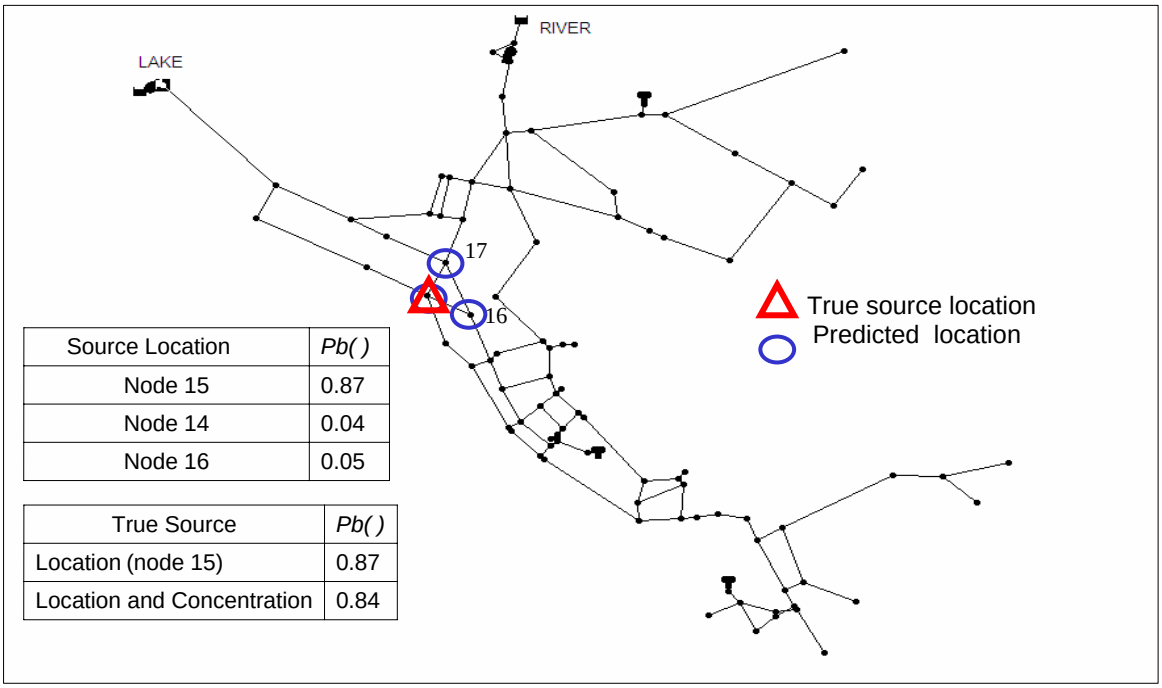
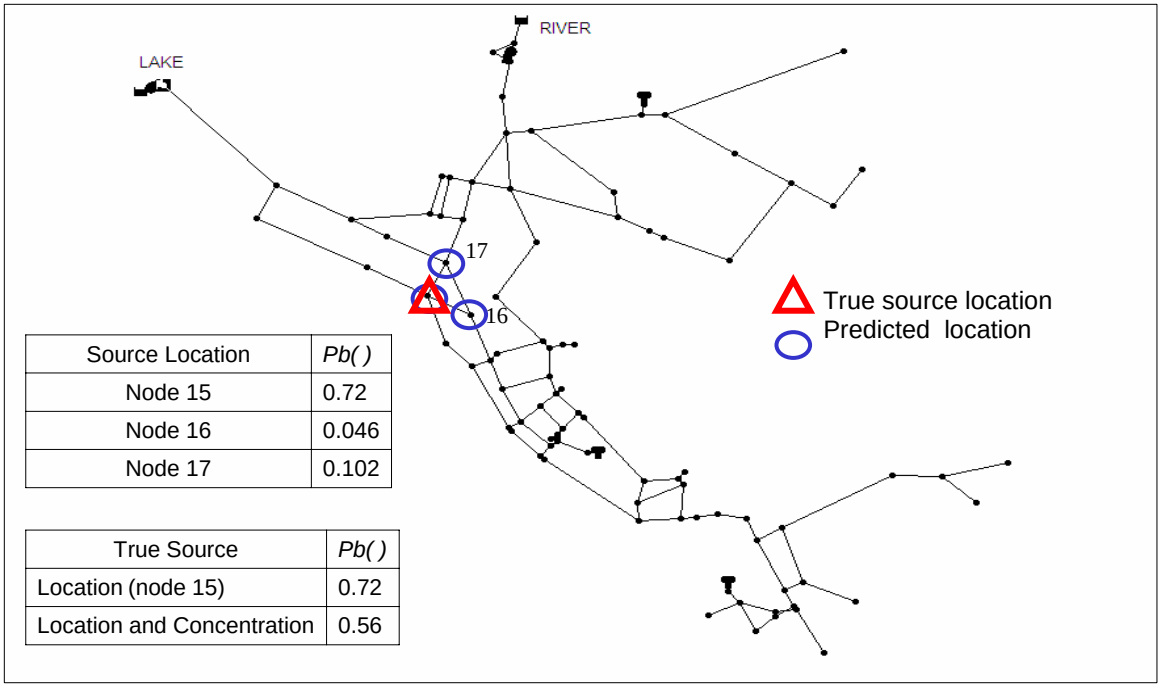


Figure 4. 6 Noisy GA – Gaussian model results

4.3 AR model: results

4.3.1 Stochastic GA approach

In this case, the hourly demands are varied using AR model and the Stochastic GA approach is used as optimization algorithm. The problem is solved for 500 different demand realizations and the results are shown in Figure 4.7. The node with highest probability represents the true source location. In this case, Node index 15 is identified as the true source location with probability 0.72. The rest of the node indexes have very less probability of being true source location. The probability of finding the true source location and the concentration (i.e. 3000) is 0.56.



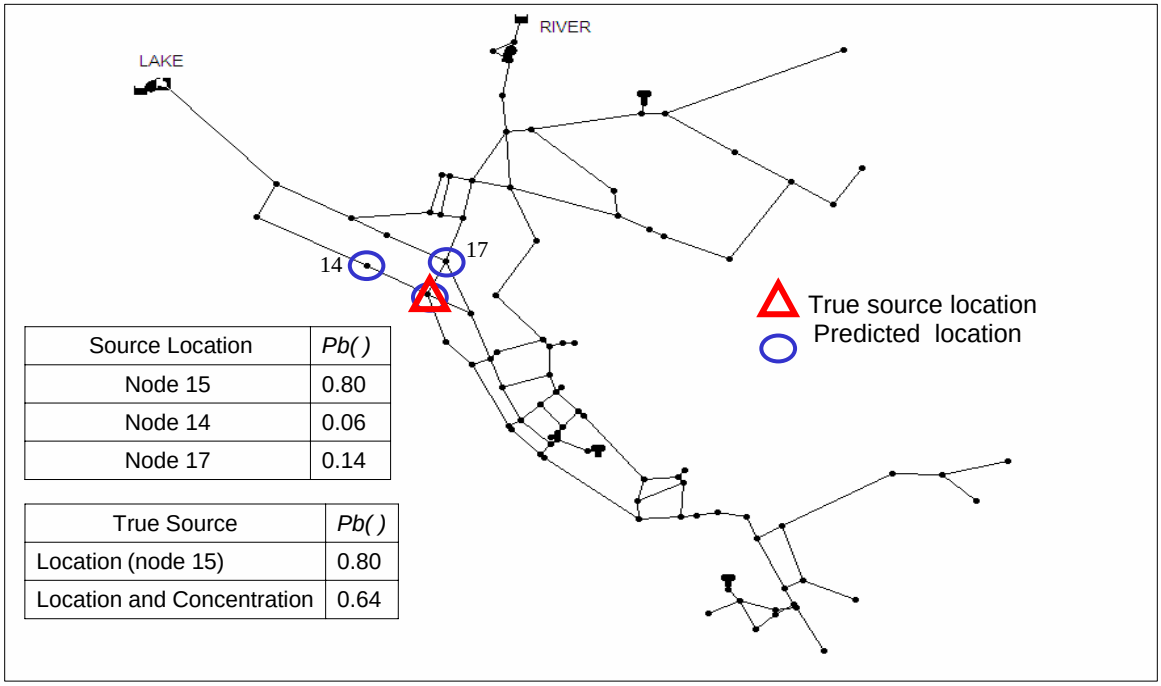
Source Location	*Pb( )*
Node 15	0.72
Node 16	0.046
Node 17	0.102

True Source	*Pb( )*
Location (node 15)	0.72
Location and Concentration	0.56

Figure 4. 7 Stochastic GA – AR model results

4.3.2 Noisy GA approach

The hourly demand parameter is varied using AR model and Noisy GA optimization approach is employed to solve problem. Multiple realizations of the uncertain variable i.e. demand are created for each generation, and the same set of realizations are used for all the individuals in a given generation. A total of eight demand realizations are generated for each GA generation in this case. The problem is solved for 100 such evaluations and the results are shown in Figure 4.8. Node 15 is the one with highest probability and it is identified as the true source location. The probability of finding the exact source location and true concentration (i.e. 3000) is 0.64.



Source Location	*Pb( )*
Node 15	0.80
Node 14	0.06
Node 17	0.14

True Source	*Pb( )*
Location (node 15)	0.80
Location and Concentration	0.64

Figure 4. 8 Noisy GA – AR model results

4.4 Discussion

The objective of the study, as stated in chapter 1, is to identify the true contaminant source location and the concentration by varying the hourly demands. A total of four cases are employed to solve the problem using simulation-optimization approach. Table 4.3 summarizes all the results obtained from four cases.

Table 4. 3 Summary of Results

Probabilities	**Gaussian model**		**AR model**	
	True source location	True source location and concentration	True source location	True source location and concentration
Stochastic GA	0.81	0.68	0.72	0.56
Noisy GA	0.87	0.84	0.80	0.64

The results indicate that both of the optimization approaches, stochastic GA and Noisy GA are able to predict the true source location with more than 50% probability. When Noisy GA is used instead of Stochastic GA, higher probabilities in predicting true source location are observed for both of the random demand models. The reason might be due to Noisy GA being more robust than stochastic GA. Also, the results show that Noisy GA approach predicts the true source location and concentration with higher probabilities than stochastic GA for both of the random demand models. 500 demand realizations are used in stochastic GA optimization approach. In Noisy GA optimization approach, only 8 different demand realizations are created for each generation and the same set of realizations are used for every

individual in a generation.

Gaussian and Autoregressive models are used to generate the demand realizations. The results indicate that Gaussian model has relatively higher probabilities in predicting the true source location and both location and concentration. This is because AR model is more complex model than Gaussian model. The demands generated using AR model are temporally correlated. On the other hand, in the Gaussian model, demands are neither spatially nor temporally uncorrelated and the demands are varied only by incorporating error term. AR model is used when demand data is available for calculating the AR parameters. This model is used when no demand data is available and one has to incorporate demand variability in source identification problem.

CHAPTER 5

5. CONCLUSION

5.1 Summary

Water distribution systems are vulnerable to deliberate biological or chemical contamination into the system and have adverse health and safety impacts on the consumers. On the other hand, the values of demands at consumer nodes in a water distribution system remain one of the major sources of uncertainty, due to their inherent variability in instantaneous water consumption levels. The above two problems are coupled together for this project and were studied. The study investigated the effects of demand variability for a source identification problem in water distribution networks. The first objective of this project is to develop a methodology to vary the consumer nodal demands temporally. Two random demand models: Gaussian model and AR model are studied and implemented in this project. The second objective is to solve the source identification problem under stochastic demand conditions. Optimization-Simulation method is employed to solve the problem and the final objective is to implement the methodology to an example network. The methodology is successfully implemented to the network provided in EPANET and the results are discussed below.

An illustrative water distribution was used to demonstrate the methodology. Four different scenarios were implemented on the example network using two random demand models and two optimization approaches. The results indicated that both Stochastic GA

and Noisy GA optimization approaches were able to locate the exact source location with more than 50% probabilities. It is observed that Noisy GA approach found the true source location and concentration with higher probabilities than Stochastic GA approach for all the scenarios. Moreover, the Noisy GA approach is less computationally expensive than Stochastic GA approach. Thus, the source identification problem is solved using simulation-optimization method under random demand conditions.

5.2 Future Work

The study presented here investigated the demand variability for a single contaminant source in the network. This can be made more complex by extending to multiple contaminant sources. Also, the demands are varied only temporally in this problem. Spatial variation of demands for can be incorporated along with temporal variation. Other uncertainty parameters such as error in measurements etc. can also be incorporated and studied.

6. LIST OF REFERENCES

[1]. Aly, A. A., and Peralta, R. C. (1999). "Optimal Design of Aquifer Cleanup Systems under Uncertainty using a Neural Network and a Genetic Algorithm." *Water Resources Research*, 35(8), 2523–2532.

[2]. Aral, M. M., Guan, J., and Maslia M. L. (2001). "Identification of Contaminant Source Location and Release History in Aquifers." *Journal of Hydrologic Engineering*, 6(3), 225-234.

[3]. Buchberger, S. G., and Wu, L. (1995). "Model for Instantaneous Residential Water Demands." *Journal of Hydraulic Engineering*, 121(3), 232–246.

[4]. Buchberger, S.G., Carter, J.T., Lee, Y., and Schade, T.G. (2003). "Random Demands, Travel Times and Water Quality in Deadends." *AWWA Research Foundation and American Water Works Association*, Denver,CO.

[5]. Buchberger, S. G., and Wells, G. J. (1996). "Intensity, Duration and Frequency of Residential Water Demands." *ASCE J. of Water Resources Planning and Management*, 122(1), 11-19.

[6]. Chan Hilton, A.B and Culver, T. B. (2005). "Groundwater Remediation Design under Uncertainty using Genetic Algorithms". *Journal of Water Resources Planning and Management*, 131(1), 25-34.

[7]. Clark, R. M., and Deininger, R. A. (2000). "Protecting the Nation's Critical Infrastructure: The vulnerability of U.S. water supply systems." *Journal of Contingency and Crisis Management*, 8(2), 73–80.

[8]. Cristo, C. D., and Leopardi A. (2006). "Uncertainty Effects on Pollution Source Location in Water Networks." Water Distribution Systems Analysis Symposium, Cincinnatti, OH, August 2006.

[9]. Gargano, R. and Pianese, D. (2000). "Reliability as Tool for Hydraulic Network Planning."*Journal of Hydraulic Engineering*, 126(5), 354-364.

[10]. Goldberg, D. E., Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, New York, NY, 1989.

[11]. Goldberg, D. E. and Deb, K. (1991). "A comparison of selection schemes used in genetic algorithms." *Foundations of Genetic Algorithms I*, 69-93.

[12]. Guan, J., Aral, M. M., Maslia, M. L., Maslia, and Grayman, W. M. (2006). "Identification of Contaminant Sources in Water Distribution Systems Using Simulation–Optimization Method: Case Study", *Journal of Water Resources Planning and Management*, 132(4) 252-262.

[13]. Haestad Methods., "Advanced Water Distribution Modeling and Management". Haestad Press, Waterbury, Connecticut (2003).

[14]. Harrell, L. J. (2001). "Evolutionary algorithm-based design of a system of wet detention basins under uncertainty for watershed management." *Proceedings of World Water and Environmental Resources Congress*, Reston, VA.

[15]. Holland, J. H., Adaptation in Natural and Artificial Systems, University of Michigan, Ann Arbor, Michigan, 1975.

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

[17]. Liu, L., Zechman, E. M., Brill, E. D. Mahinthakumar, G., Ranjithan, S. and Uber ,J (2006)."Adaptive Contamination Source Identification in Water Distribution Systems Using an Evolutionary Algorithm-based Dynamic Optimization Procedure." Water Distribution Systems Analysis Symposium, Cincinnatti, OH.

[18]. Miller B. L, *Noise, Sampling and Efficient Genetic Algorithms*, IlliGAL Report No.97001, May 1997.

[19]. Murray, R., Janke, R., and Uber, J. (2004). “The Threat Ensemble Vulnerability Assessment (TEVA) program for drinking water distribution system security.” *Proceedings of EWRI Congress*, Salt Lake City,UT.

[20]. Nilsson, K., 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.

[21]. Gopalakrishnan, Minsker, B. S., and Goldberg D. E. (2001).“Optimal Sampling in a Noisy Genetic Algorithm for Risk-based Remediation Design.” World Water and Environmental Resource Congress, Orlando, FL.

[22]. Lansley, K., E., and Mays, L., W. (1989). “Optimization model for Water Distribution System Design.” *Journal of Hydraulic Engineering*, 115(10), 1401-1418.

[23]. 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.

[24]. Preis, A., and Ostfeld, A. (2006). “Contamination Source Identification in Water Systems: A Hybrid Model Trees-Linear Programming Scheme.” *Journal of Water Resources Planning and Management*, 132(4), 263-273.

[25]. Propato, M. (2006). “Contamination Warning in Water Networks: General Mixed-Integer Linear Models for Sensor Location Design.” *Journal of Water Resources Planning and Management*, 132(4), 225-233.

[26]. Rossman, L. A. (2000). *EPANET 2 users manual*, National Risk Management Research Laboratory, U.S. EPA, Cincinnati.

[27]. Rossman, L.A. (1999). The EPANET programmer’s toolkit. In Proceedings of Water Resources Planning and Management Division Annual Specialty Conference, ASCE, Tempe, AZ.

[28]. Smalley, J. B., Minsker, B. S., and Goldberg D. E. (2000). “Risk-based in situ bioremediation design using a noisy genetic algorithm.” *Water Resources Research*, 36(20), 3043-3052.

[29]. Walski, T.M., Chase, D.V., Savic, D.A., Grayman, W.M., Bickwith, S., and Koelle, E. (2003). *Advanced Water Distribution Modeling and Management*. Haestad Press, CT.

[30]. MATLAB Genetic Algorithm and Direct Search Toolbox user guide. Webpage. http://www.mathworks.com/access/helpdesk/help/toolbox/gads/