Spring 1-1-2013

Adjoint-Based Probabilistic Method for Source Identification in Water Distribution Systems

David Erich Wagner
University of Colorado at Boulder, dewagnerusaf@gmail.com

Adjoint-Based Probabilistic Method for Source Identification in Water Distribution Systems

by

DAVID ERICH WAGNER

B.S., Bucknell University, 2000

M.S., Air Force Institute of Technology, 2006

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

Of the requirement for the degree of

Doctor of Philosophy

Department of Civil, Environmental and Architectural Engineering

2013

This thesis entitled:
Adjoint-Based Probabilistic Method for Source Identification in Water Distribution Systems
written by David Erich Wagner
has been approved for the Department of Civil, Environmental and Architectural Engineering

Roseanna M. Neupauer

R. Scott Summers

JoAnn Silverstein

Angela R. Bielefeldt

Vanja Dukic

Date

The final copy of this thesis has been examined by the signatories, and we
find that both the content and the form meet acceptable presentation standards
of scholarly work in the above mentioned discipline.

Wagner, David Erich (Ph.D., Department of Civil, Environmental, and Architectural Engineering)

Adjoint-Based Probabilistic Method for Source Identification in Water Distribution Systems

Thesis directed by Associate Professor Roseanna M. Neupauer

The events of September 11, 2001 have increased the focus on protecting utilities and infrastructure from acts of terrorism. For water utilities, this increased focus has led to researching more efficient and effective methods for finding the source of contamination in the event of contaminant intrusion. Better source identification can significantly reduce both the population affected by water contamination (with subsequent loss of service) and the resources required to mitigate the spread of contamination. Water contamination and/or loss of service have a clear impact on public welfare (both physical and psychological), whether it is due to terrorist activity or accidental contamination.

Source identification can be accomplished using system observations (i.e. the location, time, and magnitude of contamination in the system) and modeling software, such as EPANET. We develop an adjoint-based probabilistic method which uses the system observations as the input information and propagates the information in a backward simulation to determine all potential contamination node and release time scenarios for a system observation. By using multiple system observations and conditioning the results using the system uncertainty and the potential range of source masses, we probabilistically determine the true source node and contamination time.

We develop and test the adjoint-based probabilistic method for source identification in water distribution systems with pipes, nodes, tanks, and pumps; steady and transient flows; perfect and imperfect sensors; and complete and incomplete mixing at the nodes.

Acknowledgements

I am grateful to:

Roseanna Neupauer, for being my advisor. In addition to her vast knowledge of the adjoint method, she helped me to become a better researcher, technical writer, and presenter. For all of this, I will be eternally grateful.

Scott Summer, JoAnn Silverstein, Angela Bielefeldt, and Vanja Dukic for being member of my committee. Thank you for constantly opening my ideas to how my work might or might not work in a true water distribution system.

Arnold Strasser for helping me to understand how a true water distribution system works.

Mark DeMay for teaching me how to work with and manipulate some of the computer programs involved in my research.

Cody Cichowitz for laying some of the groundwork for using the adjoint-based probabilistic method in a water distribution system.

Last, but not least, my wife and two sons for giving me the freedom to work on this research at all hours of the day and night, and continuing to love and support me through the entire adventure.

My work was funded by the United States Air Force through the Air Force Institute of Technology Civilian Institutes program.

CONTENTS

TABLES

Table

FIGURES

Figure

CHAPTER 1

INTRODUCTION

1.1. Problem Statement

Water utilities are tasked with providing an uninterrupted supply of potable drinking water to their service populations. The events of September 11, 2001 have increased the focus on protecting utilities and infrastructure from acts of terrorism (DHS 2003). A malicious attack on water utilities could lead to severe health and environmental impacts, as well as loss of public confidence in the security of public water supplies (NRC 2007). Government agencies such as the EPA and water utilities are searching for better methods for preventing, detecting, and remediating contamination in drinking water distribution systems. While prevention is clearly the most effective measure against contamination, human ingenuity seems to know no bounds and impregnable water distribution systems do not exist. Given the possibility of contamination exists, it is important to develop methods to determine the source of contamination. An efficient and effective method can significantly reduce both the population affected by water contamination (with subsequent loss of service) and the resources required to truncate the spread of contamination.

1.2. Background

1.2.1. Inverse Methods for Source Identification

In the past, inverse methods have been used in conjunction with software, such as EPANET, to identify sources of contamination (e.g., Islam et al. 1997; Laird et al. 2005, 2006; Guan et al. 2006; Preis and Ostfeld 2006, 2007). Inverse methods employ advanced algorithms to determine the source of contamination. Using system observations (e.g., contaminant concentration arriving at a sensor) as simulation goals, the algorithms in conjunction with modeling software test multiple contamination scenarios (e.g., source node, release time, and source concentration) and run

multiple forward simulations in the water distribution system to determine what contamination scenario leads to the observed outcomes.

Islam et al. (1997) did not use EPANET, but they developed an inverse method to calculate the chlorine concentration needed at a source node to maintain appropriate chlorine residual concentrations throughout a water distribution system. Their method used the advection-dominated transport equation in one dimension with first-order decay. They discretized the transport equations using the four-point implicit finite-difference scheme solved simultaneously together with the complete mixing equation at the nodes. They tested their method in a system similar to EPANET Example 1 (Rossman 2000) and found that they were able to successfully replicate the results from a forward simulation.

Laird et al. (2005) developed a method for determining the source node and release time based on the location, concentration, and time contaminant was observed in the system. Their approach used non-linear programming and an origin tracking algorithm; their method did not need to discretize along the length of the pipe. They demonstrated how their method could characterize the source and time of contamination in a more complex network containing 469 nodes and 4 storage tanks.

Laird et al. (2006) developed a two-step approach to source identification. First they identified the potentials source nodes and then they used these nodes to test the potential contaminant injection scenarios using a mixed-integer quadratic program. They demonstrated how their approach was able to successfully determine the source of contamination in a water distribution system with approximately 400 nodes.

Guan et al. (2006) refined the source characterization process by using an optimization algorithm. Given contaminant observations in a water distribution system, Guan et al. (2006) used EPANET to simulate a potential contamination event at a potential source node. Based on this

simulated event, they used an algorithm and the simulated observations to determine how the contaminant mass moves throughout the system. They compared the resulting observations against the observations from the true contamination event. The optimization algorithm selected a new contamination scenario based on the results and continued the process until the difference between the simulated event and the true contamination event converged. They were able to correctly identify the source and time of contamination in the Dover Township model which has nearly 15,000 nodes and 9 storage tanks.

Preis and Ostfeld (2007) coupled a genetic algorithm with EPANET simulations to determine the source of contamination. The genetic algorithm was used to diminish the difference between the simulated results and the observed results. Using this method, they were able to determine the source characteristics in three different water distribution systems: Anytown USA (17 nodes/2 storage tanks), EPANET Example 3 (94 nodes/3 tanks), and the Richmond Water Distribution System, Yorkshire, UK (865 nodes/6 tanks). While they were able to replicate the observed results, they concluded that the genetic algorithm makes the method computationally intensive.

Research regarding inverse methods primarily focuses on developing a more efficient and accurate method for determining the potential source nodes (e.g., Laird et al. 2005; Propato et al. 2010) and the contamination scenario (i.e., the source node, release time, and source concentration; e.g., Laird et al. 2005; Guan et al. 2006; Preis and Ostfeld 2007; Vankayala et al. 2009; Zechman and Ranjithan 2009; Propato et al. 2010; Liu et al. 2011). While some inverse methods consider the entire population of nodes as potential source nodes (e.g., Preis and Ostfeld 2007), some researchers have developed methods for narrowing down the potential source nodes to a subset of the entire population. Laird et al. (2005) using an origin tracking algorithm to find hydraulically connected upstream nodes, while Propato et al. (2010) used observations of no contamination to rule out

potential source nodes using linear algebra. Once the set of potential source nodes is determined, the inverse methods depend on various methods for calculating the source characteristics. Laird et al. (2005) used non-linear programming to calculate the contamination scenario. Guan et al. (2006) used an optimization algorithm which was able to iteratively decrease the difference between the observed and simulated results, while Preis and Ostfeld (2007), Vankayala et al. (2009), and Zechman and Ranjithan (2009) used genetic algorithms for the same purpose. Propato et al. (2010) used the minimum relative entropy method to determine the source characteristics. Liu et al. (2011) developed a method that is able to provide real-time response using an evolutionary algorithm for source characterization.

Some researchers have also explored creating a model tree with pre-developed scenarios which are referenced if a contamination event occurs; the system observations are compared to the simulation results to determine which scenario is most likely (e.g., Preis and Ostfeld 2006; Shen et al. 2009). Preis and Ostfeld (2006) and Shen et al. (2009) constructed a model trees by simulating potential contamination scenarios and then storing the results. Once a sufficient number of scenarios are run, the model tree can be used in place of EPANET to predict what contamination scenario could have led to the contaminant observations. Preis and Ostfeld (2006) had 10,000 results in their model tree, thus the procedures for finding the most similar scenario are important. Preis and Ostfeld (2006) used linear programming, while Shen et al. (2009) used a data mining technique.

Research in inverse methods has also explored using observations from non-ideal sensors. Researchers have used data from fuzzy sensors (Preis and Ostfeld 2008) and binary sensors (e.g., Preis and Ostfeld 2008; Kumar et al. 2012) to determine the source of contamination using an inverse method. Preis and Ostfeld (2008) were able to determine the true contamination scenario, but found that using binary sensor data led to a much higher number of possible contamination

nodes than fuzzy sensors or perfect sensors; in one case, the number of possible source nodes using binary sensors is nearly four times as many as determined using fuzzy sensor (124 for binary; 32 for fuzzy; 19 for perfect; 10 actual source nodes). Kumar et al. (2012) found similar results; an inverse correlation exists between the information provided by the sensors and the number of possible solutions calculated by the method.

Overall, while inverse methods are becoming increasingly more efficient, these methods have an inherent inefficiency requiring that multiple forward simulations need to be run in order to determine the true contaminant source characteristics (Liu et al. 2012).

1.2.2. Adjoint Methods for Source Identification

A second class of source identification methods is the adjoint method (e.g., Neupauer et al. 2010). The adjoint method is a tool for directly calculating the sensitivity of a system state at a particular location and time to a system parameter; this sensitivity is called the adjoint state and is the state variable of the adjoint of the forward contaminant transport equations. For source identification, the system state is the concentration observed at a sensor and the system parameter is the mass released at the source. As discussed previously, inverse methods complete multiple forward simulations attempting to replicate the conditions which produced the sensor observations. However, in the source identification problem, the concentration is known and the source characteristics are not. The adjoint approach is particularly suited to solve this type of problem (Sun, 1994). The adjoint method uses the sensor observations as sources of the adjoint state, which is propagated upgradient through a water distribution system to all possible source nodes. The adjoint state is used to quantify the probability that the observed contamination could have been released from a particular potential source node. Using this adjoint state, one simulation is run for each sensor observation and provides information about all possible source scenarios; while the

standard inverse methods require one simulation for each possible source scenario and provides concentrations for all nodes in the system.

Adjoint methods have been successfully used for source identification in groundwater (Neupauer and Wilson 1999), but source identification in water distribution systems is more complicated as a result of multiple flowpaths through the system, transient flows that can lead to flow reversals in individual pipes, and the presence of storage tanks that can hold some water and contaminant while the remainder of the water and contaminant flows through the system. Prior to this work, the application of adjoint methods in water distribution systems has been limited to relatively simple systems consisting entirely of pipes and nodes (i.e., no tanks or pumps) under steady-state flow conditions (Neupauer et al. 2010). Further research is necessary to determine how the method can be modified for use in more realistic water distribution systems.

Neupauer et al. (2010) developed the adjoint method to probabilistically determine the source of contamination in simple water distribution systems. Their method uses travel time probability density functions (PDFs) to identify the most likely source of contamination in a water distribution system. These PDFs are related to adjoint states of concentration. The adjoint states are obtained by solving adjoints of forward equations.

1.2.3. Water Distribution System Modeling Software

The ability for water utilities to secure, monitor, and model water distribution systems is constantly evolving. Supervisory control and data acquisition (SCADA) systems allow utilities to monitor the system remotely and adjust parameters as needed, while modeling software allows the water system managers to explore the potential effects of system changes. The usefulness of these software packages is highly dependent upon accurate measurement and input of system parameters (e.g., flow rates, pipe characteristics, and reaction kinetics) and dynamics. While the software cannot

precisely depict how an actual system reacts, it can be used to approximate the probable outcomes of system changes.

EPANET (Rossman, 2000) is often used to simulate the movement of contaminant mass for source identification methods. This software uses forward transport equations to model hydraulic and water quality behavior in systems containing pipes, junctions, storage tanks, and pumps. By adjusting demands and supplies at the nodes, steady-state or transient flow fields, the tank mixing model, the valve type, and pump curves, EPANET can be used to reproduce the true system as closely as possible, however it cannot simulate all potential scenarios. EPANET 2.0 uses the complete mixing model at pipe junctions: the water from the input pipes intersect, mix completely, and form a homogeneous solution that flows downstream through each of the output pipes. Many researchers have demonstrated that this is an inadequate representation of the true mixing at pipe junctions leading to inaccuracies in the simulations (e.g. Romero-Gomez and Choi, 2008).

EPANET-BAM (Sandia National Lab 2008) is an extension for EPANET which can be used to simulate the movement of contaminant mass through junctions with incomplete mixing. In EPANET-BAM, the user is able to specify the mixing at any node meeting the following requirements: (1) equal-sized pipes and (2) two adjacent inflows, and (3) two adjacent outflows (Ho and Khalsa 2007). EPANET-BAM uses the bulk advective mixing model where the concentration exiting the junction depends on the flowrates in the pipes and the concentrations in the two pipes exiting the junction are not the same. EPANET-BAM allows the user to specify the mixing at the junctions by specifying the proportion of the mixing which is bulk-advective mixing versus complete mixing; however it is limited to the junctions that meet the specified criteria.

In practice, the degree of mixing depends on the Reynolds numbers for flow in the inlet pipes and on the geometry of the junction (Austin et al. 2008). Choi et al. (2008) developed AZRED, a computer program that simulates incomplete mixing at (1) cross-junctions with adjacent

inflows and outflows or opposing flows; (2) double tee junctions with various inflow/outflow configurations, and (3) one tee and one wye junction. In each of these types of junctions, the mixing is simulated using experimental results based on the Reynolds numbers for the inflow and outflow pipes. These options increase the ability to simulate the true water distribution system behaviors.

1.3. Hypotheses

Adjoint methods have already been successfully employed in groundwater systems (e.g., Neupauer and Wilson 1999) and in simple water distribution systems (e.g., Neupauer et al. 2010). The goal of this research is to develop an adjoint method for source identification in more realistic water distribution systems (i.e., including tanks, pumps, and transient flow fields) under more realistic conditions (i.e., non-ideal sensors and realistic transport). This goal leads to the following three hypotheses:

H1. The adjoint method can be used to determine the source of contamination in water distribution systems containing pumps, storage tanks, and transient flow conditions. Neupauer et al. (2010) demonstrated that the adjoint method can identify sources of contamination in simple water distribution systems (i.e., systems containing only source, pipes, and nodes) under steady-state flow conditions. This research develops an adjoint method that can be used to identify sources of contamination in more complex water distribution systems (i.e., systems containing pumps and storage tanks) under transient flow conditions.

H2. The adjoint method can be used to determine the source of contamination in water distribution systems when using realistic system sensors. All previous work on the adjoint method for source identification (both in groundwater and water distribution systems) assumed that the contaminant concentrations come from perfect sensors that provide exact concentration readings. Perfect sensors do not exist (Preis and Ostfeld 2008), so we develop a method for using data from

fuzzy sensors which only identify the approximate range of contamination (e.g., high, medium, low) and binary sensors which only identify whether contamination is present or absent (or whether it is above or below a designated threshold value). We also develop a method to use non-detect measurements, i.e., sensor measurements for which the level of contamination is below the limit of detection.

H3. The adjoint method can be used to determine the source of contamination in water distribution systems with non-ideal mixing at pipe junctions. Previous work has assumed complete mixing at pipe junctions. This is not the case in an actual distribution system leading to discrepancies between the modeled system and the true system (Austin et al. 2008). A bulk advective mixing algorithm (EPANET-BAM) has been developed to introduce more realistic mixing at pipe junctions (Ho and O'Rear 2009). We develop adjoint theory incorporating incomplete mixing at the pipe junctions.

1.3.1. Pumps, Storage Tanks, and Transient Flow Conditions

Many studies have been conducted using inverse methods to identify the source of contamination in water distribution systems with pumps, storage tanks, and transient flow conditions; one of the most widely used example networks for testing source identification methods is Example 3 in EPANET (Rossman 2000) which has a transient flow field, 3 tanks, and 2 pumps (e.g., Preis and Ostfeld 2007; Vankayala 2009; Zechman and Ranjithan 2009; Liu et al. 2012). Since inverse methods attempt to replicate system observations by running forward simulations of possible contamination events, any event that can be simulated in EPANET or other modeling software can theoretically be identified by inverse methods. Previous adjoint methods were developed for systems containing only pipes and pipe junctions under steady-state flow conditions.

In the adjoint method the adjoint state is propagated backward relative to the flow field. In a steady flow system, the flow rates do not depend on the system state (i.e., pressure), so the flow reversal can be accomplished manually by changing the positive demands to negative. This approach was used by Neupauer et al. (2010). In systems with storage tanks, the flow rate into or out of the tank depends not only on the system demands, but also on the water level in the tank and the pressure in adjacent nodes. Therefore, the flow rates in the system are dependent on the system state, and changing the signs on the demands is not equivalent to reversing the flow. We develop an approach to reverse flows and we use this approach in Chapter 2 to develop and test the adjoint method for complex water distribution systems containing pumps, storage tanks, and transient flow conditions.

1.3.2. Realistic System Sensors

The current adjoint method assumes the observed contaminant concentrations come from perfect sensors that provide exact concentration readings. Perfect sensors do not exist (Preis and Ostfeld 2008), so we develop a method for using more realistic sensors and data. We develop a method for using data from fuzzy sensors that measure concentration ranges of high, medium, and low; binary sensors which only provide presence/absence data (or above/below a threshold value); and non-detect measurements which can be used to determine nodes/times where contaminant could not have entered the system.

Since fuzzy sensors use a measurement range (e.g., 5-10 mg/L for a "medium" reading) instead of a measured concentration, we modify the probability conditioning step in the adjoint method to use a range, rather than a single measurement. We assume that the true concentration is equally likely to fall anywhere within the measured concentration range and we integrate over the entire measurement range.

Binary, or Boolean, sensors are only able to provide information regarding whether contamination is present or not (Preis and Ostfeld 2008). As with the fuzzy sensors, the binary sensor data will impact the calculation of the PDF of obtaining the measured concentration. In the case of binary sensors, we do not have a measured concentration; we only know that the concentration is above or below a threshold value. In our method, we focus on the observations indicating that contamination is present.

Non-detects are sensor measurements at contamination levels below the limits of detection. They can be used in the adjoint method to eliminate some nodes as potential sources. For instance, if contaminant observations at a sensor node lead to identifying that node X is a potential source node, a non-detect measurement could be backtracked through the system to show that contamination could not have been at node X (i.e., contaminant could not have been at a node that is hydraulically connected to a sensor node that did not see any contamination). The adjoint method uses the concentration observed at a node and propagates that information backward through the system. Previously, the method has been developed for positive concentration measurement; we develop a method for using non-detect measurements.

Using non-detect measurements presents two challenges: (1) Using the non-detect measurements in conjunction with the positive results and (2) The large number of non-detects. We answer the first challenge by using the non-detects to determine times when the contaminant could not have been at a potential source node. We then assume that the probability is 0 that the contaminant could have come from any node/time combinations that we calculated as connected to a non-detect measurement. The results can then be used directly with the positive results. The second challenge is important because the quantity of non-detects is so high. For example, if a sensor takes readings every 15 minutes for a span of 24 hours and never sees any contaminant concentration (i.e., all non-detects), 97 readings of non-detect would need to be evaluated. The

adjoint method requires one simulation for each observation, so using 97 observations would require 97 simulations, which is computationally intensive given the current method. To overcome this limitation, we treat consecutive non-detect measurements as a continuous source of the adjoint state (i.e., occurring over multiple timesteps) rather than multiple instantaneous sources. This allows the computation of all the individual PDFs in one backward simulation rather than multiple simulations; thus only one additional simulation is needed for each sensor node at which non-detects are observed.

1.3.3. Incomplete Mixing

We use EPANET-BAM to simulate the movement of contaminant mass in a system with incomplete mixing. We determine the adjoint equations for this type of mixing and develop an adjoint method which can be used in systems with incomplete mixing at the junctions.

1.4. Organization

This dissertation is organized into six chapters. Chapter 2 addresses hypothesis 1 and describes the adjoint method we developed for water distribution systems with transient flow fields and storage tanks. The results are presented for multiple scenarios including using 2, 3, or 4 observations, and using observations that do or do not contain information from contaminant mass that passed through a storage tank. Chapter 3 addresses hypothesis 2 and describes how the information from imperfect sensors (fuzzy or binary) is used in the adjoint method, and how non-detect measurements are used. The adjoint method is demonstrated using fuzzy sensor data with and without non-detect measurements and binary sensor data. Chapter 4 addresses hypothesis 3 and uses an incomplete mixing model at junctions to simulate the spread of contamination. We develop the adjoint model for incomplete mixing and develop a method for adjoint-based source

identification in water distribution systems with incomplete mixing at the junctions. We present scenarios with various degrees of incomplete mixing and demonstrate how the adjoint method is able to determine the source of contamination. Chapter 5 presents a larger water distribution system with transient flow fields, two water sources, and three storage tanks. We use this system to test the robustness of the adjoint methods developed in response to hypotheses 1 and 3. Chapters 2-4 are written as stand-alone papers which will be published in the *Journal of Water Resources Planning and Management* (Chapters 2 and 3) and the *Journal of Hydraulic Engineering* (Chapter 4).

CHAPTER 2

PROBABILISTIC SOURCE CHARACTERIZATION IN WATER DISTRIBUTION SYSTEMS WITH TRANSIENT FLOWS.

Abstract

Water system sensors are becoming increasingly more efficient and effective at discovering water quality changes in distribution systems. This paper describes a method for determining the source of the water quality change. We use publicly available software (EPANET) and a conditioning method to probabilistically locate the contamination source. Prior work has shown that a similar method is effective for a distribution system with pipes, nodes, and a steady-state flow field. In this work, we demonstrate the effectiveness of this approach in a more complex distribution system with a pump, a tank, and transient flow conditions.

2.1. Introduction

Water utilities are tasked with providing an uninterrupted supply of potable drinking water to their service populations. While this typically relates to accidental contamination, the events of September 11, 2001 shifted the focus on protecting water distribution systems from acts of terrorism (DHS 2003). Utilities are searching for better methods for preventing, detecting, and remediating contamination in drinking water distribution systems. While prevention is clearly the most effective measure against contamination, human ingenuity seems to know no bounds and the impregnable water distribution system does not exist. Given the possibility of contamination always exists, it is important to develop methods to determine the source of contamination. An efficient and effective method can significantly reduce both the population affected by water contamination (with subsequent loss of service) and the resources required to mitigate the spread of contamination.

Water contamination and/or loss of service have a clear impact on public welfare (both physical and psychological), whether it is due to terrorist activity or accidental contamination.

Contaminant detection is the initial step in contaminant source identification. The ability for water utilities to detect contamination in water distribution systems is constantly evolving. Supervisory control and data acquisition (SCADA) systems allow utilities to remotely monitor the system, while sensor technology is becoming increasingly more effective at detecting contamination. When contamination is found, modeling software packages, such as EPANET (Rossman 2000), can be used to simulate how the contaminant might move through the system and help the water utility personnel determine what portions of the system might be affected. The correlation between the simulations results and how the true system reacts is highly dependent upon accurate measurement and input of system parameters (e.g., flow rates, pipe characteristics, and reaction kinetics) and dynamics. While the software cannot precisely depict how an actual system reacts, it can be used to approximate the probable outcomes of a contaminant released in the system.

In the past, inverse methods have been used in conjunction with software, such as EPANET, to identify sources of contamination (e.g., Islam et al. 1997; Laird et al. 2005, 2006; Guan et al. 2006; Preis and Ostfeld 2006, 2007). The solution to a contaminant source identification problem is non-unique (e.g., De Sanctis, et al. 2010), so the goal of these methods is typically to select the true source node and contaminant time as a possible solution. Inverse methods employ advanced algorithms to achieve this goal. Using observations of contaminant concentration arriving at a sensor as simulation goals, the algorithms in conjunction with modeling software test multiple contamination scenarios (e.g., source node, release time, and source concentration) and run multiple forward simulations in the water distribution system to determine what contamination scenario leads to the observed outcomes. Research regarding inverse methods primarily focuses on developing a more efficient method for determining the potential source nodes (e.g., Laird et al. 2005; Propato et

al. 2010) and the contamination scenario (i.e., the source node, release time, and source concentration; e.g., Laird et al. 2005; Guan et al. 2006; Preis and Ostfeld 2007; Vankayala et al. 2009; Zechman and Ranjithan 2009; Propato et al. 2010; Liu et al. 2011). Some researchers have also explored creating a model tree with pre-developed scenarios which are referenced in the event of contamination; the system observations are compared to the simulation results to determine which scenario is most likely (e.g., Preis and Ostfeld 2006; Shen et al. 2009). Overall, while inverse methods are becoming increasingly more efficient, these methods have an inherent inefficiency requiring multiple forward simulations to determine the true contaminant source characteristics.

An alternative method for source identification is the adjoint method that has been developed by Neupauer et al. (2010) and Neupauer (2011) for water distribution systems. The adjoint method is based on the calculation of the adjoint state, which in this context is the sensitivity of the concentration at the observation node to a contaminant mass released at an upstream node. The adjoint state is backtracked through the system from locations where contamination was observed (e.g., sensors) to all possible source locations. The resulting distributions of adjoint states can be related to probabilities representing the random source node and the random release time of contamination from that source node. This probabilistic representation of the possible source characteristics is an additional benefit of the adjoint method that is not provided with all inverse methods. The adjoint method has been used for source identification in groundwater (e.g., Neupauer and Wilson 1999) and water distribution systems with simple hydraulics (Neupauer et al. 2010).

The goal of this paper is to develop an adjoint-based probabilistic method for source identification in water distribution systems containing storage tanks and transient flow conditions. The implementation of the adjoint method presented by Neupauer et al. (2010) cannot handle transient changes in storage, so we present a more robust implementation here. Specifically, the

adjoint method requires the direction of the pipe flows to be reversed relative to their actual flow direction to enable the model to backtrack the adjoint state from the observation nodes to the possible source nodes. Neupauer et al. (2010) reversed the flow directions in a simple, steady-state water distribution system manually by changing positive demands to negative demands; however this manual reversal of flows is not feasible for water distributions systems that contain tanks.

We used the EPANET Programmer's Toolkit to create a program to backtrack the adjoint state in a more complex, transient water distribution system with storage tanks. We also demonstrate how flowpaths through the storage tank dilute the information content of the sensor observations in identifying the source characteristics.

The approach that we developed to meet these goals and objectives includes the calculation of probability density functions (PDFs) of the possible contamination times for potential source nodes. These PDFs use observation data from one or multiple sensor nodes and allow us to identify the most likely source node and release time of contamination. Unlike inverse methods, however, we do not determine the release concentration of the contaminant.

In the next section, we present the theory of the adjoint method and probabilistic approach for source identification. Then, we provide examples using the adjoint method for source identification in a system similar to Example 1 from EPANET (Rossman, 2000). Finally, we offer some conclusions based on our results.

2.2. Theory

Forward transport of a conservative chemical in pipes can be modeled using

$$\frac{\partial C_i}{\partial t} + \frac{Q_i}{A_i}\frac{\partial C_i}{\partial x_i} = 0 \tag{2.1}$$

where C_i, Q_i, A_i, and x_i are the concentration, flow rate, cross-sectional area, and distance along the pipe i respectively, while t is time. Assuming complete mixing at the junctions, the concentration at any node is equal to the mass flow rate into the node (either from the upstream pipes or direct input) divided by the total flow rate out of the node, given by

$$C_j^* = \frac{\sum_{i \in d_i \in j} Q_i C_i \mid_{x_i = L_i} + U_j}{D_j + \sum_{i \in u_i \in j} Q_i} \tag{2.2}$$

where C_j^* is the concentration of water leaving node j, d_i is the downstream node of pipe i, L_i is the length of pipe i, U_j is the mass loading rate at node j, D_j is the water demand at node j, and u_i is the upstream node of pipe i. This expression defines the boundary condition for (1).

If the source characteristics (i.e., source node, release time, and source concentration) are known, (2.1) and (2.2) can be solved to obtain concentrations as a function of time at one or more nodes of interest, such as sensor nodes. In other words, for a source at a single node ℓ, the forward equations can be solved for concentration, C_j^* at all nodes j = 1, 2, … N_n, where N_n is the number of nodes. In this way, information (e.g., contaminant concentration) is propagated downstream from the source to all possible downstream nodes. If concentrations are measured at one or multiple sensors (nodes), the information from these sensors can be propagated upstream from the sensors to all possible source nodes. This upgradient propagation of information is carried out by solving the adjoints of the forward equations. Neupauer (2011) showed that the adjoint of (2.1) is

$$\frac{\partial \psi_i}{\partial \tau} - \frac{Q_i}{A_i} \frac{\partial \psi_i}{\partial x_i} = 0 \tag{2.3}$$

where ψ_i is the adjoint state of the concentration in pipe i, defined as the marginal sensitivity of the concentration in pipe i (C_i) to a source mass released at node ℓ (M_ℓ) given by $\psi_i = \partial C_i / \partial M_\ell$, and τ is backward time, defined as the time prior to a reference time (e.g. if the time of contaminant

detection is selected as $\tau = 0$ hr, then $\tau = 1.5$ hr would refer to 1.5 hour before the contamination was discovered). The advection term in the adjoint equation ($\frac{Q_i}{A_i}\frac{\partial \psi_i}{\partial x_i}$) is negative instead of positive showing that the adjoint state is propagated against the flow of water.

The adjoint state of the concentration at a node is obtained by solving the adjoint of (2.2), which is given by (Neupauer 2011)

$$\psi_\ell^* = \frac{\sum_{i \in u_i = \ell} Q_i \psi_i\big|_{x_i = 0} + U_\ell^*}{\sum_{i \in d_i = \ell} Q_i} \tag{2.4}$$

where ψ_ℓ^* is the adjoint state of concentration at Node $\boldsymbol{\ell}$, which physically represents the marginal sensitivity of nodal concentration C_j^* at a single sensor node *j* at backward time τ to a source release of mass $M\boldsymbol{\ell}$ at any potential source node ($\boldsymbol{\ell} = 1, 2, \ldots N_n$). U_ℓ^* is the adjoint state load term and is given by (Neupauer 2011)

$$U_\ell^* = \begin{cases} 0 & j \neq \ell \\ \delta(\tau - \tau_s) & j = \ell \end{cases} \tag{2.5}$$

where $\delta(\cdot)$ is a Dirac delta function and τ_s is the backward time at which contamination is observed at the sensor node (*j*). The load term is non-zero only at the sensor node *j* where contamination is observed. In this way, information enters the adjoint equation only at the sensor node and is propagated upgradient to all possible source nodes.

The system of equations in (2.3) and (2.4) is solved once for each sensor observation to obtain the temporal distribution of the adjoint state ψ_ℓ^* at all nodes $\boldsymbol{\ell} = 1, 2, \ldots N_n$. In this work, we use EPANET to determine the adjoint state, ψ_ℓ^*. The magnitude of the adjoint load, U_ℓ^*, is approximated by

$$U_{\ell}^{*} = \frac{1}{\Delta t} \tag{2.6}$$

where Δt is the length of the timestep used in the simulation. This adjoint load is inserted into the water system as a mass booster for a single timestep and propagated backward through the water distribution system (i.e. against the flow of water) to potential source nodes. In this way, the adjoint state, the sensitivity of the concentration at the observation node to a mass input at the source node, is calculated for all hydraulically connected upstream nodes.

The adjoint state can be used to obtain the backward travel time probability density function (BTTPDF), $f_T(\tau;\boldsymbol{\ell},j,\tau_S)$, that defines the random backward time τ that a contaminant particle observed at node j at backward time τ_s could have been released at node $\boldsymbol{\ell}$, which is a potential source node in the water distribution system. This BTTPDF is related to the adjoint state through (Neupauer et al. 2010)

$$f_T(\tau;\ell,j,\tau_S) = \psi_{\ell}^{*}(\tau;j,\tau_S)\sum_{i\in n_i=\ell} Q_i(\tau) \tag{2.7}$$

Since a different adjoint state is obtained for each sensor observation, (2.7) is solved once for each sensor observation. The BTTPDF is non-zero for all nodes that are upstream of and hydraulically connected to the sensor node. The BTTPDF for any node which is not hydraulically connected to the sensor node is zero at all times.

We assume the contaminant entered the system at only one source node and release time, so all sensor observations are traceable back to an instantaneous release at a single source node (which has not yet been determined). Let $\mathbf{J} = \{j_1, j_2, \ldots j_{Ns}\}$ be a set of N_s sensor nodes at which contamination is observed, and let $\mathbf{T}_s = \{\tau_{s1}, \tau_{s2}, \ldots \tau_{sNs}\}$ be the set of observation times at each sensor. We use Bayes' theorem and determine a joint BTTPDF, $f_T(\tau;\ell,\mathbf{J},\mathbf{T}_s)$, which defines the

random time τ that contamination that was observed at sensor nodes $\mathbf{J}$ at times $\mathbf{T}_x$ could have been released at node $\boldsymbol{\ell}$. This joint BTTPDF is given by (Neupauer and Records 2009)

$$f_T(\tau;\ell,\mathbf{J},\mathbf{T}_s)=\alpha_T\prod_{n=1}^{N_S}f_T(\tau;\ell,j_n,\tau_{sn}) \tag{2.8}$$

$$\alpha_T^{-1}=\int_\tau\prod_{n=1}^{N_S}f_T(\tau;\ell,j_n,\tau_{sn})d\tau \tag{2.9}$$

where $f_T(\tau;\ell,j_n,\tau_{sn})$ is the backward travel time probability density function from (2.7) for the nth sensor observation, and α_T is used to ensure that the total probability is unity.

The measured concentrations at the sensors provide additional information that can be used to determine the potential source nodes. Let $\hat{\mathbf{C}}^*=\{\hat{c}^*_{j1},\hat{c}^*_{j2},\ldots\hat{c}^*_{jNs}\}$ be a vector of N_s sensor observations, where $\hat{c}^*_{jn}(\tau_{sn})$ is the measured concentration of the nth sensor observation which occurs at node j_n at time τ_{sn}. The joint BTTPDF in (2.8) can be conditioned on these measured concentrations to obtain a conditioned BTTPDF, $f_{T|\hat{\mathbf{C}}^*}(\tau\,|\,\hat{C}^*;\ell)$, which defines the random backward time τ that contamination observed in the N_s sensor observations in $\hat{\mathbf{C}}^*$ could have been released at node $\boldsymbol{\ell}$, which is given by (Neupauer et al. 2010)

$$f_{T|\hat{\mathbf{C}}^*}(\tau\,|\,\hat{C}^*;\ell)=\beta_T^{-1}\int_m\prod_{n=1}^{N_S}f_{\hat{C}^*_{jn}|M,T}(\hat{c}^*_{jn}\,|\,m,\tau;\ell)f_T(\tau;\ell,j_n,\tau_{sn})dm \tag{2.10}$$

$$\beta_T=\int_0^\infty\int_m\prod_{n=1}^{N_S}f_{\hat{C}^*_{jn}|M,T}(\hat{c}^*_{jn}\,|\,m,\tau;\ell)f_T(\tau;\ell,j_n,\tau_{sn})dmd\tau \tag{2.11}$$

where $f_T(\tau;\ell,j_n,\tau_{sn})$ is the backward travel time probability density function for the nth sensor observation, $f_{\hat{C}^*_{jn}|M,T}(\hat{c}^*_{jn}\,|\,m,\tau;\ell)$ is the PDF of obtaining the measured concentration $\hat{c}^*_{jn}(\tau_{sn})$ for a given source node $\boldsymbol{\ell}$, release time τ, and source mass m, defined as a normal distribution with a mean

of $m\psi^*_{jn}(\tau;\ell,\tau_{sn})$, and a standard deviation σ representing model and measurement uncertainty (Neupauer et al. 2010). In (2.10) and (2.11), integration is carried out over a likely range of source masses, and β_T is used to ensure that the total probability is unity.

2.3. Examples

We tested the adjoint-based probabilistic method for source identification using the network shown in Figure 2.1, which has the same layout as EPANET Example 1 (Rossman 2000). The water distribution system contains 12 pipes, 9 nodes, a reservoir, a variable speed pump, and a fully-mixed tank which is initially 80%. With the exception of node 10, all nodes have demands which vary over time. The demands at the nodes are indicated in Figure 2.1 and the demand pattern is shown in Table 2.1.

The tank is filling from the start of the simulation, t = 0:00 (hh:mm), to 12:36 and from 23:40 until the end of the simulation (t = 24:00) and draining at all other times. The draining and filling of the tank affects the direction of flow in pipes 11, 21, and 110 (connected to the tank) and contributes to the overall transient state in the flow field for the system.

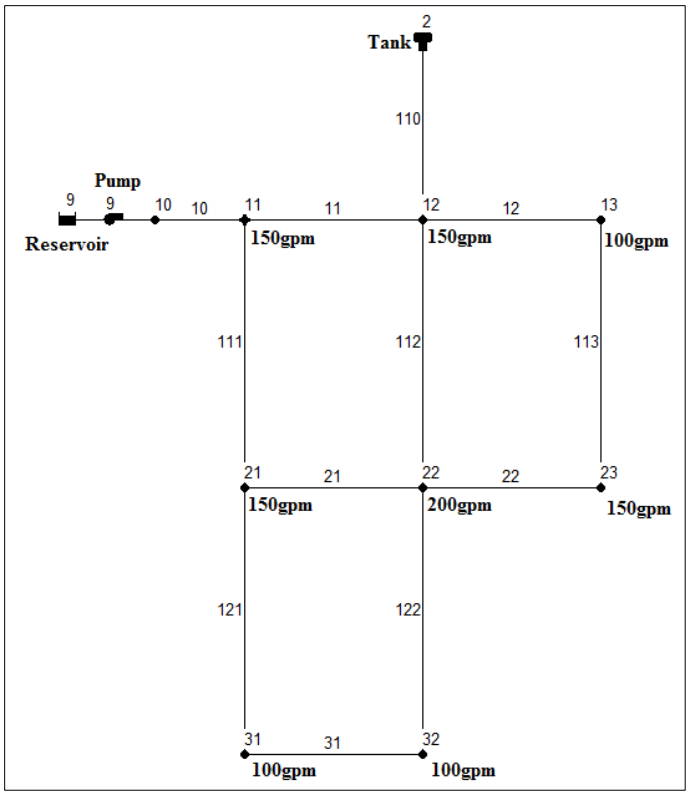

Figure 2.1. EPANET Example 1 water distribution system. Lines represent pipes. Circles represent nodes. The base nodal demands in gallons per minute are indicated at the nodes (e.g., 150 gpm). The number above the node is the node identifier. The number above the pipe is the pipe identifier.

Table 2.1. Demand Pattern.

Start Time (hh:mm)	End Time (hh:mm)	Multiplier
0:00	2:00	1.0
2:00	4:00	1.2
4:00	6:00	1.4
6:00	8:00	1.6
8:00	10:00	1.4
10:00	12:00	1.2
12:00	14:00	1.0
14:00	15:00	0.8
15:00	16:00	0.1
16:00	18:00	0.6
18:00	20:00	0.4
20:00	22:00	0.6
22:00	24:00	0.8

We use EPANET to simulate the movement of a generic non-reactive contaminant in this network to generate the sensor measurements used to test the source identification method. We use a time step of Δt = 2 minutes and we assume a contaminant source at node 11, modeled as a flow-paced booster with an input concentration of 5,000 mg/L from t = 3:58 to 4:00, with a total mass of approximately 69 kg. Nodes 23 and 32 were used as observation nodes at which sensors measured the contaminant concentration over time. Figure 2.2 shows the contaminant concentration as a function of time for observation nodes 23 and 32. Each of the observation nodes shows four distinct arrivals of the contaminant, labeled A though D in Figure 2.2, due to multiple flowpaths from the source node to the observation node and transient flow in the system. We define each of these distinct arrivals as an "observation set". Table 2.2 shows the start and end time of each observation set, and the peak concentration and the forward time at which the peak concentration occurs. The start and end times are shown in both forward and backward time, with backward time τ defined as τ = 24:00-t. The latest arrival (labeled as D in Figure 2.2) has a significantly lower concentration than the earlier arrivals because this water was previously in the tank. The contaminant enters the tank at t = 4:46 and begins exiting the tank at 12:38. Since the tank is fully-mixed, any contaminant that enters the tank is instantaneously mixed with the entire contents of the tank effectively diluting the contaminant.

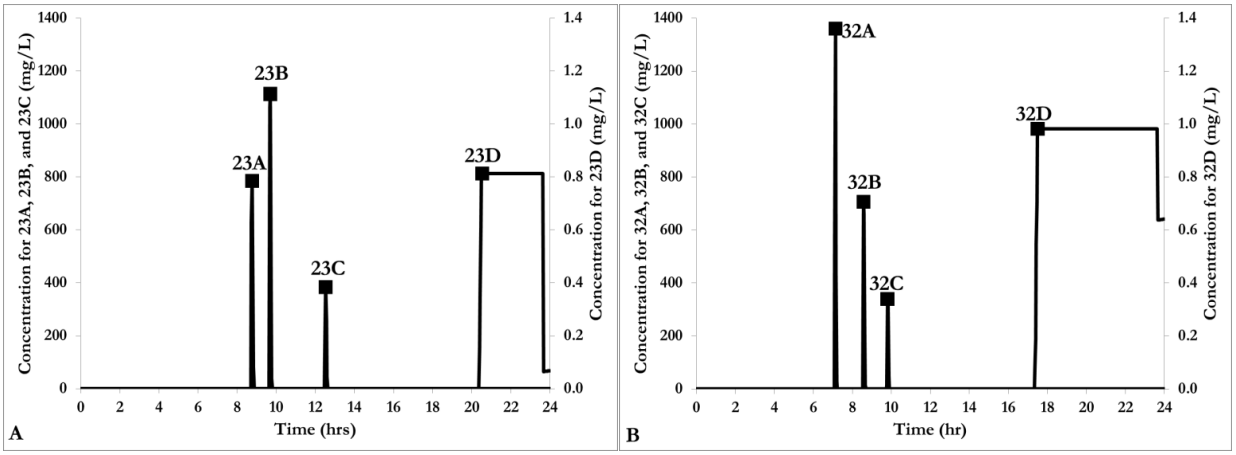

Figure 2.2. Observed concentrations at sensors at (A) node 23 and (B) node 32. The square markers in the figures indicate the observations used. The secondary axes on the right side of the each figure are for observations 23D and 32D.

Table 2.2. Data for Observations in Figure 2.2. Reference Time for the Backward Time is t = 24:00.

Observation Set	Forward Time, t (hh:mm)		Backward Time, (hh:mm)		Observation Peak Values	
	Start Time	End Time	Start Time	End Time	Concentration (mg/L)	Time (hh:mm)
23A	8:44	8:50	15:10	15:16	790	8:46
23B	9:40	9:46	14:14	14:20	1070	9:42
23C	12:30	12:36	11:24	11:30	380	12:32
23D	20:24	24:00	0:00	3:36	0.81	20:32
32A	7:06	7:12	16:48	16:54	1360	7:08
32B	8:34	8:40	15:20	15:26	710	8:34
32C	9:46	9:52	14:08	14:14	340	9:48
32D	17:22	24:00	0:00	6:38	0.98	17:30

Ideally all of the available information should be used for source identification; however, the adjoint-based approach requires one simulation for each observation; thus for a large system with many sensors and observations, running one adjoint simulation for each observation could be computationally prohibitive. Instead the user can select a small number of observations to be used. Similarly, instead of using each observation in each observation set, we use the peak concentration and time at which the peak concentration occurred (shown as squares in Figure 2.2). For instance, the peak concentration for the first observation for node 23 is 790 mg/L and the time is 8:46; this is referred to as observation 23A. We demonstrate the method using three example sets of data: (1)

two observations, neither of which passed through the tank, (2) two observations, one of which passed through the tank, and (3) three observations, one of which passed through the tank and two which did not. As we demonstrate here, reasonably accurate source identification can be achieved with a small subset of the available information.

2.3.1. Scenario 1: Observations at 23B and 32A.

In this example we use two observations, the peak concentration for observations 23B (j_1 = 23) and 32A (j_2 = 32). Both of these observations contain contaminant which never entered the storage tank. The first step in the source identification method is to calculate the adjoint states by solving (2.3) and (2.4) once for each sensor observation. As stated above, we solve these equations using EPANET. We use the EPANET Programmer's Toolkit to create a hydraulics file that contains pressure and demands for all nodes and flow rates, status, and settings for all pipes for all time steps. We then create a new hydraulics file with the information reversed in time and with the signs reversed on the flows and demands. This new hydraulics file is used to propagate the adjoint state backward through the water distribution system. The adjoint load is determined for each observation using (2.6) and is equal to 0.5 min^{-1}, producing adjoint state values of 0.009 gal^{-1} and 0.012 gal^{-1} for observation nodes 23 and 32, respectively, at the times of release. As the adjoint state is propagated through the water distribution system model, it is diluted, thus the simulated values will all be substantially less than unity, and some accuracy may be lost due to rounding the output to a fixed number of decimal places. To avoid this, we multiply the adjoint load by 10^6 and then divide the output by 10^6. Since the adjoint load is an instantaneous load (2.5), the adjoint load is inserted at a single timestep corresponding to the time at which the peak concentration was observed (e.g., 9:42 for observation 23B).

Using observations 23B and 32A in separate simulations, we obtain adjoint states for upstream nodes $\boldsymbol{\ell} = 10, 11, 12, 13, 21, 22$, and 31 and use these in (2.7) to calculate the unconditioned BTTPDFs. Figure 2.3 shows the resulting adjoint states and unconditioned BTTPDFs for upstream node $\boldsymbol{\ell} = 11$ (the true source node) for observations 23B and 32A (similar plots can be made for the other upstream nodes, but are not shown here). The unconditioned BTTPDF represents the possible times at which the observed contamination could have been at node 11. For both observations, the observed contaminant could have been at node 11 at multiple times as a result of the multiple flowpaths between the observation nodes and node 11. Table 2.3 shows the time range for which the adjoint states and unconditioned BTTPDFs are non-zero for any upstream nodes $\boldsymbol{\ell} = 10, 11, 12, 13, 21, 22$, and 31. These are the candidate source release times for each of the nodes.

The results in Table 2.3 demonstrate that the contamination from observation 23B could have originated at any node except node 31, while the contamination from observation 32A could have come from any node except node 13. In the forward simulation, water flows from the reservoir through nodes 10, 11, 12, 13, 21, and 22 prior to arriving at node 23; node 31 is not hydraulically connected to node 23, so it cannot be a source node. Similarly, water flows from the reservoir through nodes 10, 11, 12, 21, 22, and 31 prior to arriving at node 32, so all of these nodes are hydraulically connected to node 32 and capable of producing the contaminant observation 32B. We assume that the contamination occurred at a single node and time, thus the contamination observed at both 23B and 32A could only have originated at nodes for which the candidate times (Table 2.3) overlap. These results show us that the true source node and time, node 11 at t = 4:00, is a candidate source scenario for this combination of observations. These results also show that nodes 10, 12, 21, and 22 are candidate source nodes, while 13 and 31 are not.

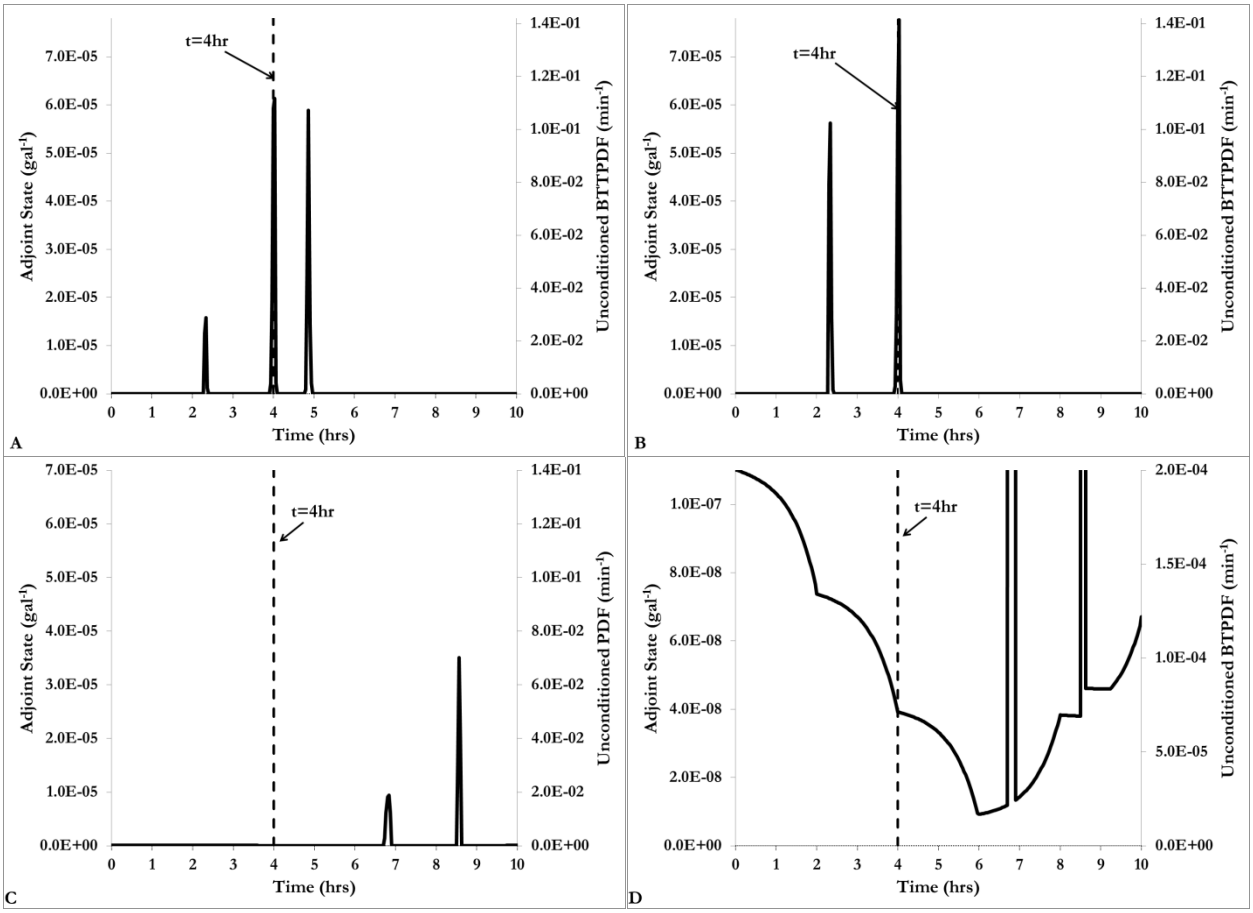

Figure 2.3. Adjoint state and unconditioned BTTPDF for potential source node 11 using (A) observation 23B, (B) observation 32A, and (C)-(D) observation 23D. The dashed line indicates the time at which the contaminant source actually entered the system (4:00).

Table 2.3. Potential Contaminant Release Times Based on Observation 23B and 32A.

Source Node	23B		32A		Overlapping Times	
	Start Time (hh:mm)	End Time (hh:mm)	Start Time (hh:mm)	End Time (hh:mm)	Start Time (hh:mm)	End Time (hh:mm)
10	1:02	1:08	1:02	2:22	1:02	1:08
	2:38	3:40	2:38	2:48	2:38	2:48
11	2:18	2:22	2:18	2:24	2:18	2:22
	3:56	4:04	3:56	4:04	3:56	4:04
	4:48	4:56	-	-	-	-
12	4:36	4:42	4:36	4:42	4:36	4:42
	5:28	5:34	-	-	-	-
13	6:58	7:00	-	-	-	-
21	3:02	3:04	3:02	3:06	3:02	3:04
22	5:56	5:58	5:56	5:58	5:56	5:56
31	-	-	4:24	4:26	-	-

The unconditioned BTTPDFs are obtained by scaling the adjoint states by the flow rates, as in (2.7), and are shown for observations 23B and 32A in Figures 2.3A and 3B respectively. The temporal changes in the flow rate across node 11 are small at times when the adjoint state is non-zero, so the differences between the adjoint state and unconditioned BTTPDF in Figure 2.3 are imperceptible. The unconditioned BTTPDF is non-zero at release times that could produce the observations. For both observations, the unconditioned BTTPDF has non-zero values for node 11 (the true source node) at time 4:00, thus the true release time is identified as a candidate release time.

We used these unconditioned BTTPDFs in (10) to calculate the conditioned BTTPDF that describes the random release time of contamination from node ℓ that could produce the measured concentrations in the two observations. In (10), we used a source mass range of 65-75 kg (the actual source mass was about 69 kg) and a standard deviation of model uncertainty equal to 10% of the measured concentration (e.g., if the measurement is 15 mg/L, then standard deviation is ±1.5 mg/L).

Figure 2.4A shows the resulting conditioned BTTPDFs for nodes 10, 11, 12, 21, and 22. Only these nodes have non-zero conditioned BTTPDFs because these are the only source nodes for which the adjoint states for the two observations have non-zero values at the same range of times (see Table 2.3). The results show various possible contamination scenarios for nodes 10, 11, 12, 21, and 22. Note that the true source node (node 11) and release time (t = 4:00) is identified as a likely source scenario.

We can estimate the likelihood of each candidate source node being the true source node. In (2.10), β_T is used to normalize the conditioned BTTPDF to ensure that the total probability is unity; thus, by Bayes' theorem, β_T is proportional to the joint probability density function of the source node and release time (Neupauer et al., 2010). A larger value of β_T indicates a greater probability that the node is the true source node. The β_T values for nodes 10, 11, 12, 21, and 22 are shown in Table 2.4. Node 11 has the largest value, thus it is the most likely source node.

Although the maximum value of β_T (Table 2.4) occurs at node 11, β_T for node 10 is only slightly lower; indicating that node 10 is nearly as likely as node 11 to be the source node. If a contaminant entered the system at node 10, the source concentration at node 10 and the concentration observed at node 11 would be very similar; the only dilution between node 10 and 11 is the demand at node 11, which is about 20% of the average flow of water from node 10 to node 11. Similarly, when the adjoint state at node 11 is backtracked to node 10, the adjoint state at node 10 will have a similar value as it did at node 11. The result is adjoint state values which are similar in magnitude for node 10 and 11. The β_T values are also similar because the PDF of obtaining the measured concentration for nodes 10 and 11 are also similar.

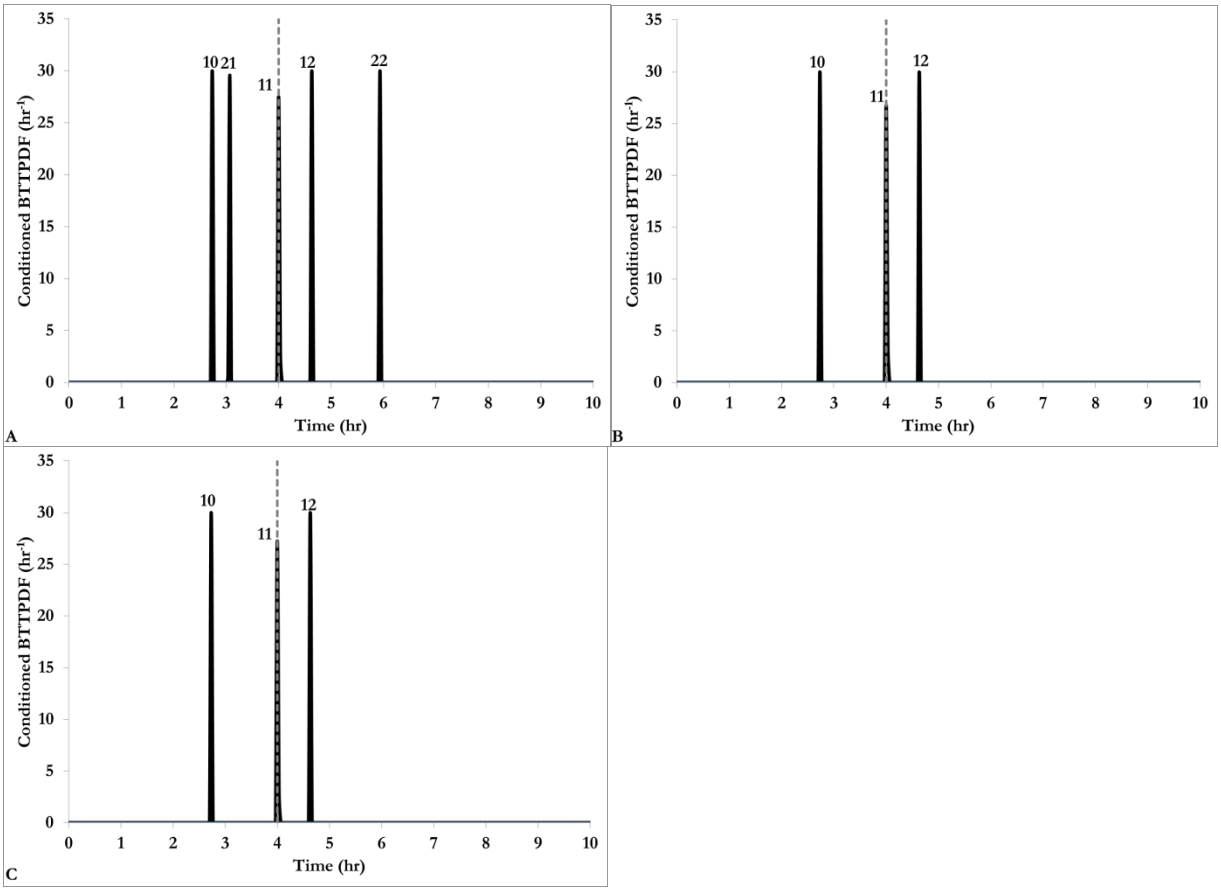

Figure 2.4. Conditioned BTTPDF for potential source nodes using (A) observations 23B and 32A (Scenario 1), (B) observations 23D and 32A (Scenario 2), and (C) observations 23A, 23D, and 32A (Scenario 3). The dashed line indicates the true contaminant release time (t = 4:00). The number above the curve identifies node number of the potential source node.

Table 2.4. β_T Values for Complex Water Distribution System in Units of $L^3/mg^2/hr^2$ (Maximum Value is Bolded).

Scenario	Node 10	Node 11	Node 12	Node 13	Node 21	Node 22	Node 31
1	4.28	**6.47**	2.10E-7	0	1.89E-82	1.37E-240	0
2	2.15	**4.09**	4.47E-7	0	0	0	0
3	1.91E-4	**3.86E-4**	2.64E-15	0	0	0	0

In summary, based on the β_T values, the adjoint method correctly identified the true source node (node 11) as the most likely source node; and based on the conditioned BTTPDF for node 11, the adjoint method correctly identified the true release time (t = 4:00) as the most likely release time.

However, node 10 is also a likely source node. If it were the source node, the most likely release time would be t = 2:44.

2.3.2. Scenario 2: Observations at 23D and 32A.

In this example we use two observations, the peak concentration from observation 23D (j_1 = 23) which contains information from contaminant that passed through the storage tank, and the peak concentration from observation 32A (j_2 = 32) which contains information from contaminant which did not enter the storage tank. Following the same method as the first example, we use (2.3) and (2.4) to calculate adjoint states, and use the results in (2.7), (2.10), and (2.11) to calculate the conditioned BTTPDFs (Figure 2.4B) and β_T (Table 2.4). Based on the β_T values, the true source node (node 11) is successfully identified as the most likely source node, and from Figure 2.4, the true release time (t = 4:00) is successfully identified as the most likely release time.

The β_T values for this scenario indicate that either node 10 or 11 is the probable source node. The β_T values for nodes 10 and 11 in this scenario are lower than in Scenario 1 because the adjoint state for observation 23D is diluted as it passes through the tank, similar to how the concentration is diluted in the forward simulation (Figure 2.2). Figure 2.3C,D shows the adjoint state and unconditioned BTTPDF for 23D. The range of possible release times is much broader for 23D than for 23B (Figure 2.3A) or 32A (Figure 2.3B), with non-zero values of the adjoint state from t = 0:00 to 12:32 (partially shown in Figure 2.3D) due to the effect of the completely mixed tank. Whenever the tank is emptying in the adjoint simulation (filling in the forward simulation), the water leaving the tank contains low values of the adjoint state. Since the adjoint state exits the tank over a wide range of times and at a low magnitude, as it is propagated through the system, it arrives at each upstream node, specifically nodes 10 and 11, over a wide range of times and at a low magnitude.

In a forward simulation, contaminant may enter a storage tank at one instant in time while the tank is filling. If the fully-mixed tank model is assumed, the contaminant is instantaneously diluted throughout the water in the tank. As the tank drains, all of the water draining from the tank contains low concentrations of the contaminant, and any information about the precise timing of the arrival of contamination at the tank is lost. All that can be determined is that contamination arrived at the tank prior to the time it started draining. Thus, many different source release times could lead to the same observed concentrations at sensors downstream of the tank. This loss of information also occurs in the adjoint simulation as the adjoint state is diluted upon entering the tank. When the water drains out of the tank in the adjoint simulation, it contains low levels of the adjoint state that is propagated upgradient to all possible source nodes. Thus, many different source release times are identified. This loss of information results in lower magnitude β_T values as compared to the β_T values obtained for the same number of observation with none having passed through the tank.

2.3.3. Scenario 3: Observations at 23B, 23D, and 32A.

In this example we use three observations, the peak concentrations at 23B ($j_1 = 23$) and 32A ($j_3 = 32$) which contain information from contaminant that did not pass through the storage tank, and the peak concentration at 23D ($j_2 = 23$) which contains information from contaminant which passed through the storage tank. Following the same method as the first example, we use (2.3) and (2.4) to calculate adjoint states, and use the results in (2.7), (2.10), and (2.11) to calculate the conditioned BTTPDFs (Figure 2.4C) and β_T (Table 2.4). Based on the β_T values, the true source node (node 11) is successfully identified as the most likely source node, and from Figure 2.4C, the true release time (t = 4:00) is successfully identified as the most likely release time.

Once again, the β_T values (Table 2.4) indicate that nodes 10 and 11 are nearly equally likely to be the true source node. This is an expected result for the same reasons elucidated in Scenario 2. The most likely release time for node 10, t = 2:44, is also the same time that was found in Scenario 2.

Note that the β_T values for this scenario, with three observations, are lower than the β_T values for the previous two scenarios, which each used only two observations. The calculation for β_T uses the product of the unconditioned BTTPDFs for all observations; since the unconditioned BTTPDFs are less than unity, the product of three unconditioned BTTPDFs will have a lower value than the product of two. Thus, the value of β_T depends on the number of observations, so it cannot be compared across different scenarios that use different observations. β_T is useful for determining which source node is the most likely within each scenario

2.3.4. Analysis

In the previous examples, we used sets of two or three observations to identify the source node and release time. Each observation contains different information content, so different results are produced with different combinations of observations. In this section, we investigate the robustness of the source identification method using different combinations of two, three, and four observations.

We tested our adjoint-based probabilistic method for all combinations of two observations from the observations denoted in Figure 2.2, for a total of 28 combinations of two observations. For all 28 observation pairs, the true source node (node 11) and release time (t = 4:00) was always identified as a possible source node and release time, indicating that the method is sufficiently robust to identify the true source node and release time for any set of two observations. Node 11 was selected as the most likely source node in 25 of the 28 cases (89%), based on β_T; node 10 was

selected as the most likely source node in the other three cases. Each of the three cases that did not produce node 11 as the true source node had at least one observation that passed through the tank and one case had two observations that passed through the tank. These results demonstrate how the loss of information when the adjoint state passes through the tank in conjunction with the similar adjoint states for nodes 10 and 11 can lead to the incorrect node being chosen as the source node; however, in another ten cases where an observation that had traveled through the tank was used node 11 was still chosen as the true source node.

We also tested our adjoint-based probabilistic method for combinations of three observations. We only consider combinations of three observations that include at least one observation from each observation node, resulting in 48 different combinations of three observations. In each case, the true source node and release time were identified as a possible source node and release time. In 46 of the 48 cases (96%), node 11 was identified as the most likely source location; node 10 was identified for the other two cases. Again, the two cases that calculated node 10 as the true source node used information from an observation that passed through the tank and one of the cases contained information from two observations that passed through the tank. As discussed previously, information is lost when that adjoint states passes through the tank which increases the likelihood of calculating node 10 as the true source node over node 11; however, in another 30 instances where an observation that had traveled through the tank was used, node 11 was still chosen as the true source node.

Finally, we tested our adjoint-based probabilistic method for the 34 combinations of four observations. In each case, the true source node and release time were identified as a possible source node and release time. Node 11 was selected as the true source node, based on the value of β_T, in 33 of the 34 cases (97%). The case that did not select node 11 as the true source node used information from two observations of contaminant that had passed through the tank and node 10

was selected as the true source node. Six scenarios using two observations of contaminant that passed through the tank still selected node 11 as the true source node, however.

In general, the true source node and release time were identified as a possible source node/release time regardless of the number of observations used. We also demonstrated that the method is more accurate when more observations are used; we correctly identified node 11 as the true source node in 89% of the scenarios using two observations, 96% with three observations, and 97% with four observations. In the cases where node 11 was not selected as the true source node, node 10, which is similar to node 11 due to the flowpaths in the system, was selected as the true source node. In all of the cases where node 10 was selected over node 11, at least one of the observations used had contamination that had passed through the tank; the dilution of the adjoint state in the tank led to the loss of information and increased the likelihood that the incorrect source node could be selected.

2.4. Conclusions

We used an adjoint-based probabilistic modeling approach to identify the source node and release time of a contaminant that is observed at one or more sensors in a water distribution system. We considered a pipe network that has transient hydraulics and a fully-mixed tank. We assumed perfect knowledge of the hydraulics, and we assumed that sensors measure concentration to within 10% of the true value. Our method successfully identified the source node and release time as a potential contamination scenario in all scenarios evaluated using two, three, or four observations. In the majority of cases, our method was able to probabilistically determine the true source node as the most likely source node. The cases that did not select the true source node as the most likely source node all had the following characteristics: (1) a node which is hydraulically similar to the true source

node was chosen as the most likely source node, and (2) at least one observation from contaminant that passed through the water storage tank was used in the calculations.

As stated previously, the solution to a source identification problem such as the one presented here is non-unique; we expect to have multiple potential source nodes and release times. Part of our method uses the following criteria to help identify the true source node: (1) Is the node hydraulically connected to the observation node(s)?, and (2) Does the contaminant mass needed to be released from the potential source node to reproduce the contaminant observations fall within the likely range of source masses? Our water distribution system model contains a node directly upstream of the true source node which meets both of these criteria and is nearly indistinguishable from the true source node. Not only is this node hydraulically connected to the observation nodes, all of the water from the node passes through the true source node, and most of the water at the true source node comes from the node; if not for the external demands at the true source node, the two nodes could almost be considered the same node. The similarities between the two nodes result in similar adjoint states between the source nodes. In addition, the source mass needed at the upstream node to replicate the concentration observations at the observation nodes is close to the true source mass released at the true source node and within the range of the potential source masses we tested, thus the two nodes also have similar probabilities of being the true source node.

It is also important to note that observations from contamination that passed through the water storage tank were used in the calculations leading to selecting the incorrect node as the true source node. The water storage tank is fully mixed, so, in the forward model, concentration of the contaminant in water is diluted when the contaminant passes through the tank. Similarly, the adjoint state is diluted as it is propagated backward through the tank resulting in a loss of information at upstream nodes. The consequence of passing through the storage tank is a wide span of potential release time that, when used in conjunction with adjoint states calculated using observations that did

not pass through the tank, does little to narrow down the number of potential release times, but does decrease the probability that any one source node is the most likely source node.

CHAPTER 3

PROBABILISTIC SOURCE CHARACTERIZATION IN WATER DISTRIBUTION SYSTEMS USING IMPERFECT SENSOR DATA

Abstract

Source characterization is important in drinking water distribution systems to find the source of contamination, discontinue the event, and prevent future contamination events. Previous work has demonstrated the effectiveness of an adjoint-based probabilistic method using contaminant concentration measurements from system sensors to probabilistically determine the source of contamination in a drinking water system. The method can be applied in systems with steady-state or transient hydraulics. The method uses publicly available software (EPANET) coupled with a conditioning method to probabilistically identify the contamination source and release time. Prior work depended on sensors capable of measuring distinct contaminant concentrations (e.g., 12.5 mg/L). In this work a method is developed to use measurements from more realistic, non-ideal sensors that identify the range of contamination at the sensor (e.g., high, medium, low) or, simply, the presence or absence of contamination, rather than the precise concentration. In addition, a method is developed to use non-detect measurements, i.e., sensor measurements for which the measured contamination is below the limit of detection.

3.1. Introduction

The events of September 11, 2001 brought to life the potential for large-scale terrorist activities on civilian infrastructure. With this has come an increased attention to enhancing the security of public utilities; water distribution system operators are seeking better ways to prevent, detect, and remediate contamination in drinking water systems (DHS 2003). The ideal course of action is to prevent contamination from entering a water distribution system, but this is not always possible. If contamination is released into the system, remediation techniques are necessary to

prevent further contamination, but contamination must be detected before remediation can take place. Contaminant detection in water distribution systems allows the purveyor to prevent the spread of contamination and attempt to identify the source of contamination. The ability to respond effectively to contaminant detection depends on the sensors that are available and the information that they are able to provide. Unfortunately, the sensors used to detect contamination are often unable to measure the specific concentration of contamination and instead measure the general contaminant level (e.g., high, medium, low) or the presence/absence of contamination. While this information is not ideal, it can still be used to determine the source of contamination in the water distribution system.

Many researchers have worked on developing methods for effective source characterization. Previous work has focused on inverse methods (e.g., Islam et al. 1997; Laird et al. 2005, 2006; Guan et al. 2006; Preis and Ostfeld 2006, 2007). Inverse methods employ advanced algorithms to determine the origin of contamination. Observations (i.e., observation node, time, and contaminant concentration) are collected from sensors in a water distribution system and used as simulation goals for the inverse methods. Algorithms in conjunction with modeling software test multiple contamination scenarios (i.e.., source node, release time, and contaminant concentration) and run multiple forward simulations in the water distribution system to determine what scenario replicates the observations, or are as close as the user dictates. While many researchers have shown that inverse methods are able to correctly determine the contamination scenario, these methods are inherently inefficient because a new simulation is required for each potential contamination scenario.

Neupauer et al. (2010) developed an adjoint-based probabilistic method for source identification that uses adjoints of the forward transport equations to probabilistically determine the node at which the contaminant entered the system and the time at which the contaminant was released from the node. Information from sensor nodes in the system is used to calculate the adjoint

state which is the sensitivity of the concentration at the sensor node to mass entering the system at another node. Running a single adjoint simulation, using a single observation location, time, and concentration, provides information about possible contamination times and locations for multiple potential source nodes. Adjoint-based probabilistic methods that have been successfully used for water distribution systems have been limited to water distribution systems with perfect sensors (e.g., Neupauer et al. 2010; Wagner et al. 2013).

Both inverse and adjoint-based probabilistic methods are highly dependent on the availability of data within the system (e.g., contaminant concentration, contaminant arrival time, etc.). Although the sensors which are currently being used by water utilities are unable to accurately determine the levels of contaminant concentrations consistently (ASCE, 2004), many researchers assume that precise contaminant concentration data are available (e.g., Laird et al. 2005; Neupauer et al. 2010; Preis and Ostfeld 2006; Wagner et al. 2013). A more likely scenario is using "fuzzy sensors" which are only able to determine the relative ranges of water quality parameters (e.g., low, medium, or high level). Preis and Ostfeld (2008) developed an inverse method which was able to find the true source node and release time using data from fuzzy sensors, which only transmitted readings as high, medium, or low. Their research showed that the fuzzy sensor data led to an increased number of potential contamination scenarios. Another type of sensor is a binary, or Boolean, sensor which only provides information regarding whether contamination is present or not (Preis and Ostfeld 2008). Researchers have used data from binary sensors to determine the source of contamination using an inverse method (e.g., Preis and Ostfeld 2008; Kumar et al. 2012). Preis and Ostfeld (2008) found that using binary sensor data led to a much higher number of possible contamination nodes than fuzzy sensors or perfect sensors; in one case, the number of source nodes determined using binary sensor data was nearly four times as many as were determined using fuzzy

sensor data. Kumar et al. (2012) also found that the quantity and quality of sensor data directly influences the ability to determine the source of contamination.

The goal of this paper is to develop an adjoint-based probabilistic method which uses imperfect sensor data to determine the contaminant release node and release time in a water distribution system. The adjoint-based probabilistic method developed by Neupauer et al. (2010) and Wagner et al. (2013) uses contaminant concentration data from ideal sensors to determine the source of contamination. They used the distinct concentration value to calculate the probability density function (PDF) of obtaining the measured concentration. We develop methods for using a concentration range observed by a fuzzy sensor and presence/absence measurements from binary sensors.

We also develop a method to use non-detect measurements (i.e. concentrations below the limit of detection for the sensors) to determine the source of contamination. These non-detects often comprise a large percentage of the observed data and were previously unused by adjoint-based probabilistic methods. All prior work on adjoint-based source identification only used non-zero contaminant concentrations for source characterization (e.g., Neupauer et al., 2010; Wagner et al. 2013).

3.2. Probabilistic Approach for Source Identification Using Perfect Sensor Data

3.2.1. Theory for Using Perfect Sensor Data in an Adjoint-Based Probabilistic Approach

Forward transport of a conservative chemical in pipes can be modeled using

$$\frac{\partial C_i}{\partial t} + \frac{Q_i}{A_i}\frac{\partial C_i}{\partial x_i} = 0 \tag{3.2}$$

where C_i, Q_i, A_i, and x_i are the concentration, flow rate, cross-sectional area, and distance along the pipe i respectively, while t is time. Assuming complete mixing at the junctions, the concentration at

any node is equal to the mass flow rate into the node (either from the upstream pipes or direct input) divided by the total flow rate out of the node, given by

$$C_j^* = \frac{\sum_{i \in d_i \in j} Q_i C_i \mid_{x_i = L_i} + U_j}{D_j + \sum_{i \in u_i \in j} Q_i} \tag{3.2}$$

where C_j^* is the concentration of water leaving node j, d_i is the downstream node of pipe i, L_i is the length of pipe i, U_j is the mass loading rate at node j, D_j is the water demand at node j, and u_i is the upstream node of pipe i. This expression defines the boundary condition for (3.1).

If the contamination scenario is known, (3.1) and (3.2) can be solved to obtain concentrations as a function of time at one or more nodes of interest, such as sensor nodes. In other words, for a source at a single node ℓ, the forward equations can be solved for concentration, C_j^*, at all nodes $j = 1, 2, \ldots N_n$, where N_n is the number of nodes. In this way, information (e.g., contaminant concentration) is propagated downstream from the source to all possible downstream nodes.

In the source identification problem, we have information about the occurrence of contamination at the sensors, and we are seeking information about the contamination sources; thus, we want to propagate information from the sensor node to all possible source nodes. The upgradient propagation of information is carried out by solving the adjoints of the forward equations. Neupauer (2011) showed that the adjoint of (3.1) is

$$\frac{\partial \psi_i}{\partial \tau} - \frac{Q_i}{A_i} \frac{\partial \psi_i}{\partial x_i} = 0 \tag{3.3}$$

where ψ_i is the adjoint state of the concentration in pipe i, defined as the marginal sensitivity of the concentration in pipe i (C_i) to a source mass released at node ℓ (M_ℓ) given by $\psi_i = \partial C_i / \partial M_\ell$, and τ is backward time, defined as the time prior to a reference time (e.g. if the time of contaminant

detection is selected as $\tau = 0$ hr, then $\tau = 1.5$ hr would refer to 1.5 hour before the contamination was discovered). The advection term in the adjoint equation ($\frac{Q_i}{A_i}\frac{\partial \psi_i}{\partial x_i}$) is negative instead of positive showing that the adjoint state is propagated against the flow of water.

The adjoint state of the concentration at a node is obtained by solving the adjoint of (3.2), which is given by (Neupauer 2011)

$$\psi_\ell^* = \frac{\sum_{i \in u_i = \ell} Q_i \psi_i\big|_{x_i = 0} + U_\ell^*}{\sum_{i \in d_i = \ell} Q_i} \tag{3.4}$$

where ψ_ℓ^* is the adjoint state of concentration at node ℓ, which physically represents the marginal sensitivity of nodal concentration C_j^* at a single sensor node *j* at backward time τ to a source release of mass M_ℓ at any potential source node ($\ell = 1, 2, \ldots N_n$). U_ℓ^* is the adjoint state load term given by (Neupauer 2011)

$$U_\ell^* = \begin{cases} 0 & j \neq \ell \\ \delta(\tau - \tau_s) & j = \ell \end{cases} \tag{3.5}$$

where $\delta(\cdot)$ is a Dirac delta function and τ_s is the backward time at which contamination is observed at the sensor node *j*. The load term is non-zero only at the sensor node *j* where contamination is observed. In this way, information enters the adjoint equation only at the sensor node and is propagated upgradient to all possible source nodes.

The system of equations in (3.3) and (3.4) is solved once for each sensor observation to obtain the temporal distribution of the adjoint state ψ_ℓ^* at all nodes $\ell = 1, 2, \ldots N_n$. The adjoint state can be used to obtain the backward travel time probability density function (BTTPDF) through (Neupauer et al. 2010)

$$f_T(\tau; \ell, j, \tau_S) = \psi_\ell^*(\tau; j, \tau_S) \sum_{i \in u_i = \ell} Q_i(\tau) \tag{3.6}$$

where $f_T(\tau;\boldsymbol{\ell},j,\tau_S)$ is the backward travel time probability density function representing the backward time τ that a contaminant particle observed at node j at backward time τ_s could have been released at node $\boldsymbol{\ell}$, which is a potential source node in the water distribution system. Since a different adjoint state is obtained for each sensor observation, (3.6) is solved once for each sensor observation. We assume that only one true source node and release time exists for the contamination, so all sensor observations are traceable back to an instantaneous release at a single source node (which has not yet been determined).

The measured concentrations at the sensors provide additional information that can be used to determine the potential source nodes. Let $\hat{\mathbf{C}}^* = \{\hat{c}^*_{j1}, \hat{c}^*_{j2}, \ldots \hat{c}^*_{jN_S}\}$ be a vector of N_s sensor observations, where $\hat{c}^*_{jn}(\tau_{sn})$ is the measured concentration of the nth sensor observation which occurs at node j_n at time τ_{sn}. The BTTPDF in (3.6) can be conditioned on these measured concentrations to obtain a conditioned BTTPDF, $f_{T|\hat{\mathbf{C}}^*}\left(\tau \mid \hat{C}^*; \ell\right)$, which defines the random backward time τ that contamination observed in the N_s sensor observations in $\hat{\mathbf{C}}^*$ could have been released at node $\boldsymbol{\ell}$, which is given by (Neupauer et al. 2010)

$$f_{T|\hat{\mathbf{C}}^*}\left(\tau \mid \hat{C}^*; \ell\right) = \beta_T^{-1} \int_m \prod_{n=1}^{N_S} f_{\hat{C}^*_{jn}|M,T}\left(\hat{c}^*_{jn} \mid m, \tau; \ell\right) f_T\left(\tau; \ell, j_n, \tau_{sn}\right) dm \tag{3.7}$$

$$\beta_T = \int_0^\infty \int_m \prod_{n=1}^{N_S} f_{\hat{C}^*_{jn}|M,T}\left(\hat{c}^*_{jn} \mid m, \tau; \ell\right) f_T\left(\tau; \ell, j_n, \tau_{sn}\right) dm d\tau \tag{3.8}$$

where $f_T\left(\tau; \ell, j_n, \tau_{sn}\right)$ is the backward travel time probability density function for the nth sensor observation, $f_{\hat{C}^*_{jn}|M,T}\left(\hat{c}^*_{jn} \mid m, \tau; \ell\right)$ is the PDF of obtaining the measured concentration $\hat{c}^*_{jn}\left(\tau_{sn}\right)$ for a given source node $\boldsymbol{\ell}$, release time τ, and source mass m, defined as a normal distribution with a mean

of $m\psi^*_{jn}(\tau;\ell,\tau_{sn})$, and a measurement uncertainty σ representing the overall uncertainty of the model, expressed as (Neupauer et al. 2010)

$$f_{\hat{C}^*_{jn}|M,T}\left(\hat{c}^*_{jn} \mid m,\tau;\ell\right)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(C-\hat{C}\right)^2}{2\sigma^2}\right\} \tag{3.9}$$

where $C=m\psi^*_{jn}(\tau;\ell,\tau_{sn})$ is the true concentration that would be expected for a release of mass m at node ℓ at release time τ. In (3.7) and (3.8), integration is carried out over a likely range of source masses, and β_T is used to ensure that the total probability is unity.

3.2.2. Example Using Perfect Sensors

Neupauer et al. (2010) demonstrated how the adjoint-based probabilistic method works in the water distribution system shown in Figure 3.1. We reproduce their simulation and results here for comparison with the fuzzy sensor results. The water distribution system contains a reservoir at node 52 which supplies water to the system and constant demands of 100 gallons per minute (gpm) at nodes 7, 45, and 51 which cause the water (and contaminant) to flow through the system.

Neupauer et al. (2010) used 5-minute time steps and no reactions in the water distribution system. The contaminant was inserted at node 11 as a flow-paced booster with an input concentration of 100 mg/L from t = 0:00 (hh:mm) to 0:05. The total contaminant mass is approximately 116 g. Nodes 21 and 47 are observation nodes which measure the concentration of contaminant. They simulated this contamination scenario using EPANET (Rossman, 2000). Figure 3.2 shows the contaminant concentration as a function of time for observation nodes 21 and 47. Using ideal sensors, Neupauer et al. (2010) found that node 21 has non-zero contaminant concentrations from 3:50 to 4:15 with the peak concentration of 12.5 mg/L occurring at 4:05. Node 47 has non-zero contaminant concentrations from 7:30 to 8:05 with two peaks (6.0 and 5.4 mg/L)

occurring at 7:50 and 8:20. The two peaks for node 47 indicate two flowpaths from the source node (node 11) to the observation node (node 47).

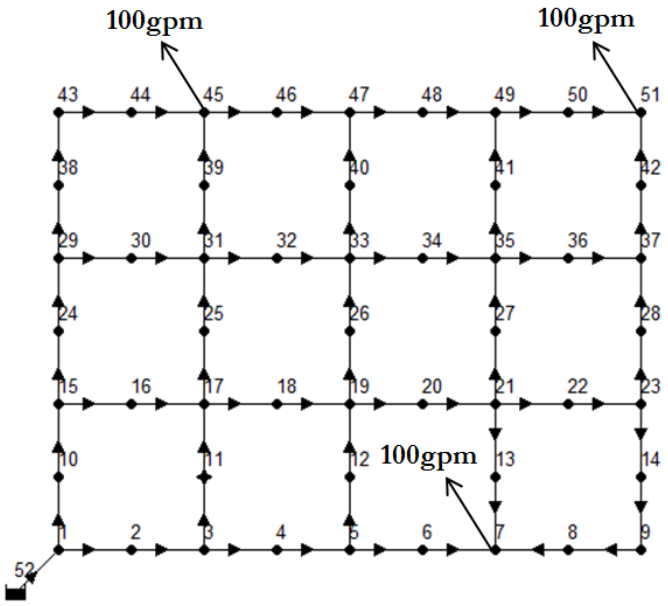

Figure 3.1. Example network [used by Neupauer et al. (2010)]: Lines represent pipes. Circles represent nodes. The arrows on the pipes indicate the direction of flow; the arrows on the nodes indicate demands at the nodes with the magnitude indicated at the end of the arrow. The number above the node is the node identifier.

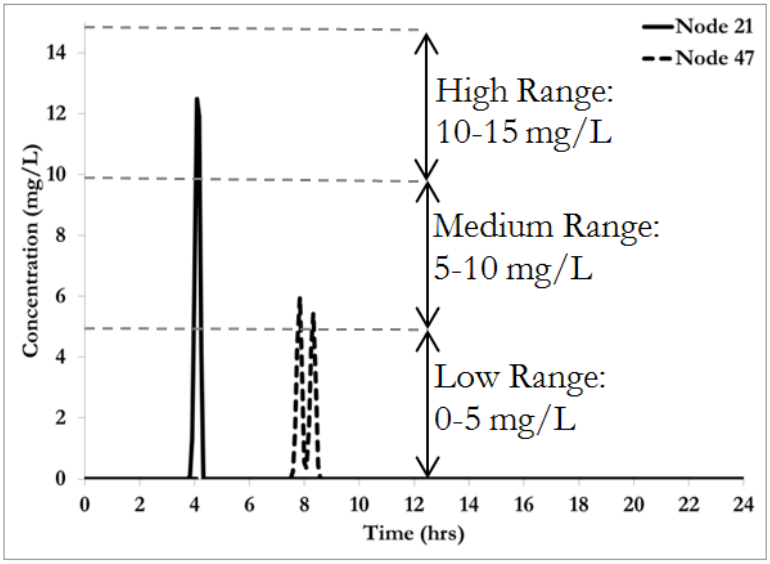

Figure 3.2. Concentration versus time plots for observation nodes 21 (solid black line) and 47 (dashed black line). The gray dashed lines represent the boundaries of the fuzzy sensor ranges for Cases 1-3.

Using the peak concentration observations (node, time, and concentration) for nodes 21 and 47, Neupauer et al. (2010) calculated the adjoint states using EPANET and (4). The adjoint states were used in (6) to calculate the BTTPDF for each observation and each potential source node, and the conditioned BTTPDFs were calculated using (7). Figure 3.3 shows the unconditioned BTTPDFs for source node 11 using each of the three observations, while Figure 3.4 shows the conditioned BTTPDFs for all of the potential source nodes evaluated using all three observations together. Note that we converted their backward time, τ, to forward time, *t*, to maintain consistency with how we will present results in this paper.

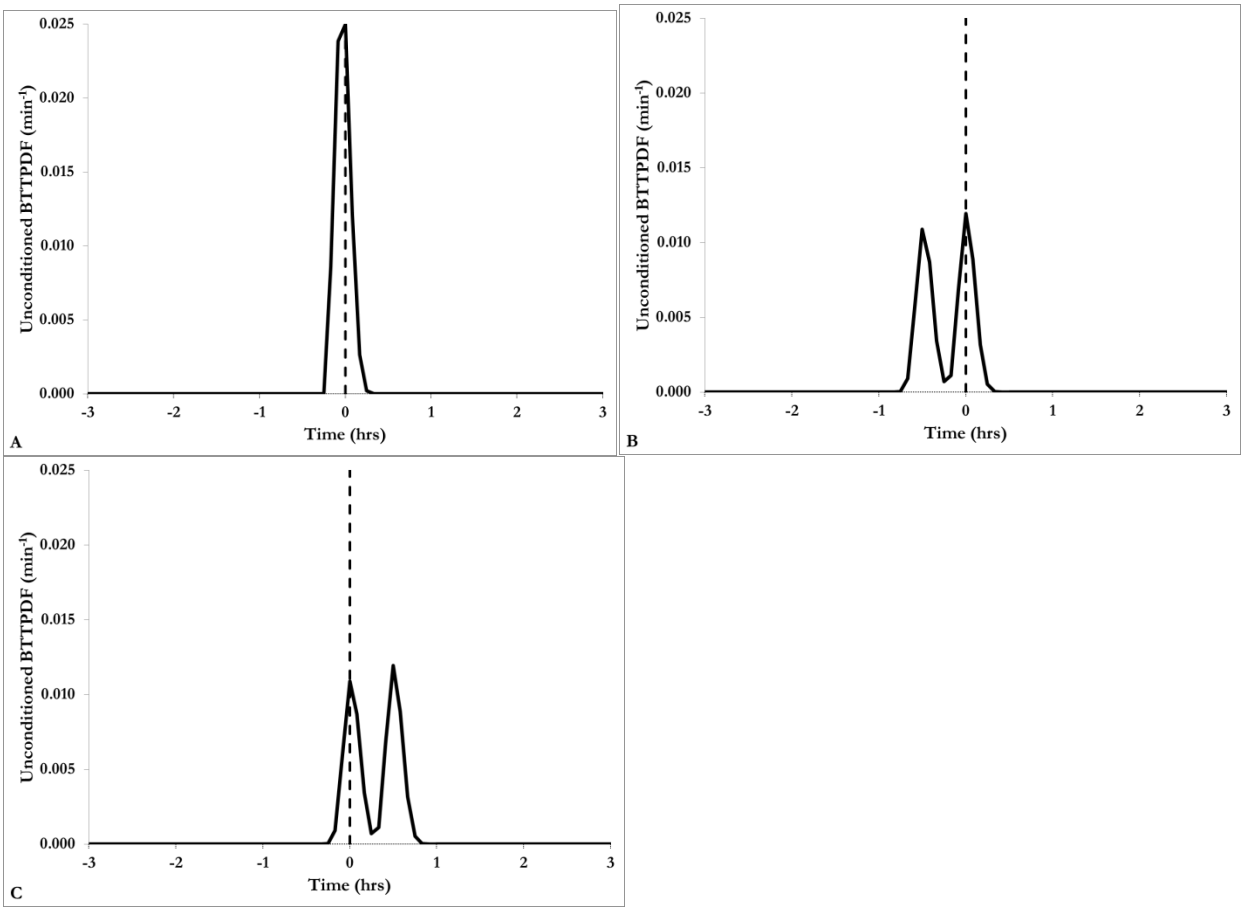

Figure 3.3. Unconditioned BTTPDF for source node 11 using (A) observation 21 at *t* =4:05, (B) observation 47A at 7:50, and (C) observation 47B at 8:20. The dashed line indicates the true release time (*t* = 0:00).

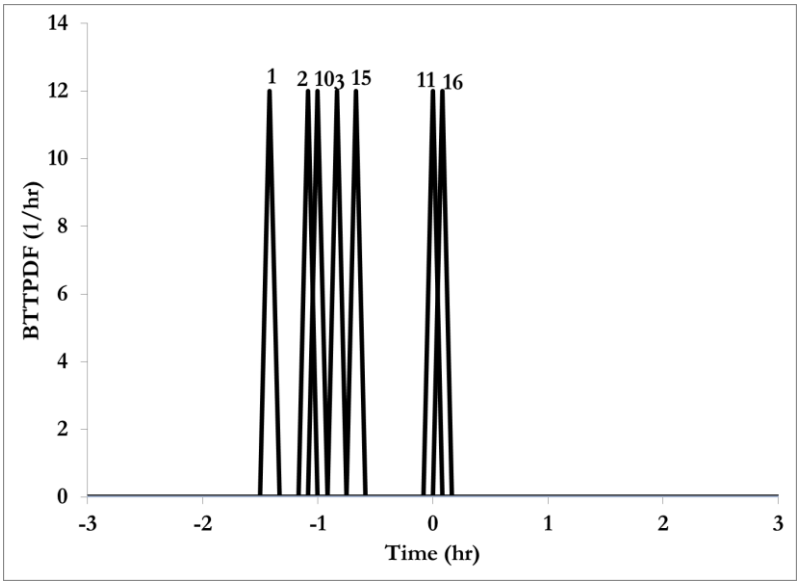

Figure 3.4. Conditioned BTTPDFs using ideal sensor data (from Neupauer et al. 2010). The number above each curve denotes the node number of the potential source node.

The non-zero unconditioned BTTPDFs in Figure 3.3 indicate the times when the contaminant that was observed at one of the sensors may have entered the system. For all three observations, the true source release time is chosen as a potential release time. Figure 3.3 also demonstrates how the adjoint state relates to the concentration observations. In the forward simulation, node 21 has one concentration peak and node 47 has two peaks; the same results are found for the adjoint states demonstrating how the flowpaths of the contaminant and the adjoint load are the same. The conditioning step combines the unconditioned BTTPDFs from all three observations into a joint BTTPDF where only the common release times are carried through; in Figure 3.3, only the adjoint states near t = 0:00 are common to all three observations.

The non-zero conditioned BTTPDFs in Figure 3.4 indicate the node/time combinations where the contaminant in the three observations may have entered the system. The true source node and release time (node 11 at t = 0:00) is selected as a potential source node, however, multiple other potential contamination scenarios (e.g., source node 16 at release time 0:05) are also identified. Using Bayes' theorem, Neupauer et al. (2010) demonstrated that β_T, which is used to normalize the

conditioned BTTPDF and ensure that the total probability is unity in (8), can be used to determine the most probable source node. The most probable source node has the largest value of β_T. Based on the β_T values in Table 3.1, node 11 is the most likely source node. Note that our method for calculating the adjoint states is different than the method used by Neupauer et al. (2010), so the β_T values are slightly different; Neupauer et al. (2010) calculated the adjoint states by simulating the adjoint load as a flow-paced booster with magnitude $\left(\Delta t \sum_{i \in \mathcal{U}_i = \ell} Q_i\right)^{-1}$ while we simulated the adjoint load a mass booster with magnitude $(\Delta t)^{-1}$.

Table 3.1: β_T Values Using Observations from Perfect Sensors (Units are $L^3/mg^2/hr^2$).

Node	**1**	**2**	**3**	**10**	**11**	**15**	**16**
β_T	2.5E-118	1.7E-156	1.1E-206	1.1E-108	**1.9E+4**	8.3E-157	7.2E-304

3.3. Probabilistic Approach for Source Identification Using Fuzzy Sensor Data

3.3.1. Theory for Using Fuzzy Sensor Data in an Adjoint-Based Probabilistic Approach

The equations for the conditioned BTTPDF in (3.7) and (3.8) are based on the assumption that the measured concentration is known exactly. With fuzzy sensors, the concentration at the sensor is only known as a specific range of concentrations, so (3.7) and (3.8) do not hold. Here we derive the equivalent expression for the conditioned BTTPDF for fuzzy sensor measurements. The approach is based on the derivation of conditioned BTTPDFs in Neupauer and Lin (2006) and Neupauer and Records (2009).

The conditioned BTTPDF can be expressed as a marginal probability density function, given by

$$f_{T|\mathbf{C}}(\tau \mid \mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b}) = \int_m f_{M,T|\mathbf{C}}(m, \tau \mid \mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b}) dm \tag{3.10}$$

where $f_{M,T|\mathbf{C}}(m,\tau \mid \mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b})$ is the joint PDF of source mass and release time, given the vector of measured concentrations being bound by a lower bound vector **a** and an upper bound vector **b**. Using Bayes' theorem, this joint PDF of source mass and release time can be expressed as

$$f_{M,T|\mathbf{C}}(m,\tau \mid \mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b}) = \frac{P\left(\mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b} \mid m,\tau\right) f_{M,T}(m,\tau)}{\iint P\left(\mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b} \mid m,\tau\right) f_{M,T}(m,\tau)\,dm\,d\tau} \tag{3.11}$$

where $P\left(\mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b} \mid m,\tau\right)$ represents the probability that the vector of measured concentrations is bound by **a** and **b**, given the source mass m and the release time τ, and $f_{M,T}(m, \tau)$ is the joint PDF of the source mass and release time. Assuming the source mass and release time are independent, this joint PDF can be expressed as

$$f_{M,T}(m,\tau) = f_M(m)\, f_T(\tau) \tag{3.12}$$

where $f_M(m)$ is the PDF of the source mass, which can be assumed to be a uniform distribution in the absence of any other information, and $f_T(\tau)$ is the PDF of the source release time in the absence of any other information, which can be taken as (Neupauer and Lin, 2006)

$$f_T(\tau) = \prod_{n=1}^{N_s} f_T(\tau; \ell, j_n, \tau_{sn}) \tag{3.13}$$

where $f_T(\tau; \ell, j_n, \tau_{sn})$ is the unconditioned BTTPDF for observation n. Assuming that each observation is independent, the probability $P\left(\mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b} \mid m,\tau\right)$ can be written as

$$P\left(\mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b} \mid m,\tau\right) = \prod_{n=1}^{N_s} P\left(a_n < \hat{c}_{jn}{}^* < b_n \mid m,\tau\right) \tag{3.14}$$

where $P\left(a_n < \hat{c}_{jn}{}^* < b_n \mid m,\tau\right)$ is the probability that the concentration of sample n is between a_n and b_n, where a_n and b_n are the boundaries of the fuzzy sensor range for sample n. Using (3.9) as the PDF of the measured concentration, the probability $P\left(a_n < \hat{c}_{jn}{}^* < b_n \mid m,\tau\right)$ is given by

$$P\left(a_n < \hat{c}_{jn}^{\,*} < b_n \mid m,\tau\right) = \int_{a_n}^{b_n} f_{\hat{C}_{jn}^*|M,T}\left(\hat{c}_{jn}^* \mid m,\tau;\ell\right) d\hat{c}_{jn}^* = \int_{a_n}^{b_n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(C-\hat{C}\right)^2}{2\sigma^2}\right\} d\hat{c}_{jn}^* \tag{3.15}$$
$$= F_{\hat{C}_{jn}^*|M,T}\left(b_n \mid m,\tau;\ell\right) - F_{\hat{C}_{jn}^*|M,T}\left(a_n \mid m,\tau;\ell\right)$$

where $F_{\hat{C}_{jn}^*|M,T}\left(b_n \mid m,\tau;\ell\right)$ is the cumulative distribution function (CDF) of the upper bound of the concentration range.

Substituting (3.11) – (3.15) into (3.10), we obtain the final expression for the conditioned BTTPDF for fuzzy sensor measurements, given by

$$f_{T|\mathbf{C}}(\tau \mid \mathbf{a} < \hat{\mathbf{c}}^* < \mathbf{b}) = \beta_{TF}^{-1} \int_m \prod_{n=1}^{N_s} \left\{ f_T\left(\tau;\ell,j_n,\tau_{sn}\right) \left[F_{\hat{C}_{jn}^*|M,T}\left(b_n \mid m,\tau;\ell\right) - F_{\hat{C}_{jn}^*|M,T}\left(a_n \mid m,\tau;\ell\right) \right] \right\} dm \tag{3.16}$$

where

$$\beta_{TF} = \int_0^\infty \int_m \prod_{n=1}^{N_s} \left\{ f_T\left(\tau;\ell,j_n,\tau_{sn}\right) \left[F_{\hat{C}_{jn}^*|M,T}\left(b_n \mid m,\tau;\ell\right) - F_{\hat{C}_{jn}^*|M,T}\left(a_n \mid m,\tau;\ell\right) \right] \right\} dm\ d\tau \tag{3.17}$$

is defined to ensure that the total probability is unity.

3.3.2. Example Using Fuzzy Sensors

We tested our adjoint-based probabilistic method for fuzzy sensors using the same system used in Section 3.2.2. The contaminant concentration observations in Figure 3.2 assume that ideal sensors are available. We convert these observations to fuzzy sensor readings by specifying the concentration ranges that the fuzzy sensors measure. For this example, we specify the low range as $0 < c \leq 5$ mg/L, the medium range as 5 mg/L $< c \leq 10$ mg/L, and the high range as 10 mg/L $< c \leq 15$ mg/L, as shown in Figure 3.2. We use the three peak concentrations to test our adjoint-based probabilistic method to obtain a reading of High at t = 4:05 for node 21 (Observation 1), a reading of Medium at t = 7:50 for node 47 (Observation 2), and a reading of Medium at t = 8:20 for node

47(Observation 3). We use the unconditioned BTTPDFs from the previous example (Figure 3.3) in (3.16) and (3.17) along with the CDF of measured concentration with a model uncertainty σ of 0.04 mg/L and lower and upper bound vectors of **a**=[10 mg/L, 5 mg/L, 0 mg/L] and **b**=[15 mg/L, 10 mg/L, 5mg/L] to obtain the conditioned BTTPDFs shown in Figure 3.5 and the β_{TF} shown in Table 3.2 as Case 1.

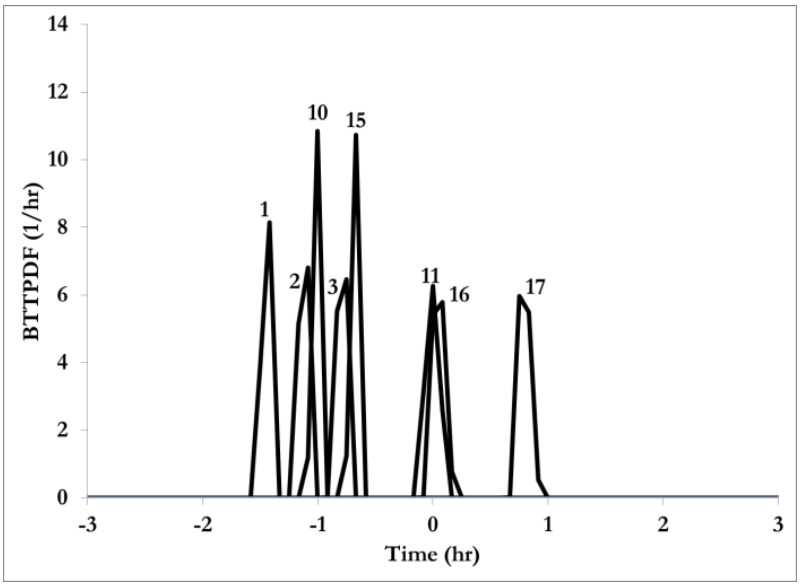

Figure 3.5. Conditioned BTTPDF for all potential source nodes (number above curve denotes the node number) using fuzzy sensor observations from observation nodes 21 and 47.

In Figure 3.5 the non-zero values of the conditioned BTTPDF shows the potential release times of contamination from the source. For the true source node (node 11), the true release time is identified as the most likely release time, but the range of possible release times is from t = -0:04 to 0:02. When perfect sensors were used, the only calculated non-zero conditioned BTTPDF value occurred at t = 0:00. In general, the range of potential release times identified using observations from fuzzy sensors is broader for each source node as compared to the potential release times obtained using perfect sensors.

Table 3.2. β_{TF} Values Using Observations from Fuzzy Sensors (Units are $L^3/mg^2/hr^2$).

Case		1	2	3	4	5
Range	Low	0 - 5 mg/L			0-5 mg/L	0-10 mg/L
	Medium	5 - 10 mg/L			5-10 mg/L	10-20 mg/L
	High	10 – 15 mg/L			10-100 mg/L	20-30 mg/L
σ (mg/L)		0.04	0.40	4.0	0.04	0.04
Node	1	6.0E+2	6.9E0	1.3E0	3.7E-83	4.1E+2
	2	1.0E+2	1.0E0	1.5E-1	0	4.4E+1
	3	1.1E+2	1.1E0	1.6E-1	0	5.0E+1
	4	0	0	1.1E-9	0	0
	10	5.0E+2	8.1E0	3.5E-1	3.5E+2	1.2E+3
	11	3.2E+2	5.7E0	2.0E-1	2.6E+2	3.2E+2
	15	5.0E+2	8.3E0	3.8E-1	3.6E+2	5.0E+2
	16	4.1E+2	6.8E0	2.7E-1	3.4E+2	4.1E+2
	17	**3.1E+3**	**5.5E+1**	**2.0E0**	**2.5E+3**	**3.1E+3**

While Figure 3.5 shows that the true contamination scenario (node 11 at t = 0:00) is selected as a potential contamination scenario, various other combinations of potential source nodes and release times are also identified. Based on the β_{TF} values in Table 3.2, node 17 is the most likely source node. While this is not the true source node, a direct flowpath exists between node 17 and node 11, the true source node (see Figure 3.1). Note also that all values of β_{TF} are within about an order of magnitude of each other, indicating that no single node stands out as the obvious source node. With the loss of information in the fuzzy sensor measurements, the method cannot distinguish between the potential source nodes.

We further tested this method using different values for the model uncertainty (σ = 0.4 mg/L and 4.0 mg/L, Cases 2 and 3 respectively). Table 3.2 presents the β_{TF} values for the potential source nodes using different values of model uncertainty. The β_{TF} values for nodes 1-3, 10, 11, and 15-17 remain within about one order of magnitude, even when the model uncertainly changes by two orders of magnitude. Neupauer et al. (2010) used perfect sensors and found that β_T values varied by over 200 orders of magnitude when the model uncertainty was 0.04 mg/L and by about

three orders of magnitude when the model uncertainty was 0.4 mg/L. This shows that with perfect sensor data, a low model uncertainty allows the method to identify a single potential source node as the most likely, but as the model uncertainty increases, the method cannot differentiate as well between several likely source nodes due to the loss of information with increasing model uncertainty. We show here, however, that model uncertainty is less important when fuzzy sensors are used; with fuzzy sensors, the more significant loss of information comes from observing the contamination only as precisely as the fuzzy sensor ranges.

We illustrate how the loss of information from the fuzzy sensor ranges affects the identified source nodes by using two different sets of fuzzy sensor ranges. In Case 4, we increase the width of the high range from 10-15 mg/L to 10-100 mg/L while keeping the low and medium ranges unchanged. In Case 5, we increase the width of all of the ranges from 5 mg/L to 10 mg/L. Table 3.2 shows that the magnitude and range of the β_{TF} values are affected by the choice concentration ranges. For three equal concentration ranges that are 5 mg/L wide (i.e., Case 1), all of the β_{TF} values are within about one order of magnitude and eight potential source nodes are identified. When we change the upper limit of the concentration range from 15 mg/L to 100 mg/L, we calculate non-zero β_{TF} values at six nodes, two fewer nodes than we found in Case 1; this demonstrates how increasing the width of a measurement range can leads to a loss of information. In Case 5, when we change the width of all of the concentration ranges to 10 mg/L, the number of potential source nodes is eight, the same number of potential source nodes as in Case 1 and the β_{TF} values calculated in Case 5 are very similar to the values in Case 1 these results demonstrate how the increasing the width of all concentration ranges does not have as great of an effect on the results as expanding just the high range did in Case 4.

3.4. Probabilistic Approach for Source Identification Using Non-Detect Measurements

3.4.1. Theory for Using Non-Detect Measurements in an Adjoint-Based Probabilistic Approach

While fuzzy sensors decrease the amount of information available for source identification, non-detect measurements, which have not been used in previous adjoint methods, provide information to help fill in the data gaps. In this section, we describe the theory behind using non-detect measurements in an adjoint-based method. Let $g(\tau; \ell, j, \tau_Z)$ be the random time τ that a water particle that was at observation node j at time τ_Z could have been at source node ℓ. If this water particle is contaminated, then this PDF also represents the random time that a contaminant could have been released at node ℓ. If this water particle is not contaminated, then this PDF represents random times when a contaminant could not have been released at node ℓ. This PDF is related to the adjoint state of concentration, as in (3.6), and is given by

$$g_T(\tau; \ell, j, \tau_Z) = \psi^*_\ell(\tau; j, \tau_Z) \sum_{i \in u_j = \ell} Q_i(\tau), \tag{3.18}$$

where $\psi^*_\ell(\tau; j, \tau_Z)$ is the adjoint state obtained by solving (3.4) with the load term given by (3.5).

If a range of sampling times exists, $\tau_{Z\min} < \tau_Z < \tau_{Z\max}$, for which the water passing through node j was not contaminated, then the probability that one of the water particles was at node ℓ at a time t is given by

$$G(\tau; \ell, j, \tau_{Z\min}, \tau_{Z\max}) = \int_{\tau_{Z\min}}^{\tau_{Z\max}} g(\tau; \ell, j, \tau'_Z)\, d\tau'_Z \tag{3.19}$$

Since none of these water particles is contaminated, this probability also represents the probability that the contaminant was not at node ℓ at a random time τ. Therefore, the complement of this probability, $G^c(\tau; \ell, j, \tau_{Z\min}, \tau_{Z\max})$ represents the probability that the contaminant could have been at node ℓ at random time τ. This complement is given by

$$G^c(\tau;\ell,j,\tau_{Z\min},\tau_{Z\max}) = 1 - G(\tau;\ell,j,\tau_{Z\min},\tau_{Z\max}) \tag{3.20}$$

We can obtain $G(\tau;\ell,j,\tau_{Z\min},\tau_{Z\max})$ directly by using an equivalent of (3.18), given by

$$G_T(\tau;\ell,j,\tau_Z) = \Psi_\ell^*(\tau;j,\tau_Z)\sum_{i\in\mu_i=\ell} Q_i(\tau) \tag{3.21}$$

where $\Psi_\ell^*(\tau;j,\tau_Z)$ is the adjoint state obtained by solving an equivalent form of (3.4) with the load term U_ℓ^* obtained by integrating the load term in (3.5) over the range of sampling times, to obtain a new load term defined as

$$U_\ell^* = \begin{cases} 0 & j \neq \ell \\ H(\tau-\tau_{Z\min}) - H(\tau-\tau_{Z\max}) & j = \ell \end{cases} \tag{3.22}$$

where $H(\tau)$ is the Heaviside function.

The conditioned backward travel time PDF, $f_{T|\widehat{C}^*}(T\,|\,\widehat{\mathbf{C}}^*;\ell)$, represents the random time at which contamination observed at nodes $\mathbf{J}$ at times $\mathbf{T}_s$ could have been at the node $\boldsymbol{\ell}$. Since $G^c(\tau;\boldsymbol{\ell},j,\tau_{Z\min},\tau_{Z\max})$ represents the probability that the contaminant could have been at node $\boldsymbol{\ell}$ at random time τ, the final conditioned BTTPDF that accounts for both the measured concentration and the non-detects is given by

$$f_{T|\widehat{C}^*,\tau_{Z\min},\tau_{Z\max}}(T\,|\,\widehat{\mathbf{C}}^*;\ell,\tau_{Z\min},\tau_{Z\max}) = G^c(\tau;\ell,j,\tau_{Z\min},\tau_{Z\max})\, f_{T|\widehat{C}^*}(T\,|\,\widehat{\mathbf{C}}^*;\ell) \tag{3.23}$$

In theory, the probability $G^c(\tau;\boldsymbol{\ell},j,\tau_{Z\min},\tau_{Z\max})$ varies between 0 and 1, where a value of 1 indicates a probability of unity that contaminant could not have been released from node $\boldsymbol{\ell}$ at time τ. This probability is related to the adjoint state through (3.21). In the pipe network, this adjoint state is propagated upstream from the observation node j through the pipe network to node $\boldsymbol{\ell}$. As it travels through the network, it passes through junctions where it can be diluted by water entering the junction from other pipes. Thus, the maximum value of the probability will often be less than unity. If the simulation accounted for all of the non-detects at all of the nodes, then the water

entering the junction from other pipes would also contain non-zero values of the adjoint state, and dilution would not occur, resulting in a probability with a maximum value of unity. However, it is not practical to run a simulation that accounts for all non-detects at all nodes because the network typically contains only a few sensors, so the nodal concentrations at most nodes are not known. Even if they were known, using all non-detects would be cumbersome, and false negatives could eliminate the true source node.

If we implement the theory as described above, using only non-detects from one node, the maximum value of the probability will be less than unity, indicating that the probability of contamination not being at node ℓ at time τ is less than unity, and therefore from (3.20) the probability that contamination could have come from node ℓ at time τ is greater than zero. Practically, this means that even though no contamination was observed at node *j* at a time for which contamination that was released from node ℓ at time τ would have arrived at node *j*, the model results would still indicate that contamination could have been released at node ℓ at time τ. To avoid this consequence, we adjust the probability, $G(\tau; \ell, j, \tau_{Zmin}, \tau_{Zmax})$, to have a value of unity wherever the value obtained from (3.21) is non-zero. Thus, this probability and its complement, $G^c(\tau; \ell, j, \tau_{Zmin}, \tau_{Zmax})$, have values of only zero or unity.

3.4.2. Example Using Fuzzy Sensors and Non-Detect Measurements

As an example, we use the system shown in Figure 3.1 with the contaminant concentration observations and fuzzy sensor ranges shown in Figure 3.2. We defined "non-detects" as any times at which the observed contaminant concentration is zero. Figure 3.2 shows that node 21 has non-detects from 0:00 to 3:45 and 4:20 to the end of the 24-hour simulation. Similarly, node 47 has non-detects from 0:00 to 7:25 and 8:40 to the end of the simulation. While non-detects are available at observation nodes 21 and 47, we only use the non-detect measurements for node 21.

We use (3.4) to calculate the adjoint states, with the adjoint load defined by (3.5), and we use this adjoint state in (3.21) and (3.20) to calculate $G^{\ell}(\tau;\boldsymbol{\ell},j,\tau_{Z\min},\tau_{Z\max})$ which represents the probability that the contaminant could have been at node $\boldsymbol{\ell}$ at backward time τ. We rescale this probability to have values of 0 and 1 (as described in 3.5.1) to obtain the probability shown in Figure 3.6 for $\boldsymbol{\ell}$ = 11. This probability is non-zero only at t = 0:00, indicating that a release from node 11 could have only occurred at t = 0:00. For all other release times from node 11, a non-zero concentration would have been observed at node 21 at some time at which a non-detect was observed. We used the probability in (3.22) with the conditioned BTTPDFs from Figure 3.5 to calculate the conditioned BTTPDFs in Figure 3.7.

Figure 3.7 shows the conditioned BTTPDF with the non-zero values indicating the potential release times for each source node. We are able to use the non-detect information to narrow down the number of potential source nodes from eight to three potential source nodes, and also decrease the range of potential release times. It is important to note that we accomplished this reduction by using non-detect measurements which were already available from the sensor measurements at the observation nodes.

In Figure 3.7, three nodes have non-zero BTTPDF values: node 11 at t = 0:00 (the true scenario), node 16 at 0:05, and node 17 between 0:45 and 0:50. The β_{TF} values in Table 3.3 indicate node 17 is the most probable source node, but the values for nodes 11 and 16 are within an order of magnitude indicating that any of these nodes might be the true source node. So, while this method is able to determine that the true contamination scenario (node 11 at t = 0:00) is a potential scenario, it is unable to determine that node 11 is the most probable source node. The benefit of using the non-detects, however, is demonstrated in the reduction of the number of potential source nodes and range of potential release times.

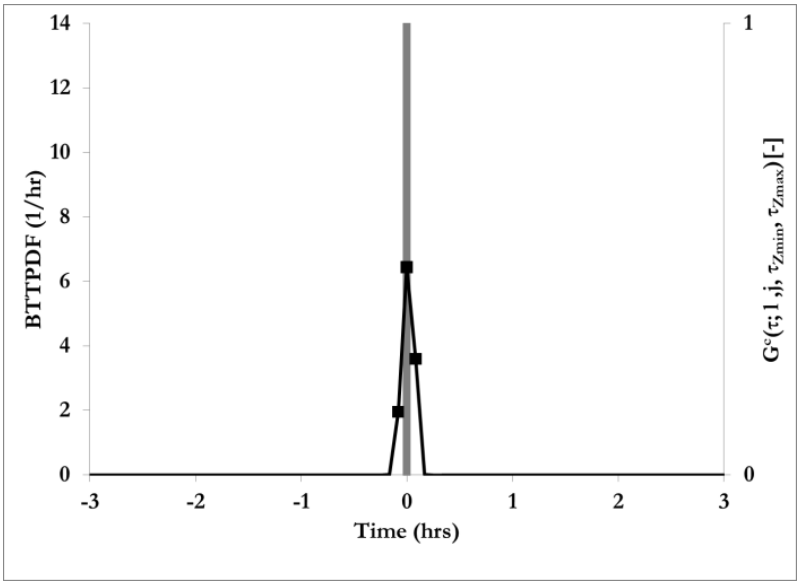

Figure 3.6. Quantities used to calculate the conditioned BTTPDF for node 11 using both fuzzy sensor data and non-detects. The black line represents the conditioned BTTPDF for the fuzzy sensor data in Figure 3.2. The gray line represents $G^c(\tau, l, j, \tau_{zmin}, \tau_{zmax})$ for the non-detects at observation node 21.

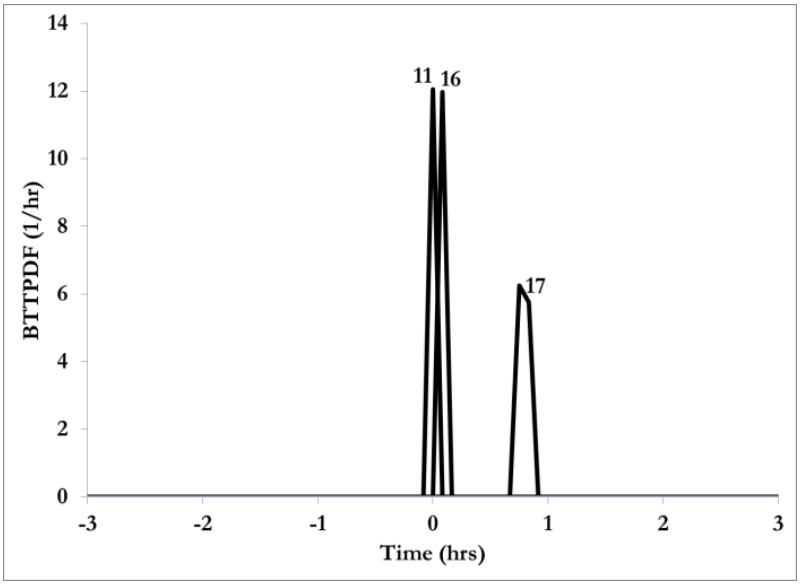

Figure 3.7. Conditioned BTTPDF for all potential source nodes (number above the curve is the node number) using fuzzy sensor observations at nodes 21 and 47, and non-detect measurements at node 21.

Table 3.3. β_{TF} Values Using Fuzzy Sensors and Non-Detects (Units are $L^3/mg^2/hr^2$).

Node	**11**	**16**	**17**
β_T	1.6E+2	2.0E+2	**3.0E+3**

3.5. Probabilistic Approach for Source Identification Using Binary Sensor Data

3.5.1. Theory for Using Binary Sensor Data in an Adjoint-Based Probabilistic Approach

For binary sensors, the theory is similar to the theory for using perfect sensors, but, we do not have a measured concentration, so we do not calculate the PDF of obtaining the observed concentration (3.9) or the conditioned BTTPDF (3.7). Let $f(\tau; \boldsymbol{\ell}, j, \tau_S)$ be the random time τ that a water particle that was at observation node j at time τ_S could have been at source node $\boldsymbol{\ell}$, which is calculated using (3.6). If this water particle is contaminated, then this BTTPDF also represents the random time that a contaminant could have been released at node $\boldsymbol{\ell}$.

Similar to the adjoint method that uses perfect sensor or fuzzy sensor data, the observations are assumed to be taken at individual times and the adjoint load in (3.4) and (3.5) is an instantaneous release of the adjoint load at the observation node and sampling time. All non-zero observations for binary sensors are the same; each non-zero observation only indicates that contaminant is present or outside a desired range of values; thus, instead of using the peak observations, we use a non-zero observation at the beginning, middle, and end of the range of non-zero observations for an observation node.

If multiple observations are used, the BTTPDF, $f(\tau; \boldsymbol{\ell}, j, \tau_S)$, can be calculated for each one. We assume the contaminant entered the system at only one source node and release time, so all sensor observations are traceable back to an instantaneous release at a single source node (which has not yet been determined). Let $\mathbf{J} = \{j_1, j_2, \ldots j_{N_s}\}$ be a set of N_s sensor nodes at which contamination is observed, and let $\mathbf{T}_s = \{\tau_{s1}, \tau_{s2}, \ldots \tau_{sN_s}\}$ be the set of observation times at each sensor. We define the joint BTTPDF, $f\left(\tau; \ell, \mathbf{J}, \mathbf{T}_s\right)$, as the random time τ that contamination that was observed at

sensor nodes **J** at times $\mathbf{T}_s$ could have been released at node $\boldsymbol{\ell}$. This joint BTTPDF is given by (Neupauer and Records 2009)

$$f(\tau;\ell,\mathbf{J},\mathbf{T}_s)=\alpha_T\prod_{n=1}^{N_S}f(\tau;\ell,j_n,\tau_{sn}) \tag{3.24}$$

$$\alpha_T^{-1}=\int_{\tau}\prod_{n=1}^{N_S}f(\tau;\ell,j_n,\tau_{sn})d\tau \tag{3.25}$$

where $f(\tau;\ell,j_n,\tau_{sn})$ is the BTTPDF from (3.7) for the nth sensor observation, and α_T is used to ensure that the total probability is unity.

3.5.2. Example Using Binary Sensors

We used the system shown in Figure 3.1 for the binary sensor method example as well. The contaminant concentration observations shown in Figure 3.2 also hold true, but all non-zero values are converted to "1," which indicates that contamination is present, or above a threshold value. In the form of binary sensor data, non-zero contaminant observations exist from 3:55 to 4:15 at node 21 (observation set 1) and from 7:35 to 8:35 at node 47 (observation set 2). We use three non-zero observation times for each node: (1)t = 3:55, 4:05, and 415 for node 21, and, (2) t = 7:35, 8:05, and 8:35 for node 47.

We use (3.4) to determine the adjoint state, with the adjoint load defined in (3.5). Figure 3.9 shows the joint BTTPDF for source nodes 1-5, 10-12, and 15-19, using observation nodes 21 and 47 is shown in Figure 3.9. Node 1, 2, 3, 10, 11, 15, 16, and 17 all have non-zero joint probabilities indicating that they are potential source nodes; the non-zero values in Figure 3.9 indicate potential release times for each node with the most likely release time occurring at the maximum value of the joint BTTPDF. The true contamination scenario, node 11 at t = 0:00 is one of the potential scenarios. Unlike perfect sensors and fuzzy sensors, we do not have a concentration measurement

(perfect sensors) or concentration range (fuzzy sensor), so we cannot condition these values to determine the most probable source node. We can, however, use the non-detect measurements to narrow down the potential scenarios.

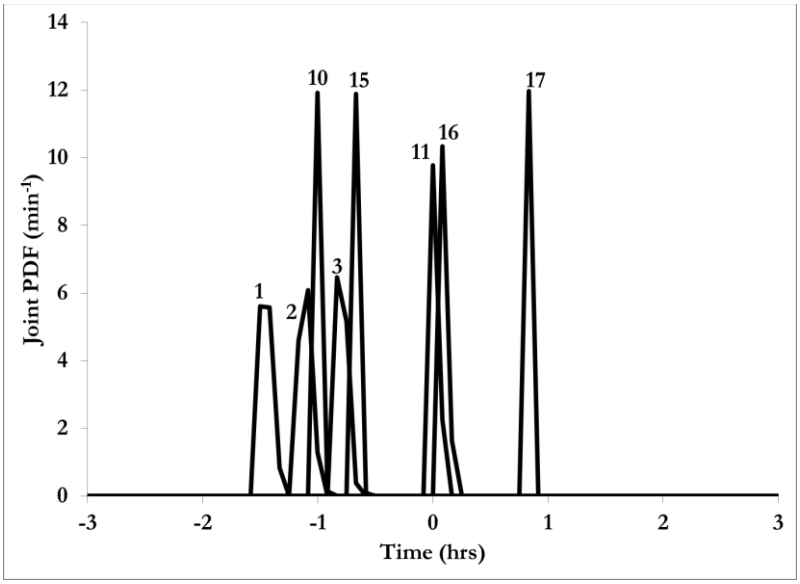

Figure 3.8. Joint BTTPDF for all source nodes using binary sensors.

We use the non-detect binary file as described in Section 3.4.2 in conjunction with joint BTTPDF in Figure 3.9. The result are shown in Figure 3.10 which has non-zero joint BTTPDF values for the all of the potential source nodes, but with potential release time ranges that are narrower than when non-detect measurements are not used. For instance, for node 2, the range of release times when non-detect measurements are used is t = -0:55 to -0:50, but the range was t = 1:30 to -0:50 when the non-detects were not used. Again, we were able to significantly reduce the number of potential source node/release time scenarios without requiring any additional information; we used non-detect information that was already available.

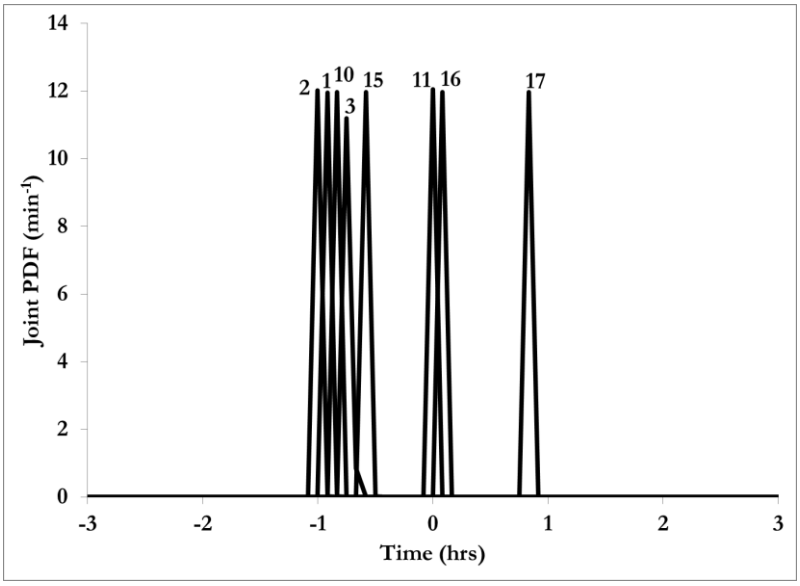

Figure 3.9. Joint BTTPDF for all source nodes using binary sensors and non-detect measurements.

3.6. Conclusions

We developed adjoint-based probabilistic methods that can be used with fuzzy sensors, binary sensors, and non-detect measurements. As expected, when the amount of information available is reduced, the number of potential source node/release time scenarios increases. The number of source nodes and the ranges of the potential release times identified using fuzzy sensors, which measure the contamination level only to within a range of concentrations, was more than what Neupauer et al. (2010) identified using perfect sensors. Similarly, the number of source nodes and the ranges of the potential release times identified using binary sensors, which provide information only about the presence or absence of contamination, was much more than the number identified using fuzzy sensors. We were, however, able to decrease the number of potential source node/release time scenarios by using the non-detect measurements.

CHAPTER 4

PROBABILISTIC SOURCE CHARACTERIZATION IN WATER DISTRIBUTION SYSTEMS WITH INCOMPLETE MIXING

Abstract

As the events of September 11, 2001 made real the threat of terrorism on public infrastructure, the priority for safer and more secure infrastructure has increased. For water distribution systems, this includes development of faster and more effective responses to drinking water distribution system contamination events. Previous work has demonstrated that adjoint methods can be used to identify the contaminant source node and release time. Adjoint methods use system sensor data (i.e., the location and concentration of contamination in the system) to backtrack information through a water distribution system and probabilistically determine the source of contamination. Previous work has assumed complete mixing at pipe junctions, but this is not the case in water distribution systems, leading to discrepancies between the modeled system and the true system. A bulk advective mixing algorithm (EPANET-BAM) has been developed that uses incomplete mixing at pipe junctions. We develop an adjoint method which incorporates incomplete mixing at the pipe junctions and test it using EPANET-BAM.

4.1. Introduction

Water utilities are tasked with providing an uninterrupted supply of potable drinking water to their service populations. The events of September 11, 2001 illuminated the potential for terrorist activities on U.S. soil and increased the focus on protecting utilities and infrastructure from such attacks (DHS 2003). Specifically, efforts have been made to harden the water infrastructure against contamination. Since the system cannot be secured well enough to remove the threat of contamination, it is important to develop ways to moderate the effects of contamination. Expedient source identification helps the water distribution system operators to respond quickly and take

countermeasures to stop the spread of the contamination and lessen the number of consumers affected.

Adjoint methods have been developed to characterize the source of contamination in water distribution systems under steady state (e.g., Neupauer et al. 2010) and transient (Wagner et al. 2013) flow conditions. Adjoint methods are able to identify multiple potential source node and release time scenarios with each observation (contaminant location, concentration, and arrival time) (Neupauer et al. 2010). Neupauer et al. (2010) demonstrated that simulations from multiple observations can be used in conjunction with the probability density function (PDF) of obtaining the observed measurement to probabilistically determine the source node. Previous adjoint methods have used EPANET (Rossman 2000) which assumes complete mixing at the nodes, but research (e.g., van Bloemen Waanders et al. 2005; Romero-Gomez et al. 2008) has shown that true water distribution systems have incomplete mixing at nodes.

van Bloemen Waanders et al. (2005) demonstrated the inaccuracy of the complete mixing model for a cross junction with two adjacent inflows and two adjacent outflows all with the same Reynolds number. A sodium chloride (NaCl) tracer was introduced into one of the inflows, while the other inflow had no tracer. If the complete mixing model is accurate, the tracer concentration in each of the outflows should be the same; however, van Bloemen Waanders et al. (2005) found that 85% of the tracer was discharged to the adjacent outflow and 15% to the opposite outflow. Romero-Gomez et al. (2008) looked at how the Reynolds number in the inflows and outflows influenced the true solute mixing at cross junctions of four pipes. They used computational fluid dynamics (CFD) simulations to evaluate at scenarios where (1) the Reynolds number is the same in all pipes, (2) the outflows are the same, but the inflows have two different Reynolds numbers, (3) the inflows are the same, but the outflows are different, and (4) all pipes have different Reynolds

numbers. In all cases, their results demonstrated that the mixing at junctions is incomplete due to limited interaction between the two inflows.

For the same scenarios simulated by Romero-Gomez et al. (2008), Austin et al. (2008) experimentally determined how mass of NaCl tracer is split between the two outflows. They found that neither the complete mixing model used in EPANET nor the CFD simulations used by Romero-Gomez et al. (2008) accurately predicted the experimental results, although the CFD simulations were closer.

EPANET assumes complete mixing at pipe junctions, so other software packages have been developed to simulate incomplete mixing. Ho (2008) developed EPANET-BAM (Sandia National Lab 2008), which uses a bulk advective model to simulate incomplete mixing at nodes that (1) are the intersection of four equal-sized pipes (i.e. cross-junctions), (2) have two adjacent inflows, and (3) have two adjacent outflows. With EPANET-BAM, incomplete mixing is simulated as a user-selected weighting between complete mixing and bulk advective mixing. Choi et al. (2008) developed AZRED, a computer program that simulates incomplete mixing at (1) cross-junctions with adjacent inflows and outflows or opposing flows; (2) double tee junctions with various inflow/outflow configurations, and (3) one tee and one wye junction. These options increase the ability to simulate the true water distribution system behaviors.

The goal of this paper is to develop an adjoint-based probabilistic method for source identification in water distribution systems that have incomplete mixing at pipe junctions. The implementation of the adjoint method presented by Neupauer et al. (2010) uses EPANET which models complete mixing at all pipe junctions. We use EPANET-BAM which allows the user to specify the mixing at some pipe junctions. While this program does not allow for incomplete mixing at all pipe junctions, we use it to show how the adjoint method performs in a water distribution system with incomplete mixing.

4.2. Theory

4.2.1. Adjoint Method

Forward transport of a conservative chemical in pipes can be modeled using

$$\frac{\partial C_i}{\partial t} + \frac{Q_i}{A_i}\frac{\partial C_i}{\partial x_i} = 0 \tag{4.1}$$

with boundary conditions given by

$$C_j^* = \frac{\sum_{i \in d_i = j} Q_i C_i \,|_{x_i = L_i} + U_j}{D_j + \sum_{i \in u_i = j} Q_i} \tag{4.2}$$

where C_i, Q_i, A_i, and x_i are the concentration, flow rate, cross-sectional area, and distance along pipe i respectively, t is time, C_j^* is the chemical concentration in the water leaving node j, d_i is the downstream node of pipe i, L_i is the length of pipe i, U_j is the mass loading rate at node j, D_j is the water demand at node j, and u_i is the upstream node of pipe i.

The adjoints of (4.1) and (4.2) are (Neupauer, 2011)

$$\frac{\partial \psi_i}{\partial \tau} - \frac{Q_i}{A_i}\frac{\partial \psi_i}{\partial x_i} = 0 \tag{4.3}$$

and

$$\psi_\ell^* = \frac{\sum_{i \in u_i = \ell} Q_i \psi_i \big|_{x_i = 0} + U_\ell^*}{\sum_{i \in d_i = \ell} Q_i} \tag{4.4}$$

respectively, where $\psi_i = \partial C_i / \partial M_\ell$ is the adjoint state of the concentration in pipe i, which represents sensitivity of the concentration in pipe i (C_i) to a source mass released at node ℓ (M_ℓ), τ is backward time, defined as $\tau = t_f - t$ where t_f is a reference time, $\psi_\ell^*(\tau; j, \tau_s) = \partial C_j^*(\tau_s) / \partial M_\ell(\tau)$ is the adjoint state of concentration at node j, which represents the sensitivity of nodal concentration at backward

time τ to a source release of mass M_ℓ at node ℓ, τ_s is the backward time at which the concentration is at node *j*, U^*_ℓ is the load term given by $U^*_\ell = \delta(\tau - \tau_s)$ for ℓ = *j* (at the sensor node), and δ(·) is a Dirac delta function.

In the adjoint method, (4.3) and (4.4) are solved once for each observation, and the resulting adjoint states are related to the conditioned backward travel time probability density function (BTTPDF) given by (Neupauer and Records 2009; Neupauer et al. 2010)

$$f_{T|\hat{C}^*}\left(\tau \mid \hat{c}^*; \ell\right) = \beta_T^{-1} \int_m \prod_{n=1}^{N_S} f_{\hat{C}^*_{jn}|M,T}\left(\hat{c}^*_{jn} \mid m, \tau; \ell\right) \left(\psi^*_\ell(\tau; j, \tau_S) \sum_{i \in u_i = \ell} Q_i(\tau) \right) dm \tag{4.5}$$

where $\hat{\mathbf{C}}^* = \{\hat{c}^*_{j1}, \hat{c}^*_{j2}, \ldots \hat{c}^*_{jNs}\}$ is a vector of N_s contaminant concentration observations, $f_{T|\hat{C}^*}\left(\tau \mid \hat{c}^*; \ell\right)$ is the conditioned BTTPDF at node ℓ that describes the random backward time τ that the contaminant that was observed in the N_s sensor observations in $\hat{\mathbf{C}}^*$ could have been released at node $\boldsymbol{\ell}$, *M* is a random source mass, *m* is a particular value of the random source mass, $f_{\hat{C}^*_{jn}|M,T}\left(\hat{c}^*_{jn} \mid m, \tau; \ell\right)$ is the PDF of obtaining the measured concentration $\hat{c}^*_{jn}(\tau_{sn})$ for a given source node ℓ, release time τ, and source mass *m*, defined as a normal distribution with a mean equal to the true concentration at node *j* for a release of mass *m* at node $\boldsymbol{\ell}$ at time τ_{sn}, and the standard deviation σ of the measured concentration. With complete mixing at all nodes, the true concentration is equivalent to $m\psi^*_{jn}(\tau; \ell, \tau_{sn})$; for incomplete mixing, this relationship is not appropriate. At this time, we have not determined the appropriate relationship, so we assume that the true concentration is equivalent to $m\psi^*_{jn}(\tau; \ell, \tau_{sn})$. In (4.5), β_T ensures that the total probability is unity, is determined using (Neupauer et al., 2010)

$$\beta_T = \int_0^\infty \int_m \prod_{n=1}^{N_S} f_{\hat{C}^*_{jn}|M,T}\left(\hat{c}^*_{jn} \mid m, \tau; \ell\right) f_T(\tau; \ell, j_n, \tau_{sn}) dm dt \tag{4.6}$$

The conditioned BTTPDF in (4.5) is calculated for each possible source node and shows the likely release time of contamination from that node. Also, β_T is calculated for each source node. It is proportional to the joint PDF of the source node and release time; thus, the node with the largest value of β_T is the most likely source node (Neupauer et al., 2010).

4.2.2. Incomplete Mixing at Pipe Junctions

We use EPANET-BAM to simulate incomplete mixing at some nodes in the water distribution system model. EPANET-BAM uses the bulk advective mixing model (Ho and Khalsa 2007) which defines the concentration in pipes downstream of a node by

$$C_D = C_A \tag{4.7}$$

$$C_C = \frac{Q_B C_B + (Q_A - Q_D) C_A}{Q_C} \tag{4.8}$$

where C_A and C_B are the contaminant concentrations at the downstream ends of pipes A and B, C_C and C_D are the contaminant concentrations at the upstream ends of pipes C and D, and Q_A, Q_B, Q_C, and Q_D are the flowrates in pipes A, B, C, and D respectively.. The pipe configuration is shown in Figure 4.1. The orientation of the figure is determined in EPANET-BAM such that $Q_A + Q_C > Q_B + Q_D$; the pipe with the highest inflow rate is always pipe A, the pipe with the highest outflow rate is always pipe C.

The mixing at the node is specified as a weighted combination of complete mixing and bulk advective mixing, controlled by a user-specified mixing parameter, s, where

$$C_{Total} = C_{Bulk} + s\left(C_{Complete} - C_{Bulk}\right) \tag{4.9}$$

where C_{Total} is concentration, C_{Bulk} is concentration using the bulk advective mixing model (4.7) and (4.8), and $C_{Complete}$ is the concentration using the complete mixing model (4.2).

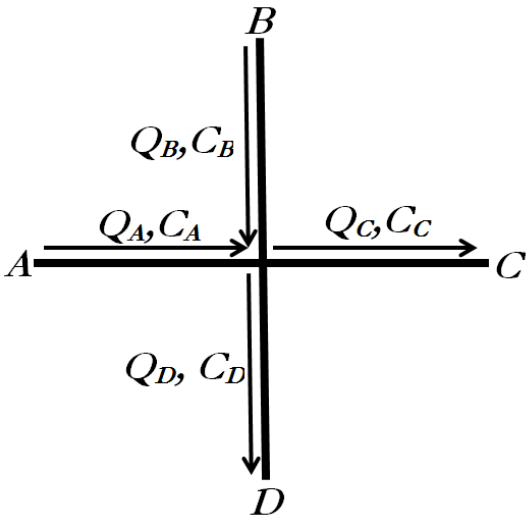

Figure 4.1. Pipe configuration for nodes with incomplete mixing. The thick lines indicate the pipes, which are labeled A, B, C, and D. The arrows indicate the direction of flow. Q_x and C_x are the flowrates and contaminant concentrations in the pipes.

The adjoints of (4.7) – (4.9) are given by (see Appendix)

$$\psi_B = \psi_C \tag{4.10}$$

$$\psi_A = \frac{Q_D\psi_D + (Q_C - Q_B)\psi_C}{Q_A} \tag{4.11}$$

$$\psi_{Total} = \psi_{Bulk} + s\left(\psi_{Complete} - \psi_{Bulk}\right) \tag{4.12}$$

The adjoint method solves (4.3) using (4.4) at nodes with complete mixing and (4.10)-(4.12) at nodes with incomplete mixing. The incomplete mixing equations are used to determine the concentration, or adjoint state, in the pipes exiting the nodes. The actual nodal concentration or adjoint state at any node is calculated using (4.2) in the forward equation or (4.4) in the backward equation

4.3. Examples

We tested our adjoint method for water distribution systems with incomplete mixing using the grid system network shown in Figure 4.2. This is the same system used by Neupauer et al. (2010). The water distribution system contains a reservoir at node 52 which supplies water to the system and demands (100 gpm) at nodes 7, 45, and 51 which cause the water (and contaminant) to flow through the system. The demands remain constant over time; the flow field is at steady-state.

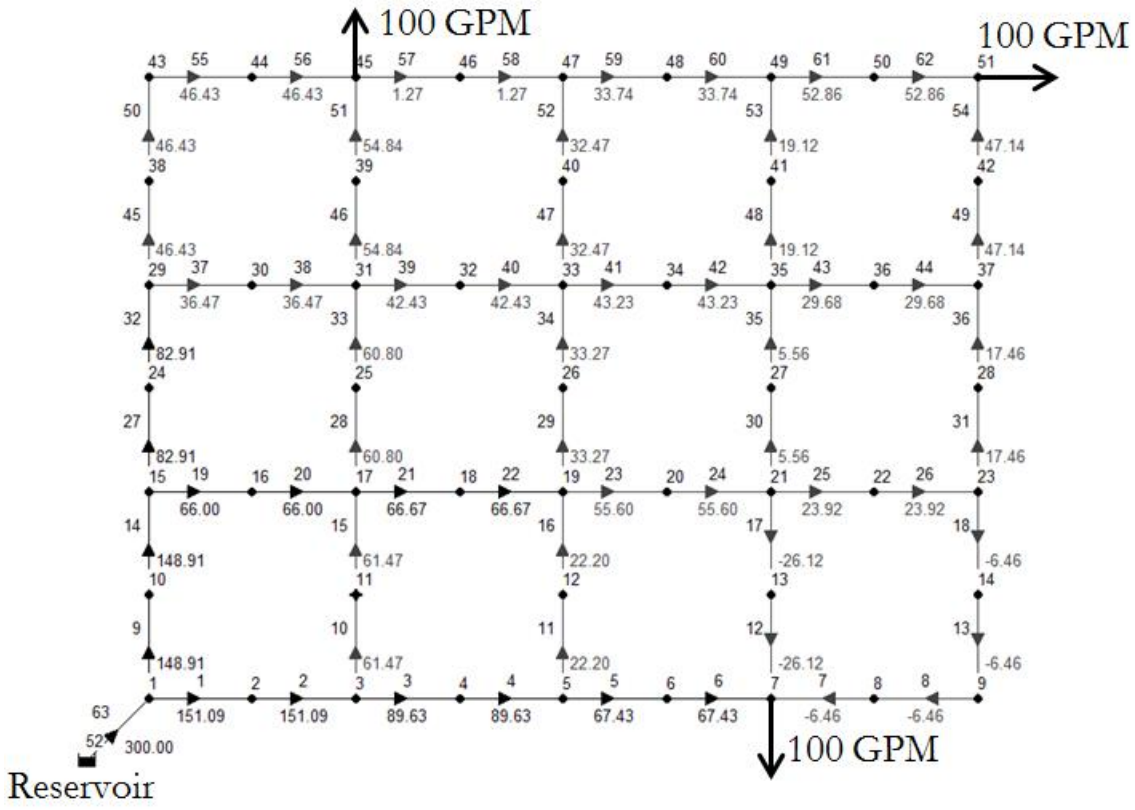

Figure 4.2. Layout of pipe network [used by Neupauer et al. 2010]. Nodes are shown as numbered circles. Pipes are shown as line segments, with pipe numbers above or to the left of the line segment. Water enters from the reservoir (Node 52) and demands of 100 gpm are placed at nodes 7, 45, and 51. Arrows denote the flow direction and the flow rates (in gpm, calculated using EPANET-BAM) are shown adjacent to the arrow.

We evaluate six different scenarios using this system, shown in Table 4.1. For each scenario, mass is released at the source node using a flow-paced booster with a concentration of 100 mg/L

between t = 0:00 and 0.05. Flow through the pipe network is simulated using EPANET-BAM. For the interior four-way nodes (nodes 17, 19, 31, 33, and 35), the mixing parameter is set as 0, 0.5, or 1, as shown in Table 4.1; we call these nodes "incomplete mixing nodes." Node 21 does not qualify as an incomplete mixing node because, while it is a junction of four equal-sized pipes, two of the pipes must be adjacent inflows and the other two pipes must be adjacent outflows to use the bulk advective mixing model; node 21 has one inflow and three outflows, therefore it does not meet the requirements. For node 21 and all other nodes which do not meet the requirements for the bulk advective mixing model, the complete mixing model (s = 1) is used.

Table 4.1. Example Scenarios for Incomplete Mixing.

Scenario	Source Node	Source Mass (g)	Mixing Parameter, *s*, at Incomplete Mixing Nodes	Mixing Parameter, *s*, Assumed in Adjoint Method
1	11	116.4	1.0	1.0
2A	11	116.4	0.0	0.0
2B	16	124.9	0.0	0.0
3A	11	116.4	0.5	0.5
3B	11	116.4	0.5	0.5
3C	11	116.4	0.5	0.75

4.3.1. Scenario 1

Scenario 1 uses complete mixing (s = 1) at all junctions. Figure 4.3A shows the contaminant concentration as a function of time for observation nodes 21 and 49 for Scenario 1. The peak contaminant concentration and observation time are used as the observations for each observation node. For node 21 the peak concentration of 12.5 mg/L occurs at 4:05. We used two of the peaks for node 49: 3.1 mg/L at 10:45 and 1.5 mg/L at 12:10. The multiple peaks for node 49 indicate multiple flowpaths from the source node (node 11) to the observation node (node 49).

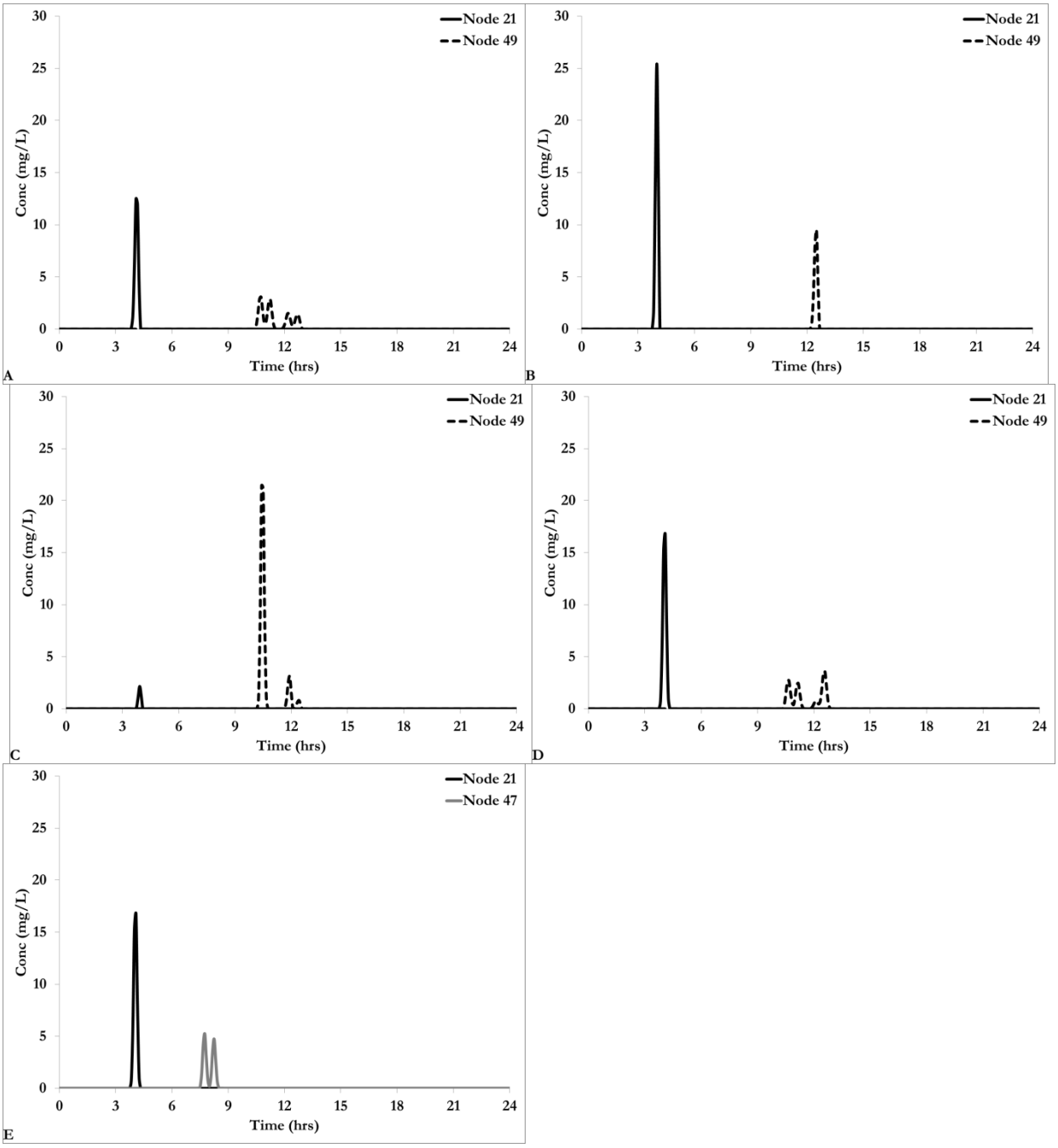

Figure 4.3. Concentration versus time plots for observation nodes 21 (solid line), 47 (gray line), and 49 (dashed line) for (A) Scenario 1, (B) Scenario 2A, (C) Scenario 2B, (D) Scenario 3A, and (E) Scenario 3B.

We used the observations at nodes 21 and 49 to calculate the adjoint states using (4.3) and (4.4). The adjoint states were used in (4.5) and (4.6) along with the PDF of measured concentration with a model uncertainty, σ, of 0.04 mg/L and a range of source masses of 10 to 500 grams to

obtain the conditioned backward travel time PDFs shown in Figure 4.4A and the β_T values shown in Table 4.2. Non-zero BTTPDFs in Figure 4.4A indicate potential source nodes and release times, with the maximum value indicating the most probable release time. For example, the maximum value of the BTTPDF for node 11 occurs at time t = 0:00 indicating that the most likely release time of contamination from node 11 is t = 0:00, the true release time. The non-zero values in Figure 4.4A indicate eight nodes are potential source nodes: 1, 2, 3, 10, 11, 15, 16, and 17. Table 4.2 shows the β_T value for each of the potential source nodes, with node 11 having the largest value and, thus, the highest probability of being the true source node. These results demonstrate that our method finds the true source node as the most likely source node and the true release time as the most likely release time.

Table 4.2. β_T Values for Incomplete Mixing Scenarios (Units are $L^3/mg^2/hr^2$).

Node	Scenario 1	Scenario 2A	Scenario 2B	Scenario 3A	Scenario 3B	Scenario 3C
***s* =**	**1.0**	**0.0**	**0.0**	**0.5**	**0.5**	**0.5**
1	1.2E-14	0	0	0	0	0
2	1.5E-39	0	0	**1.9E+4**	1.3E-107	6.4E-59
3	1.6E-31	0	0	3.3E+2	2.2E-166	**1.3E-51**
10	2.8E-27	0	0	0	0	0
11	**7.0E+3**	1.6E-76	0	1.4E-6	**3.3E+1**	3.8E-173
15	1.9E-42	0	0	0	0	0
16	9.9E-86	0	**1.2E-262**	0	0	0
17	6.0E-129	**5.0E-18**	0	0	0	0
18	0	1.2E-205	0	0	0	0

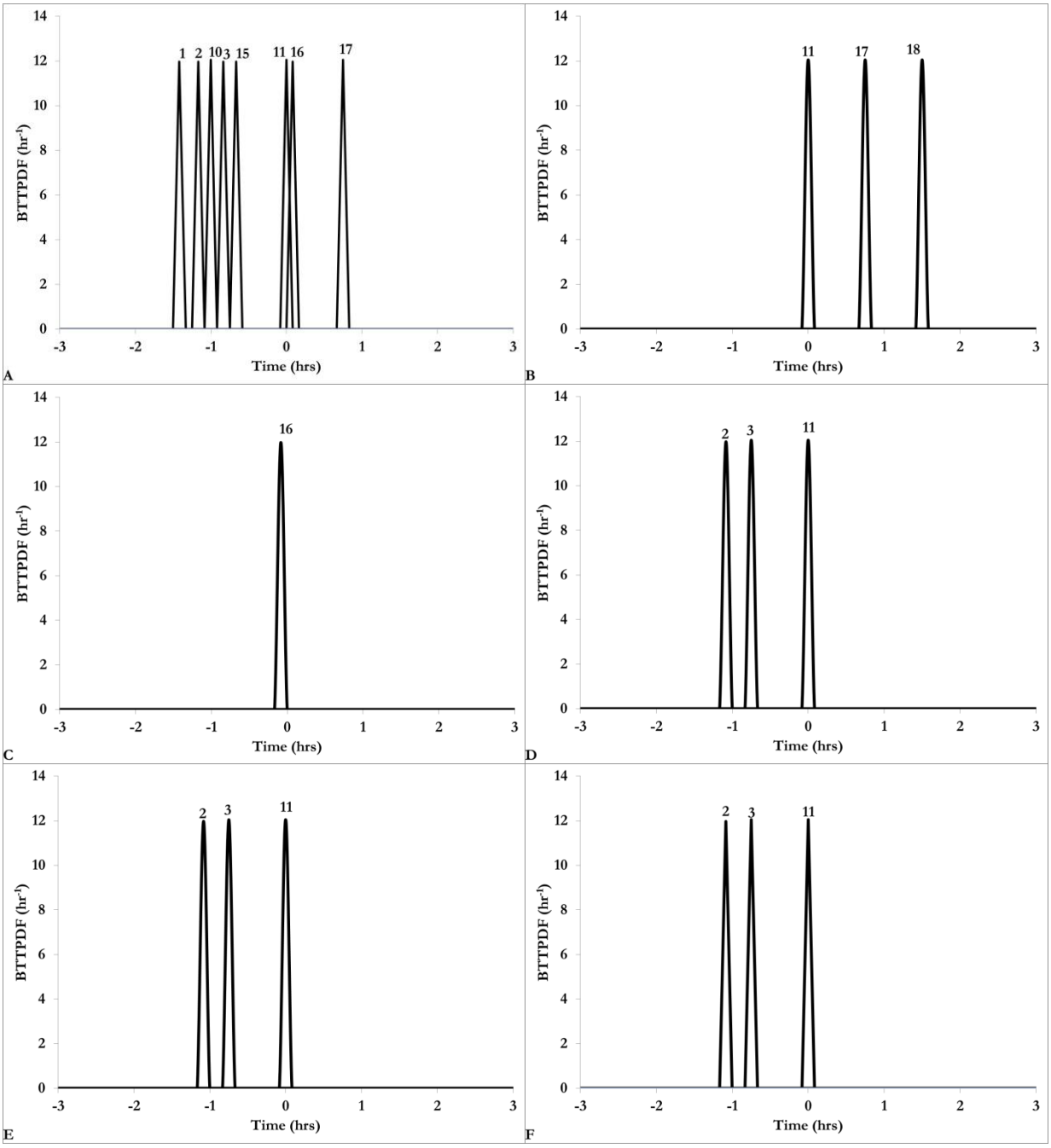

Figure 4.4. Conditioned BTTPDFs using contaminant observations from observation nodes 21 and 49 for (A) Scenario 1, (B) Scenario 2A, (C) Scenario 2B, (D) Scenario 3A, (E) Scenario 3B, and (F) Scenario 3C.

4.3.2. Scenario 2

In this scenario, the incomplete mixing nodes are assumed to have bulk advective mixing ($s = 0$) and all other nodes have complete mixing ($s = 1$). For example, node 17 in Figure 4.2 has inflows from

pipe 20 and pipe 15. The flow rate in pipe 20 is higher than pipe 15, so, using the bulk advective model, any contaminant mass that enters node 17 from pipe 15 will flow directly into pipe 21. If the mixing model was not fully bulk advective mixing, some of the contaminant mass would enter pipe 28. Node 33 has a similar scenario: the contaminant mass arrives at node 33 from pipe 34, but the flow rate through pipe 40 is great than the flow rate through pipe 34, so the contaminant mass goes directly from pipe 34 to pipe 41 and no contaminant enters pipe 47. These flowpaths are indicated in Figure 4.5. We tested two different variations using $s = 0$ at the incomplete mixing nodes: Scenario 2A with node 11 as the source node, and Scenario 2B with node 16 as the source node.

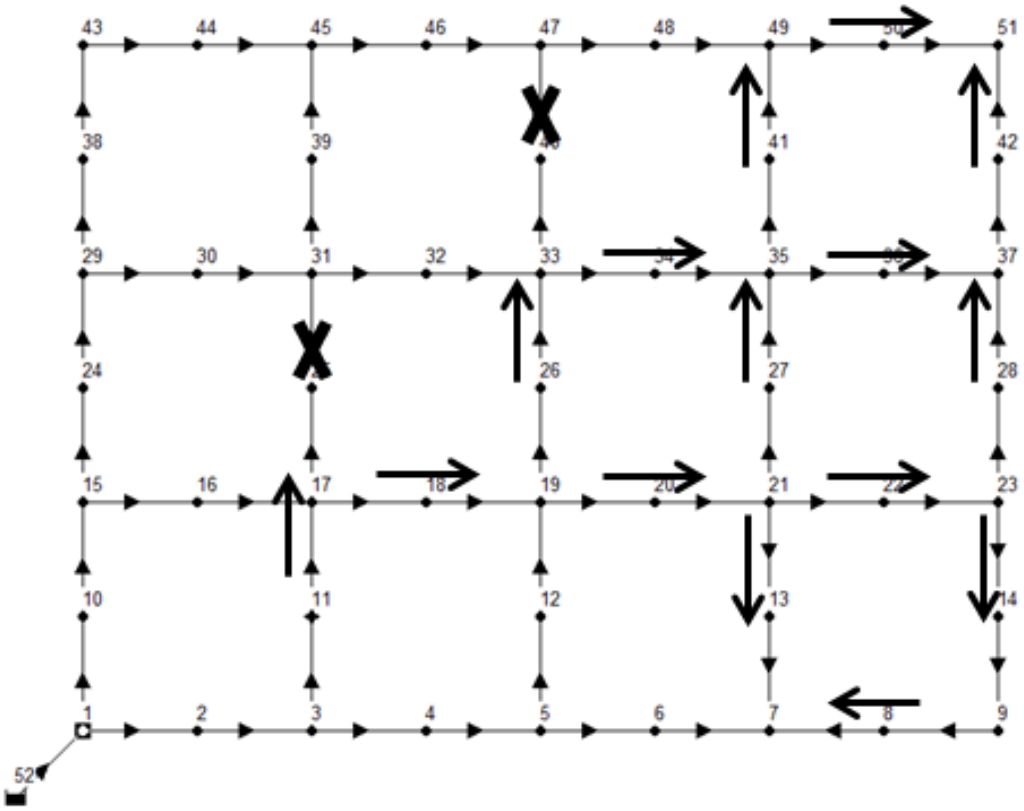

Figure 4.5. Contaminant flowpaths for Scenario 2A. The source node is node 11. The bolded arrows indicate the flowpath for the contaminant, while the crosses indicate the paths that receive water from the node, but no contaminant.

4.3.2.1 Scenario 2A

In this scenario, we evaluate the adjoint method using node 11 as the source node and observation nodes 21 and 49; the flowpath of the contaminant is indicated in Figure 4.5; some of the flowpaths observed in Scenario 1 are not an option in this scenario due to the bulk advective mixing model. These flowpaths are indicated by the "x." In addition to decreasing the number of contaminated nodes, the number of flowpaths from the source node (node 11) to the observation nodes (nodes 21 and 49), and other downstream nodes, is decreased.

Figure 4.3B shows the contaminant concentration as a function of time for observation nodes 21 and 49. We used the peak concentration observations for node 21 (25.4 mg/L at 4:00) and node 49 (9.4 mg/L at 12:30) as the observations for calculating the contamination scenarios. We can compare Figure 4.3B to Figure 4.3A to demonstrate how changing the mixing parameter at the incomplete mixing nodes affects the movement of the contaminant mass. The peak contaminant concentration at node 21 is much higher in this scenario (25.5 mg/L versus 12.5 mg/L). The contaminant concentration profile at node 49 is also much different; Scenario 1 produces four distinct peaks, with a maximum value of 3.1 mg/L, while Scenario 2A produces one peak with a maximum value of 9.4 mg/L. Both the decrease in number of peaks and the increase in maximum contaminant concentration can be attributed to the diminished number of flowpaths from node 11 to node 49 as the contaminant mass is concentrated into fewer flowpaths, as shown in Figure 4.5. The change to the flowpaths also increases the total mass of contaminant that flows through the observation nodes, because the contaminant is no longer removed from the system at node 45 due to the external demand.

We use the observations at nodes 21 and 49 to calculate the adjoint states using (3) and (4) for nodes with complete mixing and (10)-(12) for nodes with bulk advective mixing. The adjoint

states were used in (5) and (6) along with the PDF of measured concentration with a model uncertainty, σ, of 0.04 mg/L and a range of source masses of 10 to 500 grams to obtain the conditioned backward travel time PDFs shown in Figure 4.4B and the β_T values shown in Table 4.2. Figure 4.4B shows that the true contamination scenario (node 11 at t = 0:00) is selected as a potential contamination scenario, however, additional potential source nodes are also identified. Based on the β_T values, node 18, which is not the true source, is the most likely source node. Thus, the method incorrectly identifies the source node.

4.3.2.2. Scenario 2B

This scenario is similar to Scenario 2A; however, the source node is moved from node 11 to node 16, changing the contaminant flowpath, as seen in Figure 4.6. The contaminant concentration at the observation nodes for this scenario is shown in Fig 4.3C, which shows a peak concentration of 2.2 mg/L occurring at 3:55 for node 21 and two peaks for node 49: 21.5 mg/L at 10:25 and 3.1 mg/L 11:55. The two peaks for node 49 indicate two flowpaths from the source node (node 16) to the observation node (node 49). Figure 4.4C shows the BTTPDFs; the true contamination scenario (node 16 at t = 0:00) is selected as the only potential contamination scenario. This demonstrates how, under these circumstances, our method is able to determine the true source node and release time in a system with bulk advective mixing at some nodes.

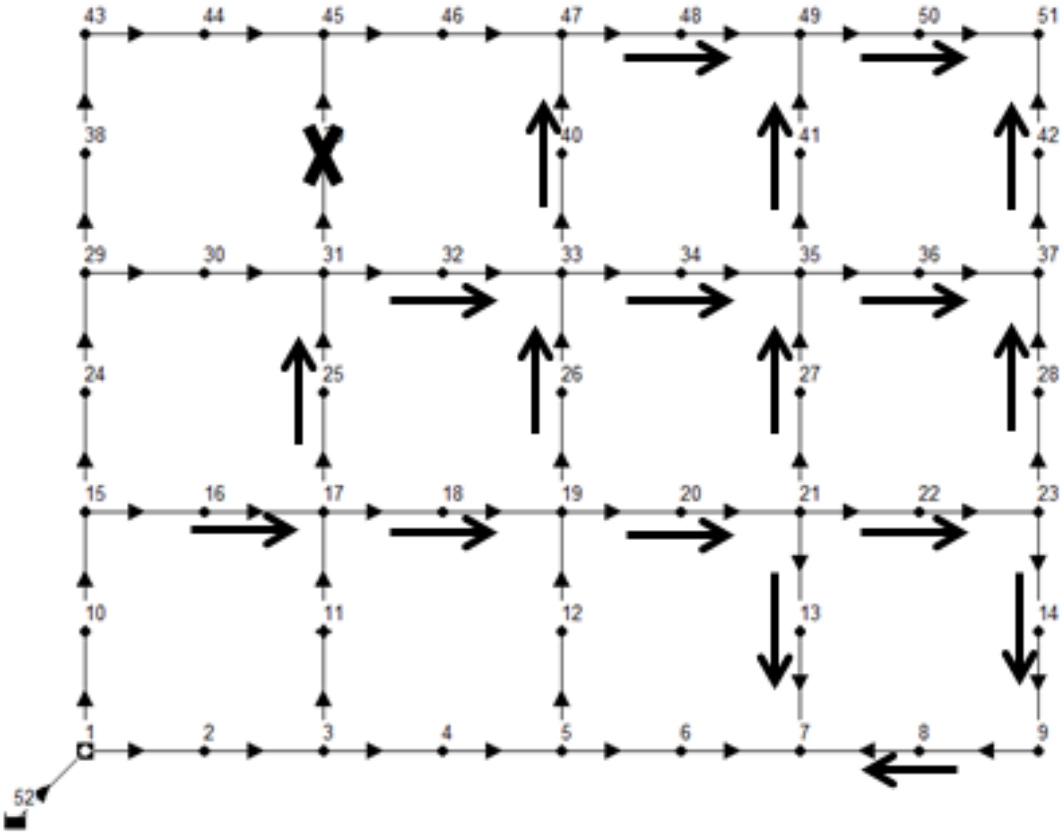

Figure 4.6. Contaminant flowpaths for Scenario 2B. The source node is node 16. The bolded arrows indicate the flowpath for the contaminant, while the x's indicate the paths that receive water from the node, but no contaminant.

4.3.2.3. Discussion

Our method is able to select the true source node as a possible source node and the true release time as a possible release time in both of the scenarios evaluated; however, we are only able to probabilistically select the correct source node in Scenario 2B. One reason why we are not consistently calculating the true source node as the most likely source nodes is because we do not know the appropriate method for calculating the true concentration for(4.5). In addition, the location of the source node is different in each scenario. Moving the source node leads to different flowpaths between the source node and the observation nodes, due to the bulk advective mixing model; this is demonstrated in Figure 4.3B and 4.3C. Figure 4.3B shows one concentration peak arriving at each observation node, indicating only one flowpath from the source node to the

observation node; Figure 4.3C, however, shows three peaks when the source node is node 16, indicating three flowpaths. When we select observations, we use the observation peaks, so only two observations were used to for Scenario 2A, while three were used for Scenario 2B. The additional information used in Scenario 2B likely contributed to the more accurate results. We can also, demonstrate the importance of the source node location by looking at how the contaminant flows through node 17, which is the first node after node 11 and node 16. All of the contaminant entering node 17 from node 11 will exit through pipe 21 (see Figure 4.2), while contaminant that enters the systems at node 16 will flow through node 17 and exit the node through pipes 21 and 28; the amount of contaminant flowing into each pipe iscalculated using (4.7) – (4.9). However, the adjoint state relationships show that, any adjoint state that enters node 17 from pipe 21 will exit through pipes leading to nodes 11 and 16, but all of the adjoint state entering node 17 through pipe 28 exits through the pipe leading to node 16; the amount of adjoint state in each pipe is calculated using (4.10) – (4.12). This behavior explains why the β_T values in Scenario 2A indicate that node 17 is a more probable node than node 11; the number of flowpaths between node 17 and the observation nodes is greater than the number of flowpaths for node 11. In comparison to Scenario 2B, node 11 has one flowpath from node 49 for the adjoint state in this scenario versus four flowpaths in the Scenario 1; for node 17, Scenario 2B has three flowpaths from node 49 to node 17, while Scenario 1 has four flowpaths. Thus, the bulk advective model simulates proportionally more adjoint state arriving at node 17 than node 11 and node 17 becomes the more probable source node. Node 16 is selected as the true source node in Scenario 2B because such a large portion of the adjoint mass flows to node 16 from node 17.

Node 16 is not a potential source node in Scenario 2A for a similar reason. The transport of contaminant when the source node is node 16 (Figure 4.2C) can be compared to the results from Figure 4.2B. When the contaminant enters node 17 from node 16, the resulting contaminant

concentration at observation node 21 is lower than if the contaminant enters from node 11, while the concentrations measured at node 49 are higher for source node 16 than the for source node 11; the result is that the calculations for the PDF of obtaining the measured concentrations (used in (5)) require a different source mass to calculate the observations for nodes 21 and 49, and the result is all zero values for the PDF of the measure calculations.

4.3.3. Scenario 3

In this scenario, we used a mixing parameter of 0.5 at the incomplete mixing nodes resulting in a combination of bulk advective mixing and complete mixing. The contaminant flowpaths are the same as in Scenario 1, but the concentration measurements are different. We evaluated three different variations using this scenario: Scenario 3A where the source node is node 11 and observation nodes are at nodes21 and 49, Scenario 3B where the source node is node 11 and observation nodes are at nodes 21 and 47, and Scenario 3C where the source node is node 11 and observations nodes are at nodes 21 and 47, but we calculate the adjoint states using an incorrect mixing parameter value (s = 0.75) at the incomplete mixing nodes.

4.3.3.1. Scenario 3A

In this scenario, we use node 11 as the source node and observation nodes 21 and 49. Figure 4.3D shows the concentration versus time profiles for nodes 21 and 49. Node 21 has a peak concentration of 16.7 mg/L at 4:05. Node 49 has multiple contaminant concentration peaks; we use the peaks at 10:40 (2.7 mg/L) and 12:35 (3.6 mg/L) as the observations.

We used the observations at nodes 21 and 49 to calculate the adjoint states using (3) and (4) for nodes with complete mixing and (10)-(12) for nodes with bulk advective mixing. The adjoint states were used in (5) and (6) along with the PDF of measured concentration with a model

uncertainty, σ, of 0.04 mg/L and a range of source masses of 10 to 500 grams to obtain the conditioned BTTPDFs shown in Figure 4.4D and the β_T values shown in Table 4.2. The true contamination scenario (node 11 at t = 0:00) is selected as a potential contamination scenario, however, other potential source nodes also have non-zero BTTPDFs. Table 4.2 shows the β_T values for all potential source nodes. Based on the β_T values in Table 4.2, node 2, which is not the true source, is the most likely source node. These results demonstrate that under these conditions we are able to determine that the true source node is a potential source node and the true release time is a potential release time, but we are not able to probabilistically determine that the true source node is the most probable node.

4.3.3.2. Scenario 3B

In this scenario, we use node 11 as the source node and observation nodes 21 and 47. The contaminant concentration measurements at the observation nodes are shown in Figure 4.3E. Node 21 has a peak concentration of 16.7 mg/L at 4:05. Node 47 has contaminant concentration peaks at t = 7:45 (5.2 mg/L) and 8:15 (4.7 mg/L). We used these observations at nodes 21 and 49 to calculate the adjoint states using (3) and (4) for nodes with complete mixing and (10)-(12) for nodes with bulk advective mixing. The adjoint states were used in (5) and (6) along with the PDF of measured concentration with a model uncertainty, σ, of 0.04 mg/L and a range of source masses of 10 to 500 grams to obtain the conditioned backward travel time PDFs shown in Figure 4.4E and the β_T values shown in Table 4.2.

The true contamination scenario (node 11 at t = 0:00) is selected as a potential contamination scenario, however, nodes 2 and 3 also have non-zero BTTPDFs. Table 4.3 shows the β_T values for the potential source nodes. Based on the β_T values in Table 4.3, node 11, the true

source, is the most likely source node. These results demonstrate that our method is able to probabilistically calculate the true source node as the most likely source node under these conditions.

4.3.3.3. Scenario 3C

In this scenario we use node 11 as the source node and nodes 21 and 47 as the observation nodes, but we calculate the adjoint states using mixing parameter of $s = 0.75$, instead of the true mixing parameter, $s = 0.5$. The concentration observations are the same as Scenario 3B (Figure 4.3E). We used the observations at nodes 21 and 49 to calculate the adjoint states using (3) and (4) for nodes with complete mixing and (10)-(12) for nodes with bulk advective mixing. The adjoint states were used in (5) and (6) along with the PDF of measured concentration with a model uncertainty, σ, of 0.04 mg/L and a range of source masses of 10 to 500 grams to obtain the conditioned backward travel time PDFs shown in Figure 4.4F and the β_T values shown in Table 4.2.

The non-zero BTTPDF values identify three potential source nodes and release time scenarios: node 2 at -1:05, node 3 at -0:45, and node 11 at 0:00, the true source node and release time. These are the same source node and release times that we found in the previous example. Table 4.2 shows the β_T values for this scenario; the β_T values indicate that node 3 is the most likely source node. So, in this scenario we are able to find the true source node and release time as a potential scenario, but we are unable to probabilistically determine that it is the most likely source node.

4.3.3.4. Discussion

Our method is able to identify the true source node as a possible source node and the true release time as a possible release time in all three of the scenarios evaluated; however, we are only able to probabilistically select the correct source node in Scenario 3B. One reason why we are not

consistently calculating the true source node as the most likely source nodes is because we do not know the appropriate method for calculating the true concentration for (4.5). In addition, the contaminant flowpaths between the source node and observation node may lead to selecting the incorrect source node as the most likely source node. In Scenario 3A, node 2, is determined to be the most probable source node instead of node 11. In this scenario one of the observation nodes is node 49. As the adjoint state is backtracked from node 49 a portion of the adjoint state will arrive at node 19. From node 19, the adjoint state takes two different flowpaths: one that goes to node 18, 17, and 11, etc., and one that goes to nodes 12, 5, 4, 3, 2, bypassing node 11. Adjoint mass that takes the first flowpath passes through the true source node and arrives at nodes 2 and 3, so all mass that passed the true source node also gets to nodes 2 and 3. Adjoint mass that takes the second flowpath arrives at nodes 2 and 3, without passing through node 11, so the unconditioned BTTPDFs at nodes 2 and 3 have a larger total probability than the BTTPDF at node 11. Figure 4.7 shows the unconditioned BTTPDFs for nodes 2 and 11. In Figure 4.7B, node 11 has four peaks; these same peaks occur in in the unconditioned BTTPDF for node 2 (Figure 4.7A) along with an additional two peaks which indicate two additional flowpaths between node 49 and node 2. The additional unconditioned BTTPDF increases the probability that node 2 is the true source node. Thus, when β_T values are calculated using the unconditioned BTTPDFs in (6), β_T is larger for nodes 2 than it is for node 11.

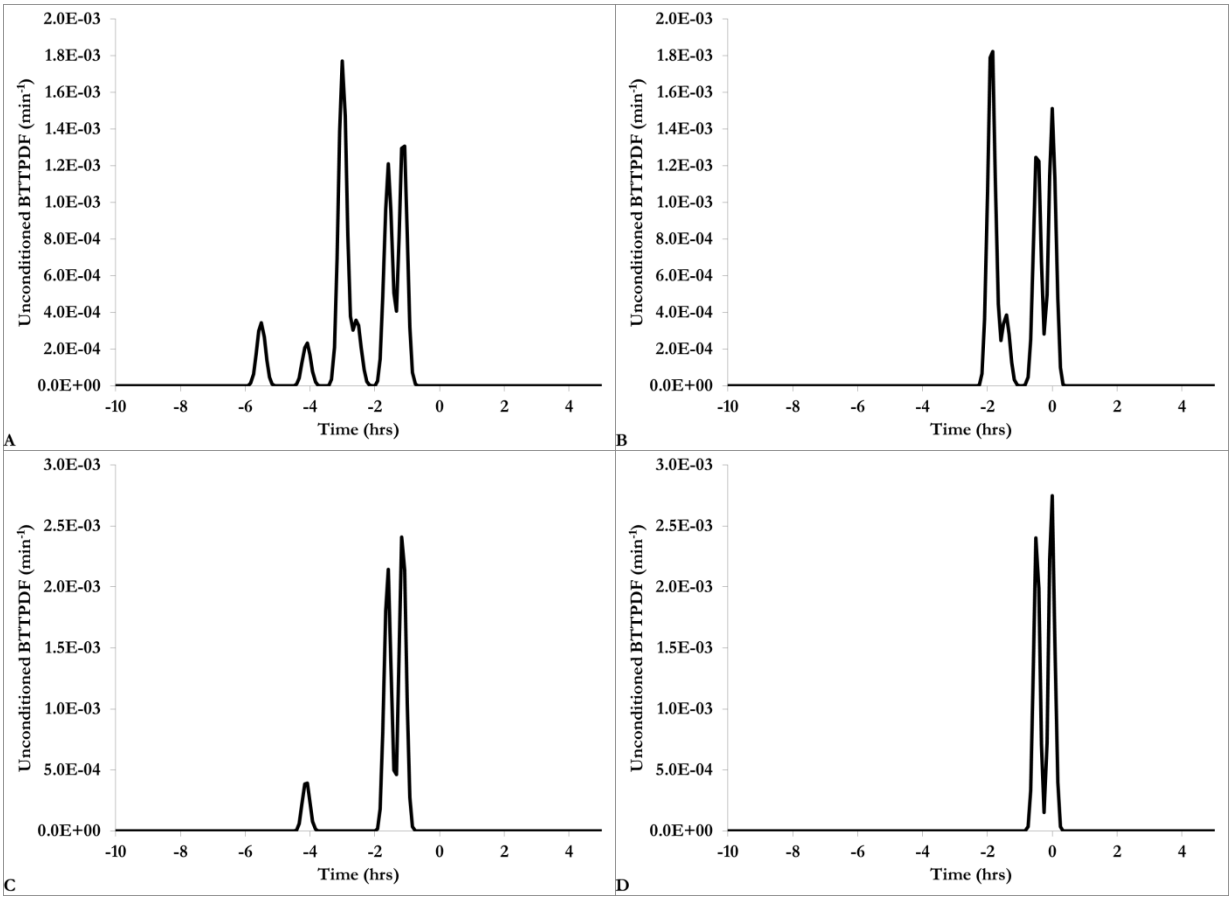

Figure 4.7. Unconditioned BTTPDF for (A) node 2 in Scenario 3A, (B) node 11 in Scenario 3A, (C) node 2 in Scenario 3B, and (D) node 11 in Scenario 3B.

In Scenario 3B, node 11, the true source node, is determined to be the most probable source node. The flowpaths in this scenario are different from Scenario 3A because the observation node has changed from node 49 to node 47. In this scenario, the adjoint state sill arrives at node 19 from node 47, but it enters only through pipe 29. Following this flowpath, a larger portion of the adjoint state will flow from node 19 toward node 18 than from node 19 toward node 12. The result is only one additional peak shown in Figure 4.7C which is not sufficient to make node 2 (or 3) the most probable source node.

In Scenario 3C, node 3 is determined to be the most probable source node instead of node 11. In this scenario, the true system used a mixing parameter of $s = 0.5$, while we calculated the

adjoint states using a mixing parameter of $s = 0.75$. Using the incorrect mixing parameter led to incorrect mixing at the nodes and, therefore, incorrect adjoint states for the potential source nodes. We would expect to get incorrect results in this scenario. We were able to select the true source node as a potential source node and the true release time as a potential release time, however.

4.4. Conclusions

Our goal was to develop a method to probabilistically determine the contaminant release node and time in water distribution systems with incomplete mixing at nodes. We looked at six different scenarios: one with complete mixing at all nodes, two with bulk advective mixing at some nodes, two with incomplete, but not bulk advective, mixing at some nodes, and one scenario where the incorrect mixing parameter was used to determine the source node and release time. We were able to find the true contaminant release node and time in all six scenarios, and probabilistically determine the true contaminant release node and time in three of the six scenarios.

In the scenarios with bulk advective mixing at some node, $s = 0$, the results were dependent on the placement of the source node. The orientation of the pipes in the incomplete mixing model is determined based on the flow rates in the pipes; the inflow pipe with the larger flow rate is labeled A and the outflow pipe with the larger flow rate is labeled C. When the source node was connected to pipe A, we were able to determine that the source node was the most probable source node. When the source node was not connected to pipe A, we were unable to determine that the true source node was the most probable source node.

In the scenarios with incomplete mixing at some nodes, $s = 0.5$, the results were similarly dependent on the observation nodes used. Just as the location of the source nodes relate to the flowpath of the contaminant mass, the location of the observations relate to the flowpath of the adjoint mass. If the adjoint mass flows through pipe B, the majority of the adjoint mass will flow

into pipe C; however, if the adjoint mass flows through pipe A, the adjoint mass will be split more evenly between the two outflow pipes. Although the method always found the true source node as a potential source node and the true release time as a potential release time, the most probable node results are dependent on the flowpath between the observation node and source node.

Appendix. Derivation of Adjoint Equations for Incomplete Mixing

The goal of the adjoint method in our context is to calculate the sensitivity of the contaminant concentration at node j (C_j^*) to a mass of contamination (M_ℓ) that enters the system at node ℓ, defined as

$$\frac{\partial C_j^*(t_s)}{\partial M_\ell} = \int_t \frac{\partial}{\partial M_\ell}\left[C_j^*(t)\delta(t-t_s)\right]dt = \int_t \phi_j^*\delta(t-t_s)dt \tag{4.13}$$

where $\delta(t - t_s)$ is a Dirac delta function, t_s is the time of the observation, and $\phi_j^* = \partial C_j^*(t)/\partial M_\ell(t_o)$ is the marginal sensitivity of concentration at node j at **any** time t to a mass released at node ℓ at a **specific** time t_o. Our goal is to replace this marginal sensitivity with the sensitivity of concentration at the **specific** time t_s to mass released at node ℓ at **any** backward time τ. To do this, we follow the adjoint derivation approach of Neupauer (2011) and present the steps that are unique to obtaining the adjoints of (4.7) – (4.9). First we differentiate (4.1), (4.2), (4.7) and (4.8) with respect to $\mathbf{M}_\ell$ to obtain

$$\frac{\partial}{\partial t}\phi_i + \left(\frac{Q_i}{A_i}\right)\frac{\partial}{\partial x_i}\phi_i = 0 \tag{4.14}$$

$$\phi_j^* = \frac{\sum_{i\in d_i=j} Q_i\phi_i\,|_{x_i=L_i} + U_j^*}{\sum_{i\in u_i=j} Q_i} \tag{4.15}$$

$$\phi_D\big|_{x_D=0} = \phi_A\big|_{x_A=L_A} \tag{4.16}$$

and

$$\phi_C\big|_{x_C=0} = \frac{\phi_B\big|_{x_B=L_B} Q_B + \phi_A\big|_{x_A=L_A} Q_A - \phi_A\big|_{x_A=L_A} Q_4}{Q_C} \tag{4.17}$$

where $\phi_i = \partial C_i / \partial M_\ell$ and the initial condition is $\phi_i = 0$ at $t = 0$.

We take the inner product of each term in (4.13) with arbitrary functions φ_i to obtain

$$\sum_{i=1}^{N_p} \iint_{x_i,t} \left[\varphi_i \frac{\partial \phi_i}{\partial t} + \varphi_i \frac{Q_i}{A_i} \frac{\partial \phi_i}{\partial x_i} \right] dx_i dt = 0 \tag{4.18}$$

where N_p is the number of pipes, which we add to the sensitivity equation (4.13) to obtain

$$\frac{\partial C_j^*(t_s)}{\partial M_\ell} = \sum_{i=1}^{N_p} \iint_{x_i,t} \left[\varphi_i \frac{\partial \phi_i}{\partial t} + \varphi_i \frac{Q_i}{A_i} \frac{\partial \phi_i}{\partial x_i} \right] dx_i dt + \int_t \phi_j^* \delta(t - t_s) dt \tag{4.19}$$

We use the product rule to change derivatives of ϕ_i in (4.18) to derivatives of φ_i and rearrange to obtain

$$\begin{aligned} \frac{\partial C_j^*(t_s)}{\partial M_\ell} = & -\sum_{i=1}^{N_p} \iint_{x_i,t} \phi_i \left(\frac{\partial \varphi_i}{\partial t_i} + \frac{Q_i}{A_i} \frac{\partial \varphi_i}{\partial x_i} \right) dx_i dt \\ & + \sum_{i=1}^{N_p} \iint_{x_i,t} \left(\frac{\partial}{\partial t} (\varphi_i \phi_i)_i + \frac{Q_i}{A_i} \frac{\partial}{\partial x_i} (\varphi_i \phi_i) \right) dx_i dt + \int_t \phi_j^* \delta(t - t_s) dt \end{aligned} \tag{4.20}$$

Since φ_i is arbitrary, we define it in such a way as to eliminate the first summation from (4.20) to obtain (4.3), and (4.20) is simplified to

$$\frac{\partial C_j^*(t_s)}{\partial M_\ell} = \sum_{i=1}^{N_p} \iint_{x_i,t} \left(\frac{\partial}{\partial t} (\varphi_i \phi_i)_i + \frac{Q_i}{A_i} \frac{\partial}{\partial x_i} (\varphi_i \phi_i) \right) dx_i dt + \int_t \phi_j^* \delta(t - t_s) dt \tag{4.21}$$

The temporal divergence terms in (4.21) can be rewritten as

$$\sum_{i=1}^{N_p} \iint_{x_i,t} \frac{\partial}{\partial t}(\varphi_i \phi_i) dx_i dt = \int_{x_i} \left[\int_t \frac{\partial}{\partial t}(\varphi_i \phi_i) dt \right] dx_i$$
$$= \int_{x_i} \left[\varphi_i \phi_i \big|_{t=t_s} - \varphi_i \phi_i \big|_{t=0} \right] dx_i = 0 \tag{4.22}$$

The terms evaluated at $t = 0$ vanish because the initial condition on ϕ_i is defined as $\phi_i = 0$ at $t = 0$. The terms evaluated at $t = t_s$ vanish if we define the final condition on φ_i as $\varphi_i = 0$ at $t = t_s$. Using this information (4.21) becomes

$$\frac{\partial C_j^*(t_s)}{\partial M_\ell} = \sum_{i=1}^{N_p} \iint_{x_i,t} \frac{Q_i}{A_i} \frac{\partial}{\partial x_i}(\varphi_i \phi_i) dx_i dt + \int_t \phi_j^* \delta(t - t_s) dt \tag{4.23}$$

The spatial divergence terms in (4.21) can be written as

$$\sum_{i=1}^{N_p} \iint_{x_i,t} \frac{Q_i}{A_i} \frac{\partial}{\partial x_i}(\varphi_i \phi_i) dx_i dt = \sum_{i=1}^{N_p} \int_t \frac{Q_i}{A_i} \left[\varphi_i \phi_i \big|_{x_i = L_i} - \varphi_i \phi_i \big|_{x_i = 0} \right] dt \tag{4.24}$$

For the pipes whose downstream ends are connected to nodes with complete mixing, we use the boundary conditions on ϕ^* defined in (4.15) and φ^* defined in (4.4) to eliminate their terms from (4.24) (Neupauer 2011). For the pipes whose downstream nodes are connected to nodes with bulk advective mixing, we use the following approach.

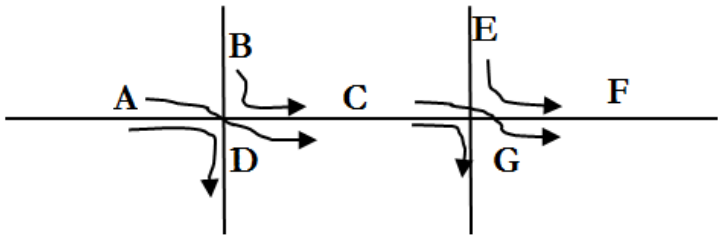

Figure 4.8. Example network for incomplete mixing derivation. The straight lines represent pipes which are labeled by the letters. The arrows indicate the flow of the contaminant mass.

Figure 4.8 shows an example network where the letters identify the pipes and the arrows indicate the flow of the contaminant mass in accordance with the bulk advective mixing model. The contaminant concentrations are known in pipes *A*, *B*, and *E*. The following relationships apply to

the flowrates in the pipes: (1) $Q_A + Q_C > Q_B + Q_D$, and (2) $Q_C + Q_F > Q_E + Q_G$. These flow rate relationships result in the following equations for contaminant concentration, based on the bulk advective mixing equations in (4.7) and (4.8):

$$C_D\Big|_{x_D=0} = C_A\Big|_{x_A=L_A} \tag{4.25}$$

$$C_C\Big|_{x_C=0} = \frac{C_A\Big|_{x_A=L_A} Q_A + C_B\Big|_{x_B=L_B} Q_B - C_D\Big|_{x_D=0} Q_D}{Q_C} \tag{4.26}$$

$$C_G\Big|_{x_G=0} = C_C\Big|_{x_C=L_C} \tag{4.27}$$

$$C_F\Big|_{x_F=0} = \frac{C_C\Big|_{x_C=L_C} Q_C + C_E\Big|_{x_E=L_E} Q_E - C_G\Big|_{x_G=0} Q_G}{Q_F} \tag{4.28}$$

Assuming the pipes in Figure 4.8 are the only pipes that are connected to nodes with bulk advection mixing, (4.24) can be expanded as

$$\begin{aligned}
&\sum_{i=1}^{N_p} \int_t \frac{Q_i}{A_i} \left[\varphi_i \phi_i \Big|_{x_i=L_i} - \varphi_i \phi_i \Big|_{x_i=0} \right] dt = \\
&\frac{Q_A}{A_A} \phi_A \varphi_A \Big|_{x_A=L_A} + \frac{Q_B}{A_B} \phi_B \varphi_B \Big|_{x_B=L_B} + \frac{Q_C}{A_C} \phi_C \varphi_C \Big|_{x_C=L_C} + \frac{Q_D}{A_D} \phi_D \varphi_D \Big|_{x_D=L_D} \\
&+ \frac{Q_E}{A_E} \phi_E \varphi_E \Big|_{x_E=L_E} + \frac{Q_F}{A_F} \phi_F \varphi_F \Big|_{x_F=L_F} + \frac{Q_G}{A_G} \phi_G \varphi_G \Big|_{x_G=L_G} \\
&- \frac{Q_A}{A_A} \phi_A \varphi_A \Big|_{x_A=0} - \frac{Q_B}{A_B} \phi_B \varphi_B \Big|_{x_B=0} - \frac{Q_C}{A_C} \phi_C \varphi_C \Big|_{x_C=0} - \frac{Q_D}{A_D} \phi_D \varphi_D \Big|_{x_D=0} \\
&- \frac{Q_E}{A_E} \phi_E \varphi_E \Big|_{x_E=0} - \frac{Q_F}{A_F} \phi_F \varphi_F \Big|_{x_F=0} - \frac{Q_G}{A_G} \phi_G \varphi_G \Big|_{x_G=0}
\end{aligned} \tag{4.29}$$

The assumptions of EPANET-BAM require that all pipes entering a node must be equally sized, so $A_A = A_B = A_C = A_D = A_E = A_E = A_F = A$. Also, each node must have two inflows in adjacent pipes, and two outflows, also in adjacent pipes, so we use the nodes at $x_A=0$, $x_B=0$, $x_D=L_D$, $x_E=0$, $x_F = L_F$, and $x_G=L_G$. The values at $x_A=L_A$, $x_B=L_B$, $x_D=0$, $x_E=L_E$, $x_F= 0$, and $x_G=0$ are at nodes with

complete mixing and are defined using (4.3) and (4.4); we eliminate these terms from (4.29) and (4.29) becomes:

$$\sum_{i=1}^{N_p}\int_t \frac{Q_i}{A_i}\left[\varphi_i\phi_i\big|_{x_i=L_i} - \varphi_i\phi_i\big|_{x_i=0}\right]dt = \frac{1}{A}\left[\begin{array}{l} Q_A\phi_A\varphi_A\big|_{x_A=L_A} + Q_B\phi_B\varphi_B\big|_{x_B=L_B} + Q_C\phi_C\varphi_C\big|_{x_C=L_C} + Q_E\phi_E\varphi_E\big|_{x_E=L_E} \\ - Q_C\phi_C\varphi_C\big|_{x_C=0} - Q_D\phi_D\varphi_D\big|_{x_D=0} - Q_F\phi_F\varphi_F\big|_{x_F=0} - Q_G\phi_G\varphi_G\big|_{x_G=0} \end{array}\right] \tag{4.30}$$

Our goals are to show that the following are true:

$$\varphi_C\big|_{x_C=0} = \varphi_B\big|_{x_B=L_B} \tag{4.31}$$

$$\varphi_A\big|_{x_A=L_A} = \frac{\varphi_C\big|_{x_C=0}Q_C + \varphi_D\big|_{x_D=0}Q_D - \varphi_B\big|_{x_B=L_B}Q_B}{Q_A} \tag{4.32}$$

$$\varphi_F\big|_{x_F=0} = \varphi_E\big|_{x_E=L_E} \tag{4.33}$$

$$\varphi_C\big|_{x_C=L_C} = \frac{\varphi_F\big|_{x_F=0}Q_F + \varphi_G\big|_{x_G=0}Q_G - \varphi_E\big|_{x_E=L_E}Q_E}{Q_C} \tag{4.34}$$

We can focus on the junction of pipes *A*, *B*, *C*, and *D* and define the boundary conditions such that the inflows minus the outflows must equal zero

$$Q_A\phi_A\varphi_A\big|_{x_A=L_A} + Q_B\phi_B\varphi_B\big|_{x_B=L_B} - Q_C\phi_C\varphi_C\big|_{x_C=0} - Q_D\phi_D\varphi_D\big|_{x_D=0} = 0 \tag{4.35}$$

Using (4.16), we can substitute in for $Q_C\phi_C\big|_{x_C=0}$

$$Q_A\phi_A\varphi_A\big|_{x_A=L_A} + Q_B\phi_B\varphi_B\big|_{x_B=L_B} - \varphi_C\big|_{x_C=0}\left(Q_B\phi_B\big|_{x_B=L_B} + Q_A\phi_A\big|_{x_A=L_A} - Q_D\phi_D\big|_{x_D=0}\right) - Q_D\phi_D\varphi_D\big|_{x_D=0} = 0 \tag{4.36}$$

If we set $\varphi_C\big|_{x_C=0} = \varphi_B\big|_{x_B=L_B}$ as a boundary condition, then (4.36) becomes

$$Q_A\phi_A\varphi_A\big|_{x_A=L_A} - Q_A\phi_A\big|_{x_A=L_A}\varphi_C\big|_{x_C=0} + Q_D\phi_D\big|_{x_D=0}\varphi_C\big|_{x_C=0} - Q_D\phi_D\varphi_D\big|_{x_D=0} = 0 \tag{4.37}$$

Since $Q_A + Q_B = Q_C + Q_D$, this is also true:

$$Q_A \phi_A\big|_{x_A=L_A} \varphi_C\Big|_{x_C=0} = Q_C \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0} + Q_D \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0} - Q_B \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0} \tag{4.38}$$

So, (4.37) becomes

$$\begin{aligned} &Q_A \phi_A \varphi_A\big|_{x_A=L_A} - \left(Q_C \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0} + Q_D \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0} - Q_B \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0}\right) \\ &+ Q_D \phi_D\big|_{x_D=0} \varphi_C\big|_{x_C=0} - Q_D \phi_D \varphi_D\big|_{x_D=0} = 0 \end{aligned} \tag{4.39}$$

We use (4.17) to cancel the 3rd and 5th terms in (4.39) and (4.39) becomes

$$Q_A \phi_A \varphi_A\big|_{x_A=L_A} - Q_C \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0} + Q_B \phi_A\big|_{x_A=L_A} \varphi_C\big|_{x_C=0} - Q_D \phi_D \varphi_D\big|_{x_D=0} = 0 \tag{4.40}$$

We differentiate (4.25) with respect to M to get

$$\phi_D\big|_{x_D=0} = \phi_A\big|_{x_A=L_A} \tag{4.41}$$

And use (4.41) in (4.40) to get

$$\phi_A\big|_{x_A=L_A} \left(Q_A \varphi_A\big|_{x_A=L_A} - Q_C \varphi_C\big|_{x_C=0} + Q_B \varphi_B\big|_{x_B=L_B} - Q_D \varphi_D\big|_{x_D=0}\right) = 0 \tag{4.42}$$

Thus, from (4.42)

$$\varphi_A\big|_{x_A=L_A} = \frac{Q_D \varphi_D\big|_{x_D=0} + Q_C \varphi_C\big|_{x_C=0} - Q_B \varphi_B\big|_{x_B=L_B}}{Q_A} \tag{4.43}$$

We use (4.31) to show that

$$\varphi_A\big|_{x_A=L_A} = \frac{Q_D \varphi_D\big|_{x_D=0} + Q_C \varphi_C\big|_{x_C=0} - Q_B \varphi_C\big|_{x_C=0}}{Q_A} \tag{4.44}$$

This is the relationship in (4.32) and can be used as a boundary condition along with (4.31). (4.33) and (4.34) can be evaluated in a similar manner. We focus on the junction of pipes *C*, *E*, *F*, and *G* and assert that the inflows minus the outflows must be zero:

$$\mathcal{Q}_C \phi_C \varphi_C\big|_{x_C=L_C} + \mathcal{Q}_E \phi_E \varphi_E\big|_{x_E=L_E} - \mathcal{Q}_F \phi_F \varphi_F\big|_{x_F=0} - \mathcal{Q}_G \phi_G \varphi_G\big|_{x_G=0} = 0 \qquad (4.45)$$

From this point on, the derivation is exactly the same, but with the following subscript changes in (4.36)-(4.43): A (in the originals) = C (in the new); B=E; C=F; and, D=G. The result is that (4.33) and (4.34) can be used as boundary conditions.

CHAPTER 5

CASE STUDY

Abstract

In Chapters 2 and 3 we developed adjoint-based probabilistic source identification methods for water distribution systems with perfect, fuzzy, or binary sensors. We tested these methods using relatively small systems to demonstrate how the methods work. In this chapter we evaluate how the methods work in a more complex system.

5.1. Introduction

We evaluate our methods using the EPANET Example 3 (Rossman 2000) water distribution system (Figure 5.1). Many researchers have used this water distribution system to test their own methods (e.g., Preis and Ostfeld 2007). The only change we made to the system is changing all time steps in the network to 2 minutes. The system contains 2 pumps, 2 reservoirs, 3 fully-mixed tanks, 92 nodes, and 117 pipes. The system is subject to transient flows due to the filling and draining of the tanks and varying demand patterns at nodes throughout the system; the demand patterns and tank characteristics are all the same as EPANET Example 3.

We use EPANET to simulate the movement of a generic non-reactive contaminant in this network to generate the sensor measurements used to test the source identification methods. We use a time step of Δt = 2 minutes and we assume a contaminant source at node 119, modeled as a flow-paced booster with an input concentration of 5,000 mg/L from t = 1:00 to 1:02, with a total mass of approximately 339 kg. Nodes 143, 181, and 213 were used as observation nodes at which sensors measured the contaminant concentration over time; these nodes were previously used by Preis and Ostfeld (2007) in the same system. Figure 5.2 shows the contaminant concentration as a function of time for observation nodes 143, 181, and 213. While node 143 only has one clear concentration

peak, indicating a single flowpath from the source node, nodes 181 and 213 have multiple flowpaths. The peak concentrations and times are defined as observations and listed in Table 5.1.

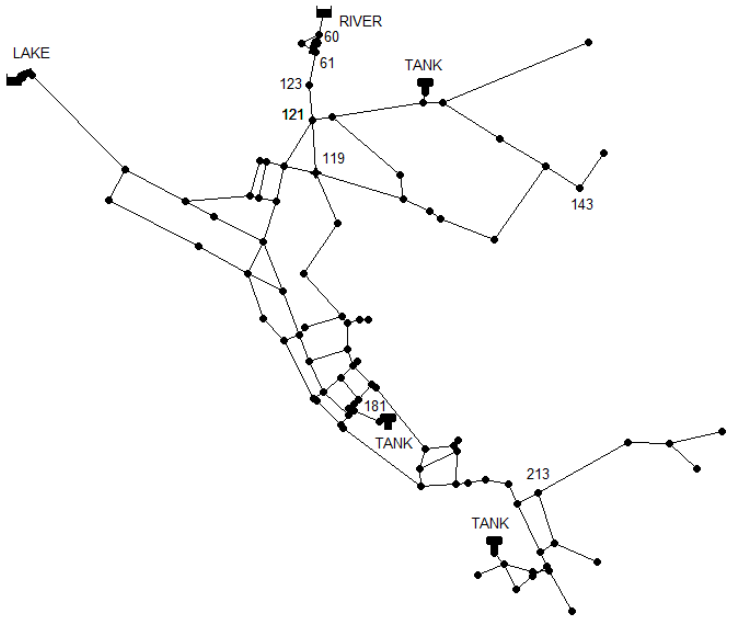

Figure 5.1: Example 3 network from EPANET; Lines represent pipes. Circles represent nodes. The number next to some of the nodes is the node identifier.

Table 5.1. Concentration Observations for Figure 5.2.

Obs.	Start Time (HH:MM)	End Time (HH:MM)	Observation Time (HH:MM)	Perfect Sensors	Fuzzy Sensors	Binary Sensor
				Peak Conc. (mg/L)	Conc. Range	Value
143A	3:18	3:30	3:26	880	Medium	Present
181A	1:54	2:22	2:02	1400	High	Present
181B	2:38	4:02	2:44	8.9	Low	Present
181C	4:08	5:32	4:38	3.2×10^{-6}	Low	Present
181D	11:46	12:04	12:00	1.3×10^{-6}	Low	Present
181E	12:44	13:00	12:44	1.1×10^{-6}	Low	Present
181F	15:06	21:24	15:56	3.5	Low	Present
181G	22:22	24:00	23:02	1.4	Low	Present
213A	2:42	16:58	3:34	290	Low	Present
213B	23:30	24:00	24:00	2.1×10^{-5}	Low	Present

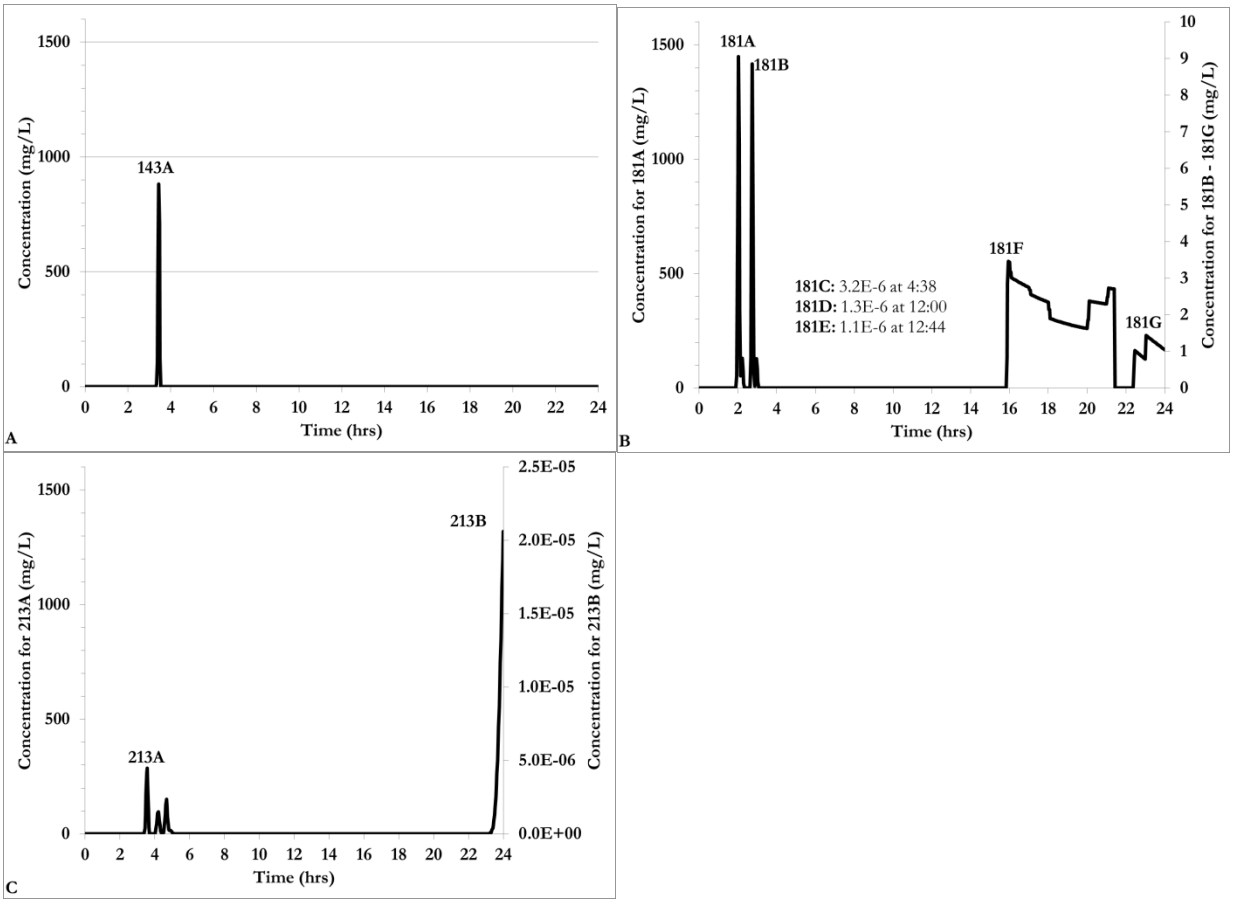

Figure 5.2. Concentration versus time plots for (A) observation node 143, (B) observation node 181, and (C) observation node 213.

5.2. Perfect Sensors

If perfect sensors are located at the observation nodes in the system, we use the peak concentration values in the adjoint method. Using observations 143A, 181A, and 213A (node, time, and concentration) we calculate the BTTPDFs for each observation and each potential source node, and the conditioned BTTPDFs using all three observations were calculated using the method described in Chapter 2. We considered all 92 nodes as potential source nodes. Using a source mass range of 300-400 kg and a model uncertainty, σ, equal to 10% of the measured concentration (e.g., if the measurement is 15 mg/L, then standard deviation is ±1.5 mg/L), we calculated the conditioned BTTPDFs in Figure 5.3 and the β_T values in Table 5.2.

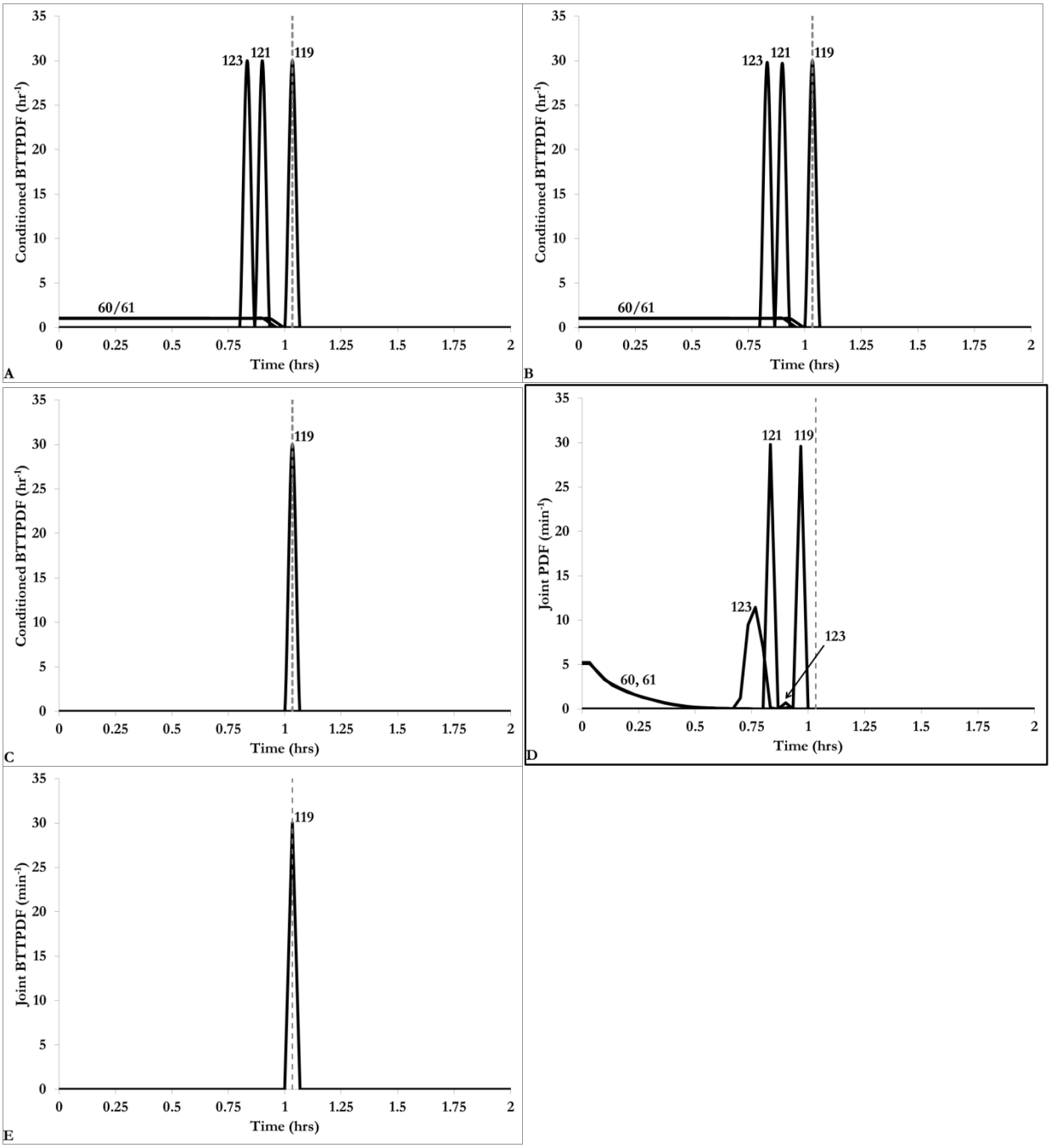

Figure 5.3. Conditioned BTTPDF using (A) ideal sensor data, (B) fuzzy sensor data, and (C) fuzzy sensor data with non-detects, and the joint BTTPDF using (D) binary sensor data and (E) binary sensor data including non-detects. The number above each curve denotes the node number of the potential source node and the dashed line indicates the true release time (1:02).

The conditioned BTTPDFs show five potential source nodes with non-zero values: (1) node 60, (2) node 61, (3) node 119, the true source node, (4) node 121, and (5) node 123. The non-zero values indicate potential release times for each potential source node. The true release time (t = 1:02) is selected as a potential release time for node 119. As discussed in previous chapters, the most probable source node is indicated by the largest value of β_T. The node with the largest β_T value is node 119, the true source node. These results demonstrate how our adjoint-based probabilistic method can be used to determine the true source node and release time in a water distribution system using perfect sensors.

5.3. Fuzzy Sensors

We used the same system to test our adjoint-based probabilistic method for using fuzzy sensor data. The contaminant concentration observations in Figure 5.2 assume that ideal sensors are available. We convert these observations to fuzzy sensor readings by specifying the concentration ranges that the fuzzy sensors measure. For this example, we specify the low range as $0 < c \leq 500$ mg/L, the medium range as 500 mg/L $< c \leq 1000$ mg/L, and the high range as 1000 mg/L $< c \leq$ 1500 mg/L. We use the same three observations as were used in the perfect sensor example to test our adjoint-based probabilistic method to obtain a reading of Medium at t = 3:26 for node 143 (Observation 1), a reading of High at t = 2:02 for node 181 (Observation 2), and a reading of Low at t = 3:34 for node 213(Observation 3). Following the theory elucidated in Chapter 3, we use an lower and upper bound vectors of **a**=[1000 mg/L, 500 mg/L, 0 mg/L] and **b**=[1500 mg/L, 1000 mg/L, 500 mg/L] to obtain the conditioned BTTPDFs shown in Figure 5.3B and the β_{TF} value shown in Table 5.2.

Table 5.2. β_T Values for Case Study using Observations from (A) Perfect Sensors, (B) Fuzzy Sensors, and (C) Fuzzy Sensor with Non-Detects (Units are $L^3/mg^2/hr^2$).

Node	60	61	119	121	123
A	3.7E-81	3.9E-81	**4.1E-36**	1.9E-44	6.7E-45
B	1.4E-87	1.5E-87	**1.1E-65**	1.5E-68	9.8E-69
C	0	0	**2.9E-66**	0	0

In Figure 5.3B the non-zero values of the conditioned BTTPDF shows the potential release time of contamination from the source. For the true source node (node 119), the true release time is identified as the most likely release time. The potential source nodes and release times identified using observations from fuzzy sensors are nearly the same as compared to the potential release times obtained using perfect sensors; while it may not be clear in the Figure 5.3B, the BTTPDF is non-zero for times from 1:00 to 1:04 for node 119, 0:00 to 0:56 for node 60, 0:00 to 0:58 for node 61, 0:52 to 0:58 for node 121, and 0:46 to 0:54 for node 123. The results are similar when perfect sensor data is used; the range for nodes 60 and 61 are the same, but the time ranges are smaller for the other nodes: node 119 is centered on one 2-minute timestep at t = 1:02, node 121 has non-zero BTTPDF values from 0:54 to 0:56, and node 123 from 0:50 to 0:52. In Table 5.2, the β_T values indicate that node 119, the true source node, is the most probable source node for this scenario. The influence of using fuzzy sensors instead of perfect sensors is apparent in the magnitude of the β_T values. The maximum β_T value for node 119 using perfect sensors is 4.1 x 10^{-36} $L^3/mg^2/hr^2$, but using fuzzy sensor data the value is 1.1 x 10^{-65} $L^3/mg^2/hr^2$. These β_T values illustrate how information is lost using fuzzy sensors because the results are spread over a measurement range (e.g., 100-500 mg/L) instead of using a single measurement (e.g., 290 mg/L). However, even with the loss of information using fuzzy sensors, we were able demonstrate that our adjoint-based probabilistic method can be used to determine the true source node and release time in a water distribution system using fuzzy sensors.

We can use non-detect measurements to improve the fuzzy sensor results. Figure 5.3C shows the conditioned BTTPDF results using the same fuzzy sensor data, but adding non-detect measurements from node 143. We used non-detect measurements for t = 0:00 to 3:16 and t = 3:30 to 24:00, using the theory described in Chapter 3. Using the non-detect measurements in conjunction with observations 143A, 181A, and 213A, we decrease the number of potential source nodes using fuzzy sensors from five potential source nodes to only the true source node, and find that the most likely release time is 1:02. These results demonstrate how our adjoint-based probabilistic method can be used to determine the true source node and release time in a water distribution system using fuzzy sensors and non-detect measurements.

5.4. Binary Sensors

We used the system shown in Figure 5.1 for the binary sensor method example as well. The contaminant concentration observations shown in Figure 5.2 also hold true, but binary sensors are only able to measure the presence or absence of contamination, thus all non-zero measurements in Figure 5.2 are converted to "1," which indicates that contamination is present, or above a threshold value. In the form of binary sensor data, we used contaminant observations at node 143 at t = 3:18, 3:24, and 3:30, at node 181 at t = 1:54, 2:08, and 2:22, and at node 213 at t = 2:42, 9:50, and 16:58. We use the method described in Chapter 3 to calculate the joint probabilities for potential source nodes 60, 61, 119, 121, and 123. The joint BTTPDFs are displayed in Figure 5.3D.

The joint BTTPDF is non-zero at times when the contamination observed at all three observation nodes could have been released from the potential source nodes. The joint BTTPDF for node 119 is non-zero at t = 1:02 (in Figure 5.3D the majority of the non-zero BTTPDF values occur at t < 1:00, however, a very small non-zero BTTPDF value of 1.6×10^{-6} min^{-1} occurs at t = 1:02) indicating that contaminant mass released at node 119 (the true source node) at 1:02 (the true

release time) would have resulted in the contaminant observations at the observation nodes and observation times. Figure 5.3D indicates that contaminant observed at the sensors at times shown in Figure 5.2 could have been released at times t = 0:00 to 0:58, or 1:02 for node 119; 0:00 to 0:50 for node 60; 0:00 to 0:52 for node 61; 0:00 to 0:56 for node 121; and 0:00 to 0:54 for node 123 (Note that many of these BTTPDF values are very small and do not appear on the figure). These ranges are much broader than the ranges calculated using fuzzy or perfect sensors; this increase in possible release times occurs because of the loss of information using binary sensors, which only indicate the presence or absence of contamination and provide no information about the specific contaminant concentrations. However, even with the loss of information using binary sensors, the adjoint-based probabilistic method determines the true source node and release time.

We can use non-detect measurements to improve the binary sensor results. Figure 5.3E shows the joint BTTPDFs using binary sensor data with the same non-detect measurements we used in the fuzzy sensor example. We decreased the number of potential source nodes using binary sensors from three to just one, node 119, which is the true source node; and the range of possible release times is decreased, with the most likely release time at t = 1:02, the true release time. These results demonstrate how our adjoint-based probabilistic method can be used to determine the true source node and release time in a water distribution system using binary sensors and non-detect measurements.

5.5. Discussion

We used an adjoint-based method with data from perfect, fuzzy, and binary sensors to identify the source node and release time and found that the amount of information provided by the sensors changed the number of potential source node and range of possible release time scenarios calculated. Perfect sensors provide the most specific information in the form of a specific

contaminant concentration observation; the binary sensors provide the least information in the form of the presence or absence of contamination; the fuzzy sensors only provide information about the range of contaminant concentrations. This difference in information is reflected in the spread of possible release times identified by the method, with the spread of the possible release times increasing as the amount of information contained in the data decreases. The spread of possible release times is narrow if perfect sensor data are used, becomes wider if fuzzy sensor data are used, and becomes even wider if binary sensor data are used.

We found that including the non-detect measurements in the calculations enabled us to narrow down the results significantly. In the case of the fuzzy sensors, using non-detect measurements decreased the number of potential source nodes from five to one, the true source node, and the range of possible release times was reduced to just the true release time. In the binary sensor example, five potential source nodes with release times spread over broad ranges were identified without using the non-detect measurements, while only one potential source node was identified when the non-detect measurements were used. These results indicate the potential power of the non-detect measurements; by using non-detect measurements, we determined that only the true contaminant source node and release time was a potential contamination scenario, even if fuzzy or binary sensors were used.

5.6. Conclusions

The goal was to test our adjoint-based probabilistic methods for using perfect, fuzzy, and binary sensors in a complex water distributions system. We used EPANET Example 3 as the basis for the water distribution system and inserted a contaminant at a node in the system to simulate contaminant observations using EPANET. We successfully calculated the true source node and release time as a potential scenario while using perfect, fuzzy, and binary sensors. We also

demonstrated how non-detect measurements can be used to decrease the number of potential source node and release time scenarios.

While we were able to determine the true source node and release time using all types of sensors, we found that the sensors which provide less information identify more potential contaminant nodes and release times. We were able to increase the accuracy of the source node and release time identification, however, by including additional information in the form of non-detect measurements. Overall, we demonstrated that our method achieves the primary goal of finding the true source node and release time as a potential source node and release time in this system using perfect, fuzzy, or binary sensors, regardless of whether we used non-detect measurements.

CHAPTER 6

CONCLUSIONS

6.1. Summary of Research Results

Previous research used adjoint methods in simple water distribution systems using ideal sensors (e.g., Neupauer et al. 2010). The goal of this research was to develop adjoint methods for source identification in more realistic water distribution system (i.e., including tanks, pumps, and transient flow fields) under more realistic conditions (i.e., non-ideal sensors and incomplete mixing at junctions). This goal led to the following three hypotheses:

H1. The adjoint method can be used to determine the source of contamination in water distribution systems containing pumps, storage tanks, and transient flow conditions.

H2. The adjoint method can be used to determine the source of contamination in water distribution systems when using realistic system sensors.

H3. The adjoint method can be used to determine the source of contamination in water distribution systems with incomplete mixing at pipe junctions.

In Chapter 2, we developed an adjoint method that can be used to identify the source of contamination in a water distribution system containing pumps, storage tanks, and transient flow conditions. Neupauer et al. (2010) demonstrated that the adjoint method can identify sources of contamination in simple water distribution systems (i.e., systems containing only source, pipes, and nodes) under steady-state flow conditions. We demonstrated a method that can determine the source of contamination in a more complex water distribution system.

In Chapter 3, we developed a method for using information from two types of realistic system sensors: (1) fuzzy sensors, which are only able to identify the approximate range of contamination (e.g., high, medium, low), and (2) binary sensors, which only identify whether

contamination is present or absent (or whether it is above or below a designated threshold value). We also developed a method for incorporating non-detect measurements (i.e., sensor measurements indicating that the level of contamination is below the limit of detection) in the calculations to determine the true source node and release time. Previous adjoint methods (e.g. Neupauer et al. 2010) were developed specifically for use with ideal sensors using only non-zero observations.

In Chapter 4, we developed an adjoint method for water distribution systems which have incomplete mixing at pipe junctions. Previous work assumed complete mixing at pipe junctions; contaminant mass enters the node through the inflows, mixes completely and all pipe outflows contain the same concentration of contamination. True water distribution systems are unlikely to experience such ideal situations; this leads to discrepancies between the modeled system and the true system (Austin et al. 2008). We used a bulk advective mixing algorithm (EPANET-BAM) to introduce more realistic mixing at pipe junctions (Ho and O'Rear 2009). We developed adjoint theory incorporating incomplete mixing at the pipe junctions and demonstrated how our new method works in an example water distribution system.

6.2. Conclusions

6.2.1. Hypothesis 1

The results show that the adjoint method can be used to identify the source node and release time of contaminant that is observed at one or more sensors in a complex system with transient flow conditions and a water storage tank. We assumed perfect knowledge of the hydraulics, assumed that sensors measure concentration to within 10% of the true value, and considered scenarios with two, three, or four observations. The true source node was identified as the most likely source node for 89% of the test cases that used two observations, 96% of the test cases that used three observations, and 97% of the cases with four observations. These results show that the method is successful even

with limited observation data. The cases that did not select the true source node as the most likely source node all had the following characteristics: (1) a node which is hydraulically similar to the true source node was chosen as the most likely source node, and (2) at least one observation from contaminant that passed through the water storage tank was used in the calculations.

Our method uses the following criteria to help identify the true source node: (1) Is the node hydraulically connected to the observation node(s)?, and (2) Does the contaminant mass needed to be released from the potential source node to reproduce the contaminant observations fall within the likely range of source masses? The water distribution system model we used contains a node directly upstream of the true source node which meets both of these criteria and is nearly indistinguishable from the true source node. Not only is this node hydraulically connected to the observation nodes, all of the water from the node passes through the true source node, and most of the water at the true source node comes from the node; if not for the external demands at the true source node, the two nodes could almost be considered the same node. The similarities between the two nodes result in similar adjoint states between the source nodes. In addition, the source mass needed at the upstream node to replicate the concentration observations at the observation nodes is close to the true source mass released at the true source node and within the range of the potential source masses we tested, thus the two nodes also have similar probabilities of being the true source node.

It is also important to note that observations from contamination that passed through the water storage tank were used in the calculations leading to selecting the incorrect node as the true source node. The water storage tank is fully mixed, so, in the forward model, contaminant concentration in water that has passed through the tank is lower than in water that did not enter the tank. Similarly, the adjoint state is diluted as it is propagated backward through the tank resulting in a loss of information at upstream nodes. The consequence of passing through the storage tank is a

wide span of potential release time that, when used in conjunction with adjoint states calculated using observations that did not pass through the tank, does little to narrow down the number of potential release times, but does decrease the probability that any one source node is the most likely source node.

Overall, we were able to determine that the true source node and release time was a potential source node and release time even when we used information that traveled through the tank. In most cases, we were also able to determine that the true source node was the most probable source node, and, in the cases where we did not find the true source node, we determined that a very similar node was the most probable source node.

6.2.2. Hypothesis 2

We successfully developed and used new adjoint methods for use with fuzzy sensor data, binary sensor data, and non-detect measurements. We were able to determine the true source node and contamination time as a possible contamination scenario for various model uncertainties and concentration ranges using fuzzy sensor data or binary sensor data. The impact of the fuzzy sensor or binary sensor data (versus ideal sensor data) was most clearly demonstrated when attempting to determine the most probable source node. Calculations for determining the most probable source node proved to be inconclusive when using fuzzy sensors, as multiple potential source nodes had similar likelihoods. Similar results were found by Preis and Ostfeld (2008) who used an inverse method to identify the source of contamination using fuzzy sensor data. When using the binary sensor data, very limited information is available and we were only able to narrow down the potential release times for each node based on the adjoint states. Both fuzzy sensors and binary sensors provide less information than perfect sensors which translates to a broader range of possible results for the potential source nodes and release times. We were able to incorporate information

from non-detect measurements to overcome the loss of information from not using perfect sensors; we accomplished this by using the non-detect measurements to find locations where the contaminant could not have entered the system.

The results show that the adjoint method can be used to identify the source node and release time of contaminant as a potential source node and release time using information from fuzzy sensors, binary sensors, and non-detect measurements. When fuzzy sensor data was used we were able to determine that the source node was the most probable source node under some conditions, but the success was dependent upon the source node used. When binary sensor data was used we did not have enough information to probabilistically determine the true source node. For both fuzzy sensors and binary sensors, the non-detect measurements allowed us to narrow the number of potential source nodes and/or release times.

6.2.3. Hypothesis 3

Our goal was to develop a method to probabilistically determine the contaminant release node and time in water distribution systems with incomplete mixing at nodes. We looked at six different scenarios: one with complete mixing at all nodes, two with bulk advective mixing at some nodes, two with incomplete, but not bulk advective, mixing at some nodes, and one scenario where the incorrect mixing parameter was used to determine the source node and release time. We were able to find the true contaminant release node and time in all six scenarios, and probabilistically determine the true contaminant release node and time in three of the six scenarios.

In the scenarios with bulk advective mixing at some node, $s = 0$, the results were dependent on the placement of the source node. The bulk advective mixing scenario has significantly different contaminant transport paths due to the bulk advective mixing model. The result is that the location of the source node has a significant impact on where the contaminant is transported in the system.

Similarly, the location of the observation node has a significant impact on our ability to probabilistically determine the true source node; the impact is based on the flowpath of the contaminant mass due to the bulk advective mixing model. We were able to successfully determine that the true source node and release time were potential source nodes and release times in all $s = 0$ scenarios, however. Since the scenarios with incomplete mixing at some nodes, $s = 0.5$, is part bulk advective mixing and part complete mixing, the results were similarly dependent on the source node and observation nodes used.

We evaluated one scenario where we used an incorrect mixing parameter in the adjoint state calculations to determine whether we could calculate the true source node. We were able to find the true source node and release time as a potential source node and release time, however, we were not able to determine the true source node as the most probable source node

Overall, we were able to find the true source node and release time as a potential source node and release time in all incomplete mixing scenarios. We were only able to determine the true source node as the most probable source node in certain scenarios, however. One reason why we could not consistently calculate the true source node as the most likely source nodes is because we do not know the appropriate method for calculating the true concentration. In addition, the success was dependent upon the observation nodes used, because the observation nodes used influenced the flow of the adjoint mass using the bulk advective mixing model.

6.3. Limitations

This research has some limitations. These limitations fall into three categories: (1) Contaminant Transport, (2) Observations, and (3) Node Mixing.

6.3.1. Contaminant Transport: Reactions

We did not account for how reactions affect the transport of contaminant mass in the water distribution system. Incorporating reactions into these methods would require calculating the degradation of the contaminant in the water distribution system, which can be accomplished in EPANET in the forward transport mode. The adjoint theory would have to be developed for incorporating the reactions into the adjoint method.

6.3.2. Observations: Number of Observations

Both inverse and adjoint methods require observations (contaminant concentration, location, and arrival time throughout the water distribution system) to identify the source of contamination. Logically, using more observations leads to a better understanding of the movement of the contaminant mass and an increased ability to determine the source node and contamination time. The ability for our methods to determine the true source node and contamination time is limited by the number of observations; this limitation is common to both adjoint and inverse methods.

6.3.3. Node Mixing: Types of Junctions for Non-Ideal Mixing

We used EPANET-BAM to simulate the flow of contaminant mass in a water distribution system with non-ideal mixing at pipe junctions. This software can simulate non-ideal mixing at junctions that meet the following criteria: (1) only four pipes entering the junction and they must be equally sized, (2) only two inflow pipes and they must be adjacent to each other, and (3) only two outflow pipes and they must be adjacent to each other. Many pipe geometries do not fit these criteria and, therefore, cannot be simulated used non-ideal mixing.

6.3.4. Node Mixing: True Concentration

We used the same method for calculating the true concentration in systems with incomplete mixing as in systems with complete mixing. This method leads to incorrect calculations for the true concentration in systems with incomplete mixing; however, we are currently unable to determine the correct method. We expect that the adjoint method for systems with incomplete mixing would calculated the true source node as the most likely source node if the true concentration calculation was determined.

6.4. Recommendations for Future Work

Here are some recommendations for future work:

6.4.1. Reactions

Water quality reactions occur in true water distribution systems, therefore a model which is intended to predict the behaviors of a true water distribution system needs to incorporate reactions. The adjoint methods used in this research did not account for reactions. These reactions will lead to contaminant mass changes in the system (e.g., loss of contaminant mass) and, if they are not incorporated in the adjoint method, will lead to erroneous results. EPANET (Rossman 2000), can be used to simulate both bulk reactions with n^{th} order kinetics and wall reactions for a single chemical, while the EPANET multiple species extension (EPANET-MSX) can be used for tracking the interaction of multiple chemicals (EPA 2013). Further research should be done to develop a method for incorporating these processes in the adjoint method.

6.4.2. Mixing Model

An adjoint method should be developed and tested using a more robust mixing model. Many pipe junction geometries do not meet the conditions for using EPANET-BAM (i.e., a junction of four equal-sized pipes with two adjacent inflows and two adjacent outflows) for incomplete mixing, thus many nodes are simulated as ideal mixing by default although this might not be the case in the true water distribution system. A more robust mixing model would increase the prediction abilities of the model by increasing the similarities between the simulation and the true system. The AZRED solute mixing model (Choi et al. 2008) is an extension for EPANET and uses experimental data to simulate incomplete mixing at various types of pipe junction geometries: cross junctions, double-tee junctions, and wye junctions. Further research would be needed to develop the adjoint theory and corresponding adjoint method which can be used in a system with these types of junctions and mixing.

Also, the appropriate method for calculating the true concentration in systems with incomplete mixing needs to be determined. Currently we use the same method as we use with complete mixing; this method is not appropriate for systems with incomplete mixing. Determining and using the appropriate method should lead to improved chances for selecting the true source node as the most likely source node.

BIBLIOGRAPHY

Anderson, M. P., and Woessner, W. W. (2002). *Applied Groundwater Modeling*, Academic Press, San Diego.

American Society of Civil Engineers (ASCE). (2004). *Interim voluntary guidelines for designing an online contaminant monitoring system*, ASCE, Reston, VA.

Austin, R. G., van Bloemen Waanders, B., McKenna, S., and Choi, C. Y. (2008). "Mixing at cross junctions in water distribution systems. II: Experimental study. *J. Water Resour. Plann. Manage.*, 134(3), 295-302.

Choi, C. Y., Shen, J. Y., and Austin, R. G. (2008). "Development of a comprehensive solute mixing model (AZRED) for double-tee, cross, and wye junctions." *Proc., 10th Annual Water Distribution Systems Analysis Conference*, ASCE, Kruger National Park, South Africa, 1004-1014.

Dawsey, W., Minsker, B., and VanBlaricum, V. (2006). "Bayesian belief networks to integrate monitoring evidence of water distribution system contamination." *J. Water Resour. Plann. Manage.*, 132(4), 234-241.

De Sanctis, A. E., Feng, S., and Uber, J. G. (2010). "Real-time identification of possible contamination sources using network backtracking methods." *J. Water Resour. Plann. Manage.*, 136(4), 444-453.

Department of Homeland Security (DHS). (2003). *Homeland Security Presidential Directive 7: Critical Infrastructure Identification, Prioritization, and Protection*, DHS, Washington, D.C.

Grayman, W. M., Ostfeld, A., and Salomons, E. (2006). "Location monitors in water distribution systems: red team-blue team exercise." *J. Water Resour. Plann. Manage.*, 132(4), 300-304.

Guan, J., Aral, M., Maslia, M., and Grayman, W. (2006). "Identification of contaminant sources in water distribution systems using simulation-optimization method: Case study." *J. Water Resour. Plann. Manage.*, 132(4), 252-262.

Ho, C. K. (2008). "Solute Mixing Models for Water-Distribution Pipe Networks." *J. Hydraul. Eng.*, 134(9), 1236-1244.

Islam, M. R., Chaudhry, M. H., and Clark, R. M. (1997). "Inverse modeling of chlorine concentration in pipe networks under dynamic conditions." *J. Environ. Eng.*, 123(10), 1033-1040.

Kumar, J., Brill, E. D., Mahinthakumar, G., and Ranjithan, S. R. (2012). "Contaminant source characterization in water distribution systems using binary signals." *J. Hydroinformatics*, 14(3), 585-602.

Laird, C., Biegler, L., van Bloemen Waanders, B., and Barlett, R. (2005). "Contamination source determination for water networks." *J. Water Resour. Plann. Manage.*, 131(2), 125-134.

Laird, C., Biegler, L., and van Bloemen Waanders, B. (2006). "Mixed-integer approach for obtaining unique solutions in source inversion of water networks." *J. Water Resour. Plann. Manage.*, 132(4), 242-251.

Liu, L., Ranjithan, S. R., & Mahinthakumar, G. (2011). "Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure." *J. Water Resour. Plann. Manage.*, 137(2), 183-192.

Liu, L., Sankarasubramanian, A., and Ranjithan, S. R. (2011). "Logistic regression analysis to estimate contaminant sources in water distribution systems." *J. Hydroinformatics*, 137(2), 545-557.

Neupauer, R. (2011). "Adjoint sensitivity analysis of contaminant concentrations in water distribution systems." *J. Eng. Mech.*, 137(1), 31-39.

Neupauer, R. M., and Lin, R. (2006). "Identifying sources of a conservative groundwater contaminant using backward probabilities conditioned on measured concentrations." *Water Resour. Res.*, 42(3), 424-436.

Neupauer, R. M., and Records, M. K. (2009). "Conditioned backward probability modeling to identify contamination sources in a water distribution system." *Proc., World Environmental and Water Resources Congress 2009*. ASCE, Reston, VA, 1-8.

Neupauer, R., Records, M., and Ashwood, W. (2010). "Adjoint model to identify contaminant sources in water distribution systems." *J. Water Resour. Plann. Manage.*, 136(5), 587-591.

Neupauer, R. M., and Wilson, J. L. (1999). "Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant." *Water Resour. Res.*, 35(11), 3389-3398.

Preis, A., and Ostfeld, A. (2006). "Contamination source identification in water systems: A hybrid model tree-linear programming scheme." *J. Water Resour. Plann. Manage.*, 132(4), 263-273.

Preis, A., and Ostfeld, A. (2007). "A contamination source identification model for water distribution system security." *Eng. Optimization*, 39(8), 941-951.

Preis, A., and Ostfeld, A. (2008). "Genetic algorithm for contaminant source characterization using imperfect sensors." *Civ. Eng. and Environ. Systems*, 25(1), 29-39.

Propato, M., Sarrazy, F., and Tryby, M. (2010). "Linear algebra and minimum relative entropy to investigate contamination events in drinking water systems." *J. Water Resour. Plann. Manage.*, 136(4), 483-492.

Romero-Gomez, P. H., and Choi, C. Y. (2008). "Mixing at Cross Junctions in Water Distribution Systems. I: Numerical Study." *J. Water Resour. Plann. Manage.*, 134(3), 285-294.

Rossman, L. A. (2000). *EPANET 2: User's Manual.* U.S. Environmental Protection Agency (EPA), Washington, D.C.

Sandia National Lab. (2008, January 29). "EPANET-BAM." Retrieved from Sandia National Lab website: http://www.sandia.gov/EPANET-BAM/download.htm (accessed 9 May 2013).

Shen, H., McBean, E. A., and Ghazali, M. (2009). "Multi-stage response to contaminant ingress into water distribution systems and probability quantification." *Canadian J. Civ. Eng.*, 36, 1764-1772.

United States Environmental Protections Agency (EPA). (2013, January 28). "EPANET." Retrieved from EPA website: http://www.epa.gov/nrmrl/wswrd/dw/epanet.html#extension (accessed 28 June 2013).

van Bloemen Waanders, B., Hammond, G., Shadid, J., Collis, S., and Murray, R. (2005). "A Comparison of Navier Stokes and Network Models to Predict Chemical Transport in Municipal Water Distribution Systems." *Proc., World Water and Environmental Resources Congress 2005*. ASCE, Anchorage, Alaska, 1-10.

Vankayala, P., Sankasubramanian, A., Ranjithan, S. R., and Mahinthakumar, G. (2009). "Contaminant source identification in water distribution networks under conditions of demand uncertainty." *Environmental Forensics*, 10(3), 253-263.

Wagner, D. E., Neupauer, R. M., and Cichowitz, C. J. (2013). "Probabilistic source characterization in water distribution systems with transient flows." *J. Water Resour. Plann. Manage.*, **In Preparation**.

Zechman, E. M., and Ranjithan, S. R. (2009). "Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system." *J. Water Resour. Plann. Manage.*, 135(5), 334-343.