2013

Development of watershed-based modeling approach to pollution source identification

Yangbin Tong
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses
Part of the Civil and Environmental Engineering Commons

Recommended Citation
Tong, Yangbin, "Development of watershed-based modeling approach to pollution source identification" (2013). LSU Master's Theses. 1830.
https://digitalcommons.lsu.edu/gradschool_theses/1830

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
DEVELOPMENT OF WATERSHED-BASED MODELING APPROACH TO POLLUTION SOURCE IDENTIFICATION

A Thesis

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Master of Science

In

The Department of Civil and Environmental Engineering

by

Yangbin Tong
B.S., Zhejiang University, 2006
M.S., Zhejiang University, 2008
August 2013
ACKNOWLEDGEMENTS

This thesis is dedicated to my parents who have walked me through all rains and winds.

First of all, I thank Dr. Zhi-Qiang Deng, who offered me the chance to study at LSU from 2010 fall and has led me to work on this topic with huge patience and responsibility ever since then. Without his continuous support and critical advices, there’s no way to finish the thesis by now. I will definitely miss meeting with Dr. Deng on Fridays, which help us to concentrate on research and school work. Dr. Deng will always be an example of hard-working person in my future career.

I also want to express my appreciation to Dr. Frank Tsai and Dr. Heather Smith for their service on the defense committee, providing keen comments and insightful suggestions. Thank Dr. D. Dean Adrian and Dr. Ronald F. Malone who gave valuable thoughts in my qualifying exam to help me better prepared.

Many thanks go to my colleagues – Shaowei, Bhuban, Zaihong, and Vahid for sharing your knowledge, thoughts and encouragement. I sincerely hope that this good ‘tradition’ will be kept among the new coming students.

Particularly, I thank those who helped, sought, and provided important research data for us, including Dr. JoAnn Burke and Dr. Andrea Bourgeois-Calvin from Lake Pontchartrain Basin Foundation, Jan Boydstun, Coty Rabalais and Melinda Molieri from Louisiana Department of Environmental Quality, and Dr. Krishna P Paudel from LSU AgCenter. This research and thesis could not be conducted without your generous assistance.

Special thank you goes to my dear friends for your precious friendship which makes my life in Baton Rouge colorful and meaningful. The names include but are not limited to Jessica McCallum, Jimmy Alford, Lindsay Masters, Catherine Bozeman, Amanda Yan, Christiana Compton, John Presswood, Ben Herbert, Sarah Farley, Xuan Kong, Xuan Chen, Jiexuan Hu, the Bloodworth’s family and the Hall’s family.

I owe my utmost gratitude to Doug, Nicole, Clay and Cole Cornelius for being my host family, loving me and supporting me unconditionally all the time. No more words can express my thankfulness to you! Thank God for this big blessing!

Finally, I would like to express the deepest love and affection to my family for their everlasting love, patience and support. Without them, I am not who I am today. No matter how difficult life was, is and will be, no matter how far I am away from home, I know they are always there waiting for me to go back and say “I’m coming home.”

God bless you all!

Yangbin Tong
August 2013 at LSU
# TABLE OF CONTENTS

**ACKNOWLEDGEMENTS** .................................................................................................................. ii

**ABSTRACT** .................................................................................................................................... viii

**CHAPTER 1** INTRODUCTION .......................................................................................................... 1
  1.1 Water Pollution ............................................................................................................................ 1
  1.2 Pollution Sources ......................................................................................................................... 2

**CHAPTER 2** LITERATURE REVIEW .............................................................................................. 3
  2.1 Introduction .................................................................................................................................. 3
  2.2 Materials and Methods ............................................................................................................... 5
    2.2.1 Inverse Modeling. .................................................................................................................. 5
    2.2.2 Bayesian Approach .............................................................................................................. 6
    2.2.3 Other Approaches ................................................................................................................ 8
    2.2.4 Source Tracking in Water Distribution Systems .................................................................. 11
  2.3 Discussion .................................................................................................................................. 12
    2.3.1 Source Tracking in Surface Water ....................................................................................... 12
    2.3.2 Identification of Multiple Point Sources .......................................................................... 13
    2.3.3 Future Perspectives on Identification of Critical Bacteria Source Areas. ...................... 13
    2.3.4 Combination of Methods. .................................................................................................... 14
    2.3.5 Application of Biosensors and Remote Sensing Technology ............................................ 16
  2.4 Conclusions .................................................................................................................................. 17

**CHAPTER 3** MOMENT-BASED METHOD FOR IDENTIFICATION OF POLLUTION SOURCE IN RIVERS .. 18
  3.1 Introduction .................................................................................................................................. 18
  3.2 Materials and Methods ............................................................................................................... 19
    3.2.1 Data Collection ..................................................................................................................... 19
    3.2.2 VART Model Descriptions ............................................................................................... 21
    3.2.3 Moment Equations .............................................................................................................. 22
  3.3 Results ......................................................................................................................................... 23
  3.4 Discussions .................................................................................................................................. 27
    3.4.1 Effect of River Reach Length on Computation Errors ....................................................... 27
    3.4.2 Sensitivity of Computation Errors to Model Input Parameters ....................................... 28
    3.4.3 Mass Loss Correction ........................................................................................................ 29
    3.4.4 Limitations .......................................................................................................................... 31
  3.5 Conclusions .................................................................................................................................. 31

**CHAPTER 4** CONCENTRATION TIME-BASED METHOD FOR WATERSHED-SCALE BACTERIA SOURCE
  AREA IDENTIFICATION ......................................................................................................................... 32
  4.1 Introduction .................................................................................................................................. 32
  4.2 Materials and Methods ............................................................................................................... 33
    4.2.1 Impaired Watershed and Data Collection ........................................................................... 33
    4.2.2 Time of Concentration ($T_c$) ........................................................................................... 35
    4.2.3 Concentration Time-Based Identification Approach ........................................................ 37
4.3 Results.........................................................................................................................39
  4.3.1 Discharge and Fecal Coliform Level.................................................................39
  4.3.2 $T_c$ for Tangipahoa River .................................................................42
  4.3.3 Rainfall-Runoff Driven Bacterial Pollution Events.................................44
4.4 Discussion..............................................................................................................48
  4.4.1 Multiple-Subbasin Case.........................................................................48
  4.4.2 Uncertainty in Concentration Time.........................................................48
4.5 Conclusion...........................................................................................................48

CHAPTER 5 GRAND CONCLUSIONS...........................................................................50
  5.1 Conclusions ........................................................................................................50
  5.2 Future Perspectives ..........................................................................................50

REFERENCES ........................................................................................................52

APPENDIX ...............................................................................................................61

VITA .......................................................................................................................65
LIST OF TABLES

Table 2.1 Summary of deterministic direct methods and probabilistic methods for pollution source identification.................................................................................................................. 14

Table 3.1 Tracer Injection Experiments.......................................................................................................................................................................................... 19

Table 3.2 Flow and VART Model Parameters for Tracer Test Reaches .......................................................................................................................... 20

Table 3.3 Estimated Distance and Total Mass.................................................................................................................................................. 24

Table 3.4 Sensitivities of Distance Errors to Model Input Parameters .......................................................................................................................... 28

Table 3.5 Correction to Total Mass Computation ............................................................................................................................................... 30

Table 4.1 Discharge and fecal coliform data for the Tangipahoa River .......................................................................................................................... 33

Table 4.2 Discharge and precipitation data for Tangipahoa River .......................................................................................................................... 35

Table 4.3 Coefficients for Different Formulas of $T_c$........................................................................................................................................ 36

Table 4.4 Concentration Time for Subbasins of Tangipahoa River (hour) ................................................................................................................ 42

Table 4.5 Comparison between $T_c$ and Time to $C_{\text{max}}$...................................................................................................................................... 43

Table 4.6 Fecal Coliform, Rainfall and Possible Sources .......................................................................................................................... 44

Table 4.7 Comparison of $T_T$ and $T_c$ of Source Unit for 03/11/1996 event ................................................................................................. 45

Table 4.8 Comparison of $T_T$ and $T_c$ of Source Unit for 11/18/1996 event ................................................................................................. 46
LIST OF FIGURES

Figure 2.1 Water quality assessment for rivers and streams.......................................................... 3

Figure 2.2 Known and unknown sources for impaired water bodies shown in Figure 2.1 ........ 4

Figure 2.3 Flowchart for watershed-scale bacterial source identification................................. 15

Figure 3.1 Location of Five Dye Test Rivers in USA .............................................................. 19

Figure 3.2 Comparison between calculated and measured distances ........................................ 25

Figure 3.3 Comparison between calculated and measured total mass ..................................... 25

Figure 3.4 Frequency distribution (histogram) of the percent error in calculated location and the probability density function (solid line) that best fits the histogram .............................................. 26

Figure 3.5 Frequency distribution (histogram) of the percent error in calculated total mass and the probability density function (solid line) that best fits the histogram ........................................ 26

Figure 3.6 Variation of computational error in total mass with injection distance ................... 27

Figure 3.7 Comparison between calculated total masses with/without correction against injected total mass .......................................................................................................................... 30

Figure 4.1 Bacteria-Related Data in Tangipahoa River watershed, where (a) is LPBF’s Monitoring Sites, (b) is watershed delineation, (c) is dairy farms, and (d) is WWTPs ............. 34

Figure 4.2 $T_c$ of a hypothetic watershed and the cumulative time-area curve ....................... 37

Figure 4.3 Discharge and LDEQ’s Fecal Coliform data in 1996 ................................................. 39

Figure 4.4 Discharge and LDEQ’s Fecal Coliform data in 1998 ................................................. 39

Figure 4.5 Discharge and LPBF’s Fecal Coliform data at Site TR6 ......................................... 40

Figure 4.6 Discharge and LPBF’s Fecal Coliform data at Site TR5 ......................................... 40

Figure 4.7 Discharge and LPBF’s Fecal Coliform data at Site TR3 ......................................... 41

Figure 4.8 Discharge and LPBF’s Fecal Coliform data at Site TR2 ......................................... 41

Figure 4.9 Radar Rainfall on 1996/3/11 04:00 CST ................................................................. 44
Figure 4.10 Potential Source Areas, Dairy Farms, and WWTPs for Case 1 .......................... 45
Figure 4.11 Radar Rainfall on 1996/11/18 09:00 CST .................................................................. 46
Figure 4.12 Potential Source Areas for case 7/24/2006................................................................. 47
Figure 4.13 Potential Source Areas and WWTPs for case 4 .............................................................. 47
ABSTRACT

Identification of unknown pollution sources is essential to environmental protection and emergency response. A review of recent publications in source identification revealed that there are very limited numbers of research in modeling methods for rivers. What’s more, the majority of these attempts were to find the source strength and release time, while only a few of them discussed how to identify source locations. Comparisons of these works indicated that a combination of biological, mathematical and geographical method could effectively identify unknown source area(s), which was a more practical trial in a watershed. This thesis presents a watershed-based modeling approach to identification of critical source area. The new approach involves (1) identification of pollution source in rivers using a moment-based method and (2) identification of critical source area in a watershed using a hydrograph-based method and high-resolution radar rainfall data. In terms of the moment-based method, the first two moment equations are derived through the Laplace transform of the Variable Residence Time (VART) model. The first moment is used to determine the source location, while the second moment can be employed to estimate the total mass of released pollutant. The two moment equations are tested using conservative tracer injection data collected from 23 reaches of five rivers in Louisiana, USA, ranging from about 3km to 300 km. Results showed that the first moment equation is able to predict the pollution source location with a percent error of less than 18% in general. The predicted total mass has a larger percent error, but a correction could be added to reduce the error significantly. Additionally, the moment-based method can be applied to identify the source location of reactive pollutants, provided that the special and temporal concentrations are recorded in downstream stations. In terms of the hydrograph-based method, observed hydrographs corresponding to pollution events can be utilized to identify the critical source area in a watershed. The time of concentration could provide a unique fingerprint for each subbasin in the watershed. The observation of abnormally high bacterial levels along with high resolution radar rainfall data can be used to match the most possible storm events and thus the critical source area.
CHAPTER 1  INTRODUCTION

1.1 Water Pollution

Water pollution existed in the human history for a long time since agricultural human activities, particularly when a habitat became over-populated. But it never got out of control like what people experienced hundreds of years ago during the process of Industrial Revolution, when water began to be so severely contaminated in some areas that it even threatened people’s daily life. Once the water is polluted, it takes much longer time to clean it. A typical example was the River Rhine in the 19th century when it was polluted by high volumes of industrial effluent, domestic waste, and nutrients that it could not clean itself. Since the River Rhine flows through six countries, Switzerland, Principality of Liechtenstein, Austria, Germany, France and the Netherlands, it takes both national level and international level to reduce the pollution and restore the river (Bernauer and Moser, 1996; Middelkoop, 2000; Zehnder, 1993). The international cooperation started with the treaty in 1887 that prohibited the discharge of wastes dangerous to fish (http://library.thinkquest.org/28022/case/rhine.html). As the situation deteriorated during the two world wars, it called for more international effort, which formed the International Commission for the Protection of the Rhine against Pollution (ICPR) in 1946 and the Rhine Action Programmed (RAP) in 1986. Now, after more than 100 years of control, the water quality of Rhine River is much better than it was in the 20th century. (http://www.iksr.org)

In the United States, the federal government passed the Clean Water Act (CWA, also known as Water Pollution Control Act Amendments) in 1972, and started a nationwide program called National Pollutant Discharge Elimination System (NPDES) to control the point sources (http://cfpub.epa.gov/npdes/). Researches started to focus on the nonpoint sources after the CWA was passed. Later the United States Environmental Protection Agency (U.S. EPA) initiated some specific programs to control the water pollution, including Total Maximum Daily Load (TMDL) program, Low Impact Development (LID) program, and Best Management Program (BMP). The implementation of these three programs was very effective and became a ‘standard’ procedure in each state to control and reduce the water pollution. The Department of Environmental Quality (DEQ) in each state was required to collect water quality data regularly and identify the water pollution type. All polluted water bodies were listed on the U.S. EPA’s 303(d) list and mandated to implement a TMDL report to control the pollution (40CFR1.130.7, 2001). In Louisiana, several rivers were on the 303(d) list for not supporting the primary usage for wildlife and fishery. Some of them are still not meeting this primary function even today (LDEQ, 2011). Tangipahoa River was among the successful stories that the river restored its primary function after a watershed scale water pollution control (U.S. EPA, 2008). The problem of Tangipahoa River began in 1980s, when a group of girl was reported to be sick after swimming in the river. Then investigations and water quality monitoring displayed that this river was seriously polluted by bacteria, mercury and other pollutants, and thus did not support the designated use of primary and secondary contact recreation, fish and wildlife propagation. After the effort of Louisiana DEQ (LDEQ), Louisiana Department of Health and Hospitals (LDHH), Louisiana Department of Agriculture and Forestry (LDAF), Lake Pontchartrain Basin Foundation (LPBF), U.S. National Resources Conservation Service (NRCS), as well as the participation of public, the water quality gradually restored (U.S. EPA, 2007). Data collected from 2004 to 2007 indicate that the upper
reach of Tangipahoa River is no longer impaired by fecal coliforms (U.S. EPA, 2008). The success in Tangipahoa River could provide researchers a good sum of information that could be used for more detailed investigations into critical pollution sources in the Tangipahoa River watershed.

1.2 Pollution Sources

Both the RAP and NPDES aimed at controlling the water pollution and reducing negative impacts of water pollution. However, before the implementation of TMDL, LID or BMP, it is necessary to find out the pollution sources. A usual category of water pollution source divides all sources into two types: Nonpoint Source (NPS) and Point Source (PS). Point source pollution (PSP) primarily refers to discharges from industrial plants, municipal sewage systems and wastewater treatment plants (WWTPs) (Parker, 2011), whereas nonpoint source pollution (NPSP) is generally driven by overland runoff from rainfall or snowmelt that carries pollutants away from the ground or subsurface porous media (Harmon and Wiley, 2011). In the USA, many efforts had been put on the harness of PSP before 1970s and it reduced water pollution significantly. After the CWA was enacted, public concerns shifted to control of NPSP, which typically requires the development of TMDL and implementation of LID/BMP for impaired water bodies. After decades of work, many rivers across the nation have been removed from the 303(d) list. However, some rivers remained on the list for not meeting the requirement. Another source that was not widely discussed is the unknown sources that might come from unreported accidental releases, illegal waste water discharges, or even terrorist activity. It is already reported by the U.S. EPA that around 20% of the impaired water body is caused by unknown pollution sources. Approaches to identifying this 20% would help to manage the water quality and reduce the risk of severe pollution.

This thesis will focus on how to identify unknown water pollution source in a river or watershed using modeling approach. It primarily comprises of three chapters.

(1) The first chapter is to provide a critical literature review of recent development in source area identification methods, both geographically and mathematically. The former mainly discusses issues of source location while the latter addresses source strength (or source history). A comparison of different mathematical modeling methods is presented and general suggestions are given for effective source identification.

(2) The second chapter presents a specific example of using a moment-based approach to identify unknown sources in a river. This approach was derived by combining moment equations with Laplace transform of a transient storage model. Equations were given to estimate the source location and total mass, assuming that they were unknown. Results from 5 rivers in Louisiana were compared to the dye test data to see how effective it is.

(3) The third chapter presents a geographical method that uses the watershed modeling tools BASINs, HSPF and ArcGIS to identify source locations in a watershed using time of concentration. This hydrology-based method first delineate a watershed into hydrologically uniform units, and then finds out the connection between fecal coliform increase in a water quality station and the catchment-scale storm events. Finally it uses time of concentration to identify the most probable hydrological unit that contains the critical pollution source such as dairy farms or WWTPs.
CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

Fecal contamination of coastal and inland waters is a serious environmental problem that affects both aquatic ecosystems and public health. In 2010, the U.S. EPA released the most recent water quality assessment for different types of water bodies. Figure 2.1 shows that about 50% of the assessed rivers and streams (26.5% of rivers and streams) met their designated uses, whereas the other half were found to be impaired or threatened. Sixty-six percent of assessed U.S. lakes, reservoirs, and ponds (42.2% assessed) and 99.9% of Great Lakes and open waters (93.7% assessed) were impaired and threatened, respectively.

![Figure 2.1 Water quality assessment for rivers and streams](image)

Figure 2.1 Water quality assessment for rivers and streams
(Courtesy of U.S. EPA, 2010a)

Water body impairments are commonly caused by both point and nonpoint sources. PSP primarily refers to discharges from industrial and wastewater treatment plants (Parker, 2011), whereas NPSP is generally driven by overland runoff from rainfall or snowmelt that carries pollutants away from the ground or subsurface porous media (Harmon and Wiley, 2011). PSP has been significantly reduced in recent decades in the United States and other developed nations through a system of laws, regulations, and judicial enforcement such as the CWA and the National Pollutant Discharge Elimination System (NPDES) (U.S. EPA, 2010b), and the European Water Framework Directive (Achleitner et al., 2005).

Effort to control water pollution has therefore shifted to NPSP, which typically includes the determination of a TMDL for an impaired water body (U.S. EPA, 2002). Next, LID and BMPs are often implemented to reduce loading from known pollutant sources (Tong et al., 2011). However, as shown in Figure 2.2, nearly 20% of U.S. water body impairments are caused by unknown sources that are difficult to control (U.S. EPA, 2010a). Clearly, the identification of unknown sources of pollutants is essential to both reducing water body impairments and restoring water quality.
A variety of methods have been presented for tracking sources of bacterial pollution, including biological methods, numerical modeling, optimization methods, probabilistic analysis, and sensor technologies. Biological source tracking methods, such as DNA fingerprinting and antibiotic resistance analysis, focus on the identification of host sources such as humans, livestock, and wildlife (Meays et al., 2004). A large body of literature exists on applications, advantages, limitations, and future development of biological methods (Ahmed et al., 2007; Bell et al., 2009; Blanch et al., 2006; Griffith et al., 2003; Gronewold et al., 2009; Lasalde et al., 2005; Schiff and Kinney, 2001). Although other approaches have been employed in bacterial source tracking (BST) (e.g., Albek, 1999; Bae et al., 2009), this study focuses on modeling-based methods for identification of critical source areas of bacteria at the watershed-scale.

A critical source area of bacteria is defined as the location where the bacterial source results in frequent violations of water quality standards in downstream water bodies. As a result of the nonpoint, distributed, and mixed nature of bacterial pollution in watersheds, it is often difficult to identify specific areas where significant bacterial sources are located because bacteria collected from different sampling sites might display similar fingerprints. Therefore, determination of critical source areas in a watershed is challenging. Several studies have been published to identify effective source area tracking methods (Boano et al., 2005; Salgueiro, 2008; Shang et al., 2012; Snodgrass and Kitanidis, 1997; Sun, 1994).

This objective of this Chapter is to identify effective methods for tracking critical source areas of bacteria at the watershed scale through use of an extensive literature review that emphasizes modeling methods.
2.2 Materials and Methods

2.2.1 Inverse Modeling.

From a mathematical perspective, source tracking is essentially an ill-posed inverse problem (Boano et al., 2005). According to different causal characteristics, inverse modeling can be categorized into boundary, retrospective, coefficient, and geometric problems (Alifanov, 1994; Zhang and Chen, 2007). Boundary problems are used to determine the boundary conditions that form a certain contaminant concentration field; retrospective problems (time-reversed problems) are used to find initial conditions; coefficient problems are used to estimate values of parameters in a governing equation; and geometric problems are used to reconstruct the geometric characteristics of a computational domain. Early studies focused on the problem of source strength estimation in which the source location was assumed to be known a priori (Yee, 2008). Research efforts also focused on identification of unknown bacteria source locations by the assumption that the source strength is known a priori (Matthes et al., 2005). The simultaneous determination of parameters for both the source location and source strength was investigated by Yee (2007, 2008) and Keats et al. (2007a, 2007b, 2010) using a Bayesian inferential approach.

The general form of the advection–dispersion equation (ADE) used in inverse modeling can be described as follows (Eq. 2.1):

$$\frac{\partial(nC)}{\partial t} = \nabla \cdot (Dn\nabla C) - \nabla \cdot (\bar{V}nC) + R$$  \hspace{1cm} (2.1)

where $n$ is the porosity, $C$ is the solute concentration, $D$ is dispersion coefficient, $\bar{V}$ is the flow velocity, and $R$ represents all other reaction-related sink/source terms such as sorption and radioactive decay. For pure physical transport problem in surface water, Eq. (2.1) can be simplified as:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left[ D_L \frac{\partial C}{\partial x} \right] - \frac{\partial}{\partial x} [\mu C]$$  \hspace{1cm} (2.2)

where $x$ is the longitudinal direction, $D_L$ is the longitudinal dispersion coefficient, and $\mu$ is the flow velocity. Several methods have been proposed for the inverse solution to the above equations given a set of concentration distributions observed downstream. Atmadja and Bagtzoglou (2001) summarized methods for inverse modeling into two broad classes, (1) a probabilistic approach to deduce the probability for the source location, and (2) the optimization approach that uses deterministic direct methods to solve the governing equations backward in time and to reconstruct the release history. For tracking release history of solute in groundwater, Snodgrass and Kitanidis (1997) used the following equation:

$$z = h(s, r) + v$$  \hspace{1cm} (2.3)

where $z$ was an $m \times 1$ vector of observations and $h(s, r)$ was the model function, $s$ was an $n \times 1$ state vector obtained from the discretization of the unknown function to estimate, and $r$ was a vector that contains other parameters such as the velocity or dispersivity of the aquifer. The measurement error is represented by the vector $v$.

In addition to the probabilistic and optimization approaches, mathematical tools for inverse problems also include artificial intelligence, stochastic simulation, and computational fluid
dynamics (CFD) modeling. Zhang and Chen (2007) classified inverse modeling methods into four groups of approaches—analytical, optimization, probabilistic, and direct. The analytical approach analytically solves the distribution of flow and contaminant concentrations and then inversely solves the characteristics of source. The analytical approach has been applied to heat conduction (Alifanov, 1994), groundwater contaminant transport (Alapati and Kabala, 2000), and atmospheric pollution (Kathirgamanathan et al., 2002). The direct approach reverses directly the governing equations for solving the problem using the regularization or stabilization technique to improve the solution stability. Additional details about inverse modeling approaches and their typical applications are provided in the sections that follow.

### 2.2.2 Bayesian Approach

A Bayesian approach is widely considered to be a branch of geostatistical approaches that combines statistics with geographic analysis and provides results in the form of a probability distribution function (PDF). The basic principle in Bayesian approaches is Bayes’ theorem, which has been used widely in geology, hydrogeology, hydrology, environmental sciences and engineering, and biology (Liu et al., 2008; Patil and Deng, 2011). Typical applications of Bayes’ theorem include pattern recognition, uncertainty analysis, and risk analysis. Because source identification is an ill-posed problem that has no unique solution, the Bayesian approach provides a rational framework for the formulation of a probabilistic solution.

#### 2.2.2.1 Bayesian Inference

The basic form of Bayes’ theorem is expressed in Equation 2.4 as (Keats et al., 2007a)

\[
P(M \mid D, I) = \frac{P(M \mid I) P(D \mid M, I)}{P(D \mid I)}
\]

where \(M\) is a vector of parameters, which describes the characters of a source including spatial location of the source in three dimensions (x, y, z) and its strength; \(D\) is the measurement (observation) data; and \(I\) represents background information. The prior distribution \(P(M \mid I)\) expresses knowledge of \(M\) before the acquisition of data \(D\). It reflects the state of ignorance if the original parameter values are unknown. The evidence \(P(D \mid I)\), also known as the marginal likelihood, is obtained by marginalizing (integrating) the likelihood over the entire space. The parameter \(P(D \mid M, I)\) is the likelihood function, whereas the posterior distribution \(P(M \mid D, I)\) represents the probability of \(M\) given \(D\), and is the complete solution to an inference problem.

#### 2.2.2.2 Applications of Bayesian Approach in Groundwater Source Identification

Snodgrass and Kitanidis (1997) applied Bayesian analysis in source characterization of groundwater pollution. Specifically, the Bayesian framework was used to determine the release history of a conservative solute and to quantify the estimation error. Instead of simply using the Gauss–Newton iteration to update and obtain a best estimate of parameters \(s\) (a vector representing the function that is to be estimated), the Bayesian framework transformed \(s\) and then solved the equations iteratively. Their results showed that Bayesian analysis produced the best estimate of the release history and a confidence interval. Other advantages of this method were that it (1) required no inversion of matrices, (2) ensured more general solutions, and (3) made no blind assumptions.

Wang and Zabaras (2006) simulated the release history of contaminant in a constant porous media flow by solving the ADE with a hierarchical Bayesian approach backward through time.
The contaminant concentration was modeled as a pair-wise Markov Random Field that regularized the prior distribution of concentration history. Unlike Snodgrass and Kitanidis (1997), Wang and Zabaras (2006) accounted for both the standard deviation of measurement errors and the scaling parameter of the prior distribution and treated all the parameters as structure variables. The hierarchical Bayesian approach allowed for the quantification of uncertainty in structure parameters and the estimation of distribution of the structure parameters simultaneously with the computation of the concentration distribution. The examples used in this study cover both homogeneous and heterogeneous porous media and used a dimensionless form for generality.

Jin et al. (2010) presented a Bayesian approach using the Markov chain Monte Carlo (MCMC) method to infer the possible location and magnitude of the groundwater contamination source. They also provided an example based on a field experiment conducted in Canada (Borden site). Because one of their main goals was to reduce uncertainty, Metropolis–Hasting samplers were applied to generate samples from the posterior distribution. The major advantage of the Jin et al. (2010) approach over other traditional inverse approaches was that it provided the distribution over estimated parameters rather than a single but unrealistic solution.

2.2.2.3 Applications of Bayesian Approach in Source Identification of Air Pollution.

The Bayesian approach has also been widely used in source identification of atmospheric pollution. Keats et al. (2007a) proposed a Bayesian inference framework that involves two major techniques—applying the adjoint method to solve advection–diffusion equation efficiently and using MCMC algorithms to provide a series of samples whose stable distribution was the target PDF. The Bayesian inference in their framework provided the posterior PDF for the source parameters, including location and strength, given a finite and noisy set of concentration measurements obtained from real-time sensors. The first case study used the Mock Urban Setting Test that provided a water-channel simulation of near-field dispersion using a large array of shipping containers (or building-like obstacles). Propylene gas was used as a tracer and released from various locations within the array, both continuously and near-instantaneously. The case experiment of Keats et al. (2007a) not only produced continuous source concentration data within the simulation of a built-up area, but also provided an opportunity to study the effect of obstacles in the release procedure of tracer. The second case used mean concentration data from the Joint Urban 2003 atmospheric dispersion study in Oklahoma City, where a sulfur hexafluoride (SF₆) tracer was released continuously for 30 minutes and sampled at 7 locations.

Both case studies involved a highly disturbed flow field in an urban area and both demonstrated the utility of the method for practical applications in environment management. Yee (2008) applied a similar method for source reconstruction in the adjoint representation of atmospheric diffusion. First, a measured mean concentration at a given location and time was assumed to be the sum of a modeled signal and noise that represented the difference between the measured and modeled mean concentration. Next, the posterior PDF was obtained. The representation of the source–receptor relationship was formulated in both Eulerian and Lagrangian descriptions of turbulent dispersion. The Bayesian inferential methodology for source reconstruction was illustrated using two real data sets. The first was the Joint Urban 2003 using SF₆ and the other from the European Tracer Experiment (ETEX) using perfluoro (methylcyclohexane) as a tracer. The Yee (2008) study showed that the Bayesian probabilistic inferential method could be applied...
to (1) source reconstruction in the case of turbulent contaminant transport and dispersion in complex urban-industrial conditions, and (2) on a continental scale.

Guo et al. (2009) studied unsteady atmospheric dispersion of hazardous materials and the likelihood needed to be deduced by considering time. Using the adjoint advection–diffusion equation proposed by Keats et al. (2007a, 2007b), the unsteady adjoint equation was solved once by each sensor to obtain the concentration at the given sensor site at a specified time, rather than \( n \) times to solve an unsteady advection–diffusion equation. In the case study, a point source of some hazardous chemical/biological/radiological materials was released in an urban environment with three buildings with a dimension of 500 \( \times \) 500 \( \times \) 100 m. Their results showed that this framework using the unsteady adjoint transportation equation with MCMC was efficient and could improve the accuracy of source location in the wind direction compared to the steady inversion model.

2.2.2.4 Applications of Bayesian Approach in Source Tracking for Surface Waters.
Although applications of Bayesian approach in surface waters have not been as extensive as for porous media or atmospheric dispersion, some applications are noteworthy. Kildare et al. (2007) used Bayes’ theorem to calculate the conditional probability for detecting human fecal contamination in a watershed in California. Following the method developed by Snodgrass and Kitanidis (1997), Boano et al. (2005) applied a geostatistical method for recovering the contaminant source at a known location using a limited number of concentration measurements along a river. The effect of dead zones on solute transport process was also considered. Several cases were investigated to recover the release history, extending from a product-type source and a single measurement point to independent point sources and multiple measurement points.

2.2.3 Other Approaches.

2.2.3.1 Optimization Approach.
Early studies in tracking pollution sources in the field of groundwater research focused on forward simulation and compared solutions with observations (Atmadja and Bagtzoglou, 2001). As a result of both the non-uniqueness of solutions and the infinite number of plausible combinations, an optimization method was applied to acquire the best solution. One of the earliest attempts to use this approach was by Gorelick (1983) who used linear programming and multiple regressions to combine source identification with an optimization model. Notably, their method assumed no uncertainty in the physical parameters for the aquifer and could only be applied to cases where data were available in the form of breakthrough curves. Another optimization approach developed by Wagner (1992) used a nonlinear maximum likelihood estimation to first depict the inverse model and then to perform simultaneous parameter estimation and source characterization.

The optimization approach typically involves the minimization or maximization of an objective function. The linear optimization methods have the following general form (Eq. 2.5):

\[
h(p) = h_0 + J_{hp}(p - p^*)
\]

where \( h \) is a vector of observations of state variables, \( p \) is a vector of model parameters, and \( J_{hp} \) is the Jacobian sensitivity matrix. Five essential steps involved in the optimization method were described by Carrera et al. (2005)

1) Initialization: read input data, set iteration counter \( i = 0 \), and initialize parameter, \( p^0 \).
2) Solve the simulation problem, \( h(P^i) \); compute the objective function \( F^i \), and possibly its gradient (assuming that it is continuously differentiable), \( g^i \), and the Jacobian matrix, \( J_{hp} \).

3) Compute an updating vector, \( d \), possibly using information on previous iterations, as well as \( g^i \) and \( J_{hp} \).

4) Update parameters \( P^{i+1} = P^i + d \).

5) If convergence has been reached, then stop; otherwise, set \( i = i + 1 \) and return to step (2).

Application of artificial intelligence techniques have increased sharply in recent decades. Mirghani et al. (2009) proposed a simulation–optimization approach based on a parallel evolutionary strategy for resolving pollution source identification problems. In their approach, the numerical pollutant–transport model was coupled with an evolutionary search algorithm and solved iteratively during each search. Three scenarios were considered in which the set of design variables could be described as \( (x_c, y_c, z_c, C_0) \), \( (x_c, y_c, z_c, s, C_0) \), and \( (x_1, y_1, z_1, x_2, y_2, z_2, C_0) \); where \( x_c, y_c, \) and \( z_c \) were the coordinates for the centroid of the contaminant source; \( x_k, y_k, \) and \( z_k \) \( (k = 1, 2) \) were the coordinates of the vertices of diagonally opposite corners of a hexahedron-shaped source; and \( C_0 \) was the initial source concentration. A three-dimensional (3D) groundwater domain with both a homogeneous and heterogeneous velocity field was considered. Mirghani and colleagues (2009) reported that the evolutionary strategy performed adequately for all scenarios, although the performance was affected by problem complexity. They also reported that the effect of non-uniqueness became more pronounced while increasing the number of design variables.

Jin et al. (2009) used a genetic algorithm-based procedure for 3D source identification for the Borden emplacement site in Canada. They considered the site as a test problem and employed a parallel hybrid optimization framework that coupled a real genetic algorithm with a local search approach (Nelder–Mead simplex). The local search results showed that one of these starting points would lead to the true solution when measurement or model errors were negligible. However, the procedure might also lead to multiple possible solutions if the errors were significant. The authors suggested use of a new selection criterion based on the metrics of mean and standard deviation of objective function values to address the non-unique solution problem.

A heuristic harmony search is a recent optimization algorithm that, like a musical process, seeks harmony through use of several improvisations. In the work of Ayvaz (2010), decision variables included locations and release histories of pollution source and were determined through the optimization model. The author used the model for two hypothetical examples that took into account the simple and complex aquifer geometries, measurement error conditions, and different heuristic harmony search solution parameter sets. Results indicated that the model was an effective tool for solving pollution source identification. One advantage of the model was that source locations and release histories, in conjunction with potential source numbers, were determined using the proposed implicit solution procedure. However, because the performance and efficiency of the model might depend on the availability of observation data to represent the transport process in the groundwater system, insufficient data can cause result in inaccurate source characteristics. Also, assumptions of no uncertainty in boundary conditions, hydraulic conductivity, and dispersivity fields were not realistic. Further investigations into uncertainty analysis might be helpful to address these issues.
2.2.3.2 Geostatistical Approach.
Geostatistical approaches belong to the probabilistic approach and are widely used in groundwater studies. The use of geostatistics is motivated by the need to address the spatial variability in hydraulic properties of aquifers (Kitanidis, 1995, 1996; Michalak and Kitanidis, 2004; Snodgrass and Kitanidis, 1997).

Bagtzoglou, Tompson, and Dougherty (1992) and Bagtzoglou, Dougherty, and Tompson (1992) were among the first studies that attempted to solve the ADE backward in time under a probabilistic framework. Those authors reversed the advection portion of the transport model, retained the dispersion portion using the random walk particle method, and employed a probabilistic framework to identify pollution sources in heterogeneous media. Although the studies were preliminary in the field, they successfully assessed the relative importance of each potential pollution source. Neupauer and Wilson (1999) proposed an adjoint method that replaced the forward governing equation with adjoint equation using the adjoint state as the dependent variable. Later, Neupauer and Wilson (2004) extended their previous work to a backward location and a probabilistic model for travel time, which could be used to quantify the release history and location of known and unknown pollution sources. Although the governing equation for their model remained the adjoint equation, new load terms were added with some approximations from a cell-centered finite difference method. Both hypothetical and real cases were simulated using MODFLOW-96 and MT3DMS.

Snodgrass and Kitanidis (1997) combined Bayesian theory with a geostatistical approach to estimate the release history of a conservative solute using available information (e.g., point concentration measurements at certain time after the release). A confidence interval for the best estimate was produced and conditional realizations of the release history were generated for visualization and risk analysis. Their method was considered to be general and included the Tikhonov regularization as a special case, which was commonly used to transform the ill-posed inverse problem into a minimization problem (Skaggs and Kabala, 1994).

Sun (2007) proposed a robust geostatistical approach (RGS) for contaminant source identification with the aim to reduce the effect of uncertainty introduced in the model-building process. The RGS is an extension of the geostatistical approach and can be used in any problem where a geostatistical approach is suitable. Through the use of a case study, the author demonstrated the ability of the RGS model to identify the pollution source release history in a two-dimensional (2D) aquifer, and reported that the overall performance of the RGS model exceeded that of the geostatistical model.

2.2.3.3 Direct Approach
The direct approach solves directly reversed governing equations that describe cause-effect relationships. However, application of the regularization or stabilization technique is needed to improve the solution stability for the direct approach. Examples of the direct approach include use of the quasi-reversibility method (Skaggs and Kabala, 1995), the minimum relative entropy (MRE) inversion (Woodbury and Ulrych, 1996), and the marching-jury backward beam equation (MJBBE) (Atmadja and Bagtzoglou, 2001).

Skaggs and Kabala (1995) was among the first attempts to apply the quasi-reversibility method for solving inverse problems in groundwater contamination. They used a moving coordinate system for the velocity term in the ADE, and solved the equation with a negative time step.
Results from the quasi-reversibility method showed less accuracy than that of Tikhonov regularization, but the method required less computation time. Zhang and Chen (2007) employed the quasi-reversibility method with an inverse CFD model to identify the location and strength of gaseous contaminant sources in a 2D aircraft cabin and in a 3D office. They reported that the method worked better for convection-dominated flows than the flows dominated by other terms. The MRE inversion was proposed by Woodbury and Ulrych (1996) for reconstructing (1) the source history with and without noise and (2) a 3D plume source within a one-dimensional constant velocity and dispersivity field. Neupauer et al. (2000) compared the performance of Tikhonov regularization and MRE and found that both methods performed well for reconstruction of a smooth source function, but the MRE performed better for an error-free step function source history.

It is important to emphasize out that most of the above studies addressed homogeneous media. For heterogeneous media, Atmadja and Bagtzoglou (2001) developed the MJBBE—a hybrid between a marching and a jury method that enhanced and altered the backward beam equation (BBE) method—to recover the time history and spatial distribution of a groundwater contaminant plume from the current position by solving the ADE with heterogeneous parameters. A subsequent study by Bagtzoglou and Atmadja (2003) compared the performance of MJBBE and a quasi-reversibility method in reconstructing spatial distributions of a conservative contaminant plume. Cases using spatially uncorrelated and correlated, stationary and nonstationary, homogeneous, and deterministically heterogeneous dispersion coefficient fields were presented for comparison purpose. Results showed that the MJBBE was superior in handling heterogeneous fields and was able to preserve the salient features of the initial input data. In contrast, the quasi-reversibility method performed better in cases with homogeneous parameters.

2.2.4 Source Tracking in Water Distribution Systems.

Tracking contamination source in a water distribution system is different than in other media. First, water distribution systems are generally closed environments driven by pressure. Second, contamination warning systems can monitor water quality in the distribution system, detect contamination autonomously, and provide support for remedial actions to minimize public health effects (De Sanctis et al., 2010). Another unique feature of water distribution systems is the use of special terms. For example, the water quality state at each sensor is either positive (abnormal state) or negative (normal state).

Di Cristo and Leopardi (2008) presented a simple method for locating the source of accidental contamination in a water distribution network. Their method first used a pathway analysis of the network and the demand coverage concept for an initial selection of possible pollution source nodes. Then, the inverse water quality problem was solved through an optimization approach using the water fraction matrix concept. Kim et al. (2008) discussed the application of artificial neural network (ANN) models in locating pollutants either accidentally or deliberately injected into a water distribution system. The authors measured the spatiotemporal distribution of Escherichia coli along the water distribution system with sensors. Using ANN models, the transport pattern of E. coli was inversely interpreted to identify the source location. Results revealed a positive correlation between the E. coli dispersion pattern and pH, turbidity, and conductivity. Based on the pre-programmed relationship between the E. coli transport pattern
and release locations, the ANN model identified the source location of *E. coli* with up to 75% accuracy.

De Sanctis et al. (2010) proposed a practical and efficient method for real-time identification of possible locations and times that were responsible for contamination incidents detected by sensors. Because the sensors used in their study could only detect qualitative concentration of a contaminant (i.e., positive or negative status), locations and times connected to positive sensor measurements were considered to be the possible sources. A contamination status algorithm was developed using results from particle backtracking algorithm to (1) update the contamination status for all candidate source nodes and time intervals, and (2) identify flow paths and travel times. A linear relationship between output node concentration and mass additions at upstream input nodes could be described as (Eq. 2.6)

\[
C_j(t) = \sum_{(i,T) \in U_j(t)} I_{ij}^T(t)u_i^T
\]

where \(C_j(t)\) = contaminant concentration at output node \(j\) and time \(t\), \(u_i^T\) = contaminant source strength at input node \(i\) during time interval \(T\), and the impact coefficient \(I_{ij}^T(t)\) = concentration response at output node \(j\) and time \(t\) to a unit source strength addition at input node \(i\) during time interval \(T\). The parameter \(U_j(t)\) represents upstream reachability sets. Each \((i,T) \in U_j(t)\) is connected to downstream output node \(j\) at time \(t\) so that source strength \(u_i^T\) has a non-zero effect on concentration \(C_j(t)\).

Propato et al. (2010) used an entropic-based Bayesian inversion technique to solve the variables after ruling out potential contaminant injections in a drinking water system through use of linear algebra. Their approach allowed for the less committed prior distribution with respect to unknown information and the incorporation of model uncertainties and measurement errors. The solution was a space-time contaminant concentration PDF that accounted for various possible contaminant injections.

2.3 Discussion

Following the work of Atmadja and Bagtzoglou (2001), the various modeling methods discussed above for pollution source identification can be grouped into two general classes of approach—optimization and probability. Both classes have advantages and disadvantages. The optimization approach solves the inverse problem by finding a unique, but possibly false, solution that minimizes differences between modeled and observed data. In contrast, the probability approach provides a set of possible solutions along with their probabilities. Commonly used probabilistic, optimization, and other approaches in contaminant source tracking, and issues associated with their use, are listed in Table 2.1.

2.3.1 Source Tracking in Surface Water.

Although there are a wide variety of investigations of source identification in terms of the location and the release history of contaminants in groundwater and atmosphere, there are only a few studies on source identification of contaminant in surface waters. Shen et al. (2006) treated load estimation as searching for a set of constant daily loads to minimize a defined goal function
(or cost function) that measured the difference between model predictions and observations. The mathematical expression could be written as follows (Eq. 2.7):

$$J(C : \beta^*) = \min[J(C : \beta)]$$

subject to

$$\beta^* \in \beta_0, F(\beta) = 0$$

where $J$ was defined as a goal function (or cost function); $\beta^* = (\beta_1, \beta_2, \ldots, \beta_m)$ was the constant loads from sub-watersheds; $m$ was the total number of sub-watersheds; and $\beta_0$ was an acceptable set of loads.

Cheng et al. (2010) used a backward location PDF method to locate point sources in surface waters. They established the relationship between a forward and backward location PDF with depth-averaged free-surface flow and mass transport models. Hypothetical cases were performed to evaluate the random error and number of observed values associated with the method. In a real case study, dye tracer was injected into a stream instantaneously. Results from the case study indicated that (1) the number of ADEs needed to solve the problem is equal to the number of observations, and (2) this method was efficient for the case of single point source and multiple observation points in the domain.

### 2.3.2 Identification of Multiple Point Sources.

Although most the preceding studies focused on a single point pollution source, the identification of multiple sources of pollution has drawn increasing attention. Yee (2007) developed a Bayesian inferential framework for the joint determination of the number of contaminant sources and the parameters for each source, given a finite number of concentration observations obtained from an array of sensors. The reversible-jump MCMC algorithm was used in cases where the number of sources was unknown a priori to ensure the simultaneous exploration of several prospective contaminant source models. The method was applied to two and three source case studies and the results showed that the accuracy of the two source case study was good whereas the three source case study was associated with large uncertainty, especially in parameters for the furthest contaminant source. Although both examples were tested with field experiments rather than through use of real concentration data sets and only atmospheric pollution was considered, the basic concept of Yee’s method could still be useful to water pollution source identification.

Hon et al. (2010) proposed a method based on Green’s function to solve specific classes of inverse source identification problems. Their method could be employed to recover both the intensities and locations of unknown point sources from scattered boundary measurements. Two assumptions were made, (1) locations of point sources were given with unknown intensities to be recovered from $N$ distinct boundary measurements, and (2) locations of point sources were not known but an estimated location was given for each unknown point source. Numerical results indicated that the proposed method was accurate and reliable for both bounded and unbounded domains under various boundary conditions.

### 2.3.3 Future Perspectives on Identification of Critical Bacteria Source Areas.

A broad spectrum of methodologies and technologies has been proposed for BST. Each method/technology is associated with advantages and disadvantages. As noted by many researchers, although no single method is generally applicable to identification of all bacterial
sources—especially nonpoint sources at the watershed-scale—combined applications of various methods and technologies have shown promise.

Table 2.1 Summary of deterministic direct methods and probabilistic methods for pollution source identification.

<table>
<thead>
<tr>
<th>Method</th>
<th>Reference(s)</th>
<th>Issue(s)</th>
<th>Media</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Optimization approach (deterministic direct methods)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear optimization model</td>
<td>Gorelick, 1983</td>
<td>Reconstruction of release history and spatial distribution</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Nonlinear optimization</td>
<td>Kathirgamanathan et al., 2002; Wagner, 1992</td>
<td>Tracking location and strength of a point source</td>
<td>Groundwater, open air</td>
</tr>
<tr>
<td>Tikhonov regularization</td>
<td>Skaggs and Kabala, 1994</td>
<td>Reconstruction of release history</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Quasi-reversibility</td>
<td>Skaggs and Kabala, 1995</td>
<td>Reconstruction of release history</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Backward beam equation</td>
<td>Atmadja and Bagtzoglou, 2001</td>
<td>Recovering time history and spatial distribution</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Quasi-reversibility with computational fluid dynamics (CFD) modeling</td>
<td>Zhang and Chen, 2007</td>
<td>Identifying location and strength of sources</td>
<td>Indoor air pollution</td>
</tr>
<tr>
<td><strong>Probabilistic approach</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random walk particle methods + geostatistics</td>
<td>Bagtzoglou, Dougherty, and Tompson, 1992; Bagtzoglou, Tompson, and Dougherty, 1992</td>
<td>Recovering release history</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Stochastic differential equations</td>
<td>Wilson and Liu, 1994</td>
<td>Recovering release history</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Adjoint method</td>
<td>Neupauer and Wilson, 1999</td>
<td>Identifying location and travel time probabilities</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Minimum relative entropy</td>
<td>Woodbury and Ulrych, 1996</td>
<td>Reconstruction of source history</td>
<td>Groundwater</td>
</tr>
<tr>
<td>Bayesian theory</td>
<td>Boano et al., 2005; Guo et al., 2009; Jin et al., 2010; Keats et al., 2007a, 2007b; Yee, 2008; Snodgrass and Kitanidis, 1997; Wang and Zabaras, 2006</td>
<td>Recovering release history</td>
<td>Groundwater, surface water, Air pollution</td>
</tr>
<tr>
<td><strong>Other approaches</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artificial neural network (ANN)</td>
<td>Kim et al., 2008</td>
<td>Tracking source locations</td>
<td>Water distribution systems</td>
</tr>
</tbody>
</table>

### 2.3.4 Combination of Methods.

An effective way to track locations of unknown sources of pollution is the combination of biological and mathematical methods. For example, a specific biological method is used to winnow potential sources to several candidate sources, then one of the inverse modeling methods
discussed above is used to identify the possible locations of contaminant sources. The general procedure includes the following steps (details provided in Figure 2.3):

1) Identify the host origin of bacteria with a biological method.
2) Collect point concentration measurements from sampling sites downstream of the release site.
3) Employ a mathematical method to identify the location and strength of unknown contaminant sources.

![Flowchart for watershed-scale bacterial source identification](image)

Figure 2.3 Flowchart for watershed-scale bacterial source identification
2.3.5 Application of Biosensors and Remote Sensing Technology.

Biosensors have become increasingly used in water quality monitoring applications. Biosensors can detect, record, and transmit information regarding a physiological change or the presence of multiple chemical and biological materials in the aquatic environment. Biosensors use the selectivity and sensitivity of a biological component coupled with an electronic component to yield a measurable signal (Batzias and Siontorou, 2007; Malhotra et al., 2005; Rodriguez-Mozaz et al., 2005). Several studies have reported the successful use of biosensors to detect bacteria in water bodies (e.g., Ivnitski et al., 1999; Lazcka et al., 2007; Rogers, 2006; Varshney and Li, 2009).

Baeumner et al. (2003) developed a highly sensitive and specific RNA biosensor for rapid detection of viable *E. coli* in drinking water. The biosensor could detect and quantify *E. coli* messenger RNA in 15 to 20 minutes. When correlates with a much more elaborate (and expensive) laboratory-based detection system, the biosensor can detect as few as 40 *E. coli* CFU/mL. Sun et al. (2006) proposed a flow-through piezoelectric quartz crystal (PQC)/DNA biosensor that combined sequential flow polymerase chain reaction (PCR) products denaturing before PQC detection via hybridization of single-stranded DNA (ssDNA). The detection limit of this device was 23 *E. coli* cells per 100 mL water. Sun et al. (2009) subsequently developed a system based on photodeposition of nano-silver at a titanium oxide-coated PQC electrode with an enhancement of 3.3 times for binding of complementary DNA onto the new biosensor, leading to a detection limit of 8 *E. coli* cells per 100 mL water. Berganza et al. (2007) developed a DNA biosensor that immobilized a ssDNA probe onto an electrochemical transducer surface to recognize a specific *E. coli* O157:H7 complementary target DNA sequence.

Nijak et al. (2011) proposed an autonomous, wireless in-situ sensor for rapid detection of *E. coli* in water. The sensor could detect low concentrations in less than 8 hours and higher concentrations within an hour. For the detection of *Streptococcus pyogenes*, Nugen et al. (2007) developed a software program that allowed the addition of oligotags as required by nucleic acid sequence-based amplification methods. They also designed a novel lateral flow biosensor, reducing detection times to 20 minutes and obtained a sensitivity of 135 ng. For detection of multiple pathogens, Langer et al. (2009) described a new ON–OFF type nanobiodetector to test for bacteria *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Escherichia coli*, and *Enterococcus faecalis*. Garcia-Aljaro et al. (2010) described the carbon nanotube-based immuno-sensors for detection of bacteria (*E. coli* O157:H7 and *E. coli* K12) and viruses (bacteriophage T7 and MS2). Vikesland and Wigginton (2010) reviewed the current literature on applications of nano-enabled biosensors to detect whole cells, particularly for waterborne pathogens. Additional studies have been published on other biosensor materials and targets (Mauter and Elimelech, 2008; Pang et al., 2007; Su et al., 2011).

Remote sensing technology provides an efficient alternative to other sampling methods like grab sampling and in-situ sensing, which are typically too expensive to implement across large spatial scales at high resolution. In addition, remote sensing is a non-intrusive measuring method that limits human exposure to pathogenic bacteria and viruses. Application of advancing water quality monitoring technology in combination with probability-based modeling tools provides an effective approach to address bacterial source identification.
Despite widespread applications of remote sensing technology in water quality monitoring, algorithms for direct measurement of bacterial level using remote sensing are still rare. Vincent et al. (2004) presented an imaging algorithm for Landsat TM data to map early blooms of cyanobacteria (blue-green algae) in Lake Erie and its tributaries. The 30-m spatial resolution of Landsat TM helped map bacteria in streams with widths ≥90 m and water depths ≥2 m. An indirect way to measure bacterial level is to first establish a functional relationship between the bacterial level and several independent surrogate variables that can be directly measured using remote sensing. Then, the bacterial level can be determined by indirectly measuring the surrogate water quality parameters such as total suspended solids, chlorophyll, colored dissolved organic matter (Hu et al., 2004; Wong, et al., 2008), and other parameters (Zhang et al., 2012).

2.4 Conclusions

A wide variety of methods and technologies have been developed for bacterial source identification. These range from biological methods for host tracking, mathematical models for source location tracking and release history reconstruction, and sensor technologies for water quality monitoring. Although some of these methods have been used independently, others are typically combined when applied to real water quality problems.

A comprehensive watershed-scale source tracking generally involves the following three tracking steps: geographical, mathematical, and biological. In terms of geographical tracking, bacterial source location must be identified to construct structural BMPs or LID for site treatments. Sensor technologies—especially remote sensing—can play an important role in locating bacterial source areas. In terms of mathematical tracking, the quantity (strength) or release history of bacterial source must be computed to develop TMDLs for bacterial load reduction and water quality restoration. Mathematically, source tracking is essentially an inverse modeling issue under uncertainty. Therefore, inverse modeling in combination with a geostatistical method or an optimization algorithm is necessary. In terms of biological tracking, the host origin of bacterial source should be identified to support sustainable management of the watershed. Consequently, a combined application of biological methods, mathematical models, and sensor technologies (including remote and in-situ sensing) provides an effective approach for the identification of critical source areas of bacteria at the watershed-scale.
CHAPTER 3  
MOMENT-BASED METHOD FOR IDENTIFICATION OF POLLUTION SOURCE IN RIVERS

3.1 Introduction

Identification of pollution source in a river is of vital importance to environmental protection and particularly emergency response in case of accidental chemical spills or terrorist attacks. Illegal discharges and storm event-induced pollutant discharges are other forms of accidental releases. According to the U.S. EPA, nearly 20% of pollution sources are unknown among the pollution sources that lead to waterbody impairments in the USA (U.S. EPA 2010a). A wide variety of methods have been proposed for identification of unknown pollution sources, ranging from biological methods, numerical modeling, optimization methods, probabilistic analysis, and sensor technologies (Tong and Deng 2012). From the perspective of mathematics, source identification is essentially an ill-posed inverse modeling problem, which could be further divided into boundary problems, retrospective problems, coefficient problems, and geometric problems (Zhang and Chen 2007). A number of methods have been proposed for solving an inverse problem, such as linear optimization (Gorelick et al. 1983), Tikhonov Regularization (TR) (Skaggs and Kabala 1994), Quasi-Reversibility (QR) (Skaggs and Kabala 1995), Backward Beam Equation (BBE) (Atmadja and Bagtzoglou 2001), Random Walk Particle methods (Bagtzoglou 1992a, 1992b), Minimum Relative Entropy (MRE) (Woodbury and Ulrych 1996; Ulrych and Woodbury 2003; Woodbury 2011), Bayesian approach (Snodgrass and Kitanidis 1997; Boano et al. 2005; Wang and Zabaras 2006; Keats et al. 2007a, 2007b; Yee et al. 2008), and Artificial Neural Network (Kim et al. 2008).

Boano et al. (2005) presented a geostatistical method, similar to the one proposed by Snodgrass and Kitanidis (1997), for recovering the release history from a number of observations. They used the transient storage model to account for the contaminant interaction between main channel and storage zones under several cases from a single measurement point to multiple measurement points. Cheng and Jia (2007) developed a probability-based method for tracking point sources in surface water. They employed a backward location probability density function (BL-PDF) which was connected with a forward location probability density function (FL-PDF) by adjoint analysis. Their relation was validated using depth-averaged free-surface flow and mass transport models. Both hypothetical and real cases were studied to identify the location of injected dye tracer with the distributions of pollutant concentration observed at downstream monitoring stations. In spite of the extensive efforts, no existing method has been generally accepted as a reliable method for identification of pollution sources primarily due to limitations of previous methods in simulation of pollutant dispersion and transport in rivers. The Variable Residence Time (VART) model, presented by Deng and Jung (2009) and further extended by Deng et al. (2010), shows great promise in simulation of dispersion and transport of various conservative and reactive pollutants (Helton et al. 2011; Jung and Deng 2011; Anderson and Phanikumar 2011; Liao and Cirpka 2011; Zahraefard and Deng 2012) in river systems.

---

1 This Chapter 3 previously appeared as [Yangbin Tong, Zhi-Qiang Deng, Moment-based Method for Identification of Pollution Source in Rivers. Preview Manuscript]. With permission from ASCE.
The overall goal of this paper is to present a simple yet effective method for identification of pollution sources in rivers using the VART model. Source tracking generally requires the identification of both source location and quantity for accidental pollutions in rivers. Therefore, the specific objectives of this paper are: (1) to find an effective method for determining the location of an accidental discharge and (2) to provide a method for estimating the total mass released from the accidental discharge. The objectives will be addressed by deriving two moment equations using the VART model and the Laplace transform.

3.2 Materials and Methods

3.2.1 Data Collection

![Figure 3.1 Location of Five Dye Test Rivers in USA](image)

To examine the performance of the moment-based method, conservative tracer injection experiments conducted by U.S. Geological Survey (USGS) in Monocacy River, Bayou Bartholomew, Tangipahoa River, Red River and Mississippi River are obtained from the USGS report by Nordin and Sabol (1974) (See Figure 3.1). Rhodamine WT was instantaneously injected into the rivers. The general information of these experiments is given in Table 3.1.

Table 3.1 Tracer Injection Experiments

<table>
<thead>
<tr>
<th>River Reach</th>
<th>Length (km)</th>
<th>Date</th>
<th>Injection Type*</th>
<th>Weight (kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monocacy River</td>
<td>34.3</td>
<td>7-Jun-68</td>
<td>1</td>
<td>1.90</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>117.5</td>
<td>25-Jun-71</td>
<td>-</td>
<td>2.81</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>94.0</td>
<td>15-Sep-69</td>
<td>-</td>
<td>3.81</td>
</tr>
<tr>
<td>Red River</td>
<td>193.1</td>
<td>7-Apr-71</td>
<td>4</td>
<td>23.42</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>294.5</td>
<td>7-Aug-68</td>
<td></td>
<td>108.86</td>
</tr>
</tbody>
</table>

(*Type of injection: 1 - slug at center; 4 - line source across width.)

The injection type 1 means injection in the center while type 4 means injection across the river width. For Bayou Bartholomew and Tangipahoa River, we assume the tracer was injected in the center because it was the most used injection type in the whole USGS dye tests. For each experimented river, there were multiple sampling sites downstream of the injection point. At
these sampling sites, where the RWT was fully mixed across the river, the water samples were taken and analyzed in the USGS lab for tracer concentration. The length in Table 3.1 denotes the total distance from tracer injection point to the last sampling site.

Table 3.2 summarizes flow and other parameters related to the dispersion and transient storage of tracer in the experimental reaches. The column ‘L’ shows the distance from the tracer injection point to the sampling site. In addition, at each sampling site, USGS also recorded information of discharge and channel cross-section geometry. Based on these data, the flow velocity was calculated. The USGS report also provided dispersion coefficient for each river reach of the tested rivers. They were used in different models by various researchers (Seo and Cheong, 1998; Deng et al., 2001, 2002), and referred to as measured dispersion coefficients. All USGS data were checked carefully by the USGS staff before they posted them in the report. They were reliable and were already used by different researchers. Other parameters, including the ratio of storage zone, effective diffusion coefficient, and residence time, are obtained from Deng et al. (2009, 2010).

Table 3.2 Flow and VART Model Parameters for Tracer Test Reaches

<table>
<thead>
<tr>
<th>River</th>
<th>L (km)</th>
<th>U (m/s)</th>
<th>Ks (m²/s)</th>
<th>A_{adv}/A</th>
<th>D_{eff}/A(1/s)</th>
<th>T_{min}(h)</th>
<th>Q(m³/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Distance from injection</td>
<td>Velocity</td>
<td>Dispersion coefficient</td>
<td>Ratio of storage zone</td>
<td>Effective diffusion coefficient</td>
<td>Residence time</td>
<td>Discharge</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>10.3</td>
<td>0.397</td>
<td>29.6</td>
<td>0.25</td>
<td>-9.2E-07</td>
<td>1.05</td>
<td>15.57</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>18.3</td>
<td>0.349</td>
<td>29.6</td>
<td>0.10</td>
<td>-2.5E-07</td>
<td>1.78</td>
<td>15.15</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>26.8</td>
<td>0.391</td>
<td>29.6</td>
<td>0.10</td>
<td>-2.3E-07</td>
<td>2.71</td>
<td>15.86</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>34.3</td>
<td>0.335</td>
<td>29.6</td>
<td>0.25</td>
<td>-3.6E-07</td>
<td>3.04</td>
<td>18.55</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>3.2</td>
<td>0.155</td>
<td>54.7</td>
<td>0.15</td>
<td>-7.6E-07</td>
<td>1.54</td>
<td>4.11</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>25.7</td>
<td>0.113</td>
<td>54.7</td>
<td>0.10</td>
<td>-2.4E-07</td>
<td>10.69</td>
<td>4.81</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>59.5</td>
<td>0.150</td>
<td>54.7</td>
<td>0.22</td>
<td>-1.6E-07</td>
<td>18.96</td>
<td>6.51</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>117.5</td>
<td>0.145</td>
<td>54.7</td>
<td>0.13</td>
<td>-2.7E-07</td>
<td>25.64</td>
<td>8.10</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>8.2</td>
<td>0.165</td>
<td>44.0</td>
<td>0.20</td>
<td>-4.3E-07</td>
<td>2.44</td>
<td>3.45</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>18.0</td>
<td>0.170</td>
<td>44.0</td>
<td>0.20</td>
<td>-3.4E-07</td>
<td>5.14</td>
<td>4.64</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>41.5</td>
<td>0.280</td>
<td>44.0</td>
<td>0.20</td>
<td>-2.4E-07</td>
<td>7.48</td>
<td>6.94</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>55.4</td>
<td>0.320</td>
<td>44.0</td>
<td>0.35</td>
<td>-7.4E-06</td>
<td>6.98</td>
<td>8.10</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>71.0</td>
<td>0.315</td>
<td>44.0</td>
<td>0.50</td>
<td>-1.5E-05</td>
<td>9.45</td>
<td>8.61</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>82.1</td>
<td>0.308</td>
<td>44.0</td>
<td>0.40</td>
<td>-1.6E-05</td>
<td>9.09</td>
<td>9.03</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>94.0</td>
<td>0.266</td>
<td>44.0</td>
<td>0.30</td>
<td>-1.7E-05</td>
<td>11.49</td>
<td>10.85</td>
</tr>
<tr>
<td>Red River</td>
<td>5.7</td>
<td>0.638</td>
<td>143.8</td>
<td>0.15</td>
<td>-7.2E-07</td>
<td>0.34</td>
<td>230.22</td>
</tr>
<tr>
<td>Red River</td>
<td>75.6</td>
<td>0.631</td>
<td>143.8</td>
<td>0.15</td>
<td>-4.8E-07</td>
<td>6.08</td>
<td>245.22</td>
</tr>
<tr>
<td>Red River</td>
<td>132.8</td>
<td>0.599</td>
<td>143.8</td>
<td>0.16</td>
<td>-1.5E-07</td>
<td>8.63</td>
<td>249.47</td>
</tr>
<tr>
<td>Red River</td>
<td>193.1</td>
<td>0.508</td>
<td>143.8</td>
<td>0.05</td>
<td>-2.4E-07</td>
<td>13.63</td>
<td>249.47</td>
</tr>
</tbody>
</table>
(Table 3.2 Continued)

<table>
<thead>
<tr>
<th>River</th>
<th>Distance from injection</th>
<th>Velocity</th>
<th>Dispersion coefficient</th>
<th>Ratio of storage zone</th>
<th>Effective diffusion coefficient</th>
<th>Residence time</th>
<th>Discharge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mississippi River</td>
<td>54.7</td>
<td>1.480</td>
<td>374.1</td>
<td>0.15</td>
<td>-3.3E-07</td>
<td>0.77</td>
<td>6824.36</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>96.6</td>
<td>1.550</td>
<td>374.1</td>
<td>0.30</td>
<td>-5.7E-07</td>
<td>1.19</td>
<td>6824.36</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>118.0</td>
<td>1.110</td>
<td>374.1</td>
<td>0.55</td>
<td>-8.1E-07</td>
<td>2.92</td>
<td>6824.36</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>294.5</td>
<td>1.574</td>
<td>374.1</td>
<td>0.05</td>
<td>0</td>
<td>6.37</td>
<td>6824.36</td>
</tr>
</tbody>
</table>

### 3.2.2 VART Model Descriptions

The moment-based method for identification of pollution sources requires a mathematical model for simulation of pollutant dispersion and transport in rivers. The VART model has been found to be relatively simple yet efficient (Deng and Jung 2009; Deng et al. 2010). The VART model simulates mass transport in streams and the exchange of mass between water column and two layers of transient storage zones. The upper layer is an advection-dominated transient storage zone which includes in-stream and shallow hyporheic storage. The lower layer is an effective diffusion-dominated storage zone that is deeper in the streambed and farther beneath the banks. If there is no lateral inflow/outflow and water loss/gain, the VART model for conservative pollutant transport can be expressed as:

\[
\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = K_s \frac{\partial^2 C}{\partial x^2} + \frac{A_{adv}}{A} \frac{1}{T_v} (C_s - C) \quad (3.1a)
\]

\[
\frac{\partial C_s}{\partial t} = \frac{1}{T_v} (C - C_s) \quad (3.1b)
\]

\[
A_{dif} = 4\pi D_s t_s \quad (3.1c)
\]

where \( C \) = solute concentration \([M/L^3]\) in main channel; \( C_s \) = solute concentration \([M/L^3]\) in storage zones; \( U \) = cross-sectionally averaged flow velocity \([L/T]\) in \( x \) direction; \( K_s \) = longitudinal Fickian dispersion coefficient excluding the transient storage effect \([L^2/T]\); \( T_v \) = actual varying residence time \([T]\) of solute; \( t_s \) = time \([T]\) since the solute release from storage zones to the main stream; \( D_s \) = effective diffusion coefficient \([L^2/T]\) in the storage zone; \( A \) = cross-sectional flow area of main channel \([L^2]\); \( A_{adv} \) = area \([L^2]\) of advection-dominated transient storage zone; and \( A_{dif} \) = area \([L^2]\) of effective diffusion-dominated transient storage zone. We consider \( t \geq T_{min} \) here, leading to \( T_v = t \) and \( t_s = t - T_{min} \) where \( T_{min} \) = minimum mean residence time \([T]\) for solute to travel through the advection-dominated storage zone. Then, the VART model can be simplified as:

\[
\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} = K_s \frac{\partial^2 C}{\partial x^2} + \frac{A_{adv} + 4\pi D_s (t - T_{min})}{A} \frac{1}{t} (C_s - C) \quad (3.2a)
\]

\[
\frac{\partial C_s}{\partial t} = \frac{1}{t} (C - C_s) \quad (3.2b)
\]
If the pollutant is instantaneously released into the river, the concentration of pollutant is a function of both time and distance: $C(x,t)=C_0 \delta(x) \delta(t)$, where $C_0 = M / A$, $M$ is the total mass released across the area $A$, and $\delta(\cdot)$ = Dirac delta function. The initial and boundary conditions for equations (3.2a) and (3.2b) include:

$$
C(x,0)=C_0 \delta(x) \tag{3.3a}
$$
$$
C_s(x,0)=0 \tag{3.3b}
$$
$$
\frac{\partial C(x,t)}{\partial x} = 0, \quad x = 0 \tag{3.3c}
$$
$$
C(x,t)=0, \quad x = \infty \tag{3.3d}
$$
$$
C_s(x,t)=0, \quad x = \infty \tag{3.3e}
$$

The VART model is utilized to derive the moment equations for pollutant source identification.

### 3.2.3 Moment Equations

The basic Laplace transform of concentration $C(x,t)$ with respect to time $t$ is defined as:

$$
\bar{C}(x, p) = \int_0^\infty e^{-pt} C(x,t) dt \tag{3.4}
$$

A similar definition for Laplace transform of concentration $C(x,t)$ with respect to distance $x$ is also employed. Applying Laplace transform to equations (3.2a) and (3.2b) with respect to time and making some necessary arrangements results in the following transformed concentration distribution:

$$
\bar{C}(x, p) = \frac{C_0}{\sqrt{4 \pi D_s}} \exp\left(-\frac{U - \sqrt{U^2 + 4K_s \left(\frac{4\pi D_s}{A} + p\right)}}{2K_s} x\right) \tag{3.5}
$$

More details about the derivation of Eq. (3.5) can be found in the Appendix. The inverse Laplace transformation of the above equation gives

$$
C(x,t) = \frac{C_0}{2\sqrt{K_s \pi t}} \exp\left[\frac{U x}{2K_s} - \left(\frac{U^2}{4K_s} + \frac{4\pi D_s}{A}\right) t - \frac{x^2}{4K_s t}\right] \tag{3.6}
$$

Setting $p = 0$ in equation (3.4) yields the zeroth moment:

$$
m_0 = \bar{C}(x,0) = \int_0^\infty C(x,t) dt = \frac{M}{Q} \tag{3.7}
$$

where $Q$ is the discharge across the area $A$. Equation (3.7) reflects the relationship between the total mass of released pollutant and the zeroth moment. The $i^{\text{th}}$ temporal central moment of the concentration distribution may be found by evaluating the $i^{\text{th}}$ derivative of $\bar{C}(x, p)$:

$$
m_i = (-1)^i \frac{d^i}{dp^i} \bar{C}(x, p) \bigg|_{p=0} \tag{3.8}
$$
\[ m_i = -\frac{d}{dp} \ln C(x, p) \bigg|_{p=0} = \frac{1}{U^2 + 4K_s A} \left( \frac{2K_s}{U^2 + 4K_s A} + x \right) \]  

(3.9)

In order to track the total mass and location of an unknown source, Eq. (3.9) can be employed in combination with concentration distributions \( C(x, t_i) \) (\( i = 1, \ldots, n \), and \( j = 1, \ldots, p \)) observed at downstream monitoring stations located at \( x_j \). While the absolute value of \( x_j \) is an unknown to determine, the distance between \( x_i \) and \( x_j \) (\( j = 2, \ldots, p \)) or the distance between two monitoring stations is known. With the observed data, the first moment could be estimated by the following equations:

\[ t_c \approx \frac{\sum_{i=0}^{\infty} Ct}{\sum_{i=0}^{\infty} C} \]  

(3.10)

in which \( t \) is the time since injection. Since both Eq. (3.10) and Eq. (3.9) are estimates of the centroid time, they should be identical, leading to the following equation for estimation of the source location:

\[ x = \frac{\sum_{i=0}^{\infty} Ct \sqrt{U^2 + 4K_s A}^{4\pi D_S}}{\sum_{i=0}^{\infty} C \sqrt{U^2 + 4K_s A}^{4\pi D_S}} - \frac{2K_s}{U^2 + 4K_s A} \]  

(3.11)

Eq. (3.11) is the moment equation derived and used for identification of unknown pollution source location \( x \) in a river using measured time-concentration data and calculated parameters. To estimate the total mass of released pollutant, Eq. (3.7) is rearranged using observed discrete time-concentration data as follows:

\[ M_{est} = Q \sum C(t_i) \Delta t_i, \quad i = 1, 2, \ldots, n \]  

(3.12)

Eq. (3.12) is the moment equation derived for estimation of the total mass \( M_{est} \) of unknown pollution source in a river. Therefore, Eqs. (3.11) and (3.12) are utilized in this paper for identification of the pollution source in terms of release location and total mass of pollutant.

### 3.3 Results

The distance \( x \) and total mass \( M_{est} \) calculated from Eqs. (3.11) and (3.12) are listed in columns (4) and (7) of Table 3.3 and compared with corresponding observations in columns (3) and (6), respectively. The comparison results are described using the percent error (= (estimated value – observed value)/observed value) and listed in columns (5) and (8) of Table 3.3.

It can be seen from Table 3.3 that percent errors \( \Delta x \) (%) in the estimated distance \( x \) are generally less than 18% with one exceptionally high percent error of -32.74%. It appears that the percent errors \( \Delta M \) (%) in the estimated mass are significantly higher than those in \( \Delta x \) with the maximum percent error \( \Delta M \) being as high as 41.38%. The results suggest that the estimation of total mass involves a higher error due possibly to reactions, such as photolysis and sorption, even conservative tracers were employed.
### Table 3.3 Estimated Distance and Total Mass

<table>
<thead>
<tr>
<th>River Reach (1)</th>
<th>$t_c$ (h)</th>
<th>$L$ (km)</th>
<th>$x$ (km)</th>
<th>$\Delta x$ (%)</th>
<th>$M_{inj}$ (kg)</th>
<th>$M_{est}$ (kg)</th>
<th>$\Delta M$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monocacy River</td>
<td>7.91</td>
<td>10.3</td>
<td>11.10</td>
<td>7.79</td>
<td>1.90</td>
<td>2.183</td>
<td>14.65</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>14.25</td>
<td>18.3</td>
<td>17.71</td>
<td>-3.23</td>
<td>1.90</td>
<td>2.137</td>
<td>12.21</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>20.36</td>
<td>26.8</td>
<td>28.48</td>
<td>6.25</td>
<td>1.90</td>
<td>1.970</td>
<td>3.45</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>26.56</td>
<td>34.3</td>
<td>31.77</td>
<td>-7.37</td>
<td>1.90</td>
<td>2.384</td>
<td>25.20</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>6.30</td>
<td>3.2</td>
<td>2.62</td>
<td>-18.07</td>
<td>2.81</td>
<td>2.284</td>
<td>-18.85</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>61.82</td>
<td>25.7</td>
<td>23.51</td>
<td>-8.51</td>
<td>2.81</td>
<td>2.369</td>
<td>-15.82</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>124.41</td>
<td>59.5</td>
<td>65.78</td>
<td>10.56</td>
<td>2.81</td>
<td>2.377</td>
<td>-15.55</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>239.55</td>
<td>117.5</td>
<td>122.07</td>
<td>3.89</td>
<td>2.81</td>
<td>3.952</td>
<td>40.40</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>14.98</td>
<td>8.2</td>
<td>8.20</td>
<td>-0.04</td>
<td>3.81</td>
<td>3.897</td>
<td>2.33</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>31.11</td>
<td>18.0</td>
<td>18.27</td>
<td>1.49</td>
<td>3.81</td>
<td>3.709</td>
<td>-2.61</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>48.67</td>
<td>41.5</td>
<td>48.58</td>
<td>17.05</td>
<td>3.81</td>
<td>3.055</td>
<td>-19.76</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>61.22</td>
<td>55.4</td>
<td>64.33</td>
<td>16.12</td>
<td>3.81</td>
<td>3.278</td>
<td>-13.93</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>75.02</td>
<td>71.0</td>
<td>69.54</td>
<td>-2.06</td>
<td>3.81</td>
<td>2.822</td>
<td>-25.88</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>85.04</td>
<td>82.1</td>
<td>74.44</td>
<td>-9.33</td>
<td>3.81</td>
<td>2.516</td>
<td>-33.94</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>97.83</td>
<td>94.0</td>
<td>63.22</td>
<td>-32.75</td>
<td>3.81</td>
<td>2.653</td>
<td>-30.34</td>
</tr>
<tr>
<td>Red River</td>
<td>3.01</td>
<td>5.7</td>
<td>6.43</td>
<td>11.85</td>
<td>23.42</td>
<td>17.373</td>
<td>-25.82</td>
</tr>
<tr>
<td>Red River</td>
<td>34.11</td>
<td>75.6</td>
<td>76.68</td>
<td>1.37</td>
<td>23.42</td>
<td>17.334</td>
<td>-25.98</td>
</tr>
<tr>
<td>Red River</td>
<td>61.25</td>
<td>132.8</td>
<td>131.41</td>
<td>-1.02</td>
<td>23.42</td>
<td>16.274</td>
<td>-30.51</td>
</tr>
<tr>
<td>Red River</td>
<td>94.20</td>
<td>193.1</td>
<td>171.11</td>
<td>-11.40</td>
<td>23.42</td>
<td>13.741</td>
<td>-41.33</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>11.11</td>
<td>54.7</td>
<td>58.76</td>
<td>7.42</td>
<td>108.86</td>
<td>105.641</td>
<td>-2.96</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>19.58</td>
<td>96.6</td>
<td>108.57</td>
<td>12.39</td>
<td>108.86</td>
<td>91.453</td>
<td>-15.99</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>24.97</td>
<td>118.0</td>
<td>98.21</td>
<td>-16.78</td>
<td>108.86</td>
<td>121.119</td>
<td>11.26</td>
</tr>
<tr>
<td>Mississippi River</td>
<td>57.09</td>
<td>294.5</td>
<td>323.00</td>
<td>9.68</td>
<td>108.86</td>
<td>63.815</td>
<td>-41.38</td>
</tr>
</tbody>
</table>

The distance calculated from Eq. (3.11) is also compared with the known distance from the injection site to the sampling site in Figure 3.2. It is clear from Figure 3.2 that the calculated distance varies around the perfect line, indicating that there are no systematic errors in the proposed equation (3.11). Figure 3.2 also shows that the error increases with increasing distance. Figure 3.3 shows a comparison between computed and actually released log-transformed total mass. Overall, the total mass computed from Eq. (3.12) varies around the perfect line, implying that the predictions from Eq. (3.12) are reasonable.
Figure 3.2 Comparison between calculated and measured distances

Figure 3.3 Comparison between calculated and measured total mass
Figure 3.4 Frequency distribution (histogram) of the percent error in calculated location and the probability density function (solid line) that best fits the histogram

To better understand the errors involved in the predictions of Eqs. (3.11) and (3.12), histograms for the percent errors are plotted in Figure 3.4 and Figure 3.5, respectively. The x axis is percent error of calculated distance in Figure 3.4 and percent error of calculated total mass in Figure 3.5. The f(x) axis is the probability of percent error for both figures. The error histograms are fitted using various probability density functions (PDFs) by means of the EasyFit software, such as Cauchy distribution, Student’s t distribution, and Hypersecan distribution. Among all these PDFs, the errors in Eq. (3.11) can be well fitted with the normal distribution with a mean $\mu = -0.200$ and a standard deviation $\sigma = 12.184$, indicating again there is no systematic error in Eq.
Figure 3.4 shows that the errors of the estimated source location are mostly within 15%. Since the source locations in our tested cases range from 3.2km to 294.5km, this percent error is reasonably low.

The errors in Eq. (3.12) are found to be best fitted with a 3-parameter Weibull distribution with the shape parameter $\alpha = 1.5515$, scale parameter $\beta = 36.593$ and location parameter $\gamma = -43.951$. Figure 3.5 reveals that the error distribution for the estimated total mass is highly skewed toward the left. It means that the total mass, which is based on observed concentrations and calculated using Eq. (3.12), is mostly underestimated. In other words, the total mass observed at a downstream site is generally less than the total mass released at the source location. This is easily understood due to mass losses along a river. The positive errors may be caused by random errors in sampling. It is clear from a comparison between Figure 3.4 and Figure 3.5 that the estimation of total mass from Eq. (3.12) involves higher errors than does Eq. (3.11) for estimation of the source location. It means that the source location determined from Eq. (3.11) is more reliable as compared with the total mass from Eq. (3.12).

3.4 Discussions

3.4.1 Effect of River Reach Length on Computation Errors

As seen from Figure 3.2 and Figure 3.4 that the errors involved in the source location computations are primarily random and thus independent of the distance between the source location and the sampling location. However, there is a systematic error in Eq. (3.12) in addition to random errors since the error distribution in Figure 3.5 is highly skewed toward the negative error side. To understand how the error from Eq. (3.12) varies with the distance, the computation errors in column (8) of Table 3.3 are plotted in Figure 3.6 against the distance between the source location and the sampling location, listed in column 4 of Table 3.3.
It is clear from Figure 3.6 that the negative errors in the computed total mass increase monotonically with increasing distance from the source or the injection location except a couple of outliers in the blue ellipse. Theoretically, the basic assumption of the model is that it is used to identify the source location and total mass of conservative contaminants, which should be stable and constant. However, in the dye tests, even if the conservative tracer was used, there was still mass loss.

The computational errors for the tracer test conducted in four reaches of the Monocacy River are unanimously positive. It means that the total mass collected at a downstream sampling station is greater than the total mass actually released. This is extremely unlikely in reality but this was the result obtained by the USGS for the tracer test conducted in the Monocacy River. The observed tracer concentrations were reduced or divided by a factor of 1.23 to achieve tracer mass conservation (Nordin and Sabol 1974). The positive errors may be caused by systematic errors in sampling and data analysis. The two points in the blue ellipse in Figure 3.6 are considered as outliers due possibly to some human errors in sampling or laboratory analysis.

### 3.4.2 Sensitivity of Computation Errors to Model Input Parameters

To further understand how our results are sensitive to parameters in Eqs. (3.11) and (3.12), additional analysis is performed looking at the impacts of dispersion coefficient, effective diffusion coefficient, and cross-sectional velocity on calculated distance. The cross-sectional velocity is included because it is not a measured, but calculated variable. There’s no sensitivity analysis of the total mass, because it is a function of discharge and concentration, both of which are measured directly.

Table 3.4 summarizes the new computation errors of distance due to the change of three parameters. It shows that the increase of dispersion coefficient increases the maximum negative error from -32.75% to -41.34%, and decreases the maximum positive error from 17.05% to 16.82%. The decrease of dispersion coefficient changes the error range from -32.75% ~ 17.05% to -25.14% ~ 17.29%. In other words, positive change in dispersion coefficient reduces the calculated source distance, while negative change of dispersion coefficient generally increases the calculated distance. Also, comparison of the 5 different rivers indicates that the dispersion coefficient has a bigger influence on small rivers than on large rivers. The change of effective diffusion coefficient has a similar impact on the distance calculation. However, the velocity has obviously an opposite yet more significant influence on the distance calculation than the other two parameters. The 20% and -20% changes lead to the maximum error of 43.58% and -69.21%, respectively. Thus, it is vitally important to acquire the mean velocity data as accurately as possible to reduce the error involved in the identification of source location.

<table>
<thead>
<tr>
<th>River</th>
<th>Δx (%)</th>
<th>Change in $K_s$</th>
<th>Change in $D_s/A$</th>
<th>Change in $U$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monocacy River</td>
<td>7.79</td>
<td>7.41</td>
<td>8.18</td>
<td>30.06</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>-3.23</td>
<td>-3.45</td>
<td>-3.02</td>
<td>16.52</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>6.25</td>
<td>6.12</td>
<td>6.39</td>
<td>27.76</td>
</tr>
<tr>
<td>Monocacy River</td>
<td>-7.37</td>
<td>-7.52</td>
<td>-7.22</td>
<td>11.43</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>-18.07</td>
<td>-23.95</td>
<td>-12.28</td>
<td>8.87</td>
</tr>
<tr>
<td>Bayou Bartholomew</td>
<td>-8.51</td>
<td>-9.82</td>
<td>-7.22</td>
<td>12.16</td>
</tr>
</tbody>
</table>
### 3.4.3 Mass Loss Correction

In order to incorporate the mass loss effect shown in Figure 3.6 on the total mass computation, it is important to find a quantitative relationship between the percent error and the distance from the mass injection point. To that end, the negative errors in Figure 3.6 are employed to generate a regression curve. We tried different types of regression model and the $R^2$ values for linear model, polynomial model and exponential model are 0.5008, 0.5064, and 0.8547, respectively. So we chose the exponential regression curve because it gave the best $R^2$ value. The regression curve shown in Figure 3.6 can be best described by the following regression relationship:

$$ CF(\%) = -0.4512 \times [1 - \exp(-0.00939 \times \text{distance})] $$  \hspace{1cm} (3.13)

where $CF$ represents a distance correction factor (%) for the total mass computed from Eq. (3.12). Regression analysis results in Eq. (3.13) in which the dependent variable is referred to as ‘Correction factor’. The USGS dye tests used conservative tracer but mass loss still occurred in some tests. The correction factor $CF$ could be viewed as the compensation for mass loss along the river. Based on Eq. (3.13), the adjusted moment equation for the total mass is obtained by incorporating the correction factor (Eq. (3.13)) into equation (3.12):

$$ M_{\text{cor}} = M_{\text{est}} - M_{\text{inj}} \times CF \approx M_{\text{est}} \times (1 - CF) $$  \hspace{1cm} (3.14)

in which $M_{\text{cor}}$ denotes the corrected total mass of the pollution source. Table 3.5 shows new results for the total mass estimated using Eq. (3.14) from the small distance to large one. The percent errors in the total mass are now reduced to ±20% with one exception. This is a significant improvement to the overall total mass estimation. However, further discussions are still needed in the future to control the positive errors.
Table 3.5 Correction to Total Mass Computation

<table>
<thead>
<tr>
<th>$L$ (km)</th>
<th>$M$ (kg)</th>
<th>$M_{est}$ (kg)</th>
<th>$\Delta M$ (%)</th>
<th>$CF$ (%)</th>
<th>$M_{cor}$ (kg)</th>
<th>Error after correction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2</td>
<td>2.81</td>
<td>2.284</td>
<td>-18.85</td>
<td>-1.34</td>
<td>2.315</td>
<td>-17.77</td>
</tr>
<tr>
<td>5.7</td>
<td>23.42</td>
<td>17.373</td>
<td>-25.82</td>
<td>-2.37</td>
<td>17.785</td>
<td>-24.06</td>
</tr>
<tr>
<td>18.0</td>
<td>3.81</td>
<td>3.709</td>
<td>-2.61</td>
<td>-7.02</td>
<td>3.969</td>
<td>4.23</td>
</tr>
<tr>
<td>25.7</td>
<td>2.81</td>
<td>2.369</td>
<td>-15.82</td>
<td>-9.67</td>
<td>2.598</td>
<td>-7.68</td>
</tr>
<tr>
<td>41.5</td>
<td>3.81</td>
<td>3.055</td>
<td>-19.76</td>
<td>-14.56</td>
<td>3.500</td>
<td>-8.08</td>
</tr>
<tr>
<td>54.7</td>
<td>108.86</td>
<td>105.641</td>
<td>-2.96</td>
<td>-18.12</td>
<td>124.786</td>
<td>14.63</td>
</tr>
<tr>
<td>55.4</td>
<td>3.81</td>
<td>3.278</td>
<td>-13.93</td>
<td>-18.30</td>
<td>3.877</td>
<td>1.82</td>
</tr>
<tr>
<td>59.5</td>
<td>2.81</td>
<td>2.377</td>
<td>-15.55</td>
<td>-19.31</td>
<td>2.836</td>
<td>0.76</td>
</tr>
<tr>
<td>71.0</td>
<td>3.81</td>
<td>2.822</td>
<td>-25.88</td>
<td>-21.95</td>
<td>3.442</td>
<td>-9.61</td>
</tr>
<tr>
<td>75.6</td>
<td>23.42</td>
<td>17.334</td>
<td>-25.98</td>
<td>-22.94</td>
<td>21.310</td>
<td>-9.00</td>
</tr>
<tr>
<td>82.1</td>
<td>3.81</td>
<td>2.516</td>
<td>-33.94</td>
<td>-24.25</td>
<td>3.126</td>
<td>-17.92</td>
</tr>
<tr>
<td>94.0</td>
<td>3.81</td>
<td>2.653</td>
<td>-30.34</td>
<td>-26.45</td>
<td>3.354</td>
<td>-11.91</td>
</tr>
<tr>
<td>193.1</td>
<td>23.42</td>
<td>13.741</td>
<td>-41.33</td>
<td>-37.76</td>
<td>18.929</td>
<td>-19.17</td>
</tr>
<tr>
<td>294.5</td>
<td>108.86</td>
<td>63.815</td>
<td>-41.38</td>
<td>-42.28</td>
<td>90.793</td>
<td>-16.60</td>
</tr>
</tbody>
</table>

Figure 3.7 shows a comparison between the total mass calculated from Eq. (3.12) and that from Eq. (3.14). As seen from Figure 3.7, the corrected total mass from Eq. (3.14) matches the injected total mass much better than does the total mass calculated from Eq. (3.12), demonstrating the efficacy of the proposed mass correction equation (3.14).
3.4.4 Limitations

The moment-based method proposed in this paper has several limitations to its applications or requires some conditions for its applications. First, the time-concentration data should be collected from two downstream sampling stations, one for the location identification and the other for the total mass determination. The source location should be identified using Eq. (3.11) if the required data are collected from only one downstream sampling station. Second, the moment equation for the total mass should be used for relatively conservative pollutants though a mass correction method is proposed. For reactive pollutants, reaction terms should be added to equations (1a) and (1b) and new moment equations should be derived, depending on specific pollutants. While reactions may change the total mass of pollutant observed at downstream stations, they will not change the location predicted using the moment equation (3.11). Therefore, Eq. (3.11) is applicable to both conservative and non-conservative pollutants. Third, the model should be used to rivers where the contaminants should be fully mixed across the channel cross-section. In the USGS dye test, the samples were collected at the sampling sites where the tracers were fully mixed. That’s how we could apply this 1D model to a 3D river.

Every model has uncertainties that come from model structure, measurement errors, and computational errors. In this Chapter, the biggest uncertainty may be the uncertainty related to the regression analysis, which is an empirical equation from scattered total mass errors. However, the most significant contribution of Chapter 3 is the source location identification. The uncertainty involved in the total mass correction factor does not affect the identification of source locations.

3.5 Conclusions

This paper presented a moment-based method for identification of source location and quantity of accidental pollution in rivers. Based on the results from this study, the following conclusions can be drawn:

1. An accidental pollution source can be identified by its location and total mass released.
2. Eq. (3.11) in combination with time-concentration data observed at a downstream station can be utilized to identify the unknown pollution source location, defined by the distance between the source release site and the downstream monitoring site. The errors involved in the location identification follow the normal distribution with a mean $\mu = -0.200$ and a standard deviation $\sigma = 12.184$.
3. Eq. (3.14) in combination with time-concentration data observed at a downstream station can be employed to determine the total mass released from an accidental pollution event. The percent error in the total mass estimation is generally within $\pm 20\%$.
4. Eq. (3.11) is generally applicable to identification of the release location of both conservative and reactive pollutants. Therefore, the release location determined from Eq. (3.11) is the most important parameter for pollution source identification.

Although the moment-based model presented here was tested on 5 rivers across the USA, it could also be applied to other rivers. To apply this model to a specific river, we need river specific data including discharge, channel cross-section geometry, pollutant concentration against time data at least from one sampling site.
CHAPTER 4  CONCENTRATION TIME-BASED METHOD FOR WATERSHED-SCALE BACTERIA SOURCE AREA IDENTIFICATION

4.1 Introduction

Identification and evaluation of critical source areas (CSAs) of water pollution is vitally important to integrated watershed management and restoration. Critical source areas (CSAs) are defined as the areas with disproportionately high pollutant losses and have been widely recognized as priority areas for nonpoint source pollution control. The identification and evaluation of CSAs at the watershed scale allows state and federal programs to implement soil and water conservation measures where they are needed most (Pionke et al. 2000; Hughes et al. 2005; White et al. 2009). The proportion of their contribution is arbitrarily set, usually, larger than or equal to 50% (Shang et al. 2012). Recently, studies have been conducted to identify CSAs at watershed scale with quantitative programs (Srinivasan et al. 2005; White et al. 2009; Shang et al. 2012). Srinivasan et al. (2005) discussed the usefulness of SMDR and SWAT models to identifying CSAs of phosphorus (P) loss and also evaluated the capability of the watershed models to provide maps for field usage. However, even with watershed modeling tools, their researches were still limited to identification of runoff generation areas, not P transport areas. Srinivasan and McDowell (2009) compared 5 approaches, including the curve number method, the phosphorus index, the drainage density index, the topographic index (TI) and the combination model that combines saturation-excess (SE) and infiltration-excess (IE) surface runoff, to map and validate the transport areas. The process-based approach gave the best overall performance and could be applied for predicting CSAs of contaminant loss in the future. Shang et al. (2012) extended the work of previous researches to a basin scale with the usage of SWAT model. Their CSAs identification framework used the pollutant load releasing into a lake as the criterion and included the river migration process. In order to draw the cumulative curve of pollutant load, the sub-basins or hydrology units were ranked from high to low pollution intensities, which were calculated by an empirical equation.

As mentioned by Srinivasan et al. (2005), CSAs for a nutrient contains two parts: source areas and transport areas. In practical cases, researchers often treat the areas that generate runoff as CSAs of that watershed, and turn the problem of CSAs identification into the identification of runoff source areas. In this chapter, the problem is to identify the source area(s) in a watershed if the detected bacterial counts at a downstream sampling station exceed the water quality standard. There are 3 steps for this study: 1) divide targeted watershed into a few subbasins according to the locations of gage stations on both the main river and tributaries; 2) find a specific hydrological parameter, such as concentration time, peak flow, and time to peak flow, as a fingerprint for each subbasin; And 3) given the fingerprints, radar rainfall data, and bacterial concentration, identify the possible rainfall (storm event) that occurred a certain time period ago and is responsible for the bacterial pollution at a downstream site.
4.2 Materials and Methods

4.2.1 Impaired Watershed and Data Collection

4.2.1.1 Tangipahoa River Watershed

The Tangipahoa River was found seriously polluted due to high level of fecal coliforms in the 1980s, which caught the attention from LDEQ (Louisiana Department of Environmental Quality), LDHH (Louisiana Department of Health and Hospitals), LDAF (Louisiana Department of Agriculture and Forestry), LPBF (Lake Pontchartrain Basin Foundation), U.S. NRCS (National Resources Conservation Service), as well as local residents and community. In 2000, the whole Tangipahoa River was on the 303 (d) list for fecal coliforms and other contaminations as well. Investigations reported that there were 250 dairy farms in the watershed. Discharge permit or no-discharge animal waste management system was required to be installed for these dairies. In the following decade, the NRCS along with the Louisiana Department of Agriculture’s Office of Soil and Water Conservation helped the dairies to design and build 158 no-discharge waste systems in the watershed. Besides, inspection of home sewage systems was mandated for the newly settled residents. In 2005, LPBF also started a water quality monitoring program that collected water samples every two weeks at both the mainstream and tributary stations. After decades of public efforts, the water quality data from 2004 to 2007 indicated that the river was no longer impaired by fecal coliforms (EPA, 2008) and was removed from the 303(d) list for fecal coliforms.

4.2.1.2 Fecal Coliform Data Collection

The fecal coliform data used here come from two sources: LDEQ and LPBF. Bi-weekly data at 10 monitoring sites were collected by LBPF (Targeted Tangipahoa Watershed Monitoring Program) from 2006 to 2010, while monthly data at 4 LDEQ stations were collected irregularly from 1978 to 2011. Table 4.1 shows the data information:

<table>
<thead>
<tr>
<th>Site</th>
<th>Data type</th>
<th>Date range</th>
</tr>
</thead>
<tbody>
<tr>
<td>TR1-TR10 by LPBF</td>
<td>Bi-weekly fecal coliform</td>
<td>1/9/2006-3/7/2010</td>
</tr>
<tr>
<td>LA 040701 #34 by LDEQ</td>
<td>Monthly fecal coliform</td>
<td>03/1978-05/1998</td>
</tr>
<tr>
<td>LA 040701 #33 by LDEQ</td>
<td>Monthly fecal coliform</td>
<td>03/1978-12/2011</td>
</tr>
</tbody>
</table>

Using BASINs (the software developed by U.S. EPA), the Tangipahoa River watershed was delineated into 30 subbasins, each of which contained no more than one LPBF monitoring site. Other important attributes of subbasin, including the reach name, river length, slope, area, etc. were also obtained from the software. Figure 4.1 shows data related to Tangipahoa River watershed, in which (a) the blue upward triangles are Tangipahoa River sampling sites, the green downward triangles are tributary sampling sites; (b) the redlines are the subbasin border lines; (c) the purple solid circles are Dairy farms; (d) the blue solid squares are WWTPs. With the help of some GIS software, the following pairs of site are found to be located at the same place (or the same town): (1) TR10 and #34 near Kentwood; (2) TR8 and #108 at Arcola; (3) TR3 and #33 west of Robert; and (4) TR1 and #1104 near Lake Pontchartrain.
Figure 4.1 Bacteria-Related Data in Tangipahoa River watershed, where (a) is LPBF’s Monitoring Sites, (b) is watershed delineation, (c) is dairy farms, and (d) is WWTPs
4.2.1.3 Hydrologic Data Collection

Hourly discharge data are available at two USGS gage stations on the Tangipahoa River, but only one is located inside Louisiana State, which is USGS 07375500 starting from October 1\textsuperscript{st} 1995, as shown in Table 4.2. Hourly precipitation data at a few rainfall stations in the watershed are collected from USGS and/or NOAA website. Additionally, NOAA’s National Climate Data Center (http://gis.ncdc.noaa.gov/map/viewer/#app=cdo) provided most of the necessary data with GIS platform. The radar data are distributed data while the recorded precipitation data are point data. The radar data could catch storm event in a relatively smaller area that would otherwise missed by the station precipitation data. The radar data used here is generated by the Next Generation Weather Radar system (NEXRAD), which comprises 159 Weather Surveillance Radar-1988 Doppler (WSR-88D) sites both in and outside of USA. The National Climatic Data Center (NCDC) made comparison of NEXRAD rainfall estimates with recorded amounts and found that the NEXRAD data were very reliable (NCDC, 1996). NEXRAD radar mosaic maps show the hourly/sub-hourly radar reflectivity around the country since 1995/01/01. Using the radar maps, we could identify short-duration storm events that might produce the surface runoff carrying high concentration of bacteria from dairy farms and/or WWTPs to the river.

Table 4.2 Discharge and precipitation data for Tangipahoa River

<table>
<thead>
<tr>
<th>Site</th>
<th>Data type</th>
<th>Date range</th>
</tr>
</thead>
<tbody>
<tr>
<td>USGS 07375500 at Robert</td>
<td>Hourly discharge</td>
<td>10/1/1995-1/8/2013</td>
</tr>
<tr>
<td>LA160205 at Amite</td>
<td>Hourly precipitation</td>
<td>12/31/1947-12/31/2006</td>
</tr>
<tr>
<td>LA163165 at Folsom 6 S</td>
<td>Hourly precipitation</td>
<td>08/31/1993-12/31/2006</td>
</tr>
<tr>
<td>LA163331 at Franklinton 5 SW</td>
<td>Hourly precipitation</td>
<td>03/1/1984-01/31/2004</td>
</tr>
<tr>
<td>LA164859 at Kentwood</td>
<td>Hourly precipitation</td>
<td>12/31/1947-12/31/2006</td>
</tr>
<tr>
<td>LA166466 at Mount Hermon 2 W</td>
<td>Hourly precipitation</td>
<td>02/28/1995-12/31/2006</td>
</tr>
<tr>
<td>LA167425 at Ponchatoula 4 SE</td>
<td>Hourly precipitation</td>
<td>04/30/1988-12/31/2006</td>
</tr>
<tr>
<td>LA168945 at Tickfaw 3 ENE</td>
<td>Hourly precipitation</td>
<td>03/31/2000-12/31/2006</td>
</tr>
<tr>
<td>MS225614 at Mccomb Airport</td>
<td>Hourly precipitation</td>
<td>09/30/1948-12/31/2006</td>
</tr>
</tbody>
</table>

4.2.2 Time of Concentration ($T_c$)

Time of concentration (or concentration time, $T_c$) is the time for the flow to travel from the most hydraulically remote point in a watershed to an outlet. It is one of the timing parameters of a watershed for hydrologic analysis and design.

The research on $T_c$ has been abundant and varied (Fang et al, 2005). Generally speaking, there are three components of flow that contribute to $T_c$: overland flow (sheet flow); concentrated flow; and channel flow. The overland flow occurs at the upper portion of a basin and exists over short distances like a few hundred feet (McCuen 1989, Fang et al, 2005). The concentrated flow and channel flow refer to flow in gullies and channels. There are a number of methods for estimating $T_c$, and they could be grouped into two categories: empirical or regression-based methods such as rational method, Kirpich formula and Kerby formula, and hydraulic-based methods (Kinematic wave formula and the NRCS equation) (Roussel et al, 2005; Fang et al, 2008). McCuen (1989) summarized the parameters and variables needed to calculate the travel time and suggested four types of input: flow resistance, watershed size, slope and water input. Although many methods use three or four of these inputs, they individually use a specific name for an input. For example, the Rational Method uses Rational Coefficient for flow resistance,
while the Kerby formula uses Kerby retardance roughness. A method may be applicable for a particular size of watershed, based on how a specific parameter is derived. Fang et al (2005) and Roussel et al (2005) discussed these methods in details, tested them on scores of watersheds, and recommended the usage of Kirpich-inclusive approach.

The formulas are presented here and the coefficients are listed in Table 4.3:

1) FAA equation (Rational Method):

\[
T_c = 1.8 \left(1 - C\right) \frac{L^{0.5}}{S^{1/3}}
\]

2) Kirpich equation:

\[
T_c = 0.0078 \left(\frac{L}{S^{0.5}}\right)^{0.77}
\]

3) Kerby equation:

\[
T_c = 0.83 \left(\frac{Lr}{S^{1/2}}\right)^{0.467} = \left(\frac{0.67Lr}{S^{1/2}}\right)^{0.467}
\]

4) Bransby Williams equation:

\[
T_c = 21.3 \frac{L}{5280A^{0.1}S^{0.2}}
\]

where

- \(T_c\) = Time of concentration, minutes.
- \(L\) = Longest watercourse length in the watershed, ft.
- \(S\) = Average slope of the watercourse, ft/ft or m/m.
- \(C\) = Rational Method runoff coefficient. See Table 4.3 below.
- \(k\) = Kirpich adjustment factor. See Table 4.3 below.
- \(r\) = Kerby retardance roughness coefficient. See Table 4.3 below.
- \(A\) = Watershed area, sq. mi.

<table>
<thead>
<tr>
<th>Ground Cover</th>
<th>Coefficient Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rational Runoff Coefficient for FAA Method, (C) (Corbitt, 1999; Singh, 1992)</td>
<td></td>
</tr>
<tr>
<td>Lawns</td>
<td>0.05 - 0.35</td>
</tr>
<tr>
<td>Forest</td>
<td>0.05 - 0.25</td>
</tr>
<tr>
<td>Cultivated land</td>
<td>0.08-0.41</td>
</tr>
<tr>
<td>Meadow</td>
<td>0.1 - 0.5</td>
</tr>
<tr>
<td>Parks, cemeteries</td>
<td>0.1 - 0.25</td>
</tr>
<tr>
<td>Unimproved areas</td>
<td>0.1 - 0.3</td>
</tr>
<tr>
<td>Pasture</td>
<td>0.12 - 0.62</td>
</tr>
<tr>
<td>Residential areas</td>
<td>0.3 - 0.75</td>
</tr>
<tr>
<td>Business areas</td>
<td>0.5 - 0.95</td>
</tr>
<tr>
<td>Industrial areas</td>
<td>0.5 - 0.9</td>
</tr>
<tr>
<td>Asphalt streets</td>
<td>0.7 - 0.95</td>
</tr>
<tr>
<td>Brick streets</td>
<td>0.7 - 0.85</td>
</tr>
<tr>
<td>Roofs</td>
<td>0.75 - 0.95</td>
</tr>
<tr>
<td>Concrete streets</td>
<td>0.7 - 0.95</td>
</tr>
</tbody>
</table>
(Table 4.3 Continued)

<table>
<thead>
<tr>
<th>Ground Cover</th>
<th>Coefficient Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kirpich Adjustment Factor, $k$ (Chow et al., 1988; Chin, 2000)</td>
<td></td>
</tr>
<tr>
<td>General overland flow and natural grass channels</td>
<td>2.0</td>
</tr>
<tr>
<td>Overland flow on bare soil or roadside ditches</td>
<td>1.0</td>
</tr>
<tr>
<td>Overland flow on concrete or asphalt surfaces</td>
<td>0.4</td>
</tr>
<tr>
<td>Flow in concrete channels</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kerby Retardance Coefficient, $r$ (Chin, 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conifer timberland, dense grass</td>
</tr>
<tr>
<td>Deciduous timberland</td>
</tr>
<tr>
<td>Average grass</td>
</tr>
<tr>
<td>Poor grass, bare sod</td>
</tr>
<tr>
<td>Smooth bare packed soil, free of stones</td>
</tr>
<tr>
<td>Smooth pavements</td>
</tr>
</tbody>
</table>

These formulas will be used to calculate $T_c$, and then we will compare the result with our existing data to decide which one gives the best estimation. The Kinematic equation was not considered here because we don’t have information about the cross-section of the Tangipahoa River, which is required in Manning’s equation. Figure 4.2 is a map of isolines that are drawn to show points with the same $T_c$ in a watershed.

![Figure 4.2](image)

Figure 4.2 $T_c$ of a hypothetic watershed and the cumulative time-area curve (courtesy of NOAA, http://www.nohrsc.noaa.gov/technology/gis/uhg_manual.html)

4.2.3 Concentration Time-Based Identification Approach

The core idea here is to track fecal coliforms back to a storm event that washes away the bacteria, brings them to the river and carries them to the downstream sampling stations. The storm could have occurred a few hours or a couple of days ago, depending on the range and intensity. This time interval is named as Target Time ($T_T$) here. In hydrologic analysis, there are
three common timing parameters: $T_c$, lag time ($T_L$) and time to peak ($T_P$). However, we would like to use $T_c$ to approximate $T_T$ and check if a related rainfall could be identified. The following are the steps:

(1) *List all pollution events of high fecal coliform level or increased fecal coliform level between two sampling sites;*

The graph showing the variation in fecal coliform level with discharge is plotted annually to determine if there exists a positive correlation between them. There are two possible situations: 1) an increase in fecal coliform level from an upstream site to a downstream site may indicate a local pollution storm event in the drainage areas between these two sites; 2) significantly elevated fecal coliform levels could also indicate pollution storm events somewhere upstream.

(2) *Identify possible pollution storm events;*

Step (2) will go over the NEXRAD map and site-based precipitation data for up to a week before the detection of elevated fecal coliform levels in order to include any possible storm events. Similarly, any storm that lasts longer than one week or covers the whole Tangipahoa River watershed will be ruled out. In other words, only the local storm events that last from a few hours to a couple of days will remain for further analysis after the first two steps.

(3) *Identify subbasins of possible fecal coliform source;*

In step (3), the NEXRAD map will be compared with the Tangipahoa River watershed map in BASINS to highlight subbasins that are affected by the local storm events. Also, the $T_c$ of each subbasin should be consistent with $T_T$. These subbasins are potential bacterial source areas.

(4) *Identify the smaller critical area containing the pollution source (a dairy farm or WWTP).*

Step (4) further identifies the bacteria source by first dividing each subbasin into smaller hydrologic units—the possible CSAs in the Tangipahoa River watershed. The ideal division would be that each unit contains only one dairy farm or WWTP. For each unit, the $T_c$ will be estimated from the location of the dairy farm or WWTP to the location of downstream water quality station. Then, the error between $T_c$ of each unit and $T_T$ of that storm event should be compared to find out the CSA. The CSA is the one that satisfies:

$$\min \{T_{c,i} - T_T \} \quad i = 1,2,...,n$$

where $n$ is the number of hydrological units.

It is also necessary to note that the first flush from a storm usually carries away the most amounts of pollutants on the surface. Therefore, we would calculate the $T_T$ from the beginning of the storm to the time of peak discharge at the downstream gage station.
4.3 Results

4.3.1 Discharge and Fecal Coliform Level

Figure 4.3 Discharge and LDEQ’s Fecal Coliform data in 1996

Figure 4.4 Discharge and LDEQ’s Fecal Coliform data in 1998
Figure 4.5 Discharge and LPBF’s Fecal Coliform data at Site TR6

Figure 4.6 Discharge and LPBF’s Fecal Coliform data at Site TR5
Figure 4.7 Discharge and LPBF’s Fecal Coliform data at Site TR3

Figure 4.8 Discharge and LPBF’s Fecal Coliform data at Site TR2
Figure 4.3-4.8 shows how fecal coliform levels change with the discharge, using both LDEQ and LPBF data. Since the data range differently for the 4 sampling sites, the LDEQ figures may contain one site, two sites or three sites. Here we only showed the data collected in 1996 and 1998 because in these two years we have identified some events and sources. These figures indicate that the fecal coliform levels can vary from hundreds MPN/100ml to hundreds of thousands MPN/100ml usually due to the result of intense, wide-range, long-term rainfall over the whole watershed. But a high fecal coliform level does not necessarily correspond to a high flow, and vice versa. A further confirmation needs to be made to decide which rainfall event is responsible for the elevated fecal coliform levels.

4.3.2 \( T_c \) for Tangipahoa River

The land use data is downloaded in BASINs as ‘National Land Cover Data 2001’ by default. Based on this, the proper coefficient was chosen for each formula by incorporating the land use data with the subbasins’ geographic location, and the concentration time was calculated. The result is listed in Table 4.4.

Table 4.4 Concentration Time for Subbasins of Tangipahoa River (hour)

<table>
<thead>
<tr>
<th>River Name</th>
<th>Subbasin #</th>
<th>Length (ft)</th>
<th>Slope</th>
<th>FAA</th>
<th>Kirpich</th>
<th>Kerby</th>
<th>B-W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tangipahoa River</td>
<td>8</td>
<td>200032</td>
<td>0.0013</td>
<td>24.07</td>
<td>20.50</td>
<td>17.66</td>
<td>16.98</td>
</tr>
<tr>
<td>Bala Chitto Creek</td>
<td>6</td>
<td>101246</td>
<td>0.0017</td>
<td>15.36</td>
<td>10.70</td>
<td>11.91</td>
<td>8.62</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>5</td>
<td>50314</td>
<td>0.0009</td>
<td>13.45</td>
<td>8.03</td>
<td>10.00</td>
<td>5.23</td>
</tr>
<tr>
<td>Terrys Creek</td>
<td>7</td>
<td>99381</td>
<td>0.0019</td>
<td>14.85</td>
<td>10.26</td>
<td>11.61</td>
<td>8.35</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>28</td>
<td>27853</td>
<td>0.0007</td>
<td>10.90</td>
<td>5.62</td>
<td>8.06</td>
<td>3.23</td>
</tr>
<tr>
<td>Beaver Creek</td>
<td>29</td>
<td>78000</td>
<td>0.0023</td>
<td>12.28</td>
<td>7.86</td>
<td>9.88</td>
<td>6.45</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>9</td>
<td>44549</td>
<td>0.0011</td>
<td>11.88</td>
<td>6.79</td>
<td>9.04</td>
<td>4.51</td>
</tr>
<tr>
<td>Big Creek</td>
<td>27</td>
<td>72673</td>
<td>0.0023</td>
<td>9.30</td>
<td>7.52</td>
<td>6.95</td>
<td>6.08</td>
</tr>
<tr>
<td>East Fork Big Creek</td>
<td>25</td>
<td>42817</td>
<td>0.0019</td>
<td>7.54</td>
<td>5.33</td>
<td>5.64</td>
<td>3.90</td>
</tr>
<tr>
<td>Big Creek</td>
<td>26</td>
<td>27433</td>
<td>0.0013</td>
<td>6.84</td>
<td>4.37</td>
<td>5.00</td>
<td>2.82</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>10</td>
<td>31052</td>
<td>0.0010</td>
<td>6.95</td>
<td>5.45</td>
<td>5.00</td>
<td>3.36</td>
</tr>
<tr>
<td>4 Tangipahoa River</td>
<td>24</td>
<td>39176</td>
<td>0.0007</td>
<td>13.16</td>
<td>7.46</td>
<td>9.56</td>
<td>4.44</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>11</td>
<td>31091</td>
<td>0.0002</td>
<td>14.36</td>
<td>9.74</td>
<td>9.03</td>
<td>4.55</td>
</tr>
<tr>
<td>Sweetwater Creek</td>
<td>12</td>
<td>39911</td>
<td>0.0024</td>
<td>7.25</td>
<td>4.64</td>
<td>5.76</td>
<td>3.51</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>23</td>
<td>22900</td>
<td>0.0007</td>
<td>7.65</td>
<td>4.81</td>
<td>5.30</td>
<td>2.70</td>
</tr>
<tr>
<td>Chappepeela Creek</td>
<td>14</td>
<td>93776</td>
<td>0.0017</td>
<td>13.14</td>
<td>10.09</td>
<td>10.05</td>
<td>8.04</td>
</tr>
<tr>
<td>Little Chappepeela Creek</td>
<td>15</td>
<td>77124</td>
<td>0.0020</td>
<td>11.48</td>
<td>8.31</td>
<td>8.93</td>
<td>6.60</td>
</tr>
<tr>
<td>Chappepeela Creek</td>
<td>13</td>
<td>49783</td>
<td>0.0007</td>
<td>13.26</td>
<td>9.02</td>
<td>9.39</td>
<td>5.53</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>31</td>
<td>39861</td>
<td>0.0003</td>
<td>14.95</td>
<td>9.93</td>
<td>9.95</td>
<td>5.20</td>
</tr>
<tr>
<td>Skulls Creek</td>
<td>22</td>
<td>56390</td>
<td>0.0007</td>
<td>13.84</td>
<td>9.72</td>
<td>9.82</td>
<td>6.12</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>30</td>
<td>4318</td>
<td>0.0015</td>
<td>3.32</td>
<td>1.00</td>
<td>2.82</td>
<td>0.52</td>
</tr>
<tr>
<td>Sims Creek</td>
<td>1</td>
<td>64759</td>
<td>0.0018</td>
<td>12.12</td>
<td>7.47</td>
<td>9.57</td>
<td>5.72</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>2</td>
<td>14307</td>
<td>0.0005</td>
<td>9.02</td>
<td>3.97</td>
<td>6.53</td>
<td>1.94</td>
</tr>
<tr>
<td>Sims Creek</td>
<td>4</td>
<td>24873</td>
<td>0.0011</td>
<td>9.01</td>
<td>4.41</td>
<td>6.96</td>
<td>2.70</td>
</tr>
</tbody>
</table>
(Table 4.4 Continued)

<table>
<thead>
<tr>
<th>River Name</th>
<th>Subbasin #</th>
<th>Length (ft)</th>
<th>Slope</th>
<th>FAA</th>
<th>Kirpich</th>
<th>Kerby</th>
<th>B-W</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tangipahoa River</td>
<td>3</td>
<td>24875</td>
<td>0.0001</td>
<td>20.47</td>
<td>11.38</td>
<td>12.36</td>
<td>4.41</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>21</td>
<td>33947</td>
<td>0.0001</td>
<td>23.35</td>
<td>14.07</td>
<td>5.32</td>
<td>5.75</td>
</tr>
<tr>
<td>Bedico Creek</td>
<td>19</td>
<td>73527</td>
<td>0.0008</td>
<td>16.97</td>
<td>11.29</td>
<td>12.30</td>
<td>7.55</td>
</tr>
<tr>
<td>Bedico Creek</td>
<td>17</td>
<td>21218</td>
<td>0.0001</td>
<td>18.91</td>
<td>10.07</td>
<td>11.48</td>
<td>3.82</td>
</tr>
<tr>
<td>Tangipahoa River</td>
<td>16</td>
<td>47648</td>
<td>0.0001</td>
<td>30.98</td>
<td>20.82</td>
<td>17.83</td>
<td>8.35</td>
</tr>
</tbody>
</table>

From Table 4.4, the FAA equation gives the largest $T_c$, while the Brandy-Williams equation has the smallest value. To find out which equation is better for this study, we compare $T_c$ with the time to peak concentration that was recorded during the USGS dye test. We have checked the physical locations of each sampling site in the USGS dye test, and compared them with those of the LPBF's water quality sampling site, and determined the start point and end point in Table 4.5. The $T_c$ value used in Table 4.5 is the sum of all subbasins (from Table 4.4) that are included between the start point and end point.

Table 4.5 Comparison between $T_c$ and Time to $C_{max}$

<table>
<thead>
<tr>
<th>Start point</th>
<th>End point</th>
<th>$L$ (km)</th>
<th>$T_p$ (hour)</th>
<th>$C_{max}$ (hour)</th>
<th>FAA (hour)</th>
<th>Kirpich (hour)</th>
<th>Kerby (hour)</th>
<th>B-W (hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>State line</td>
<td>TR10</td>
<td>8.2</td>
<td>13.75</td>
<td>13.45</td>
<td>8.03</td>
<td>10.00</td>
<td>5.23</td>
<td></td>
</tr>
<tr>
<td>TR10</td>
<td>TR9</td>
<td>9.8</td>
<td>16.25</td>
<td>10.90</td>
<td>5.62</td>
<td>8.06</td>
<td>3.23</td>
<td></td>
</tr>
<tr>
<td>TR9</td>
<td>TR7</td>
<td>23.5</td>
<td>17</td>
<td>18.83</td>
<td>12.24</td>
<td>14.04</td>
<td>7.87</td>
<td></td>
</tr>
<tr>
<td>TR7</td>
<td>TR6</td>
<td>13.9</td>
<td>12</td>
<td>13.16</td>
<td>7.46</td>
<td>9.56</td>
<td>4.44</td>
<td></td>
</tr>
<tr>
<td>TR6</td>
<td>TR4</td>
<td>15.6</td>
<td>14.5</td>
<td>22.01</td>
<td>14.55</td>
<td>14.33</td>
<td>7.25</td>
<td></td>
</tr>
<tr>
<td>TR4</td>
<td>TR3</td>
<td>11.1</td>
<td>9.5</td>
<td>18.27</td>
<td>10.93</td>
<td>12.77</td>
<td>5.72</td>
<td></td>
</tr>
<tr>
<td>TR3</td>
<td>TR1</td>
<td>11.9</td>
<td>13.5</td>
<td>52.84</td>
<td>29.42</td>
<td>24.21</td>
<td>12.1</td>
<td></td>
</tr>
</tbody>
</table>

It shows that the FAA method could give the closest $T_c$ values for most river reaches. In the second row, the FAAs’ $T_c$ is only 10.9 hours, much less than 16.25 hours. One main reason is that the stream length of the subbasin is shorter than the real distance from TR10 (Kentwood) to TR9 (Tangipahoa). Generally speaking, for the upper part of Tangipahoa River, the FAA method gives good estimation of $T_c$, but for the lower part of Tangipahoa River, the $T_c$ values are too high. One reason for the overestimation is the particular small slope in the lower part of Tangipahoa River. From Table 4.4, it’s not difficult to find that if the slope is smaller than 0.0001, the calculation of $T_c$ might be too large. For the lower part, the Kirpich and Kerby equations overestimate $T_c$ too. So a recommendation is that the FAA method may not be a good choice for a well-developed channel in the flat area. When it’s possible, the Kinematic equation (Manning’s equation) is suggested to calculate the travel time in the channel. In this thesis, the FAA method will be the primary method for the calculation of $T_c$. 

43
4.3.3 Rainfall-Runoff Driven Bacterial Pollution Events

The LDEQ criteria for fecal coliforms are 2000 MPN/100ml for primary contact and 400 MPN/100ml for secondary contact. We first focused on the high concentration events, and found that all extremely high concentration events were associated with intensive watershed wide rainfall events. In such events, almost all dairy farms and WWTPs are responsible for the high bacterial levels, and it is impossible to know which subbasin is the primary source area. Then we skipped the highest events and connected several medium-level events with precipitation. We summarized a few cases in Table 4.6.

Table 4.6 Fecal Coliform, Rainfall and Possible Sources

<table>
<thead>
<tr>
<th>Date</th>
<th>Fecal Coliform (MPN/100ml)</th>
<th>Rainfall</th>
<th>Source Areas (Subbasins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/11/1996</td>
<td>500 2400 1300 _ 1104</td>
<td>03/11 3:30-4:00</td>
<td>28, 29, 9</td>
</tr>
<tr>
<td>11/18/1996</td>
<td>80  - 500 _</td>
<td>11/18 7:50-8:20</td>
<td>28, 29</td>
</tr>
<tr>
<td>4/13/1998</td>
<td>80  - 500 _</td>
<td>no rainfall</td>
<td>unknown</td>
</tr>
<tr>
<td>TR6 TR5 TR3 TR2</td>
<td>30 80 50 230</td>
<td>2/3 21:00-23:00</td>
<td>unknown</td>
</tr>
<tr>
<td>7/24/2006</td>
<td>130 2300 130 _</td>
<td>no rainfall</td>
<td>11, 12</td>
</tr>
</tbody>
</table>

4.3.3.1 Case 1: 03/11/1996

The fecal coliform level increased from 500 MPN/100ml at site #34 to 2400 MPN/100ml at site #108. There was no heavy rainfall in recent 3 days before the water sample was taken. The only possible rainfall occurred between 03/11 3:30 to 4:00 in this area and upstream watershed. After checking on the NERAD map (Figure 4.9), this short-term storm moved from northeast towards southwest, and passed three subbasins, namely subbasin #28, #29, and #9 (Figure 4.10). The target time is about 8.5 hours, while the concentration time for these three subbasins are 11-12 hours. So the identified subbasins are very likely to be the source areas. And potential fecal coliform sources like dairy farms and WWTPs could also be identified.

Figure 4.9 Radar Rainfall on 1996/3/11 04:00 CST
Figure 4.10 Potential Source Areas, Dairy Farms, and WWTPs for Case 1

Table 4.7 Comparison of $T_T$ and $T_c$ of Source Unit for 03/11/1996 event

<table>
<thead>
<tr>
<th>Source Unit</th>
<th>Subbasin</th>
<th>$T_c$</th>
<th>% error</th>
</tr>
</thead>
<tbody>
<tr>
<td>WWTP_20, Dairy_87</td>
<td>9</td>
<td>52.74</td>
<td>520.50</td>
</tr>
<tr>
<td>WWTP_53, Dairy_85</td>
<td>9</td>
<td>48.57</td>
<td>471.41</td>
</tr>
<tr>
<td>Dairy_13</td>
<td>9</td>
<td>36.50</td>
<td>329.45</td>
</tr>
<tr>
<td>WWTP_5, Dairy_18</td>
<td>9</td>
<td>27.40</td>
<td>222.37</td>
</tr>
<tr>
<td>WWTP_9</td>
<td>9</td>
<td>4.67</td>
<td>-45.01</td>
</tr>
<tr>
<td>WWTP_205</td>
<td>9</td>
<td>8.95</td>
<td>5.35</td>
</tr>
<tr>
<td>WWTP_0</td>
<td>9</td>
<td>10.10</td>
<td>18.87</td>
</tr>
<tr>
<td>Dairy_45</td>
<td>9</td>
<td>4.87</td>
<td>-42.65</td>
</tr>
<tr>
<td>WWTP_144, WWTP_156</td>
<td>9</td>
<td>4.14</td>
<td>-51.24</td>
</tr>
<tr>
<td>Dairy_8</td>
<td>29</td>
<td>40.46</td>
<td>376.03</td>
</tr>
<tr>
<td>Dairy_14</td>
<td>29</td>
<td>31.71</td>
<td>273.10</td>
</tr>
<tr>
<td>Dairy_23</td>
<td>29</td>
<td>26.40</td>
<td>210.61</td>
</tr>
<tr>
<td>Dairy_19</td>
<td>29</td>
<td>18.92</td>
<td>122.59</td>
</tr>
<tr>
<td>WWTP_59</td>
<td>28</td>
<td>30.35</td>
<td>257.06</td>
</tr>
<tr>
<td>Dairy_11</td>
<td>28</td>
<td>35.12</td>
<td>313.20</td>
</tr>
</tbody>
</table>

Table 4.7 showed that the WWTP #205 and WWTP #0 are the most probable source of fecal coliforms for the event of 03/11/1996. For other possible sources, their $T_c$ values are too large to be influential.

4.3.3.2 Case 2: 11/18/1996

The fecal coliform increased from 80 MPN/100ml at site #34 to 300 MPN/100ml at site #108. Even if 300 MPN/100ml is within the LDEQ criteria for secondary contact, the percent increment is big. By checking the local rainfall data and radar map (Figure 4.11), we found that the only possible rainfall occurred at 11/18 8:50-9:20. The target time is about 3 hours. Then, the subbasins identified for this increase are #28 and #29 (Figure 4.10).
Table 4.8 Comparison of $T_T$ and $T_c$ of Source Unit for 11/18/1996 event

<table>
<thead>
<tr>
<th>Source Unit</th>
<th>Subbasin</th>
<th>$T_c$ (h)</th>
<th>% error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dairy_8</td>
<td>29</td>
<td>28.58</td>
<td>852.74</td>
</tr>
<tr>
<td>Dairy_14</td>
<td>29</td>
<td>19.83</td>
<td>561.13</td>
</tr>
<tr>
<td>Dairy_23</td>
<td>29</td>
<td>14.52</td>
<td>384.06</td>
</tr>
<tr>
<td>Dairy_19</td>
<td>29</td>
<td>7.04</td>
<td>134.65</td>
</tr>
<tr>
<td>WWTP_59</td>
<td>28</td>
<td>18.47</td>
<td>515.67</td>
</tr>
<tr>
<td>Dairy_11</td>
<td>28</td>
<td>23.24</td>
<td>674.72</td>
</tr>
</tbody>
</table>

Table 4.8 did not show a very probable fecal coliform source, the closest one is Dairy farm # 19, but still the $T_c$ is much larger than $T_T$.

4.3.3.3 Case 3: No Rainfall
On 4/13/1998 and 7/24/2006, the fecal coliform level increased significantly, leading the downstream site water quality not satisfying the criteria. However, there had been no rainfall for a few days. For the case of 4/13/1998, possible sources are WWTPs between Kentwood and Robert.

For the case of 7/24/2006, TR6 and TR5 are only a few miles distance, covering subasin #11 and #12 (Figure 4.12). There are a couple of WWTPs inside these two subbasins which could be the potential fecal coliform sources.

4.3.3.4 Case 4: 2/6/2006
The fecal coliform count increased from 50 MPN/100ml at TR3 to 230 MPN/100ml at TR2. Although this level is allowed by the standard, the increase is large enough to notice. Radar map and local rainfall data showed that the storm at 2/3 21:00-23:00 might be responsible for this increase. This storm moved from northwest to southeast in two hours and covered almost 2/3 of the subbasins in Tangipahoa River watershed. There is no easy way to tell which subbasin is the main source of bacteria. Meanwhile, a close look at the upstream sites showed that the water
quality was already very good. A hydrograph during this time also showed that it was in the recession phase. There is enough reason to believe that this increase is not caused by the upstream dairies. Possible sources could be the WWTPs between TR3 and TR2 that include subbasin #1, #2, #3 and #4 (Figure 4.13). Besides, it is critical to notice that TR3 overlaps #33 on the west of Robert, where there are many WWTPs distributed in the area. The WWTPs are more likely to be the source of fecal coliforms detected at TR2.

However, the target time is about 62 hours, while the $T_c$ from the most remote possible source, WWTP #75, is less than 48 hours. Thus, it is not likely for the fecal coliforms to be released synchronously as the rainfall.
4.4 Discussion

Four scenarios in the previous section showed how to identify bacteria source areas in a watershed using concentration time-based method. However, there are a few problems to address.

4.4.1 Multiple-Subbasin Case

This method could be useful when there’s an increase in bacterial level between two neighboring sampling sites, which minimize the searching area. But in reality, situations are always more complex when many subbasins are involved and it becomes very hard to isolate the primary source from secondary and tertiary sources. One possible solution to this problem is to add several sampling sites at the tributaries as the LPBF did. In this way, a low bacterial level at tributary site could exclude that subbasin from the potential source areas. Nevertheless, the reality could also be that many subbasins instead of one are the true source areas for fecal coliform contamination. Particular effort should be put into the investigation before excluding any possible source area.

4.4.2 Uncertainty in Concentration Time

The concentration time used in this chapter is based on the FAA method. This is an empirical equation developed to calculate the $T_c$ that lumps all effects including overland flow and channel flow. We cannot use the rational method to distinguish from overland flow and channel flow. However, the FAA method had been used for both rural and urban areas, particularly for small watersheds (Fang et al., 2005). Since each $T_c$ method might give a very different result, the one we picked may be good, but is not perfect. Thus, it is imperative to validate the chosen method. In our case, comparisons of our concentration time with the time to peak in dye test are very close. Considering that the two dye tests were conducted in relatively low flow period, it is reasonable that our values are usually small. However, one should note that the flow travel time in a real case can be very different from the concentration time, especially during a flood event when flow rate is extremely high and travel time becomes very short. That’s also a reason we excluded almost all high flow cases.

Another issue worthy to be discussed is the distribution model of concentration time. In this chapter, the concentration time is assigned to each subbasin, which is delineated from Tangipahoa River watershed. There are a few publications that use distributive hydrological model to simulate the concentration time in a watershed. In these studies, concentration time is assigned to each grid in a watershed instead of each subbasin. In such a case, it should be possible to identify the concentration time for every single dairy farm and WWTP. However, there will always be overlaps of concentration time from different grids. So the first step is to exclude all impossible areas.

4.5 Conclusion

Monitoring fecal coliform levels and finding potential unknown sources is very important to public health and watershed management. This chapter presented a concentration time-based method for identifying bacterial source areas in a watershed. The fecal coliform data were plotted with hydrographs of Tangipahoa River to eliminate high flow events. The Tangipahoa River watershed was delineated into 30 subbasins and each subbasin has a concentration time. The time and site location were marked when the fecal coliform level rose largely from upstream
to downstream sites. Then, the NEXRAD radar map were used to see if there’s a rainfall involved. The lag between the rainfall and fecal coliform detecting time must be as close as the concentration time. Four cases were presented to explain how to identify the source area using this method. The first case successfully identified two probable sources, WWTP #205 and WWTP #0, while the other cases involved some difficulty to give a very certain result. This means that we need more detailed data to find the source.

However, this method is a preliminary attempt to identify source area with concentration time. The current method could only identify the smallest hydrologic units –catchment. For each catchment, we have one value of $T_c$. All possible WWTPs and Dairy farms located in the same catchment have the same $T_c$, and thus equal possibility of being the source. In the case of one WWTP/Dairy in one catchment, it is reasonable to say this WWTP/Dairy is the source. However, for the case of multiple WWTPs/Dairies in one catchment, so far we could not identify which one is the source. Many works still need to be done and validated before we use it to address a practical issue.
CHAPTER 5 GRAND CONCLUSIONS

5.1 Conclusions

This thesis presents a watershed-based modeling approach to identification of pollution sources with emphasis on pollution source location in a river. Though a number of methods have been developed for bacterial source identification, they are limited to a specific usage. Biological methods are widely used for microbial host tracking, while mathematical models are typically applied for source location tracking and release history reconstruction. Some of these methods have been used independently, while others are usually combined when applied to real water quality problems. In this era of rapid information explosion and technology breakthrough, it is so normal to use a comprehensive watershed-scale source identification method that is a combination of biological methods, mathematical models, and sensor technologies (including remote and in-situ sensing).

In the presentation of the moment-based method, it is not difficult to find out that this method can be utilized to identify the unknown pollution source location and total mass of accidental pollution in rivers. Generally speaking, the error in the source location estimation is smaller than that of the total mass estimation. This method has showed a potential usage for identification of the release location of both conservative and reactive pollutants. However, when applied to a reactive pollutant, it is suggested that the reaction terms be added to the transient storage model.

The concentration time-based approach showed that it is possible to identify some fecal coliform sources in a watershed, given that radar rainfall maps, water quality monitoring data, and geographical data of the targeted watershed are available. Four cases presented here explained how to identify the source area using the concept of time of concentration. One case successfully identified two probable sources, while the other cases still need more information to give a more certain conclusion.

5.2 Future Perspectives

The methods and particularly the concentration time-based method discussed in this thesis are preliminary attempts to identify pollution source areas using either mathematical modeling or geographical approach. Many investigations still need to be done before applying them to a practical problem.

In terms of the moment-based method, though it is probable to use it for conservative pollutants, more efforts could be put into extending this method to reactive pollutants. In that case, the reaction term(s) will be added to the right hand side of the Variable Residence Time (VART) model. Then the Laplace transform should be applied to the model and the theoretical solutions will be derived. However, the more challenging issue would be the acquisition of time-concentration data of a specific type of bacteria, no matter fecal coliform or E. coli. The most complete data we could get so far is the bi-weekly observation of fecal coliforms on Tangipahoa River from 2006 to 2010 by the LPBF. Two weeks still seem too long for the time-concentration distribution and reconstruction of release history.

In terms of the concentration time-based method, it is obvious to notice that this method depends on largely how many catchment-scale storm events we can isolate from the radar rainfall data.
pool. This process is time consuming but necessary. Many uncertainty factors are inhabited in this method, including the estimation of time of concentration, the target time, and the watershed delineation. The rational method itself may contain some structural uncertainty, while the rational coefficient is a very sensitive factor to the time calculated. It might help if different methods are used and compared. Another way to reduce this error is to conduct several field measurements to establish a good range of rational coefficient values.

Moreover, there’re other methods for identification of the source area in a watershed. One possible method is to compare the hydrographs of hydrologic units with the observed hydrograph for a specific rainfall event. However, this method requires a good model and calibration of the watershed model parameters.
REFERENCES


APPENDIX

Applying Laplace transform to Eq. (3.2a) with respect to time and using Eq. (3.4) yields

\[
p\overline{C} - C(x,0) + U \frac{\partial \overline{C}}{\partial x} = K_s \frac{\partial^2 \overline{C}}{\partial x^2} + L\left\{ \frac{A_{adv} + 4\pi D_s}{A} \left( t - T_{\text{min}} \right) \frac{1}{t} \left( C_s - C \right) \right\}, \quad (3.15a)
\]

The last term on right hand side of Eq. (3.15) could be simplified as

\[
L \left\{ \frac{A_{adv} - 4\pi D_s T_{\text{min}}}{A} \left( \frac{1}{t} \left( C_s - C \right) + \frac{4\pi D_s}{A} \left( C_s - C \right) \right) \right\} = \frac{A_{adv} - 4\pi D_s T_{\text{min}}}{A} \left( \int_p^\infty \overline{C}_s dp - \int_p^\infty \overline{C} dp \right) + \frac{4\pi D_s}{A} \left( \overline{C}_s - \overline{C} \right) \quad (3.15b)
\]

Letting \( \frac{A_{adv} - 4\pi D_s T_{\text{min}}}{A} = X_1 \) and \( \frac{4\pi D_s}{A} = X_2 \), applying initial condition (3.3a) and rearranging equation (3.15a) leads to

\[
K_s \frac{\partial^2 \overline{C}}{\partial x^2} - U \frac{\partial \overline{C}}{\partial x} - p \overline{C} + X_1 \left( \int_p^\infty \overline{C}_s dp - \int_p^\infty \overline{C} dp \right) + X_2 \left( \overline{C}_s - \overline{C} \right) = -C_0 \delta(x) \quad (3.16)
\]

Laplace transform of equation (3.2b) under the initial condition (3.3b) gives

\[
p\overline{C}_s = \int_p^\infty \overline{C} dp - \int_p^\infty \overline{C}_s dp \quad (3.17)
\]

where \( \overline{C}_s(x, p) = \int_0^t e^{-pt} C_s(x, t) dt \). Reorganizing equation (3.2b) results in

\[
\frac{t}{\partial t} \frac{\partial C_s}{\partial t} = (C - C_s) \quad (3.18)
\]

Laplace transform of equation (3.18) gives

\[
-\frac{d}{dp} \left[ p\overline{C}_s - C_s(x,0) \right] = \overline{C} - \overline{C}_s \quad (3.19a)
\]

\[
\frac{d\overline{C}_s}{dp} = -\frac{1}{p} \overline{C} \quad (3.19b)
\]

The analytical solution to Eq. (3.19b) is

\[
\overline{C}_s = \int -\frac{1}{p} \overline{C} dp + I_1 \quad (3.20a)
\]

where \( I_1 \) is an integral constant. Substitution of \( \overline{C}_s(0, p) = 0 \) and \( \overline{C}(0, p) = 0 \) into Eq. (3.20a) results in \( I_1 = 0 \).
Substitution of Eq. (3.17) into Eq. (3.16) produces

\[ K_s \frac{\partial^2 C}{\partial x^2} - U \frac{\partial C}{\partial x} - pC + X_1(-pC_s) + X_2(C_s - \bar{C}) = -C_0 \delta(x) \]  

(3.21a)

Substituting Eq. (3.20b) into Eq. (3.21a) yields

\[ K_s \frac{\partial^2 C}{\partial x^2} - U \frac{\partial C}{\partial x} - (X_2 + p)\bar{C} + (X_2 - X_1)p \int \frac{1}{p} \bar{C} \, dp = -C_0 \delta(x) \]  

(3.21b)

Laplace transform of equation (3.21b) with respect to length \( x \) and application of the initial condition gives

\[ K_s \left[ s^2 \bar{C} - sC(0, p) - \frac{\partial C}{\partial x}(0, p) \right] - U \left[ s\bar{C} - C(0, p) \right] - (X_2 + p)\bar{C} - (X_2 - X_1)p \int \frac{1}{p} \bar{C} \, dp = -C_0 \] 

(3.22a)

where \( \bar{C}(s, p) = \int_0^\infty e^{-sx}\bar{C}(x, p) \, dx \). Rearranging Eq. (3.22a) using initial conditions gives

\[ s^2 K_s \bar{C} - sU\bar{C} - (X_2 + p)\bar{C} - (X_2 - X_1)p \int \frac{1}{p} \bar{C} \, dp = -C_0 \] 

(3.23a)

Rearrangement of (3.23a) leads to

\[ \left[ s^2 K_s - sU - (X_2 + p) \right] \bar{C} - \int \frac{1}{p} \bar{C} \, dp = \frac{-C_0}{(X_2 - X_1)p} \] 

(3.23b)

Taking derivative with respect to \( p \) on both sides of Eq. (3.23b) and making some arrangements yields

\[ \frac{\partial \bar{C}}{\partial p} + \frac{X_1}{\left[ s^2 K_s - sU - (X_2 + p) \right] (X_2 - X_1)} \bar{C} = \frac{-X_1C_0}{\left[ s^2 K_s - sU - (X_2 + p) \right] (X_2 - X_1)} \] 

(3.24)

Letting

\[ \frac{X_1}{\left[ s^2 K_s - sU - (X_2 + p) \right] (X_2 - X_1)} = M_1 \] and

\[ \frac{-X_1C_0}{\left[ s^2 K_s - sU - (X_2 + p) \right] (X_2 - X_1)} = M_2 \]

leads to the solution to (3.24) as

\[ \bar{C} = e^{-\int M_1 \, dp} \left[ \int M_2 e^{\int M_1 \, dp} \, dp - I_2 \right] \] 

(3.25)

where \( I_2 \) is a constant. The Laplace transformed initial and boundary conditions for the concentration are

\[ \bar{C}(x, 0) = 0 \] 

(3.26a)
\[
\frac{\partial \overline{C}(x, p)}{\partial x} = 0, \quad x = 0 \quad (3.26b)
\]
\[
\overline{C}(x, p) = 0, \quad x = \infty \quad (3.26c)
\]
\[
\overline{C}_s(x, p) = 0, \quad x = \infty \quad (3.26d)
\]
\[
\overline{C}(x, p) = \int_0^\infty e^{-pt} C(x, t) dt = \int_0^\infty C(x, t) dt, \quad p = 0 \quad (3.26e)
\]
\[
\overline{C}(x, p) = \int_0^\infty e^{-pt} C(x, t) dt = 0, \quad p = \infty \quad (3.26f)
\]

The first integration part in Eq. (3.25) can be expressed as
\[
\int \frac{X_1 \left[ s^2 K_s - sU - (X_2 + p) \right] - (X_2 - X_1 p)}{s^2 K_s - sU - (X_2 + p) (X_2 - X_1 p)} dp = \ln \frac{|s^2 K_s - sU - (X_2 + p)|}{|X_2 - X_1 p|}
\]

Then, Eq. (3.25) becomes
\[
\overline{C}(s, p) = \frac{|X_2 - X_1 p|}{s^2 K_s - sU - (X_2 + p)} \left[ \int \frac{-X_1 C_0}{s^2 K_s - sU - (X_2 + p) (X_2 - X_1 p)} \frac{|s^2 K_s - sU - (X_2 + p)|}{|X_2 - X_1 p|} dp + I_2 \right]
\]
\[
= \frac{I_2 (X_2 - X_1 p) - C_0}{s^2 K_s - sU - (X_2 + p)} \quad (3.27)
\]
\[
\overline{C}(s, p) = \frac{I_2 (X_2 - X_1 p) - C_0}{s - \left( \frac{U}{2K_s} \right)^2 - \left( \frac{U^2 + X_2 + p}{K_s} \right)^2} \quad (3.28)
\]

The inverse Laplace transform of \( \overline{C} \) can be found by the complex variable theory as
\[
\overline{C}(x, p) = \left[ \frac{I_2 (X_2 - X_1 p) - C_0}{s - U^2 X_2 + p K_s} \right] e^{\frac{Ux}{2K_s}} \exp \left[ \frac{U^2 X_2 + p - x}{4K_s} \right] - \exp \left[ - \frac{U^2 X_2 + p - x}{4K_s} \right] \quad (3.29)
\]

The term that generates negative concentrations is eliminated. Thus, the final form of the Laplace transformed solution is
\[
\overline{C}(x, p) = \frac{-I_2 (X_2 - X_1 p) - C_0}{\sqrt{U^2 + 4K_s (X_2 + p)}} \exp \left[ \frac{U - \sqrt{U^2 + 4K_s (X_2 + p)}}{2K_s} \right] x \quad (3.30)
\]
\[ \bar{C}(x, p) = \frac{C_0}{2\sqrt{K_s}} \exp\left( \frac{Ux}{2K_s} \right) \exp\left( - \frac{x}{\sqrt{K_s}} \sqrt{p + \frac{U^2}{4K_s} + X_2} \right) \] (3.31)

By performing the inverse Laplace transform and applying the initial conditions, Eq. (3.6) in the main text is obtained.
VITA
Yangbin Tong received his bachelor’s degree in Hydrology at Zhejiang University in 2006 and the master’s degree in Hydrology at Zhejiang University in 2008. Thereafter, he worked as a research associate in Hydrology and Water Resources Engineering Institute of Zhejiang University. One year later, he worked in Zhejiang Institute of Hydraulics and Estuary in Hangzhou, China. As his interest in research increased, he decided to enter graduate school in the Department of Civil and Environmental Engineering at Louisiana State University. He will receive his master’s degree in August 2013.