Development of Integrated Methodologies for Optimal Monitoring and Source Characterization in Contaminated Groundwater Systems under Uncertainty

Thesis submitted by

Mahsa AMIR ABDOLLAHIAN

May 2016

For the degree of Doctor of Philosophy

College of Science and Engineering

James Cook University

Acknowledgments

Firstly I want to thank my principal advisor Dr Bithin Datta. It has been an honor to be his student and benefit from his guidance throughout the course of this study. I am particularly thankful to him for his encouragement to complete this study and pursue my ideas. The submission of this thesis has become possible because of his guidance and support.

I wish to express my gratitude to the Graduate Research School and Professor Helene Marsh, Dean of Graduate Research Studies, James Cook University for technical, emotional and financial support. I am also thankful to Dr Vincent Z. Wang and Dr Christa Placzek for their support as my co-supervisors.

My sincere thanks to CRC-CARE Pty Ltd, Australia and the Graduate Research School, James Cook University, Australia, for financial support for my research, living expenses and professional development.

Special thanks to my husband, Mr Morteza Mokhtari whose unconditional love and support make the completion of this study possible and enjoyable. My sincere thanks to my parents, family and friends for all their support throughout my life and their encouragement to pursue my dreams. Finally, special thanks to my little daughter, Deanna, who fills my life with joy and hope and by entering my world made the completion of this study more enjoyable and fulfilling.

Statement of Contribution of Others

All the methodologies, concepts and results reported in this thesis are developed and written by Mahsa Amir Abdollahian under the overall supervision of principal supervisor, Dr Bithin Datta. Model conceptualization, development and modification of programs were done by Mahsa Amir Abdollahian, with Dr Bithin Datta providing advice on various aspects of the work. The text of the under review, and published articles in journals and conference proceedings and the thesis were written by Mahsa Amir Abdollahian, with Dr Bithin Datta reviewing the manuscripts. The calibrated flow and contaminant transport model utilized in chapter 7 was developed by Prakash and Datta (2015).

Financial support was received from CRC-CARE Pty Ltd, Mawson Lakes, South Australia and James Cook University, Townsville Australia. Financial support included living expenses and covered research costs including participating in national and international conferences.

Assistance was received from professional editor Dr John Cokley of EduPreneur Services International, who provided copyediting and proofreading services according to the guidelines laid out in the university-endorsed *National Guidelines for Editing Research Theses*."

ABSTRACT

Groundwater is a major source of water in many parts of the world. Due to anthropogenic activities it is subjected to various sources of contamination. Effective groundwater pollution management and remediation relies on accurate identification of unknown pollution source characteristics. The pollution source should be defined in terms of location, flux magnitude, and time of release. The source identification procedure is a challenging task due to uncertainties in model definition, parameter estimation, hydrogeologic parameters and field measurements. Availability of adequate data is vital due to the nature of the contaminant source identification procedure; however, acquiring accurate and extensive field measurement data is a very costly and time intensive task. The source identification procedure remains a challenging task due to sparse measurement data, uncertainties in model definition, parameter estimation, and field measurements. Therefore the source identification problem is often characterized by very little information, and is considered an ill-posed, complex, and non-unique inverse problem.

In the recent past, the contaminant source identification problem has been addressed using different approaches. The linked simulation-optimization approach is capable of incorporating complex real-life scenarios and large study areas. In this approach, the numerical flow and transport simulation models are linked to an optimization model. A typical optimal source identification model minimizes the difference between estimated pollutant concentration and measured contamination values at monitoring locations. Due to the complex nature of the contaminant source identification process, various issues are needed to be addressed in the proposed

methodologies in order to achieve an acceptable level of accuracy in recovering contaminant source histories. This study focuses on improving the accuracy and efficiency of contaminant source identification procedures in the presence of hydrogeologic parameter and measurement, uncertainties. This study specifically focuses on the following issues:

i. The utilized simulation model should include an accurate description of the study area in terms of hydrogeologic parameters. The uncertainty in hydrogeologic parameter values must be quantified and incorporated into the contaminant source identification methodology.
ii. Often the recorded concentration measurements are erroneous. The measurement error/uncertainty is required to be explicitly incorporated into the optimal source identification objective functions.
iii. Selection of the monitoring locations for measuring concentrations is vital for accurate source identification, and efficient selection of these locations can increase the accuracy of recorded source histories. Therefore, an optimal monitoring network is required to be designed, intended to reduce uncertainty and increase efficiency in recovering source release histories.

This study focuses on improving the accuracy and efficiency of contaminant source identification/characterization procedures in the presence of hydrogeologic parameter, and measurement, uncertainties. The main features of this study are:

i) The hydrogeological parameter uncertainty is incorporated in the optimal source identification methodology, using a new parameter for uncertainty quantification based on various realizations of the flow field. The new

parameter, called the Coefficient of Confidence (COC), is estimated for each available contaminant observation data and incorporated into the optimal source identification model. Although such quantification cannot be designated as fuzzy quantification, in a strict sense, it differs from the crisp approach generally adopted in earlier developed methodologies for source characterization. This approach can actually eliminate the possibility of incorporating additional uncertainties due to inappropriate use of so called expert judgments. Adaptive simulated Annealing (ASA) is used as the optimization algorithm.

ii) A two-objective contaminant source identification model is developed which can improve the accuracy of recovered source histories using erroneous contaminant measurement data. Non-dominated Sorting Genetic Algorithm (NSGA) II is used as the optimization algorithm.

iii) A new optimal monitoring network design methodology is developed using redundancy reduction and uncertainty reduction objective functions. This new two-objective monitoring network design model is solved using NSGA-II and integrated into the source identification methodology by sequential implementation of the optimal monitoring network design, and solution of the optimal source identification model, in order to increase the efficiency of both methodologies.

This study includes the following steps. An ASA based simulation-optimization source identification model is developed which is externally linked with numerical flow and transport simulation models. To address the hydrogeologic parameter uncertainty in

the source identification model, an uncertainty quantification method is developed. The hydraulic conductivity uncertainty is quantified using a quantification parameter based on various realizations of the flow field, and using a new parameter called the Coefficient of Confidence (COC). Incorporating the estimated COC values in the contaminant source identification methodology results in the uncertainty-based linked simulation optimization model for contaminant source identification. The performance of this proposed methodology is evaluated for both illustrative and experimental contaminated aquifers. Obtained solution results show that the proposed methodology is capable of recovering source release histories more accurately compared with those obtained using earlier developed crisp methodologies in which the hydrogeologic parameter uncertainty is not considered explicitly.

In the next stage of this thesis, a two-objective optimal source identification methodology is developed which focuses on measurement error/uncertainty. Two-objective optimal contaminant source identification models can improve the accuracy of estimating the recovered source histories using erroneous contaminant concentration measurement data. NSGA-II is linked to the flow and transport simulation models to determine the contaminant source characteristics.

When the measurements are erroneous, the source identification problem becomes non-unique. Therefore, various solutions with a possibility of being the true source characteristics may be achieved. When the contaminant concentrations are erroneous, it is not possible to match all the observed and simulated concentrations. In the two-objective approach, the first objective function is normalized using observed concentrations. Therefore, this objective function emphasizes matching smaller

observed concentrations (for larger objective function improvement); however, the second objective function, which is not normalized, tries to match the high concentrations. Therefore, the two-objective approach generally finds the possible solutions as a Pareto-front.

The performance of the developed methodology using two different objectives for optimal source characterization is evaluated for an illustrative study area considering both point and distributed contaminant sources. The obtained solution results demonstrate that the two-objective model is capable of finding more accurate source characteristics in the presence of measurement error, compared with those obtained using each objective function separately in a single-objective model.

In the next step, a new sequential monitoring network design methodology is developed which selects new monitoring locations based on sequential characterization of pollutant sources and feeds back new measurement information from the newly designed and implemented monitoring network. Integrated source identification and monitoring network design sequences are carried out to reach an acceptable level of accuracy in characterization of the contaminant source properties. A new two-objective monitoring network design model is proposed which minimizes the uncertainty in recovered source histories and minimizes the redundancy in measured contaminant concentrations.

The performance of the integrated sequential contaminant source identification and monitoring network design methodologies are evaluated using illustrative data, and data from real-life contaminated groundwater aquifers. Performance evaluation results show that the developed methodologies can improve the accuracy of recovered

contaminant source histories in the presence of uncertainty, and provides insight into the reliability of recovered source histories. The new sequential monitoring network design methodology can reduce the cost and time required to achieve relatively accurate characterization of polluted aquifers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Widespread contamination is a major threat to beneficial use of groundwater. Often contamination of groundwater aquifers results from failure to adequately safeguard stored chemicals, and is caused by the impact of these chemicals released into the water bodies. Other sources and pathways of groundwater contamination also exist, including industrial wastes, mine wastes, pesticides, etc.

In polluted groundwater systems, the existing contamination released by past activities must be identified and characterized. The characterization process includes finding the source locations among potential sources and retrieving their pollutant release histories. This approach allows "Potential Responsible Parties" (PRVs) to use the identified contaminant source characteristics along with flow and transport simulation models to (1) convince regulators that the existing contamination does not exceed the regulation thresholds and no remediation is required; or (2) develop effective remedial plans that satisfy the cost constraints and have appropriate reliability and consideration of risk. Therefore, the accuracy in characterization of contaminant sources is of great interest to PRVs and regulators; however, groundwater systems are subject to various sources of uncertainty, and acquiring data in polluted aquifers is a costly and time-consuming procedure. Therefore, this study aims to improve the accuracy in recovered contaminant source histories in existing polluted aquifers using available hydrogeological and contaminant concentration measurement information with their inherent uncertainties.

1. 1. Unknown groundwater contaminant source characterization

The identification of the contaminant sources plays an important role in modeling of subsurface water and reduces the long-term remedial cost. The source identification problem deals with the spatial and temporal variations of the location, activity duration, and the injection rate of the pollutant sources and is mostly inferred from the available sparse and sometimes erroneous concentration measurements at the site. Mainly source identification includes a simulation problem, such as groundwater flow and pollutant transport models, used to estimate past phenomena or predict future scenarios. The estimated values are then compared with observed values. Availability of adequate and accurate measurement data is vital for determining the structure of source identification procedure; however, acquiring adequate and reliable field measurement data is a cost- and time-intensive task, often making the source characterization process difficult.

The lack of sufficient hydrogeologic parameter information results in uncertain groundwater flow and solute transport models. The hydrogeological properties of a polluted aquifer have significant influence on the temporal and spatial properties of contamination plumes. Therefore, this study focuses on addressing hydrogeological parameter value uncertainty in polluted aquifers. In most of the previous contamination source identification methods (Jha & Datta, 2013; Mahar & Datta, 2001), the effect of hydrogeological parameter uncertainty has been incorporated in the performance evaluation data; however, these uncertainties were not formally or explicitly quantified in these proposed methodologies.

The other important source of uncertainty in contaminant source identification problems is the measurement error/uncertainty associated with the measured

contaminant concentrations over space and time. In most of the previous studies, measurement error was not considered explicitly in the methodology. Jha and Datta (2013) synthetically added noise to the measured concentrations, to test the method for realistic scenarios. A random quantity based on a specified uniform distribution represented the measurement error. They demonstrated that their single-objective optimal source identification methodology performed satisfactorily when up to 10% measurement error exists.

Most of previously developed optimal groundwater contaminant source identification models are single-objective optimization algorithms. As a result of erroneous measurement data, and the necessity of using optimization procedures, the optimal source identification problem may be ill-posed and non-unique (Datta, 2002). Therefore, smaller objective function values do not always mean accurate identification of contaminant sources and accurate reconstruction of release histories. A more robust approach would be to incorporate parameter and measurement uncertainties in the proposed methodologies themselves by using appropriate objective functions.

1. 2. Monitoring network design dedicated to contaminant source identification

The contaminant source characteristics are identified by minimizing the difference between the observed and estimated contaminant concentrations at monitoring locations and times. Effective selection of observation points plays a critical role in accurate source identification. An improper monitoring network will result in wasted time and money utilized for site data collecting, providing inaccurate or irrelevant information for the optimal source identification.

Optimal design of the monitoring network is necessary due to uncertainty in predicting the movement of plumes in the groundwater system, and budgetary limitations. Often the contaminated study area is large and there are many potential locations to drill and collect groundwater samples. On the other hand, installation and sample collections are very costly and this limits the number of monitoring locations. Moreover, groundwater contamination problems are mostly long-term problems. The release of contaminants can begin many years before actual detection of contaminants in groundwater aquifer, while the source release may have stopped or continued at the time of measurement data collection. Therefore, groundwater contamination monitoring can be a very long and time-consuming process.

An efficient contaminant monitoring network can substantially decrease the number of required monitoring wells based on budgetary limitations. It can also reduce the required contaminant monitoring locations by selecting appropriately located monitoring wells.

In the past various optimal monitoring network design objective functions have been proposed. The monitoring wells can be selected in order to reduce uncertainty and increase the accuracy in the contaminant source characterization process, considering budgetary limitations as constraints.

1. 3. Sequential monitoring network design and optimal contaminant source identification

Satisfactory characterization of contaminant sources is difficult without a sufficient and appropriately designed monitoring network. Design of groundwater contamination detection monitoring networks ideally requires information about the sources and the

expected distribution of contaminant plumes corresponding to location and time. Therefore, a coupled and iterative sequential source identification and dynamic monitoring network design framework is required to improve the efficiency of the designed networks. Preliminary identification of unknown sources, based on limited concentration data from initially existing arbitrarily located wells, provides initial rough estimates of the source fluxes. These preliminary identified source fluxes are then utilized for designing an optimal monitoring network for the first sequence. Both processes are repeated i.e., the sequential identification of sources and the subsequent design of the monitoring network and this provides the new concentration measurement feedback information for improved source identification in the following sequence. This sequence may be repeated few times depending on the characteristics of the study area. In the absence of adequate and reliable measurements from the field, this sequential process may be the most appropriate.

Sequential optimal monitoring network design, coupled with the optimal source identification process can provide vital feedback information for the contaminant source identification model and efficient source characterization process. Moreover, this sequential process can decrease the required contaminant monitoring time which reduces the associated costs and efforts as well. As a consequence, the decision makers and environmental managers may be able to decide and implement management plans and remedial strategies faster, utilizing reliable information.

1. 4. Research objectives

Most of the previously proposed methodologies for groundwater contaminant source characterization have paid very limited or, no consideration to the specific sources of

uncertainty in the groundwater contamination modeling and prediction. This study aims to incorporate the hydrogeologic parameter uncertainty (specifically hydraulic conductivity), and contaminant measurement error explicitly in the contaminant source identification and monitoring network design methodologies. The ultimate aim is to increase the accuracy in recovered contaminant source histories using available sparse and erroneous/ uncertain information about the contaminated aquifer study area.

The specific objectives of this study are:

i. Development and evaluation of a methodology for quantification of hydrgeologic uncertainties in a contaminated aquifer
ii. Development of a linked simulation-optimization based contaminant source characterization methodology, incorporating quantification of uncertainties in estimation of hydraulic conductivity, utilizing Adaptive Simulated Annealing (ASA) as the optimization algorithm.
iii. Performance evaluation of the developed methodology using synthetic aquifer data (using simulated measurements)
iv. Performance evaluation of the developed source characterization methodology using measurement data from an experimental aquifer study area for quantification of hydrogeologic uncertainty in groundwater source characterization
v. Development of a two-objective simulation-optimization based contaminant source characterization methodology, utilizing Non-dominated Sorting Genetic Algorithm II (NSGA-II) considering measurement error

vi. Development of a two-objective optimal monitoring network design model dedicated to reduce uncertainty in the contaminant source identification process and to reduce redundancy in contaminant monitoring

vii. Development of a methodology integrating sequential monitoring network design and optimal contaminant source characterization incorporating uncertainty, and monitoring redundancy reduction objectives

viii. Performance evaluation of the integrated sequential monitoring network design and source characterization methodology using synthetic (simulated) hydrgeologic and concentration data.

ix. Performance evaluation of the sequential monitoring network design and source characterization methodologies utilizing hydrgeologic and concentration data from a real-life contaminated aquifer.

1. 5. Organization of the thesis

This thesis comprises of eight chapters including an introduction. Chapter 2 is devoted to the review of literature related to the areas discussed in this study.

Chapter 3 presents a new developed methodology to quantify hydrogeologic uncertainty in contaminant source identification methodology. It is followed by the performance evaluation of the methodology in an illustrative study area.

Chapter 4 includes the performance evaluation of the methodology discussed in Chapter 3 for a real experimental contaminated study area. Chapter 5 introduces a new two-objective contaminant source identification methodology using a NSGA-II-based linked simulation-optimization model. This new proposed source identification methodology aims to reduce the effect of contaminant measurement error on the

accuracy of recovered source histories. The performance evaluation of this developed methodology for an illustrative study area is also presented in this chapter.

Chapter 6 presents a new integrated sequential monitoring network design and source characterization methodology aimed at: reducing uncertainty in contaminant source identification methodology, and also reducing the redundancy in the designed contaminated monitoring network. Performance evaluation of this developed methodology for an illustrative study area is presented.

Chapter 7 includes the performance evaluation of the integrated sequential monitoring network and source identification methodologies for a real-life contaminated urban aquifer. Chapter 8 presents a summary and conclusions of the study contained in this thesis.

The next chapter contains a brief review of some of the representative and important literature relevant to this study.

2. LITERATURE REVIEW

This chapter provides a brief review of the representative literature relevant to contaminant source identification, monitoring network design and uncertainty in these two areas. A part of this chapter has been published in the following journal paper:

Amirabdollahian, M., & Datta, B. (2013). Identification of contaminant source characteristics and monitoring network design in groundwater aquifers: an overview. *Journal of Environmental Protection*, 4(5A), 26-41.

To generate information about the pollutant source characteristics, when groundwater contamination is detected at user-end wells, three main issues should be studied. First, the optimization algorithm utilized to minimize the objective function representing the differences in measured and simulated concentrations should be evaluated. Second, measurements to be obtained, and the most appropriate monitoring locations should be decided. Design of an efficient monitoring network instead of arbitrary selection of monitoring locations plays a critical role in the efficiency and reliability of source characterization. Finally, as the actual observation values are collected in the field, where as simulated values are obtained as solution results of mathematical or numerical ground water flow and solute transport models, the uncertainty in hydrogeological parameter estimates should be studied in detail.

This Chapter discusses the background and review of literature on identification of source characteristics, design of monitoring network, the integration of source characterization with monitoring network design along with incorporation of uncertainties.

2. 1. Identification of contaminant source characteristics

The most important pollutant source properties which should be addressed in an identification procedure are: The source release history; source locations; and, source fluxes.

One approach for source identification consists of solving the differential equations backwards in time (inverse problem). The random walk particle method (Bagtzoglou et al., 1991, 1992), the quasi-reversibility technique (Skaggs & Kabala, 1995), the minimum relative entropy method (Woodbury & Ulrych, 1996), the Bayesian theory and geostatistical techniques (Snodgrass & Kitanidis, 1997) and genetic algorithm (Aral et al., 2001; Mahinthakumar & Sayeed, 2005), are some examples of this approach.

Wagner (1992) developed an inverse model for simultaneous parameter estimation and contaminant source characterization. Skaggs and Kabala (1994), by using the Tikhonov Regularization (TR), changed the ill-posed problem of contaminant source identification to a well-posed minimization problem solved backwards in time. Liu and Ball (1999) tested Skaggs and Kabala (1994)'s method at a low permeability site at Dover Air Force Base, Delaware. Skaggs and Kabala (1995) applied a more computationally efficient and easier to use method called Quasi-Reversibility (QR) to the previous problem; however, the results showed that the advantages of the QR method come at the expense of accuracy.

An inverse problem approach based on statistical inference, was applied to the same problem as Skaggs and Kabala (1994), by Woodbury and Ulrych (1996). Neupauer et al. (2000) evaluated the relative effectiveness of TR and Minimum

Relative Entropy (MRE) methods in reconstructing the release history of conservative contaminant in a one-dimensional domain. Snodgrass and Kitanidis (1997) developed a probabilistic method for source release history estimation that combines the Bayesian theory with geostatistical techniques.

A different approach for identification of source characteristics is simulation-optimization. It couples the forward time contaminant simulation model with optimization techniques. This approach avoids the problem of stability associated with formally solving the inverse problem but the iterative nature of simulation models usually requires increased computational effort. Many techniques were proposed utilizing coupled simulation-optimization, and a few representative ones are discussed here.

2.1. 1. Response matrix

The response matrix approach utilized unit responses of system in the form of a response matrix. Gorelick et al. (1983) used the response matrix approach in the identification of pollution source models using a linear programming optimization model. Datta et al. (1989) developed an expert system using statistical pattern recognition techniques to identify groundwater pollution sources. Stochastic dynamic programing was used for optimal pattern classification. The flow and transport processes were simulated using the response matrix approach.

The two limitations of the response matrix approach are: it is based on the premise that the superposition principle is approximately valid in terms of flow and contaminant transport in the aquifer; and the aquifer parameters must be known and the

simulation model must be used to generate the response matrix prior to running the source identification model (Mahar & Datta, 2001).

2.1. 2. Embedded optimization

This approach identified and characterized a pollutant source by solving an optimization model that embeds the governing equations of the physical process of flow and transport as constraints. Mahar and Datta (2001) used a nonlinear optimization model embedding the governing equations for flow and transport processes, to identify pollutant source characteristics as well as to estimate aquifer parameters. Finite difference discretization of flow and solute transport equations formed the binding constraints. They showed that the embedding methods need large computer storage and computational time, for large aquifers. Gorelick et al. (1983) concluded that numerical difficulties are likely to arise for large-scale problems using the embedding technique.

2.1. 3. Linked simulation-optimization

To conduct unknown pollutant source characterization in large-scale aquifers and real areas, linked simulation-optimization methodology was proposed. In this methodology the numerical models for simulation of the flow and transport process are internally linked to the optimization algorithm. Chadalavada et al. (2011a) presented an overview of the pollution source identification simulation-optimization approaches.

The linked simulation-optimization model is an efficient and effective technique to characterize the contaminant sources by internal linkage between flow and contaminant transport simulation models and the selected optimization technique (Aral et al., 2001; Datta et al., 2009; Mahar & Datta, 2001). This methodology can solve

contaminant source problems in fairly large study areas. Moreover, utilizing the evolutionary optimization algorithms simplifies the linking process. Examples of the evolutionary optimization algorithms include: Genetic Algorithm (GA) (Singh & Datta, 2006), Tabu Search (TS) (Yeh et al., 2007), Simulated Annealing (SA), Adaptive Simulated Annealing (ASA) (Amirabdollahian & Datta, 2014) and differential evolution algorithm (Gurarslan & Karahan, 2015). Yeh (2015) and Datta and Kourakos (2015) presented an overview on various optimization methods coupled with simulation techniques utilized for groundwater quantity management, and quality management, respectively.

Aral et al. (2001) formulated a source identification model which minimized the residuals between the simulated and measured contaminant concentrations at observation sites. They observed that the effect of measurement errors on identification of source locations is very small but it highly affects the accuracy of recovered release histories. Singh and Datta (2006) used GA for unknown source characterization in the case of different levels of data availability and concentration measurement errors. The utilized GA objective function was to minimize the weighted sum of absolute differences between observed and simulated concentrations subject to upper and lower bonds for source fluxes. To increase the computational efficiency of GA, Mahinthakumar and Sayeed (2006) combined GA with local search approaches. Results indicated that hybrid optimization methods, combining an initial global heuristic approach with a subsequent gradient-based local search approach, are very effective in characterizing contaminant sources.

Datta et al. (2009, 2011) were able to combine linked simulation-optimization with classical nonlinear optimization. Datta et al. (2009) were successful in combining the identification of an unknown pollution source and estimation of hydrogeologic parameters. Jha and Datta (2011) proposed a linked simulation-optimization method which incorporated Simulated Annealing (SA) based methodology to characterize pollutant sources and used SA to define the source locations, fluxes and duration. They compared the SA results with the GA based solution. Jha and Datta (2013) demonstrated that an ASA-based contaminant source identification algorithm is computationally efficient in terms of execution time and accuracy. Jha and Datta (2015b) presented an application of the linked simulation-optimization based methodology to estimate the release history from spatially distributed sources of pollution at an illustrative abandoned mine-site.

In recent years, surrogate models have been used to approximate the original, computationally expensive simulation models. A surrogate model is simpler, considerably faster to run, and is generally derived from the response surface generated from the original simulation model. A popular technique to generate surrogate models is based on Artificial Neural Network (ANN). Singh et al. (2004) used the feed forward multilayer ANN to identify unknown pollution sources and simultaneously estimate aquifer parameters. The proposed methodology was also tested for the missing data scenario where part of the concentration observation data was missing (Singh & Datta, 2007). Moreover, Genetic Programming (GP) can be used as a surrogate model for simulation of groundwater and contaminant transport processes. GP-based models are simpler and more efficient compared with conventional surrogate models and ANN

models (Sreekanth & Datta, 2011). Datta et al. (2014) improved the efficiency of the linked simulation-optimization methodologies using GP-based surrogate models. Datta et al. (2013) addressed the use of GP surrogate models for identifying unknown contaminant source characteristics and monitoring network design feedback methodology. Koohpayehzadeh Esfahani and Datta (2015) developed a trained GP-based surrogate model to replace the flow and biochemical transport simulation models in order to reduce computation time. In addition to surrogate models, parallel processing capabilities have been utilized in a number of applications to solve large-scale optimization models for groundwater source identification problems (Mirghani et al., 2009).

2. 2. Monitoring network design

Optimal design of the monitoring network is necessary due to uncertainty in predicting the movement of plumes in the groundwater system, and budgetary limitations. A comprehensive review of monitoring network design is reported in Loaiciga et al. (1992), ASCE Task Committee (2003) and U.S. EPA. (2005).

A range of different considered objectives of monitoring network design models reported in the literature include: maximizing detection (Meyer & Brill, 1988; Meyer et al., 1994; Yenigül et al., 2005); Minimizing the number of wells (Li & Chan Hilton, 2007; Meyer et al., 1994); minimizing undetected concentrations (Cieniawski et al., 1995; Datta & Dhiman, 1996; Dhar & Datta, 2007; Mahar & Datta, 1997); minimizing the estimation variance (Bogardi et al., 1985; Loaiciga, 1989; McKinney & Loucks, 1992; Nunes et al., 2004; Passarella et al., 2003; Woldt et al., 1992); minimizing the

uncertainty in terms of square root of estimation variance (Kollat & Reed, 2007b; Reed & Minsker, 2004); variance reduction with Kalman filter approach (Herrera & Pinder, 2005; Zhang et al., 2005); minimizing the monitoring cost (Dhar & Datta, 2007; Kollat & Reed, 2007b; Reed et al., 2000; Reed & Minsker, 2004; Wu et al., 2005; Wu et al., 2006); squared deviation of estimated concentration from actual measured concentrations (Kollat & Reed, 2007b; Reed et al., 2000; Reed & Minsker, 2004); minimizing mass estimation error (Montas et al., 2000; Reed & Minsker, 2004; Wu et al., 2005; Wu et al., 2006) and minimizing the error in locating plume centroid (Montas et al., 2000; Wu et al., 2005; Wu et al., 2006).

Different optimization algorithms used earlier include: integer programming (Hudak, 1998; Hudak et al., 1995; Mahar & Datta, 1997); mixed integer programming (Datta et al., 2008; Dhar & Datta, 2007; Loaiciga, 1989); simulated annealing (Mugunthan & Shoemaker, 2004; Nunes et al., 2004); simple genetic algorithm (Cieniawski et al., 1995; Mugunthan & Shoemaker, 2004; Reed et al., 2000; Wu et al., 2005; Wu et al., 2006; Zhang et al., 2005) and ant colony optimization (Li & Chan Hilton, 2007).

Meyer et al. (1994) used simulated annealing to solve the multi-objective integer programming of optimal monitoring network design. The system uncertainty was incorporated using Monte-Carlo simulation. Datta and Dhiman (1996) used a mixed-integer programming algorithm which was linked to a flow and transport model using a response matrix approach. To incorporate the uncertainty in solute transport simulation, a random error was added to response matrix elements. These random errors follow a uniform distribution where the variance is controlled by a degree of uncertainty.

Hudak (2001) tested the mass transport simulation model for seven contaminant detection-monitoring networks under a 40-degree range of groundwater flow directions. Reed and Minsker (2004) demonstrated the use of high-order Pareto optimization on a Long-Term Monitoring (LTM) application. Mugunthan and Shoemaker (2004) identified a cost-effective sampling design for long-term monitoring of groundwater remediation under multiple monitoring periods and uncertain flow conditions. Dhar and Datta (2007) proposed chance-constrained single and multi-objective nonlinear optimization models which were capable of designing an optimal time-variant groundwater quality monitoring network. Probability theory was used to define chance constraints with associated reliabilities. Chadalavada and Datta (2008) utilized two-objective functions to design an effective monitoring network. The first objective function aims to minimize the summation of all positive deviations between simulated contaminant concentrations and specified low threshold. The second one minimizes the estimated variance of pollutant concentrations at various unmonitored locations. Dhar and Datta (2009) presented a methodology based on a linear mixed-integer formulation incorporating kriging spatial interpolation technique for global optimal design of water quality monitoring. Dhar and Datta (2010) formulated a logic-based mixed-integer linear optimization model to develop a model solution for optimal design of groundwater monitoring network. In the developed methodology the monitoring redundancy reduction has been considered.

Bashi-Azghadi and Kerachian (2010) located monitoring wells in order to identify pollution sources using a multi-objective method. The objective functions of the multi-objective optimization model were minimizing the monitoring cost, i.e. the

number of monitoring wells, maximizing the reliability of contamination detection and maximizing the probability of detecting an unknown pollution source.

An uncertainty-based optimization model was used by Chadalavada et al. (2011b) to design an optimal monitoring network in a real aquifer in South Australia. The model located the monitoring wells at locations where the spatial estimation variance is large.

Prakash and Datta (2014) and Datta et al. (2013) presented GP-based monitoring factors for design of an optimal monitoring network to improve the accuracy of pollutant source identification. These methodologies used trained GP models to calculate the impact factor of the sources on the candidate monitoring locations.

Jha and Datta (2014, 2015a) developed a methodology using Dynamic Time Warping (DTW) distance as a cost function in the linked simulation–optimization model to design a monitoring network. The objective of monitoring network design was to efficiently estimate contaminant source characteristics including the starting time of release of an unknown magnitude of pollution from sources. Performance of the developed methodology was evaluated with data obtained from a real aquifer.

2. 3. Integration of sequential contaminant source characterization and monitoring network design

Satisfactory characterization of contaminant sources is difficult without concentration measurement data from an efficient monitoring network. Determination of well-located monitoring positions requires information about the sources, and the distribution of the contaminant plume corresponding to location and time. Therefore coupled and iterative

sequential source identification and dynamic monitoring network design framework is required. The preliminary identification of unknown sources, based on limited concentration data from existing arbitrary located wells, provides the initial rough estimation of the source fluxes. These identified source fluxes are then utilized for designing an optimal monitoring network for the first stage. Both processes are repeated by the sequential identification of sources, and design of the monitoring network which provides feedback information in the form of new concentration measurements.

Mahar and Datta (1997) presented a methodology combining an optimal ground-water quality monitoring network design and an optimal source-identification model. In the first step, using an embedding nonlinear optimization model, preliminary identification of contaminant sources was obtained. In the next step, an integer programming formulation-based monitoring network design model selected the monitoring wells in subsequent time periods. In the last step, simulated concentrations at new monitoring wells were also perturbed to show measurement errors. The new concentration measurement data obtained from the designed monitoring well locations were utilized for more accurate identification of source characteristics in sequential steps. Comparing results shows that where there is no uncertainty, using an existing monitoring network for source identification results in acceptable source flux estimates, but updating the monitoring network is preferable. In the presence of uncertainty, the existing monitoring network did not show acceptable results and update of the network was found to be beneficial.

Datta et al. (2008) proposed a methodology which is an improvement on the combined source identification and monitoring network design model presented by Mahar and Datta (1997). They used the trimmed mean concentration incorporated in the monitoring network design. Dokou and Pinder (2009) addressed the issue of identifying and delineating Dense Non-aqueous Phase Liquids (DNAPLs) at their sources. In this research the iterative process of source identification and monitoring network design was proposed, based on combining water quality information (hard data) with expert knowledge (soft data). The authors assumed the hydraulic conductivity as an uncertain model parameter. The utilized search strategy includes a Monte-Carlo stochastic groundwater flow and transport model, a predetermined set of potential source locations, and a Kalman filter that updates the simulated contaminant concentration field using contaminant concentration data. The updated plume was compared with the simulated contaminant concentration fields that emanate from each individual potential source using a technique rooted in fuzzy set theory. The three-dimensional extension and field application of this method were tested by Dokou and Pinder (2011).

Singh and Datta (2010) proposed a kriging linked SA model for the spatial and temporal estimation of contaminant plumes. In their proposed sequential simulation optimization model, the optimal monitoring network is designed based on the objective function of minimizing the contaminant mass estimation error. The new selected monitoring wells generate feedback information for the SA optimization model to estimate the pollutant concentration plume more accurately. Prakash and Datta (2015) demonstrated the application of an optimization-based methodology for

characterization of unknown pollution sources, integrating sequential and iterative design of the optimal monitoring network. The performance of the methodology was evaluated by application to a contaminated aquifer site in New South Wales, Australia.

2. 4. Parameter uncertainty in ground water systems

Water resources systems are inherently uncertain and relevant available information is mostly imprecise. Solution of groundwater flow and transport models requires the knowledge of various soil hydraulic parameters as well as the determination of boundary conditions, which are subject to different sources of uncertainty. Spatial variations in hydraulic conductivity play a critical role in transport of contaminants in groundwater systems, and heterogeneity controls the spread of the contamination plume. Considerable hydraulic conductivity data are required to obtain a reasonable degree of confidence in simulation of site behaviour (Smith & Schwartz, 1981). In real-life, limited hydraulic conductivity measurement data are generally available as inputs to the flow and transport simulation models. This lack of data may lead to substantial non-random uncertainty in the source identification process.

In water resources systems two distinct sources of uncertainty exist. Aleatory uncertainty, which describes the inherent variation associated with the physical system or the environment under consideration which is studied by the stochastic approach. This source of uncertainty is presented as random imprecision and the mathematical representation commonly used is a probability distribution; however, not all uncertainties are random and can be objectively quantified. Besides randomness, there exists another kind of uncertainty called fuzziness or epistemic uncertainty due to

imprecise or incomplete data, and subjectivity of opinion and judgment. This non-random uncertainty can be reduced with the acquisition of additional information (Ross 2005).

The main approach in quantifying uncertainty in scientific models, for a long time, has been the probabilistic approach. This approach assumes the input parameters to be random variables having a specific probability distribution function (PDF). Thus the outcome from the model becomes a random variable having its own PDF. Various researchers have used this method to quantify uncertainty, e.g. air quality modeling (Sax & Isakov, 2003; Smith et al., 2000; Yegnan et al., 2002), surface water quality modeling (Carroll & Warwick, 2001; Warwick & Cale, 1986), and ground water quality modeling (Copty & Findikakis, 2000; Russo & Bouton, 1992).

Freeze et al. (1992) considered two sources of uncertainty in the hydrologic site investigation framework: 1) geological uncertainty; and 2) parameter uncertainty. They utilized the search theory and Bayesian updating to calculate the data worth and regret objective functions for prior and preposterior evaluations, respectively. Tiedeman and Gorelick (1993) studied the model parameter uncertainty in the design of a clean-up strategy in a contaminated system. Using nonlinear simulation-regression analysis which linked the ground water flow model to a nonlinear weighted least-square regression model, the parameter values and their uncertainty (mean, standard deviation, and coefficient of variation) were estimated. Abrishamchi et al. (2005) incorporated model parameter uncertainty in river water quality modeling using first order reliability analysis. They also determined the key factors affecting water quality modeling parameters.

Geostatistical methods also were widely used to incorporate hydrogeologic parameter uncertainty. Sequential Gaussian Simulation (SGS) is one of the geostatistical methods to simulate multivariate Gaussian random fields (Deutsch & Journel, 1998). Mugunthan and Shoemaker (2004) generated multiple realizations of equally probable stochastic hydraulic conductivity fields using SGS. Moreover, the Monte-Carlo simulation, as a stochastic approach, has been used for groundwater contamination risk assessment (Li et al., 2007). The Monte- Carlo method was used by Dokou and Pinder (2009, 2011) to incorporate hydraulic conductivity uncertainty in characterization of Dense Non-aqueous Phase Liquids (DNAPLS) sources.

Within the stochastic approach, where uncertain parameters are treated as random variables, it is necessary to obtain a PDF or a Cumulative Distribution Function (CDF) for the uncertain input parameters. First of all, an appropriate type of PDF function has to be chosen. There are quite a lot and not all will be suitable for the treated problem. Other sources of uncertainty in using probabilistic method are: lack of data; and, inaccuracy in derivation of mean, variance and (auto-) correlation structure of data (Schulz & Huwe, 1997). Variability of the random kind cannot be reduced with additional information, although it can be quantified. Besides randomness, there exists another kind of uncertainty referred to as fuzziness, epistemic, and non-random uncertainty due to imprecise or incomplete data, and subjectivity of opinion and judgment.

Parameter uncertainty is generally epistemic or non-random. This uncertainty is due to lack of knowledge about the correct values for the input parameters, and is reducible in groundwater contamination problems by acquiring more data which is a

costly and time consuming process. OHagan and Oakley (2004) argued that although probability theory is perfect, it is not adequate for characterizing non-random uncertainty. The elicitation of non-random uncertainty using PDF tools introduces more uncertainty into the system. Alternative methods have been introduced to characterize this type of uncertainty in water resources problems. Random set and fuzzy set theories and fuzzy rule-based theories are some of the approaches available for subjective knowledge characterization.

Zadeh (1965) introduced fuzzy set theory which provides a mathematical framework to deal with uncertainty that is caused by imprecise information rather than by randomness. Incorporation of fuzzy information in the geostatistic field was first introduced by Bardossy et al. (1988). Fuzzy kriging (Bardossy et al., 1990a; Bardossy et al., 1990b) and using variogram parameters and fuzzy regression techniques (Bardossy et al., 1990c) are helpful tools to overcome the problem of insufficient numbers of measurements as well as the combination of hard and soft data.

In the fuzzy uncertainty evaluation approach, membership functions are used instead of PDFs. Schulz and Huwe (1997) utilized fuzzy set theory to express uncertainty of model parameters of steady-state water flow in unsaturated soils and to incorporate this uncertainty in the modeling procedure. Dou et al. (1997) introduced uncertainty in dispersivity and flow velocity in analytical and numerical 1D and 2D solute transport process, using fuzzy membership functions.

Schulz and Huwe (1999) mapped uncertainty in soil hydraulic proprieties and boundary conditions on soil water pressure output values. This model used the fuzzy membership function concept and nonlinear optimization method to construct output

membership functions to estimate the degree of uncertainty in water pressure values in different depths in a multi-layer soil. In the work of Abebe et al. (2000), PDFs of input parameters were used by two uncertainty analysis methods, Monte-Carlo Simulation (MCS) and Fuzzy Alpha-Cut (FAC), to construct the distribution of the output (contamination concentration) and the corresponding uncertainty. Results show that in the case of output cumulative frequency distributions, the two methods are identical. While comparing PDFs, more variability in MCS and more consistency in FAC approaches were observed; however, when enough data for building PDF of uncertain parameter is not available, considering a probabilistic method, much available information may be lost too. For instance for hydraulic conductivity with a trapezoidal membership function (Figure 2.1), assuming uniform distribution function would waste some of the expert knowledge.(Schulz & Huwe, 1997).

The fuzzy rule-based approach has been used in solute transport studies (Dou et al., 1999) and for assessing an aquifers' pollution potential (Cameron & Peloso, 2001; Dixon et al., 2002). In the work of Dixon (2005), the Geographic Information System (GIS), Global Positioning System (GPS), remote sensing data and fuzzy-rule based models were incorporated to generate the sensitivity and vulnerability of contaminant potential maps. Results showed that fuzzy-rule based models are capable of producing comparable results with 40% fewer variables than crisp logic- based method.

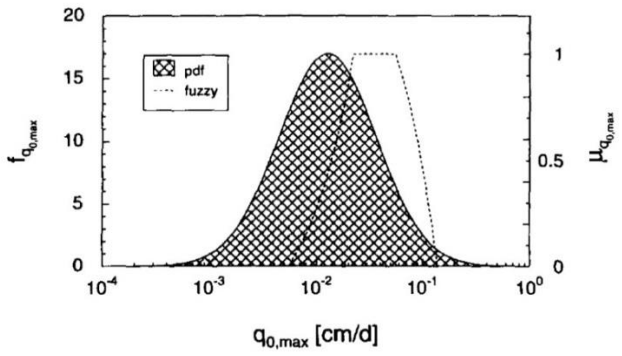


Figure 2.1 Trapezoidal fuzzy membership function compared with normal distribution function

Li et al. (2007) developed an Integrated Fuzzy-Stochastic Risk Assessment (IFSRA) approach to quantify the probabilistic and fuzzy uncertainties associated with general risk in subsurface contaminant concentration in a petroleum-contaminated site in North Canada. In this study the hydrologic input parameters were considered as probabilistic uncertain parameters. The environmental and health risk levels were assumed to be fuzzy in nature. Ayvaz (2007) identified the zone structure and associated transmissivity values for a 2D heterogeneous isotropic aquifer using Fuzzy C-Means (FCM). Yang et al. (2010) developed an Integrated Simulation-Assessment Approach (ISAA) to systematically tackle multiple uncertainties associated with hydrocarbon contaminant transport in subsurface areas and assessment of carcinogenic health risk. The fuzzy vertex analysis technique and the Latin Hypercube Sampling (LHS) based stochastic simulation approaches were combined into a fuzzy-Latin Hypercube Sampling (FLHS) simulation model. FLHS was used to predict contaminant transport in the subsurface under coupled fuzzy and stochastic uncertainties.

Many previous contaminant source identification practices do not formally address the uncertainty in the information, models and solutions, and many considered these problems as deterministic ones. In contrast to the traditional view of science where uncertainty represents an undesirable state, a state that must be avoided at all costs, engineers must accept and incorporate some level of uncertainty. One reason should be obvious: achieving high levels of accuracy costs significantly in terms of time or money, or both. Also the more complex a system is, the more uncertain or inexact is the information that is used to characterize that system. It seems intuitive that engineers should balance the degree of accuracy in a problem with the associated uncertainty in that problem.

An appropriate approach must be selected for incorporation of input parameter uncertainty in a contaminant source identification process. In the absence of adequate measurement data, the probabilistic approach is suitable for treating input parameter uncertainty (epistemic uncertainty). Utilizing fuzzy and random set theories requires reliable expert judgment to construct the input parameter membership functions; however, three challenges may limit the use of these methods. First, access to reliable sources of expert judgment is not always available, and reliability evaluation of available expert knowledge is difficult. Second, in practice the elicitation of expert judgment into membership functions by analysts adds more uncertainty. The expert true function is unknown to the analyst, and usually experts are unfamiliar with the process of representing their knowledge in the form of functions. Finally, many of the challenging problems feature information from several experts that might be

contradictory in places. Combining several judgments and eliciting a consensus from them can add more uncertainty (OHagan & Oakley, 2004).

In most of the available studies, different aspects of model parameter uncertainty are analyzed separately. Because of existing interactions among different pollutants and groundwater flow, this may cause considerable bias in model calibration and groundwater flow and contaminant transport simulations. Recently some researchers have paid great attention to simultaneous analysis of different sources of uncertainty in modeling and simulation. Markov chain Monte-Carlo analysis is one of the probabilistic methods used to consider different sources of uncertainty and their interactions. Improvements in Markov chain Monte-Carlo methods for posterior exploration include the use of several chains to better sample the full posterior distribution and evolutionary algorithms that include some degree of selection among chains. Shojaei et al. (2015) simultaneously analyze the uncertainties in inputs and parameters of a river water quality simulation model using Markov chain Monte-Carlo method. They show the interactions among the input parameters should be taken into account for uncertainty analysis of water quality simulations.

In groundwater problems, model parameters are measured at a limited number of locations of the aquifer study area and an interpolation method is required to estimate the parameters for the entire study area. Selection of appropriate interpolation methods for the conversion of discrete samples into continuous maps is a controversial issue in environmental research and is one of the sources of uncertainty in groundwater flow and contaminant simulations. Mirzaei and Sakizadeh (2016) analyzed the suitability of three interpolation methods for the discrimination of groundwater with respect to the

Water Quality Index (WQI). Three spatial interpolation methods including Ordinary Kriging (OK), Empirical Bayesian Kriging (EBK), and Inverse Distance Weighting (IDW) were compared for modeling the groundwater contamination. The selection of a suitable interpolation technique, and interpolation parameters, are problem specific and require special attention to reduce uncertainty.

2. 5. Measurement error/ uncertainty

The characteristics of contamination sources, including location, duration of activity and release fluxes, are identified using available sparse and often erroneous contaminant concentration measurements collected in the field. Generally, for the purpose of source characterization, the simulated concentrations (using candidate or possible source characteristics) are compared with uncertain field measurements to obtain the best fit. Therefore, the reliability of the information required for accurate estimation of unknown contaminant source characteristics becomes questionable.

The analysis and incorporation of uncertainty is gaining extensive attention, since uncertainties can affect decisions made regarding policies, management, regulations and program evaluations in Hydrogeologic and Water Quality (H/WQ) models (Harmel & Smith, 2007). Measurement uncertainty or error is one of the sources of uncertainty associated with H/WQ models. The measurement error/uncertainty relates to the errors associated with the information collected from the field (Beven, 2006).

The measurement data are used for calibration, validation and evaluation of model performance. The presence of measurement uncertainty has been commonly acknowledged in the evaluation of model performance. Common sources of errors in

measuring streamflow and water quality data in small watersheds were studied by Harmel et al. (2006). Various field-scale methods have been proposed in the literature for reduction and quantification of measurement uncertainty related to flow and transport characteristics, such as Integral Pumping Tests (IPTs) (Jarsjö et al., 2005); however, errors inherently exist in measured data even when strict assurance and quality control guidelines are followed (Beven et al., 2012). In groundwater contamination problems, measurement uncertainty may exist in collecting the field parameter values or observed contaminant concentrations.

The goodness of fit for calibrated H/WQ models is estimated using various coefficients of efficiency indicators such as Nash-Sutcliff coefficient, index of agreement, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Each indicator is designed to measure particular properties of the fitted H/WQ models (Harmel & Smith, 2007). Therefore, in groundwater contaminant source identification problems, it is important to select the appropriate goodness of fit indicator, to be able to find accurate source characteristics. Mahar and Datta (2001) found optimal source characteristics by minimizing the difference between estimated and observed temporal and spatial contaminant concentrations collected at monitoring locations. Their selected objective function, which was analogous to the goodness of fit indicator, was normalized using measured concentration values. Many subsequent studies have utilized the same objective function to find optimal source characteristics (Amirabdollahian & Datta, 2014; Jha & Datta, 2011; Mahar & Datta, 2001; Singh & Datta, 2006). Sum of the squared differences between the estimated and observed concentrations, and contamination mass estimation error, can be used as goodness of fit

indicators (Kollat & Reed, 2007a). The first is the non-normalized version of the one used by Mahar and Datta (2001), and the second involves comparing the total mass release by candidate source characteristics and the total mass estimated by observed concentrations at monitoring locations. Another strategy involves finding the degree of similarity between the estimated and measured concentration plumes (Dokou & Pinder, 2009). In this method, the observed contaminant concentrations are interpolated and/or extrapolated to find the observed contaminant plume, and then it is compared with the simulated plume obtained using candidate source characteristics.

Jha and Datta (2015b) use Dynamic Time Warping (DTW) distance as an objective function in the linked simulation–optimization model to design a monitoring network to efficiently estimate contaminant source characteristics. Using DTW, dissimilarity or distance between a test pattern and a set of reference patterns was used as a measure of comparison.

In addition to goodness of fit indicator, the utilized search algorithm capabilities have substantial effect on the effectiveness of the source identification procedure. In the previous studies various optimization methods/algorithms have been applied, such as: Nonlinear optimization model (NLP2) (Mahar & Datta, 2001), Genetic Algorithm (GA) (Singh & Datta, 2006), Simulated Annealing (SA) (Jha & Datta, 2011), and Adaptive Simulated Annealing (ASA) (Amirabdollahian & Datta, 2014; Jha & Datta, 2013).

3. HYDROGEOLOGICAL UNCERTAINTY QUANTIFICATION IN THE GROUNDWATER SOURCE IDENTIFICATION PROCEDURE

3. 1. Introduction

The contents of this chapter have been submitted for review and possible publication in *Water Resources Management* journal. Part of the literature review in general relevant to this chapter has been covered in Chapter 2 under the general heading of Identification of Contaminant Source Characteristics and Parameter uUcertainty in Groundwater Systems.

Effective groundwater pollution management and remediation depends on accurate and realistic characterization of the unknown groundwater pollution sources. This remains a challenging task. One issue which needs attention is the uncertainties associated with the source characterization process. The characteristics of contaminant sources are identified using available field hydrogeologic parameter values and sparse and sometimes erroneous contaminant concentration measurements collected at the monitoring locations (Chadalavada & Datta, 2008; Datta et al., 2013). In this chapter, a new methodology based on uncertainty quantification is proposed to incorporate the uncertainty in available field hydraulic conductivity values in the contaminant source identification process.

In the linked simulation-optimization approach, the simulation models (flow and transport) are linked externally to the optimization model. Solution of flow and transport processes requires knowledge of various soil hydraulic parameters including

hydraulic conductivity. Spatial variations in hydraulic conductivity play a critical role in transport of contaminants in groundwater systems, and heterogeneity controls the spread of the contaminant plume. Considerable hydraulic conductivity data are required to obtain a reasonable degree of confidence in simulation of site behaviour (Smith & Schwartz 1981). In real life, limited hydraulic conductivity measurement data are generally available as input to the flow and transport simulation models. This lack of data may lead to substantial non-random uncertainty in the source identification process.

An appropriate approach must be selected for incorporation of input parameter uncertainty in contaminant source identification process. Therefore, in this study a new methodology is introduced to quantify non-random uncertainty in a hydraulic conductivity field using sparse available data without fitting a PDF, or incorporating subjective expert judgment. A new parameter (Coefficient of Confidence) is introduced to represent quantification of parameter uncertainty based on random field realizations. Although such quantification cannot be designated as fuzzy quantification, in a strict sense, it differs from the crisp approach generally adopted in earlier developed methodologies for source characterization. This approach can eliminate the possibility of incorporating additional uncertainties due to inappropriate use of so-called expert judgments.

The aim of this research is to present a new methodology to quantify existing uncertainty in the available hydraulic conductivity data. The results of the quantification process are to be incorporated into the linked simulation-optimization source identification process to improve the accuracy of the estimated contaminant

source characteristics. Throughout this chapter, “uncertainty” refers to the “non-random or epistemic uncertainty”.

In this study, first, the ASA optimization technique is linked to the flow and transport simulation models. Then a progressive method is introduced to quantify the level of confidence for each simulated concentration value based on the uncertainty in the available hydraulic conductivity values. An illustrative application of the proposed methodology using a heterogonous flow field is presented. The performance of the proposed methodology is also evaluated incorporating contaminant measurement error, to establish the potential applicability of the proposed methodology. Then, the reliability of contaminant source identification procedure is discussed. Finally, the proposed methodology is tested in the light of the SGS method, which is commonly used in uncertain groundwater contamination problems. Some of the issues associated with the SGS approach, in the context of this chapter, are also discussed.

3. 2. Contaminant source identification methodology

The pollutant source characteristics that must be addressed in an identification procedure are: (1) source locations; (2) source release history (initiation and duration); and, (3) source flux magnitudes. The proposed methodology has two major components: Linked simulation-optimization model, and uncertainty quantification module. These components are described below.

3.2 .1 . Linked simulation-optimization model

The linked simulation-optimization model is solved using ASA algorithm. SA approaches the optimization model like a bouncing ball which bounces over mountains

from valley to valley. The SA controlling parameter is Temperature (T), which mimics the effect of fast-moving particles in a hot object like molten metal. The initial high T value allows the ball to bounce over all mountains to access any valley. Like the cooling process of metal, as the T decreases and reaches relatively colder states, the height of the ball bounce decreases and it settles gradually in the deepest valley. To reach the optimal solution, many parameters must be tuned, such as probability density function, acceptance probability density function and the reannealing temperature schedule. ASA is a variant of SA in which the automated reannealing temperature schedule and random step selection make the algorithm less sensitive to the user-defined parameters. Using ASA also eliminates the need for several trial executions of the model, to adjust the optimization parameters (Ingber, 1996).

For completeness, we briefly describe the formulation of the linked-simulation optimization source identification framework. More details can be found in Mahar and Datta (2001). Based on the available background information about the site, the set of potential contaminant source locations is assumed to be known. The optimization model estimates the optimal contaminant fluxes associated with each potential source location at each stress period. The objective function minimizes the weighted sum of normalized differences between temporal and spatial observed and simulated concentrations subject to upper and lower bounds on source fluxes. The optimal source identification model is defined by the objective function 3-1, subject to the constraints 3-2 and 3-3.

$$Min\,F = \sum_{k=1}^{nk} \sum_{iob=1}^{nob} \mu_{iob}^{k} (Cest_{iob}^{k} - Cobs_{iob}^{k})^2 / (Cobs_{iob}^{k} + \alpha)^2 \qquad (3\text{-}1)$$

Subject to:

$$Cest = f(D, HC, \theta, x_i, y_i, z_i, q_i) \quad i = 1, ..., N \tag{3-2}$$

$$0 \le q_i \le q_{\max} \quad i = 1, ..., N \tag{3-3}$$

Where *nk*, *nob* and *N* are the total number of concentration observation time periods, available monitoring locations and candidate source locations, respectively. $Cest_{iob}^{k}$ and $Cobs_{iob}^{k}$ are the concentrations estimated by the simulation model and observed concentrations at observation location *iob* and at the end of time period *k*, respectively. *D*, *HC* and θ are the dispersion coefficient, hydraulic conductivity and porosity, respectively. x_i, y_i, z_i and q_i are the Cartesian coordinates of candidate contaminant source *i*, and the contaminant release flux for candidate location *i*, respectively. $q_{\max}$ is the upper bound for contaminant release fluxes. $\mu_{\text{iob}}^{\text{k}}$ is the Coefficient of Confidence (COC) assigned to monitoring location *iob* at the end of time period *k*. COC is estimated based on the level of uncertainty in the hydraulic conductivity field.

The constraint set 3-2 represents the flow and contaminant transport simulation models and it couples the simulation model and optimization algorithm. Eq. 3-3 limits the candidate contaminant flux values, at each potential location, to an upper bound. In Eq. 3-1, α is a constant which is sufficiently large, to prevent any individual term in Eq. 3-1 becoming indeterminate due to the observed value of concentration becoming very small. Adding this parameter variable also prevents domination of the obtained solution by deviation between measured and simulated concentrations corresponding to low concentration measurement values (Mahar & Datta, 2001). Figure 3.1 shows a

schematic representation of the linked simulation-optimization source identification process using an uncertainty-based ASA optimization algorithm.

3.2.1. 1 . Groundwater Flow and Transport Simulation Models

Groundwater Flow simulation model used in this study is MODFLOW-2000 (Zheng et al. 2001). MODFLOW is a computer program that numerically solves the three-dimensional ground water flow equation for a porous medium by using a finite-difference method. The partial differential equation for transient groundwater flow utilized in MODFLOW is given by the following equation (Zheng & Bennett 2002):

$$\frac{\partial}{\partial x}(K_{xx}\frac{\partial h}{\partial x})+\frac{\partial}{\partial y}(K_{yy}\frac{\partial h}{\partial y})+\frac{\partial}{\partial z}(K_{zz}\frac{\partial h}{\partial z})+W=S_s\frac{\partial h}{\partial t} \quad (3\text{-}4)$$

where K_{xx}, K_{yy}, and K_{zz} are the hydraulic conductivities (L/T) along the x, y, and z coordinate axes which are assumed to be parallel to the principal axes of hydraulic conductivity, respectively. H, S_S and t are the potentiometric head (L), the specific storage of the porous material (L^{-1}), and time, respectively. W is the volumetric flux per unit volume representing sources and/or sinks of water, where $W<0$ for flow moving out of the ground-water system, and $W>0$ for flow moving in (T^{-1}). When combined with boundary and initial conditions, Eq. 3-4 describes transient three-dimensional groundwater flow in a heterogeneous and anisotropic medium, assuming the principal axes of hydraulic conductivity are aligned with the coordinate directions.

The contaminant transport simulation model which is used in this study is chosen as MT3DMS (Zheng & Wang 1999). The partial differential equation describing three-dimensional transport of contaminants in groundwater can be written as follows (Zheng & Bennett 2002):

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x_j}(D_{j,k}\frac{\partial C}{\partial x_k}) - \frac{\partial}{\partial x_j}(u_j C) + \sum_{i=1}^{N}\frac{q_i}{\theta} + \sum_{p=1}^{NR} R_p \quad (3\text{-}5)$$

where C is the solute concentration in groundwater (ML^{-3}); t is the time (T); j, k are the Cartesian coordinates along axes; u_j is the groundwater velocity in three dimensions (LT^{-1}); $D_{j,k}$ is the dispersion coefficient tensor (L^2T^{-1}); q_i is the flux of contaminant concentration for source i (MT^{-1}); and $\sum_{p=1}^{NR} R_p$ are chemical reaction terms ($ML^{-3}T^{-1}$).

The MT3DMS transport model uses a mixed Eulerian-Lagrangian approach to the solution of the three-dimensional advective-dispersive-reactive equation (Zheng & Wang 1999). The groundwater velocity values (u_j), estimated by the flow model, are used by the transport model to estimate concentration values. The estimated concentrations (C) are transferred to the optimization model to evaluate the objective function value.

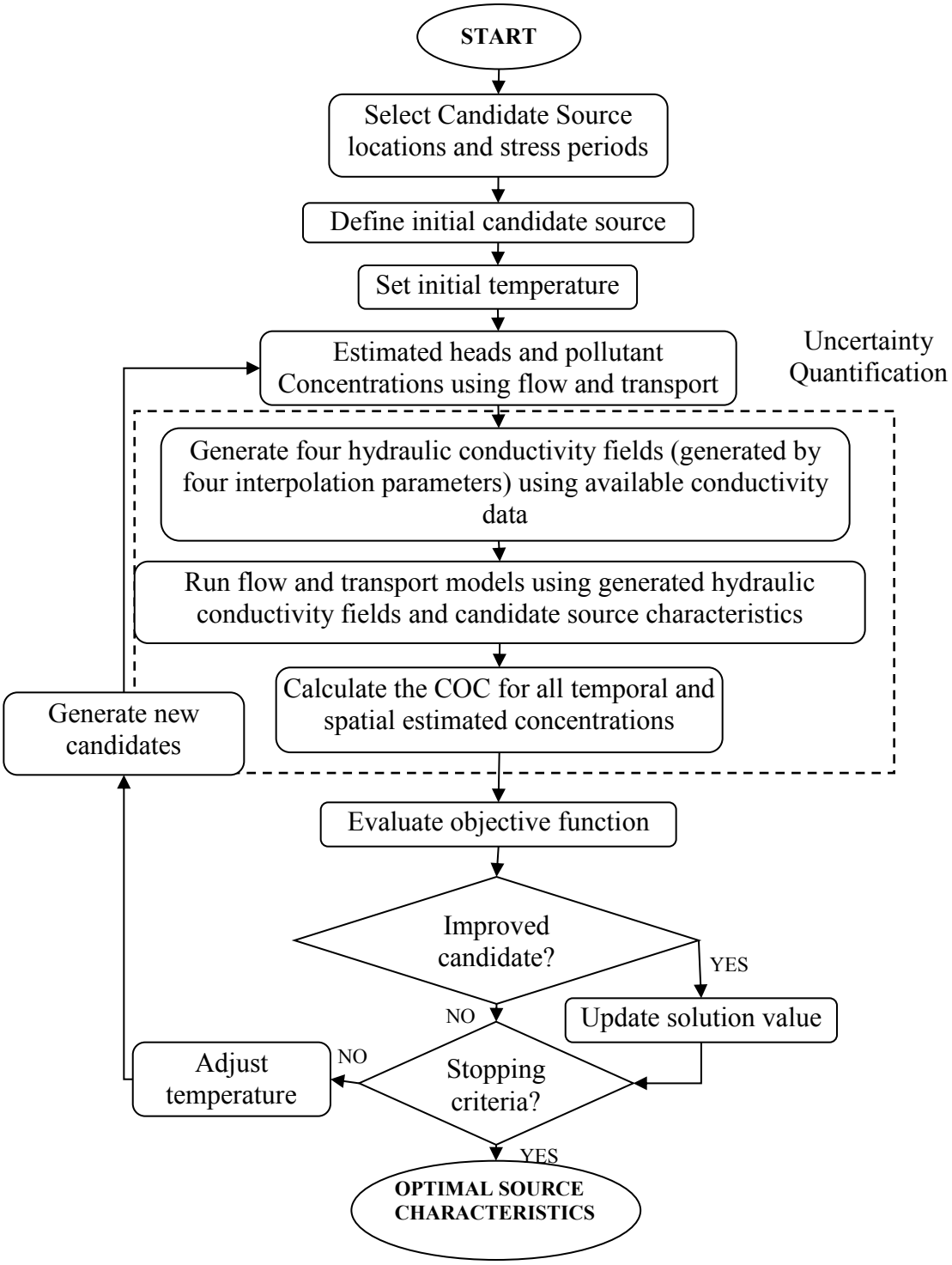


Figure 3.1 Schematic diagram of uncertainty-based contaminant source identification methodology

3.2 .2 . Uncertainty quantification

Generally, a predesigned or arbitrary sampling network is utilized to collect field hydrogeologic data for groundwater management purposes. The constructed contaminant source histories are sensitive to hydraulic conductivity values (Datta et al., 2009) and spatial variation in hydraulic conductivity is a critical factor controlling contaminant mass transport process (Smith & Schwartz, 1980). Confidence in predicted spatial and temporal contaminant concentration can be evaluated considering this variability. In real-life scenarios, there may be no other option except to use the available sparse hydraulic conductivity data. Therefore, to describe mass transport adequately, this spatial and temporal uncertainty must be quantified. Smith and Schwartz (1981) suggested that there is a complex relationship between the uncertainty in transport predictions, and the number of available hydraulic conductivity measurements. In order to quantify uncertainty in predicted spatial and temporal contaminant concentrations, the following items must be addressed.

1- Number and spatial distribution of available field measurements;
2- The hydraulic conductivity field variability and spatial correlation between available measurements;
3- The flow field properties including boundary and initial conditions and properties of available sinks and sources;
4- The location, initial time, activity duration, and fluxes of contaminant sources.

In contaminant source identification procedure, the characteristics of contaminant sources are unknown (item 4 in the above list). Therefore, a coupled procedure is required to simultaneously quantify the uncertainty in contaminant concentration

predictions, and identify sources of contamination. In the proposed methodology, as shown in Fig. 3.1, the computation of the COC values relating to each estimated spatial and temporal contaminant concentration is incorporated in the optimal source identification model.

3.2.2. 1 . Interpolation Technique

The MODFLOW simulation model numerically simulates flow process in the aquifer. In this simulation the aquifer study area is discretized into cells, and a hydraulic conductivity value must be specified for each cell; however, the hydraulic conductivity measurements are available at a finite number of locations only. Therefore, an interpolation technique is required to interpolate available measurement values through the entire aquifer.

The Inverse Distance Weighting (IDW) method is utilized in this study for interpolation of available hydraulic conductivity measurements. Many available interpolation techniques require prior information about the statistical properties of data. For instance, Kriging requires a carefully selected sample variogram and appropriate log-transformation of the data. To acquire accurate statistical properties of data, a substantial number of measurements is required. In this study the assumption is that the number of available hydraulic conductivity measurements is limited and not adequate to accurately estimate the statistical properties of hydraulic conductivity distribution for the entire aquifer. Therefore, using statistical interpolation methods with limited amounts of available data may lead to an inaccurate interpolated field.

In the IDW method, the value of variable Z at an unsampled location x_0, $Z^*(x_0)$, is estimated based on data from the surrounding locations, $Z(x_i)$, as Eq. 3-6.

$$Z^*(x_0) = \sum_{i=1}^{n} w_i Z(x_i) \tag{3-6}$$

where w_i are the weights related to each $Z(x_i)$ value and n is the number of closest sampled data points used for interpolation purposes. The weights are estimated using Eq. 3-7.

$$w_i = \frac{1/d_i^p}{\sum_{i=1}^{n} 1/d_i^p} \tag{3-7}$$

where d_i is the distance between the estimated point and the sample point, and p is the exponent parameter. Prior decisions must be made about the optimal exponent, and the number of closest neighbours is to be specified in the interpolation. Using the IDW method facilitates interpolation of the hydraulic conductivity field, specifically if limited measurement data are available. Cross-validation is used to estimate the optimal exponent value. Each value from the measurement data set is eliminated in turn, and estimated utilizing information regarding the rest of the data (Goovaerts, 1997). The exponent value that results in minimum discrepancy between available measurement data and the interpolated values at corresponding locations is selected as the optimal exponent value.

The IDW method uses a predefined number of closest sample points for estimation. The weightage assigned to each sample point is proportional to the distance between the estimated point and sample points. In this way the quantity and spatial

variability of available sample data are considered in the interpolation method without any prejudgment about the spatial-correlation between sampled data. To select suitable number of closest points (n), spatial structure of data is considered. The spatial auto-correlation of available measurement data is the correlation among values of single variables attributable to their relatively close locational positions. *Moran's I* (Moran, 1950) is an indicator of spatial auto-correlation that is applied to zones or points with continuous variable associated to them. It compares the value of the variables at any location with the value at other locations using Eq. 3-8.

$$I = \frac{N\sum_{i=1}^{m}\sum_{j=1}^{m}W_{i,j}(X_i - \bar{X})(X_j - \bar{X})}{(\sum_{i=1}^{m}\sum_{j=1}^{m}W_{i,j})(\sum_{i=1}^{m}(X_i - \bar{X})^2} \tag{3-8}$$

where X_i is the variable at the particular location. X_j, $\bar{X}$ and m are the variable values at another location, the mean of all available data, and total number of available data, respectively. $W_{i,j}$ is the weight applied based on distance between locations i and j and is estimated using Eq. 3-9.

$$W_{i,j} = \begin{cases} 1 & if\ i = j \\ 1/d_{ij} & if\ i \neq j \end{cases} \tag{3-9}$$

where d_{ij} is the distance between points i and j. The spatial auto-correlation varies between the limits, -1.0 and +1.0. If nearby areas are more alike, the spatial auto-correlation is positive. Negative autocorrelation, on the other hand, describes patterns in which neighbouring areas are not alike, and random patterns exhibits zero spatial auto-correlation. Therefore, when *Moran's I* is large, the suitable number of closest points utilized in IDW interpolation would be small; however, for field data with a small value

of spatial auto-correlation, a high value of *n* needs to be considered to maintain randomness.

3.2.2. 2 . Coefficient of Confidence (COC)

In the crisp contaminant source identification model, the model parameters are generally assumed to be known without any associated error (Jha & Datta, 2012; Sun, 1994). In field applications, however, uncertainty is associated with the hydrogeologic parameter values. Even with the idealized assumption that the numerical groundwater flow and contaminant transport simulation models are able to provide accurate solutions of the equations governing the physical processes, uncertain hydrogeologic input values result in inaccurate concentration predictions.

In the uncertainty-based source identification model, Eq. 3-1 shows the formulation of objective function where $\mu_{\text{iob}}^{\text{k}}$ is the COC. It represents the degree of confidence assigned to the predicted (estimated) contaminant concentration at each monitoring location, at each time period. In crisp formulation all monitoring locations and time periods are treated equally, therefore the μ_{iob}^{k} values correspond to a value of one; however, in the uncertainty-based formulation, less weight is assigned to locations or time steps at which the predicted concentration is less accurate due to uncertainty in the hydraulic conductivity field. Eq. 3-10 defines the coefficient of uncertainty for monitoring location *iob* at time period *k*.

$$F_{iob}^{k} = \left| \frac{Cest_{iob}^{k} - Cestn_{iob}^{k}}{Cest_{iob}^{k} + nn} \right| \qquad (3\text{-}10)$$

$$where \quad Cestn_{iob}^{k} = average(Cest_{iob,r}^{k} = f(D, K_r, \theta, x_i, y_i, z_i, q_i) \quad i = 1,...,N,\ r = 1,...,R)$$

where $Cest_{iob}^{k}$ is the concentration estimated by the simulation model at monitoring location *iob* and at time period *k*. This is the same value utilized in the objective function (Eq. 3-1) as the predicted concentration. At this stage, the hydraulic conductivity field generated by the IDW interpolation method with the *n* number of closest points is utilized in the flow simulation model. *n* is selected based on the estimated sample spatial auto-correlation value (Eq. 3-8). $Cestn_{iob}^{k}$ is the average concentration estimated at the same location *iob* and time period *k*. For the purpose of estimating $Cestn_{iob}^{k}$, R realizations of the hydraulic conductivity field are considered. Each field realization (K_r) is generated utilizing the IDW interpolation algorithm and n_r number of closest neighbours included in the interpolation ($n_r<n$). Then, the flow and transport models utilize the generated hydraulic conductivity realization to estimate $Cest_{iob,r}^{k}$. The variable *nn* is specified as a small numerical value which prevents division by zero at the locations where the estimated concentration is zero or very near to zero.

In hydraulic conductivity fields with high values of spatial auto-correlation, the selected number of closest points (n) is small. Therefore, the field realizations utilized in Eq. 3-10 will be nearly identical. In random fields with low spatial auto-correlation, the coefficient of uncertainty is able to adequately quantify the uncertainty in predicted concentrations. The COC (μ_{iob}^{k}) is estimated as follows.

$$\mu_{iob}^{k} = \begin{cases} 0.1 & if\ (1 - U * F_{iob}^{k} \le 0.1) \\ 1 - U * F_{iob}^{k} & else \end{cases} \tag{3-11}$$

where U is a normalizing factor selected based on trial and error. The U value is selected in such a way that the estimated COC values will be distributed over the range [0.1,1]. The quantity $1-U*F_{iob}^{k}$ would be very small or negative when the coefficient of uncertainty is very large. Since, it is useful to use all available measurement data for the purpose of characterizing sources, the minimum COC is considered to be 0.1; however, other suitable values may be specified.

3.2.2. 3 . Reliability of Source Identification Results

The reliability of the identified source characteristics is an important issue that needs to be addressed for estimations with uncertainty. The reliability of an uncertainty-based model can be estimated using the calculated COC values. Using the identified optimal contaminant source fluxes obtained as solutions, the COC values corresponding to available contaminant monitoring data are estimated. The larger the COC value, the smaller the uncertainty level. Therefore, reliability in an ordinal scale can be estimated using obtained COC values. Eqs. 3-12, 3-13 and 3-14 are utilized to estimate this reliability.

$$\mathrm{Re}liability = 0.5(\eta_1 + \eta_2) \tag{3-12}$$

where

$$F(\eta_1) = p(COC \le \eta_1) = \varphi_1 \quad ; \quad \eta_1 = F^{-1}(\varphi_1) \tag{3-13}$$

$$F(\eta_2) = p(COC \le \eta_2) = \varphi_2 \quad ; \quad \eta_2 = F^{-1}(\varphi_2) \tag{3-14}$$

where F represents the CDF and p is the PDF of estimated COC values. Φ_1 and Φ_2 are values in the range (0.5,1). $p(COC \le \eta_1)$ and $p(COC \le \eta_2)$ represent the probabilities that the COC value is less than or equal to η_1 and η_2, respectively.

The reliability value represents the proportion of contaminant monitoring data that has a higher contribution to the identification of contaminant source characteristics in relation to all available observed contaminant data.

3. 3. Performance evaluation

The performance of the proposed methodology is evaluated for a three-dimensional illustrative groundwater aquifer study area shown in Fig. 3.2. The study area is 1500 m by 1000 m and the aquifer depth is 30 m. It is discretized into 30 rows, 20 columns and two layers. The top, bottom and left side boundaries have a specified head, and the right-hand side boundary has variable head boundary conditions. The triangular signs show the location of active extraction wells (sinks). The candidate contaminant source locations are shown by square signs. Two of them are actual and one is a dummy. The numbers show the location of nine monitoring wells. It is assumed that the contaminant fluxes are constant in every stress period. The study period is divided in to five stress periods. Table 3.1 shows the length of stress periods and the extraction wells and contaminant sources' properties.

For this evaluation purpose, it is assumed that all the aquifer hydrogeologic parameter values are known without any error except hydraulic conductivity (K). The aquifer porosity, longitudinal dispersivity, ratios of horizontal to longitudinal and vertical to longitudinal dispersivity, and specific storage are 0.25, 35 (m), 0.1, 0.01, and 0.2 (m^{-1}), respectively.

3.3 .1 . Hydraulic conductivity field

Performance of the developed methodology is evaluated using synthetic hydraulic conductivity data. The advantage of using synthetic data is that the actual source characteristics and hydrogeologic data are known for evaluation purpose, which allows for reliable testing of uncertainty-based source identification methodology.

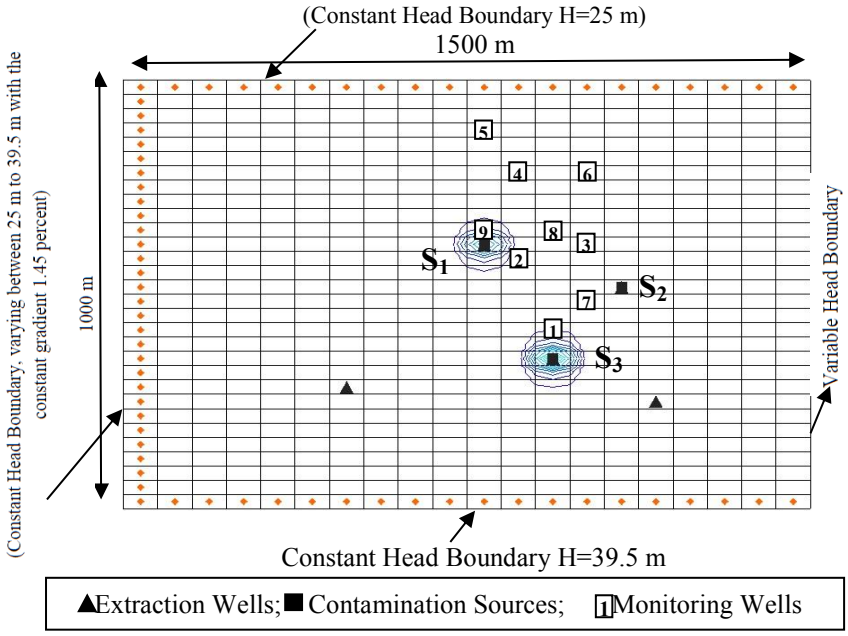


Figure 3.2 Plan view of the illustrative groundwater aquifer study area

This information is, however, not known to the source identification model. First, the actual hydraulic conductivity field is generated, which represents the real field with no associated hydrogeologic uncertainty. Second, a limited number of hydraulic conductivity measurement points are selected and utilized to generate the uncertain hydraulic conductivity field.

Table 3.1 Characteristics of the Contamination Sources and extraction wells.

	Location			Stress Period				
	Row	**Column**	**Layer**	**1** **183 days**	**2** **183 days**	**3** **183 days**	**4** **183 days**	**5** **2196 days**
Contamination Source	12	11	1	70	90	35	20	20
	15	15	1	Dummy Source				
Flux (kg/day)	20	13	1	95	85	75	50	0
Extraction well	22	7	1	100				
Flow rate (L/day)	23	16	1	500				

3.3.1. 1 . Actual Flow Field

A realistic presentation of the porous medium can include a hydraulic conductivity field distributed as a Log-normal function through space (Freeze, 1975). The methodology performance evaluation is carried out using randomly generated hydraulic conductivity values throughout the study area. The study area is divided into grids, each 50 m by 50 m for the purpose of specifying hydraulic conductivity values (including points located on the boundaries). For each location, two *HC* values corresponding to 10 m and 20 m depth are generated to produce a three-dimensional model. Truncated Latin Hypercube Sampling (LHS) is utilized to produce more efficient estimates than those obtained from random sampling of the distribution function. In the LHS, the probability distribution function is divided into non-overlapping, equal-probability intervals.

The sample is taken from each interval and permuted in a way that the correlation of the field is accurately presented (Dokou & Pinder, 2009). To be more realistic and avoid the values that are located at the tails of the frequency distribution function, the sampling is truncated to the values which are in the $(0.6\,\mu_Y,\ 1.4\,\mu_Y)$ range. μ_Y is the

mean value. For the values located at a depth of 10 m, the mean and standard deviation of the log-normal distribution function are 20 m/day and 15% of the mean (3 m/day), respectively. The log-normal distribution function utilized for 20 m depth has a mean value of 15 m/day, and the standard deviation is 0.15 times the mean value (2.25 m/day). The actual hydraulic conductivity field is generated using Ordinary Kriging (OK) interpolation technique (Cressie, 2015). Fig. 3.3, shows the *HC* values for the aquifer first layer.

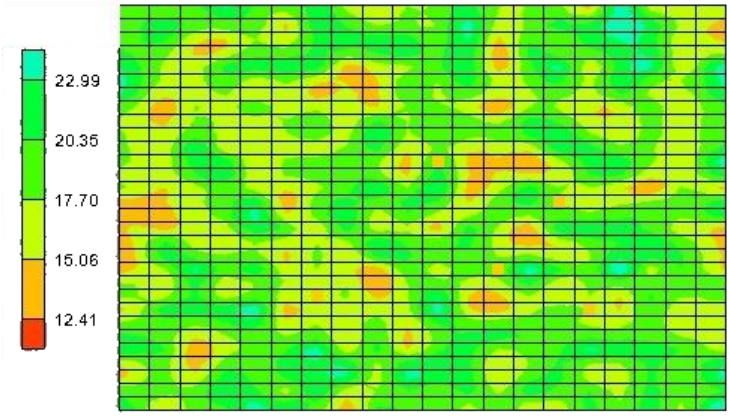


Figure 3.3 The actual hydraulic conductivity field, first layer (unit is m/day).

3.3.1. 2 . Uncertain Flow Field

In most field applications, a limited number of measurements are available. In this illustrative example, to show the uncertain flow field, hydraulic conductivity values are assumed known at locations along the horizontal and vertical axes, at intervals of 300 m and 250 m (including locations situated on the boundaries) respectively, and at depths of 10 m and 20 m. These 60 locations represent actual sampling points. The corresponding *HC* value is assigned to these points using the actual hydraulic conductivity field generated in the previous section. Reducing the number of points

with known *HC* values by 95 percent, the available hydraulic conductivity field becomes uncertain. Using the cross-validation technique, the optimal IDW exponent value is 2. The low spatial auto-correlation (I = 0.052) shows a lack of autocorrelation and randomness in the available measurement data. Therefore, a large number of closest points must be utilized in the IDW method. Fig. 3.4 shows the interpolated hydraulic conductivity field for the first layer using IDW method and 60 closest points (n=60).

To identify the characteristics of contaminant sources, nine contaminant concentration monitoring locations are selected. Only for performance evaluation purposes, monitoring concentration data are generated using the flow and transport simulation model and the actual hydraulic conductivity field (Fig. 3.3) to represent the distribution of the actual contamination plume; however, in field applications these observed concentration measurement data are to be collected from the site.

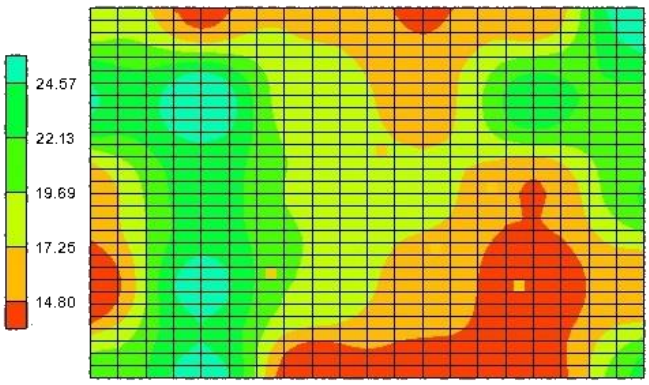


Figure 3.4 The uncertain hydraulic conductivity field generated by limited number of measurement data, first layer (unit is m/day).

For objective function evaluation (Eq. 3-1), the flow simulation model has only limited hydraulic conductivity data and therefore, the uncertain flow field (Fig. 3.4) is

utilized. In order to estimate the coefficient of uncertainty and COC assigned to each spatial and temporal monitoring datum, four realizations of the hydraulic conductivity field are generated using 15 (25%), 30 (50%), 45 (75%) and 60 (100%) values as the number of closest neighbors included in the IDW interpolation.

3. 4. Results and discussion

The characteristics of contaminant sources are identified using both crisp and uncertainty-based models, for comparison purposes. Tables 3.2-3.4 show some of the solution results. In the following tables, for presentation purpose only, the uncertainty-based models utilizing the uncertainty quantification parameter (COC) are designated 'Fuzzy model' to distinguish it from the Crisp model (without any quantification of uncertainty). As shown in Table 3.3, both uncertainty-based and crisp contaminant source identification models are able to identify the location of the dummy contaminant source by estimating zero flux values for all stress periods. To examine the capability of both models in terms of accuracy in estimating source fluxes, the Normalized Absolute Error of Estimation (*NAEE%*) is estimated using Eq. 3-15 (Jha & Datta, 2013).

$$NAEE(\%) = \frac{\sum_{i=1}^{N} \left| q_{est}^{i} - q_{act}^{i} \right|}{\sum_{i=1}^{N} q_{act}^{i}} \times 100 \qquad (3\text{-}15)$$

where N is the number of stress periods and q_{est}^{i} and q_{act}^{i} are the estimated and actual source fluxes for stress period i, respectively. Table 3.5 shows the *NAEE%* values obtained for sources 1 and 3. The uncertainty-based model improves the accuracy of results at both source locations 1 and 3 by 7.9 and 48.8 percent, respectively. Utilizing

the uncertainty-based model, substantial improvement is observed at source location 3, where inadequate hydraulic conductivity information leads to approximately 22 percent error in identified source characteristics when using the crisp model. Table 3.6 shows the estimated COC values corresponding to nine monitoring locations at three sampling times, 366, 732 and 1830 days, utilizing the final uncertainty-based model estimated source fluxes. The utilized *U* value (Eq. 3-11) is 20.

Table 3.2 Characterization of the Contamination Source 1

Contamination Flux (kg/day)	**Stress Period**				
	1	2	3	4	5
Actual	70.0	90.0	35.0	20.0	20.0
Estimated using Crisp Model	77.4	89.8	31.2	20.9	18.4
Estimated using Fuzzy Model	76.2	89.9	30.6	19.8	18.1

Table 3.3 Characterization of the Contamination Source 2

Contamination Flux (kg/day)	**Stress Period**				
	1	2	3	4	5
Actual	0.0	0.0	0.0	0.0	0.0
Estimated using Crisp Model	0.0	0.0	0.0	0.1	0.0
Estimated using Fuzzy Model	0.0	0.0	0.0	0.0	0.0

Table 3.4 Characterization of the Contamination Source 3

Contamination Flux (kg/day)	Stress Period				
	1	2	3	4	5
Actual	95.0	85.0	75.0	50.0	0.0
Estimated using Crisp Model	89.9	61.6	54.0	32.1	0.0
Estimated using Fuzzy Model	90.0	73.6	72.1	34.8	0.0

Table 3.5 Normalized Absolute Error of Estimation (%NAEE) and improvement

	Crisp Model	**Fuzzy Model**	**Improvement %**
Source 1	5.9	5.4	7.9
Source 3	22.1	11.3	48.8

Table 3.6 Coefficient of Confidence (COC)

Monitoring Location	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
366 day	0.83	0.98	0.71	0.72	0.83	0.48	0.47	0.72	0.90
732 day	0.71	0.90	0.10	0.78	0.82	0.44	0.10	0.92	0.83
1830 day	0.10	0.10	0.10	0.10	0.92	0.10	0.10	0.10	0.88

Table 3.6 shows that it is difficult to identify a pattern in the assigned COC values based on the relation among the active source locations, contaminant source fluxes, simulation time step, and uncertain hydraulic conductivity fields. These relations appear to be complex. For example, monitoring location 9 is very near to source 1. Based on the two generated hydraulic conductivity fields (Figs. 3.3 and 3.4), the uncertainty in the vicinity of monitoring location 9 is small. Both fields show

approximately the same hydraulic conductivity value for the given distance between monitoring location 9 and contaminant source 1. Furthermore, the contaminant plume at this location is mainly resulting from source 1, and no superimposition of contamination injected by other sources has occurred at this location. Therefore, the COC assigned to this location is near to 1.00 for all time steps.

For monitoring location 1, which is close to source 3, a similar explanation for COC is valid; however, since this location is about 100 m away from source 3 and the hydraulic conductivity field is uncertain between these two locations (Figs. 3.3 and 3.4) the COCs are smaller than the values assigned to location 9. The contaminant injection from source 3 stopped after two years. Therefore the contaminant concentration at monitoring location 1, after 1830 days is very small and near to zero. To prevent domination of the model by uncertainty at this location, the computed COC (at 1830 days) is the smallest possible value (0.10).

3.4 .1 . Error in concentration measurement data

In field applications, a level of measurement error is often associated with contaminant concentrations measured at monitoring wells. To evaluate the performance of uncertainty-based source identification models for realistic scenarios, different levels of random concentration measurement errors are incorporated. Synthetically generated noise is added to the available error free monitoring data using Eq. 3-16.

$$C_{pert} = C_{obs} + S_{ud} \times \beta \times C_{obs} \tag{3-16}$$

where C_{pert}, C_{obs} and S_{ud} are perturbed concentration values, error free concentration at monitoring locations and a uniform random number between -1 and +1, respectively. β is a fraction between 0 and 1.0 and illustrates the level of measurement error.

Perturbed monitoring data, which represent the erroneous contaminant concentration measurement data are utilized by the uncertainty-based as well as the crisp models to characterize sources of contamination. Model performance is evaluated using three different levels of monitoring measurement error as low (β=0.10), moderate (β=0.15), and high (β=0.20). Five sets of statistically uniform random noises are added to error-free contaminant measurement data for each level of monitoring measurement uncertainty. Table 3.7 shows the average *NAEE%* values associated with the solution results obtained by both uncertainty-based and crisp models considering uncertain hydraulic conductivity field and erroneous monitoring data.

Table 3.7 Performance Evaluation for uncertain monitoring measurement data

NAEE(%)	Low (β=0.10)		Moderate (β=0.15)		High (β=0.20)	
	Crisp Model	**Fuzzy Model**	**Crisp Model**	**Fuzzy Model**	**Crisp Model**	**Fuzzy Model**
Source 1	19.4	14.2	19.8	20.0	22.2	23.5
Source 3	27.2	18.8	27.6	21.6	30.1	23.3

Using erroneous measurement data, both the models are able to identify the source locations accurately by specifying zero and near zero fluxes for the dummy source location. The crisp model *NAEE%* values for all three levels of measurement errors are relatively similar; however, for source 1, the uncertainty-based solution results show smaller *NAEE%* values in the case of small contaminant concentration measurement errors. Reconstructed source release histories with erroneous

measurement data illustrate that for low, moderate and high levels of measurement error, *NAEE%* associated with the source 3, obtained by using the uncertainty-based model, are smaller than the corresponding values estimated using the crisp model.

3.4 .2 . Reliability estimation

To demonstrate the reliability of estimating the unknown source characteristics, three hydraulic conductivity sampling scenarios are considered as described below.

Scenario 1- Sampling locations are specified along the horizontal and vertical axes at intervals of 600 m and 250 m respectively, and at depths of 10 m and 20 m. The points that are located on the right and left aquifer boundaries are not selected; however points on the top and bottom boundaries are included in the sampling network. Hydraulic conductivity values are specified at 20 locations.

Scenario 2- The sampling locations are specified along the horizontal and vertical axes at intervals of 300 m and 250 m respectively, and at depths of 10 m and 20 m. Hydraulic conductivity values are known at 60 locations. The locations situated on aquifer boundaries are selected. This is the same scenario discussed earlier in the performance evaluation section.

Scenario 3- The sampling locations are specified along the horizontal and vertical axes at intervals of 300 m and 200 m respectively, and at depths of 10 m and 20 m. The locations situated on the aquifer boundaries are selected. Hydraulic conductivity values are specified at 72 locations.

The contaminant source fluxes are estimated using the uncertainty-based source identification model. Selected *n* values (Eq. 3-6) for scenario 1, 2 and 3 are 20, 60 and 72, respectively. The *n* values are selected based on the computed value of *Moran's I*

(Eq. 3-8). Table 3.8 shows the estimated *NAEE*% values for scenario 1, 2 and 3 using crisp and uncertainty-based source identification models. The uncertainty-based source identification model results in improvement of the *NAEE*% values in nearly all considered scenarios. Only in scenario 3 for source 3, no improvement is achieved using an uncertainty-based model and nearly the same contaminant fluxes are estimated. In all scenarios, both models were able to correctly identify the dummy (not actual) source.

Table 3.8 Normalized Absolute Error of Estimation (*%NAEE*) for three scenarios

		Crisp Model	**Fuzzy Model**	**Improvement %**
Scenario 1	**Source 1**	7.2	6.1	17.7
	Source 3	24.1	9.6	60.1
Scenario 2	**Source 1**	5.9	5.4	7.9
	Source 3	22.1	11.3	48.8
Scenario 3	**Source 1**	6.2	5.3	13.5
	Source 3	6.5	6.6	-

Figure 3.5 shows the CDF for estimated COC values using the optimal source characteristics obtained as solutions. Notably the CDF for scenario 3 differs substantially from CDFs for scenarios 1 and 2. For example, for specific Φ_1 and Φ_2 values of 0.6 and 0.8, the η_1 and η_2 values are (0.39, 0.72, 0.87) and (0.66, 0.86, 0.95), respectively for the three scenarios. These η_1 and η_2 values are utilized in Eq. 3-12 to estimate the reliability of the source identification model results. Therefore, it is evident that scenario 3, which has the largest values, results in the largest reliability. The reliability values are 0.52, 0.79 and 0.91 for scenarios 1, 2 and 3, respectively (Eqs. 3-12, 3-14). Therefore, identified results estimated for scenario 3 have the highest

reliability and the ones estimated in scenario 1 have the lowest reliability. In real-life problems, the actual contaminant source fluxes are unknown. Therefore, the *NAEE%* values are only relevant to the performance evaluation purpose and cannot be estimated in real contamination problems where source characteristics are not known; however, the measure of reliability introduced in Eq. 3-12 is estimated using outcomes of the uncertainty-based source identification model. Thus it can be used to estimate a measure of source characterization reliability, in real-life groundwater contamination problems where the actual contaminant source characteristics are unknown.

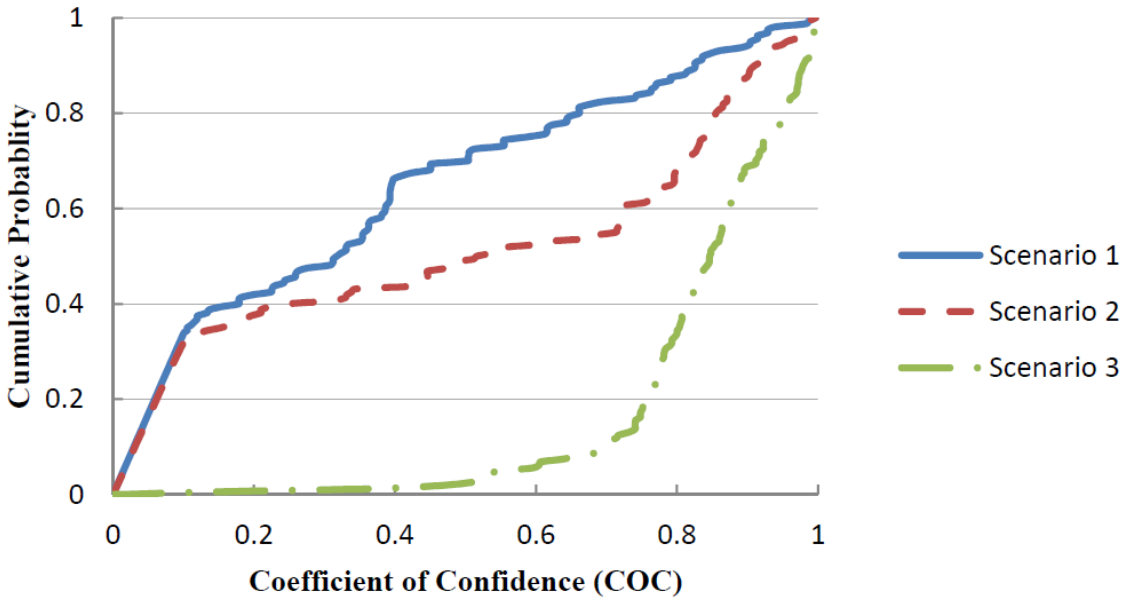


Figure 3.5 CDF of estimated COC values

3.4 .3 . Sequential Gaussian simulation method

The Sequential Gaussian Simulation (SGS) method is a geostatistical method for generation of multiple equally likely realizations of spatially correlated data (Deutsch & Journel, 1998). SGS has been widely used in uncertain groundwater contamination

problems. This method transforms data into a Gaussian distribution, and then it performs variogram modelling on the data. It selects the grid nodes at random, and krigs the value at that selected location. In the case of a log-normally distributed hydraulic conductivity field, the logs of conductivity values are Gaussian. Therefore SGS is applicable for generation of multiple realizations. Note that the longstanding log-normality assumption of hydraulic conductivity may not be correct in all cases (Mathon et al., 2010).

In this section we aimed to use multiple realizations generated by SGS to quantify uncertainty in the field hydraulic conductivity parameter, and ultimately to decrease the uncertainty in identified source characteristics. SGS requires the sample variogram to be estimated using available sparse data points. Also an adequate number of realizations must be considered to realistically represent actual uncertainty. Utilizing 60 hydraulic conductivity sample locations (scenario 2), Figure 3.6 shows the optimized sample and model variograms (Deutsch & Journel, 1998) estimated using natural logarithm of available sparse hydraulic conductivity data. The nugget, sill, and range values of the selected exponential model variogram are 0.00, 0.07, and 546.65, respectively. Considering the complexity of flow and transport simulation models and that a single run may take several seconds, 20 hydraulic conductivity realizations are generated in this study using open source MATLAB code mGstat V 0.99 (2004). The coefficients of uncertainty are estimated using Eq. 3-17 and Eq. 3-11 calculates corresponding COCs.

$$F_{i,j} = \left| \frac{Cest^{k}_{iob} - Cestnr^{k}_{iob,nr}}{(Cest^{k}_{iob} + nn)} \right| \qquad (3\text{-}17)$$

$$where \;\; Cestnr^{k}_{iob} = average(Cest^{k}_{iob,nr} = f(D, K_{nr}, \theta, x_i, y_i, z_i, q_i) \;\; i = 1,...,N, \; nr = 1,...,NR)$$

here $Cest^{k}_{iob,nr}$, and NR are the estimated concentration at monitoring location iob and time period k using nrth SGS generated hydraulic conductivity realization, and total number of realizations (here equals 20), respectively. $Cestnr^{k}_{iob}$ is the average concentration estimated at the same location iob and time period k using all SGS generated realizations. The variable nn is specified as a numerical value which prevents division by zero at the locations where the estimated concentration is zero or very near to zero. The appropriate value for *nn* depends on the relative magnitudes of the actual concentration values.

The performances of the methodology based on the SGS generated realizations and uncertainty-based source identification are analyzed for two different aspects as discussed below. First, SGS utilized the variogram selected based on sparse available information. Fig. 3.7 shows the selected variogram based on the actual hydraulic conductivity field (Fig. 3.3). The selected variogram based on the natural logarithm of actual data is exponential and the nugget, sill, and range values are 0.00, 0.077, and 190.93, respectively. Note that this evaluation is not possible in real-life problems, when the actual field is unknown. Comparing Figs. 3.6 and 3.7 shows that the model variogram used in the SGS method is not an accurate estimation of actual field characteristics. Therefore utilizing SGS for uncertainty quantification adds extra uncertainty to the model. Solution results obtained using the SGS uncertainty-based source identification model and error-free concentration measurements, show that

NAEE% for source 1, and 3 are 8.3, and 37.0 percent, respectively. In comparison, corresponding *NAEE%* values using SGS uncertainty-based source identification, and crisp source identification indicate 40.6 and 67.1 percent increase in estimation errors. Therefore using SGS-uncertainty-based source identification model in fact introduces additional uncertainty to the already uncertain groundwater contaminant source identification problem.

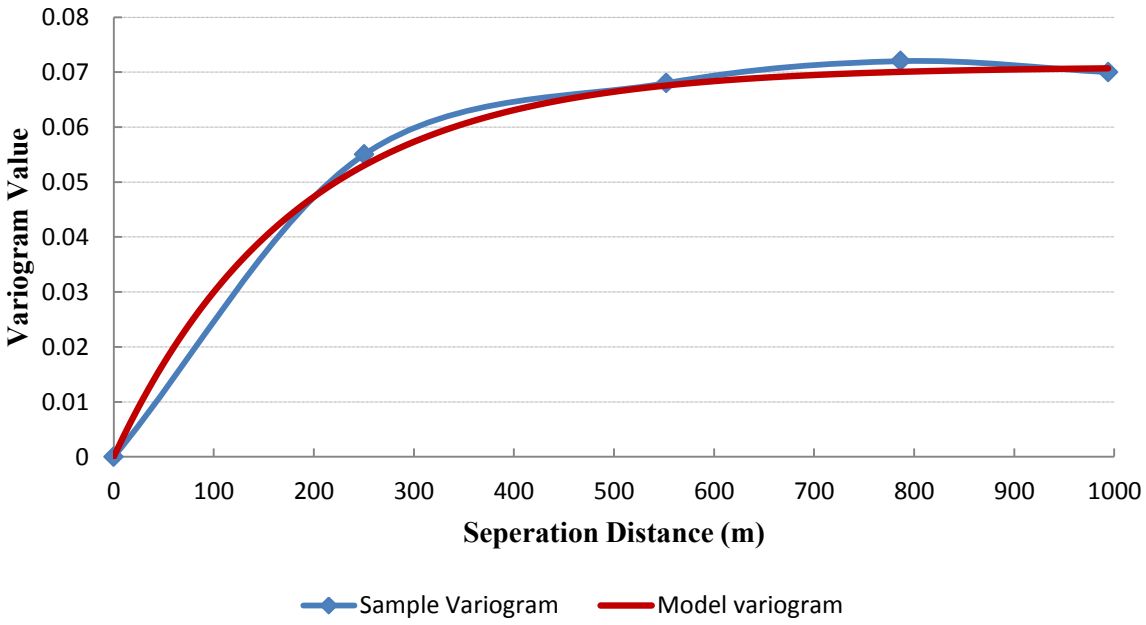


Figure 3.6 Scenario 2's Variogram

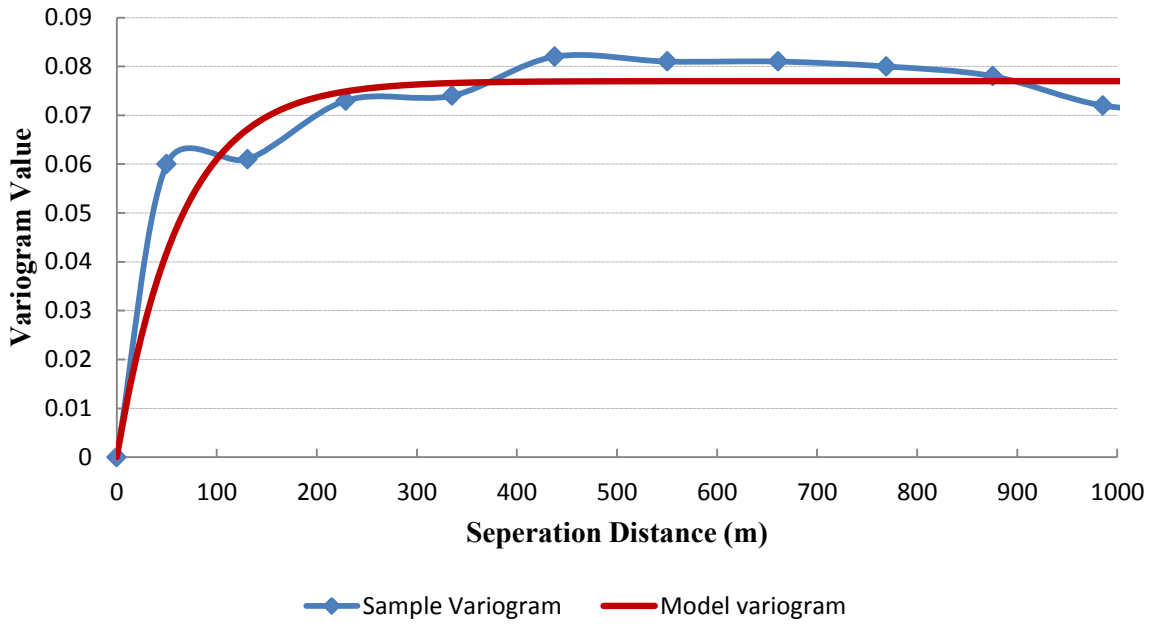


Figure 3.7 Actual hydraulic conductivity field variogram

Second, in general the computational cost associated with evolutionary optimization techniques, e.g. ASA, is high. In linked simulation-optimization source identification models, complex flow and transport simulation models must be executed for each iteration of the optimization procedure and repeated many times. This increases the computational cost substantially. Due to enormous computational costs associated with the solutions of the model, it is vital to limit the number of simulation model executions and incorporate a limited number of realizations coupled with the optimization model; however, using the SGS method, multiple realizations must be evaluated. In this study, 20 realizations are utilized for the characterization of contaminant sources, which is a small number compared with earlier applications of SGS (Mugunthan & Shoemaker, 2004). For the illustrative study area, the running times of crisp, uncertainty-based, and SGS uncertainty-based source identification

models are approximately 27, 57, and 560 hours, respectively. Therefore, because of computational costs using the SGS approach becomes prohibitive.

The SGS method is a useful tool for risk reduction in contaminant plume characterization and delineation, and for monitoring network design problems. Utilizing SGS may be preferred where the aim is to design a model that shows acceptable performance in all perturbations of the model, the spatial correlation is known, and a limited number of model simulations is required.

The performance evaluation results presented here are based on a limited set of uncertainty scenarios. Moreover, there are other possible sources of uncertainty in relation to the flow and transport simulation processes, i.e. hydrogeologic conceptualization (recharge, inter-basin flows, degrees of confinement, fractures and faults, and so on). Finally, the effect of uncertainty in hydraulic conductivity, combined with the other hydrogeologic parameter uncertainties, must be explicitly studied.

Moreover, the developed methodology and the performance evaluation consider uncertainties in only one hydrogeological parameter, i.e. hydraulic conductivity. These evaluation results demonstrate the relevance of the proposed methodology; however, this methodology must further include other hydrogeological parameter uncertainties, and other types of uncertainties in modelling and predicting the contaminant transport process in an aquifer, to improve the efficiency of contaminant source identification process.

3. 5. Conclusion

Lack of information about field hydrogeologic parameters results in non-random uncertainty in estimated contaminant concentrations using flow and transport simulation models. Inaccurate simulation results decreases the efficiency of the linked simulation-optimization based contaminant source identification process. In this study, the uncertainty in hydraulic conductivity parameter values is quantified. The proposed measure of uncertainty is incorporated in the linked simulation-optimization contaminant source identification models. The aim is to improve the accuracy of estimated contaminant source characteristics by considering the uncertainty in the available hydraulic conductivity data. Multiple realizations of hydraulic conductivity fields were generated using the IDW method with different sets of interpolation parameters. Utilizing simulated concentrations in these realizations, the COC is estimated for each available spatial and temporal contaminant concentration monitoring data. Incorporating the COC in the source identification process increases the accuracy of recovered contaminant source characterization. The efficiency of uncertainty-based source identification compared with the crisp model is more evident in locations with large hydraulic conductivity uncertainty. This efficiency also appears to prevail even with errors in measured contaminant concentration monitoring data.

For evaluation purposes, the reliability of the source characterization procedure is obtained using the estimated CDF of COC values. This reliability measure can be utilized in assessing the accuracy of recovered contaminant source characteristics in polluted groundwater aquifers.

The accuracy of solution results and the computational efficiency of the proposed methodology are compared with those obtained using the SGS method, which is a probabilistic approach. In contaminant source identification problems, probabilistic methods result in a substantial increase in computational costs. Moreover, the variogram selected based on available sparse hydraulic conductivity data, utilized in the SGS method, may not be a suitable representation of the entire field variogram. Therefore, the SGS method may add more uncertainty to the already uncertain source identification model. Performance evaluation results provide insight into the inter-relation between errors in the source identification process and available spatial hydraulic conductivity values. The proposed methodology is potentially useful in quantifying parametric uncertainty, when precise information is not available. The scope of this study was limited to only one hydrogeologic parameter. It only illustrates the utility of the proposed uncertainty quantification method, without complicating the basic issues involved. The potential applicability of the proposed methodology was also demonstrated.

In the next chapter, the performance of the proposed uncertainty-based contaminant source identification model is evaluated for an experimental contaminant study area.

4. VERIFICATION OF DEVELOPED METHODOLOGY FOR QUANTIFICATION OF HYDROGEOLOGIC UNCERTAINTY IN GROUNDWATER SOURCE IDENTIFICATION PROCEDURE USING DATA FROM AN EXPERIMENTAL SITE

4. 1. Introduction

In this chapter the application of the methodology presented in chapter 3 for hydrogeologic uncertainty quantification in contaminant source identification procedure is tested for a real-life contaminated aquifer study area.

The performance of the proposed methodology is evaluated using data recorded in an experimental site set up in Botany Sands contaminated aquifer located in New South Wales (NSW), Australia (Beck, 2000). Bromide was utilized as the conservative tracer for measuring the resulting spatial and temporal concentrations.

First a brief review of the background of the problem and the calibration of the flow and transport simulation models are presented. Next, the hydraulic conductivity uncertainty is quantified using the proposed methodology and the uncertainty-based contaminant source identification methodology is applied for this site. Finally the solution results are presented and the potential applicability of the proposed methodology in a real-life contaminated aquifer is discussed.

4. 2. Methodology

The pollutant source characteristics which are required to be addressed in a contamination source identification procedure are: (1) source locations, (2) source release history (time), and (3) source injection concentration magnitudes. The proposed methodology for characterization of unknown groundwater contamination sources using uncertain hydrogeological parameter values has two major components: a linked simulation-optimization contamination source identification model, and an uncertainty analysis module. These components are described below.

4.2 .1 . Linked simulation-optimization contamination source identification model

Application of an effective and efficient optimization model requires careful definition of objective function and possible constraints, and selection of an appropriate optimization algorithm. Therefore, in this section, first the objective function and constraints are defined. Then, the selected optimization algorithm and the linked simulation-optimization method are discussed.

4.2.2 .1 . Contamination Source Identification Formulation

The formulation of the contaminant source identification model is described in section 3.2.1. The objective function is defined by Eq. 3-1, and the constraint sets are defined by Eqs. 3-2 and 3-3. The utilized groundwater flow and pollutant transport simulation models are MODFLOW and MT3DMS.

4.2.2 .2 . Linked Simulation-Optimization Algorithm

The linked simulation-optimization model is solved using the ASA. ASA is a variant of SA in which, the automated reannealing temperature schedule and random step selection make the algorithm less sensitive to the user-defined parameters. Unlike SA, using ASA eliminates the need for several trial executions of the model, to adjust the parameters (Ingber, 1996).

In this study, the flow and transport simulation models are externally linked to the optimization algorithm, constituting the linked simulation-optimization source identification model (Datta et al., 2009). First, based on the available site information, potential contaminant source locations are selected. Second, the optimization algorithm generates the candidate contaminant concentrations associated with each potential source location at each stress period. Third, the simulation models estimate the contaminant concentration ($Cest_{iob}^{k}$) at monitoring locations (*iob*) and time steps (*k*). Then, for each set of candidate source characteristics, the estimated contaminant concentration values are transferred back to the optimization model to calculate the corresponding objective function. Finally, this process evolves to reach the optimal contaminant release histories for the potential source locations (Amirabdollahian & Datta, 2013).

4.2 .2 . Hydrogeological parameter uncertainty analysis

Generally, a predesigned sampling network is utilized to collect field hydrogeological data for groundwater management purposes. The constructed contaminant source histories are sensitive to Hydraulic Conductivity (HC) values (Datta et al. 2009).

Spatial variations in HC values are a critical factor controlling the contaminant mass transport. Therefore, this study focuses on the HC uncertainty in contaminated groundwater aquifers.

In real-life polluted sites, to accurately describe the mass transport process, the spatial and temporal uncertainties in available measured HC values must be quantified. The confidence in the estimated spatial and temporal contaminant concentration values can be evaluated considering the variation and availability of the field HC values and using the flow and transport simulation models.

As discussed in Chapter 3, in a crisp contamination source identification model, it is generally assumed that the model parameters are known without any associated uncertainty (Sun, 1994). Therefore all the estimated or predicted concentrations ($Cest_{iob}^{k}$ in Eq. 3-1) are assumed to be accurate and can have the same contribution to the estimation of the contamination source characteristics. In the crisp source identification model, μ_{iob}^{k} is implicitly assumed as 1.0; however, uncertain hydrogeological input values result in inaccurate concentration estimations. When the hydrogeological parameter uncertainty exists, $\mu_{\mathrm{iob}}^{\mathrm{k}}$ is the Coefficient of Confidence (COC). It represents the degree of confidence assigned to the predicted (estimated) contaminant concentration at each monitoring location, and time and estimated using Eqs 3-10 and 3-11.

The U in Eq. 3-11 is a relaxation factor. When U is very small, the estimated COC values are always one or very near to one. Therefore, the objective function (Eq. 3-1) would be the same as the crisp situation and the effect of parameter uncertainty

would not be properly considered. On the other hand, when U is large, the optimization algorithm would try to find the source release concentrations which result in the larger coefficient of uncertainty (Eq. 3-10) values and smallest COC values to minimize the objective function. In other words the priority of the optimization algorithm would not be the matching of the estimated and observed concentrations. Thus the suitable value for U should be selected based on the range of estimated coefficients of uncertainty values, by trial and error. In this chapter this trial and error procedure is discussed in detail.

4. 3. Performance evaluation

The performance of the proposed methodology has been evaluated using measurement data from a contaminated experimental site set up in the Botany Sands Aquifer (BSA) located to the south of Sydney CBD, New South Wales (NSW), Australia (Beck, 2000).

Although the BSA is considered to be homogenous on a regional scale, on the micro scale the aquifer can be highly heterogeneous. In contaminated sites, detailed HC and contaminant distribution information are necessary to delineate solute movement (Fu et al., 2015). Heterogeneity becomes increasingly important as the scale of the system decreases and more attention must be given to the micro hydrogeological studies.

Various hydrological, hydrogeological and hydrochemical data were collected at this experimental site, during a small-scale natural gradient experiment carried out to examine the effect of heterogeneity on the solute transport (Jankowski & Beck, 2000). The details of the study area and the tracer test experiment are described below;

however, few obvious inconsistencies in the recorded measurement data were corrected before the raw data were used for this evaluation purpose.

4.3 .1 . Study area

The northern part of the BSA is located approximately 3-5 km south of the Sydney CBD, Australia (Fig. 4.1). The BSA was the earliest groundwater resource used to supply water for Sydney and has been utilized since the 19th century (Beck, 2000). In the early 20th century commercial and industrial developments on the northern shore of Botany Bay commenced. These developments, which include oil storage and refinery facilities, airport and a variety of chemical, industrial and commercial manufacturing and storage facilities, resulted in contamination of the groundwater. As the result of present and past land uses, lack of effluent management, treatment and disposal, statutory controls (in the early years after World War II), a long history of contamination exists in this area. Since detection of groundwater contamination, various combinations of investigations, management, remediation and groundwater consumption restrictions have been utilized to control the negative effects of contamination in this area.

The BSA mainly consists of unconsolidated sands, clays and peaty sediments. The thickness of the sediments varies from zero at the northern rim of the basin, in Centennial and Moore Parks (Fig 4.1), to approximately 80 m in southern parts. Distinct hydraulic boundaries between different geological units produce variations in HC values from 1.8 m/day to 50 m/day. Variations in hydraulic properties are closely related to lithological units including quartz sand, silty/peaty sand, and sandy/peaty clay.

Figure 4.1 Location of Botany Sands Aquifer and Eastlakes Experimental Site (Jankowski & Beck, 2000)

4.3 .2 . The experimental site

Spatial heterogeneity associated with site physical properties, especially hydraulic conductivity (HC), dominates the solute and contaminant transport processes. Therefore, numerical modelling, tracer experiments and laboratory experiments are required to study the effect of the HC heterogeneity in micro scale. A tracer test was carried out at the Eastlakes experimental (ELE) site, located adjacent to the Lachlan Ponds in the middle of the northern part of the BSA, in Daceyville, NSW, Australia (Beck, 2000). The site was first established in 1992 with a three-dimensional network of piezometers installed on a 7m × 11m grid. The hydrogeological and chemical heterogeneity was measured at 815 sampling points (Evans, 1993) with a horizontal and vertical spacing of 1m and 150mm-200mm, respectively (Fig 4.2).

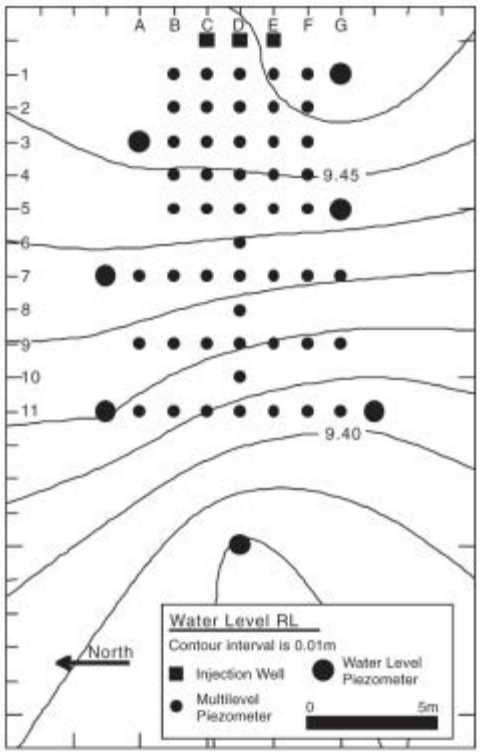


Figure 4.2 Location of piezometers, injection wells and sampling points in the Eastlakes Experimental site (Jankowski & Beck, 2000)

The geological investigations of the ELE site were carried out by two continuous sand cores. Using the geological information gathered during 1992-93 by Evans (1993), the composite geological section through the ELE site was prepared. Fig. 4.3 shows the composite geological cross-section along line D (Fig. 4.2) through the site. Five distinct lithological units were discovered as Sand, Waterloo Rock, Organic Silt and Sand Bands, Peat, and Silty Sand. Since the low permeable peat layer is located between 5m to 6m above sea level (asl), the ELE area is an unconfined aquifer. Existence of various soil types and grain size distribution should result in variation in the HC. Moreover, the deposition environment also has significant influence over the HC distribution.

The aquifer is in hydraulic continuity with the pond, which provides fixed head boundaries for the experiments. Therefore, the east (at injection wells) and west

boundaries are fixed head and the north and south ones are variable head boundaries. The initial head distribution follows the contours in Fig. 4.2 with the regional hydraulic gradient of 1:240 (from east to west).

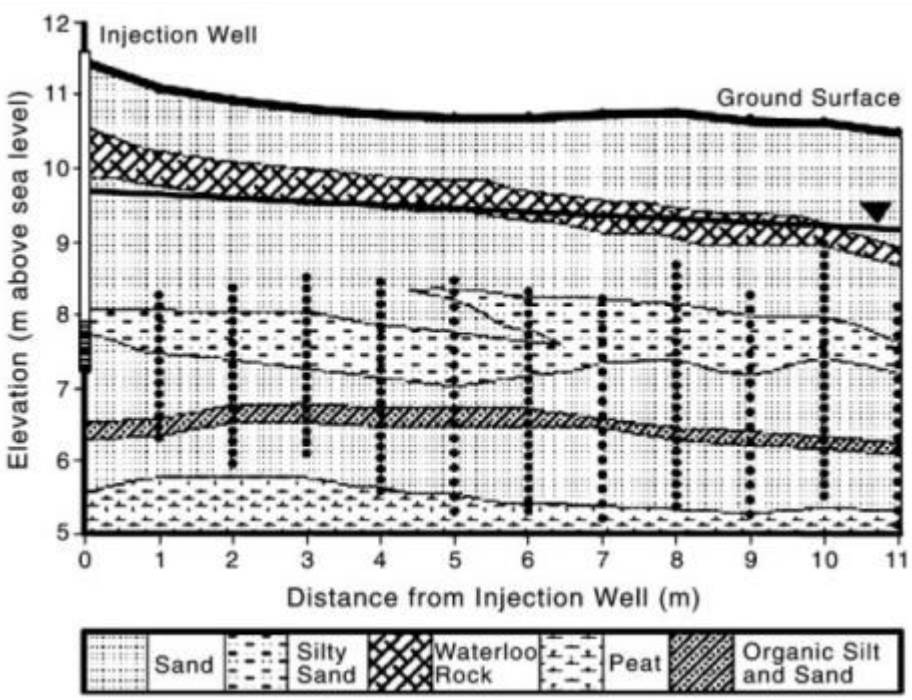


Figure 4.3 Composite geological cross-section along line D through the Eastlakes Experimental site (Jankowski & Beck, 2000)

The main source of recharge to the ELE site is direct rainfall, as it is located in the open-space area. The average annual rainfall recorded at Centennial Park and Sydney Airport are 1236mm and 1083mm, respectively. Yu (1994) estimated the monthly average recharge for the northern part of the BSA between 1986 and 1993. These values vary between 0 mm and 515 mm per year with a mean average monthly recharge of 68.4 mm. Therefore, for the ELE site flow model, the average recharge is considered to be 68.4 mm/month.

The variability of HC was measured by multilevel piezometers in lines C, D and E with vertical spacing of 150mm-200mm (Evans, 1993). HC was recorded at 455 locations and the minimum, mean and maximum values were 1.5 m/day, 14.6 m/day, and 50 m/day respectively. The collected data is not normally distributed and is skewed towards the lower conductivity values. Variations of more than 50 m/day over less than 50 cm, were present in some parts of the study area.

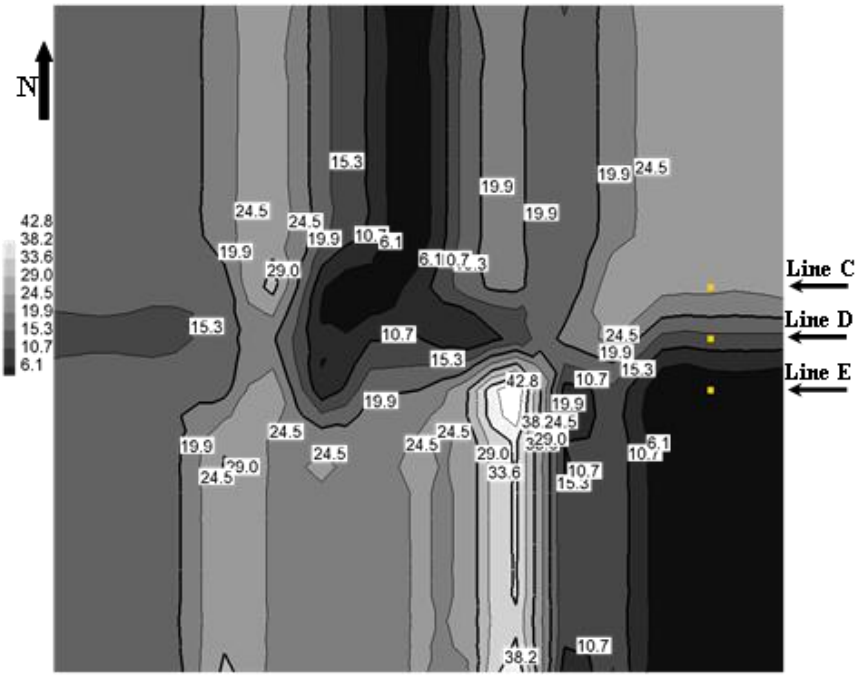
In order to find the HC values for the entire study area, the IDW interpolation method is utilized. The suitable number of the closest sampled data points used for the interpolation purpose (n) is selected using the spatial structure of data.

The spatial auto-correlation of available measurement data is the correlation among values of a single variable attributable to their relatively close locational positions. *Moran's I* (Moran, 1950) is an indicator of spatial auto-correlation which is applied to zones or points with continuous variable associated to them and estimated using Eq. 3-8.

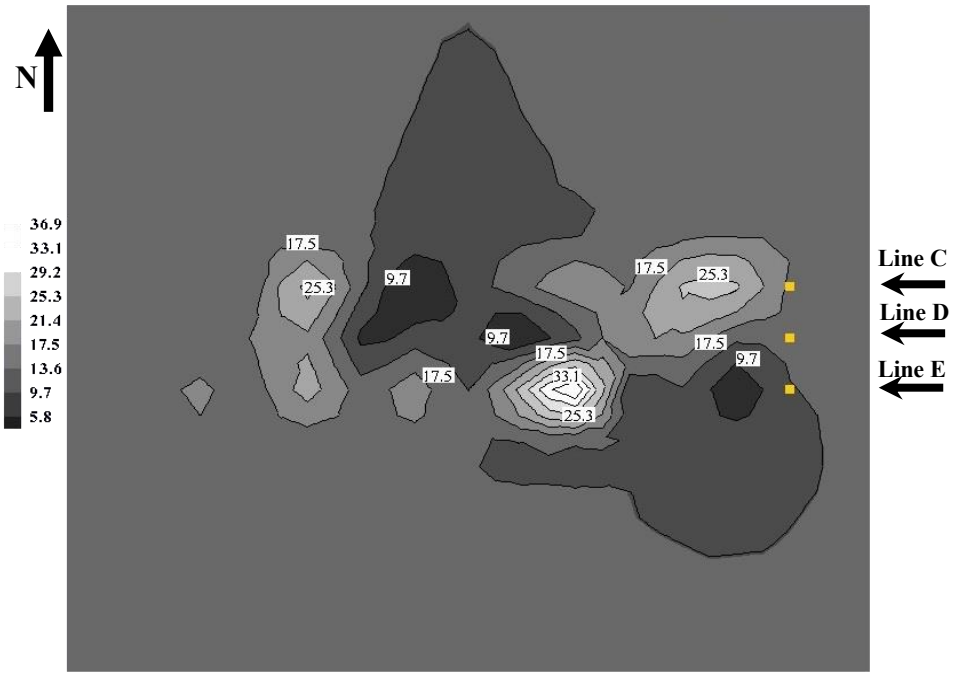
The spatial auto-correlation varies between -1.0 and +1.0. If nearby areas are more alike, the spatial auto-correlation is positive. The negative autocorrelation describes patterns in which neighbouring areas are unlike and random patterns exhibit zero spatial auto-correlation. The *Moran's I* coefficient for the available HC values from the ELE site is 0.021 which is small and close to zero. Therefore, although BSA is homogenous on a regional scale, the ELE site is a highly heterogeneous system on the micro scale. The data are not spatially auto-correlated, thus n=3 (Eq. 3-6) is selected for the IDW interpolation method. Figs. 4.4-a and 4.4-b shows the variation in HC in the third layer of the ELE site. These distributions are estimated using the IDW method

with n=3 and n=455. Here 455 is the total number of available HC measurements. The heterogeneous nature of the study area is evident in the contrast between the interpolated HC fields, as shown in figs. 4.4a and 4.4b.

Evans (1993) had measured the HC using constant-head method at three locations within the ELE site. Comparing the averaged value measured by the multilevel piezometers and the constant-head method, demonstrates that the multilevel piezometers underestimate the true HC as they were measured in non-equilibrium conditions. Therefore, by calibration of the flow model, the values shown in Fig. 4.4 are increased by 20% for the modelling purposes, to better represent the real flow field. In the calibration process all the HC values were increased uniformly without changing the heterogeneity pattern of the study area. The resulted HC field is used to estimate the head distribution and to ultimately estimate the contaminant concentrations ($Cest_{iob}^{k}$) at different monitoring locations and times. Since the HC measurements were taken along lines C, D, and E, in Fig. (4.4), the heterogeneity is presented along these lines; however, there is no measurement out of the central region of the site. Therefore, the distribution shown in Fig. (4.4) may not be a good representation of the real field conditions, where no measurement is available. The background hydrochemical conditions of the natural groundwater prior to tracer injection was recorded using 88 groundwater samples collected along the central line D.



(a)



(b)

Figure 4.4 Hydraulic Conductivity distribution in Eastlakes Experimental site using a) n=3 and b) n=455. (unit is m/day)

4.3 .3 . Tracer test

This section briefly outlines the details of the tracer test conducted in the site (Beck (2000). A total of 300 litres of tracer solution was prepared which included Boron (B), Bromide (Br), Chloride (Cl), and Lithium (Li) as the conservative tracers and Cadmium (Cd), Copper (Cu), Potassium (K), Nickle (Ni), Lead (Pb) and Zinc (Zn) as the reactive solutes. Table 4.1 summarises the injection concentrations in the tracer solution.

Table 4.1 Injection concentration of the tracer solution solutes

Solute	**B**	**Br**	**Cd**	**Cl**	**Cu**	**K**	**Li**	**Ni**	**Pb**	**Zn**
Concentration (mg/l)	110.1	186.0	53.2	741.8	56.4	106.9	123.7	54.8	51.3	51.4

Three injection wells (sources C, D, and E) were developed for one day using a combination of pumping, surging and recharging methods, to ensure that a good interaction between the wells and the aquifer was achieved. Five 20-litre batches of solution were injected in each well over a half-hour period starting at 13:00 on 2 July 1996. Care was taken to maintain low injection flow rates into the wells to ensure significant increases in the hydraulic heads of the injection wells did not occur. Excessive hydraulic heads in the injection wells would force some of the tracer up-gradient of the injection wells and cause higher hydraulic gradient than would occur under natural gradient conditions.

The contaminant concentrations sampling was started two days after tracer injection (4 July 1996). This was followed by sampling solute concentrations every two days after injection. The samples were collected along lines B, C, D, E and F (Fig. 4.2) at different elevations.

4.3 .4 . Bromide (Br) transport simulation model implementation

In this study, the Bromide (Br) concentration measurement data from the experimental test site are utilized to validate the proposed methodology. The Br concentrations were analyzed using ion selective electrodes and ion chromatography (Beck, 2000). Chemical analysis of the natural groundwater prior to tracer test did not show any detectable background concentration of Br. Therefore, the background concentration was specified as zero in the transport simulation model.

Br is classified as a conservative tracer. Therefore, the two important components of the transport simulation are advection and dispersion. The dispersion coefficient was estimated by Beck (2000) using graphical method based on the concentration versus time plots. In this study, an averaged value of 0.03 m is selected for longitudinal dispersivity and the ratio of horizontal transverse dispersivity to longitudinal dispersivity, and ratio of vertical transverse dispersivity to longitudinal dispersivity are 0.4 and 0.1, respectively.

For the flow and transport simulation purposes, the study area is divided into four layers. Each layer has different hydrogeologic properties and therefore, the study area incorporates hydrogeologic heterogeneity. Layer one is between the groundwater level and the top of the silty sand layer (Fig. 4.3). Layer two is from the top of the silty sand soil layer to 7.6 m asl. Layer three is between 7.6 m asl to 7 m asl. Layer four is between 7 m asl and the study area bed which is specified as the peat layer. The cell size on the plan view is 1 m by 1 m. Fig. 4.5 shows the three-dimensional view of the ELE site model. The three injection sources are located in layer three. Owing to the small scale of this study area, the parameter estimation and numerical computation

errors have noticeably negative effects on the accuracy of the flow and transport simulation models. In order to reduce these negative effects in this study, layer three is dedicated to the contaminant monitoring procedure and the inter-layer hydraulic conductivity is not modeled. The injection sources are located within this layer and the observed concentrations located within this layer are utilized for the source identification procedure. Layers 1-4 porosities are 0.39, 0.41, 0.36, and 0.41, respectively.

In the numerical flow simulation model MODFLOW, the Layer Property Flow (LPF) package is used. Unlike the Block Center Flow (BCF) method, the LPF model calculates the conductance for a water table bearing model cell based on the water table position instead of the center of the cell. Dynamically following the water table can substantially increase the accuracy of flow calculations (Clemo, 2003). The advection term in the numerical transport simulation model MT3DMS is solved using Method Of Characteristics (MOC). The MOC method uses a conventional particle tracking technique based on a mixed Eulerian- Lagrangian method for solving the advection term, whilst the dispersion and sink/source mixing terms are solved by finite difference. The main advantage of the MOC technique is that it is virtually free of numerical dispersion; however, other methods such as the Third-order total Variation Diminishing (TVD) method exhibits minor numerical dispersion and minor oscillation in the concentration fronts (Schlumberger Water Services, 2011). As a result of the small scale of the ELE site, minimizing the numerical dispersion has a drastic effect on the estimated contaminant concentrations in the site.

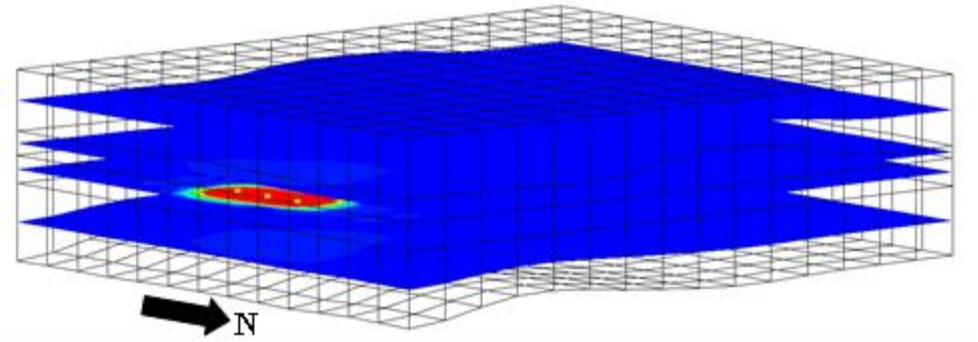


Figure 4.5 The Discretized model and contaminant injection locations for the ELE site

4.3 .5 . Recovering the contamination source histories

In this section the proposed method is utilized to find the contamination source injection concentrations using the Br concentrations collected 2, 4, 6, and 8 days after the tracer injection. In total, 19 Br concentration measurements are used for the performance evaluation. The 100 litres of tracer is injected in each well over a half-hour period. Therefore, in the flow model, the injection locations are specified as flow injection wells with 4.8 m^3/day flow rate. Although care was taken to maintain a very low flow rate during tracer injection, because of the small scale of the study area even this flow rate changes the hydraulic field. Therefore, it must be considered in the flow model.

In this performance evaluation exercise, the specified potential contamination source locations include the three actual source locations and another dummy source location. One potential source location is added to the source identification model which is not actual (dummy), and is located along line G aligned with other three actual sources. A dummy source location is introduced as a potential source location in order to examine the performance of the methodology in finding the location of actual

sources among specified potential locations. The injection flow rate in the other three source locations changes the flow system, therefore for the dummy source, a lower flow rate (1 m^3/day) is specified in the MODFLOW. This will diminish the effect of non-actual source on the hydraulic head distribution of the study area. The decision variables in the source identification problem are the four Br injection rates at four potential locations, and the duration of the study period is 8 days. The specified upper and lower bounds for contaminant release concentrations, $q_{\max}$ and $q_{\min}$ in Eq. 3-3, are 1000 mg/l and 0 mg/l, respectively.

This experimental tracer test may appear to be simple compared with most real-life aquifer contamination problems; however, since the study area is small (7m × 11m) and the study period is short (8 days), the HC uncertainty has a substantial negative effect on the accuracy of the contaminant source identification procedure. To demonstrate this, the crisp source identification procedure is used to retrieve the source release histories. In the crisp method, it is assumed that there is no uncertainty associated with the HC distribution. Therefore, in the objective function, μ_{iob}^{k} value is 1.0 for all the monitoring locations and times. The retrieved Br injection concentrations are presented in Table 4.2.

In order to quantify the HC uncertainty, multiple realizations of the flow field are required. In this study, three more realizations (in total R=4) are generated. The IDW method with n=3, 6, 9, and 12 is utilized to interpolate the available measured HC values and generate realizations. The number of realizations and corresponding *n* values are selected based on the nature of available uncertain hydrogeological

parameter values and computational resources. In this study area, the estimated spatial auto-correlation of measured HC values is small. The scale of the study area is small too. Therefore, the selected *n* values for realization generation are small, compared with the number of available HC measurements. For each realization, the transport simulation model must be executed to find the contaminant concentrations at monitoring locations and times ($Cest_{iob}^{k,r}$) for each iteration of the optimization algorithm. Thus the selected number of realizations depends on the available computational resources and also the degree of heterogeneity and uncertainty in the system. More realizations will result in better quantification of the available uncertainty, while substantially increasing computational time. The selected *nn* value (Eq. 3-5) is 0.001.

4. 4. Results

The estimated Br injection rates at four potential source locations using the uncertainty based (loosely designated as Fuzzy in the following tables) source identification methodology with different *U* values are presented in Table 4.2. The Br release concentrations at sources are shown in mg/l. These concentrations must be multiplied by a constant volume flow rate (4.8 m^3/day) specified as a given magnitude in the flow model to obtain the source fluxes. Evaluating the performance of the proposed methodology in recovering accurate source injection histories, the Normalized Absolute Error of Estimation *(%NAEE*) is determined using Eq. 3-15.

Smaller *NAEE%* values mean that the utilized source identification algorithm is able to recover source injection concentration histories with smaller associated

estimation errors. The *NAEE%* values corresponding to each source identification method are presented in Table 4.2. Without considering the effect of HC uncertainty, the *NAEE%* associated with the crisp methodology is 54%. Therefore, the crisp method exhibits the largest error compared with the HC uncertainty quantification based (fuzzy) models. This demonstrates the necessity in considering the effect of hydrogeological uncertainty for recovering source injection contamination concentrations.

Table 4.2 Retrieved Br injection concentrations at potential source locations

	Br Injection Concentration (mg/l)										
	Actual	**Crisp**	**Fuzzy *U=0.1***	**Fuzzy *U=0.3***	**Fuzzy *U=0.5***	**Fuzzy *U=0.7***	**Fuzzy *U=0.8***	**Fuzzy *U=0.9***	**Fuzzy *U=1***	**Fuzzy *U=5***	**Fuzzy *U=10***
Source C	186	127	111	8.2	11	8	9	10	65	126	137
Source D	186	89	135	163	160	171	170	160	179	91	93
Source E	186	39	40	162	170	171	170	186	128	30	40
Source G	0	0	0	0	0	0	0	0	0.1	0	0
%NAEE	-	54%	48%	40%	39%	37%	37%	36%	33%	55%	51%

In this proposed methodology, in order to compute the COC values Eq. 3-11 is utilized. In this equation the appropriate *U* value should be selected based on trial and error. Therefore, the uncertainty based method with various *U* values is utilized to find source injection histories. In Table 4.2, the smallest estimated *NAEE%* values are associated with *U=0.3-1*. In the ELE site, Br contamination problem, utilizing the uncertainty based (fuzzy) source identification methodology with appropriate *U* value results in about 37% error, and 32% improvement in accuracy compared with the crisp

methodology. All methods were able to identify the non-actual (dummy) source location.

4. 5. Discussions

In this study, the tracer test in the ELE site is utilized for performance evaluation of the proposed methodology. Therefore, the actual location and injection histories ($(q_i^t)_{est}$) are known and the estimated *NAEE%* values using Eq. 3-15 can be utilized to find the appropriate *U* value; however, in real-life aquifer contamination problems, the source locations and the injection rates associated with them are unknown and it is not possible to estimate and use *%NAEE* values. In application to a real contaminated aquifer, the suitable *U* value is to be identified using estimated COC (μ_{iob}^k) values. Fig. 4.6 presents the estimated COC associated with different *U* values. For each *U*, 19 COCs are estimated corresponding to 19 Br concentration measurements (monitoring data). The source injection concentrations, estimated as the optimal solutions of the uncertainty-based source identification procedure, presented in Table 4.2 are used to estimate COC using Eqs. 3-10 and 3-11. When *U=0.1*, a large number of COC values are close to 1. Therefore, 0.1 seems too small for this study area. On the other hand, when *U* is large, in this study area *U=5* and 10, the optimization algorithm tries to find source histories which can minimize the estimated COCs. Therefore, in Fig. 4.6 for *U=5* and *10* (shown along the top axis) a large number of COC values are 0.1, which corresponds to the smallest possible value in Eq. 3-11. Therefore, these *U* values are too large for the uncertainty quantification purposes.

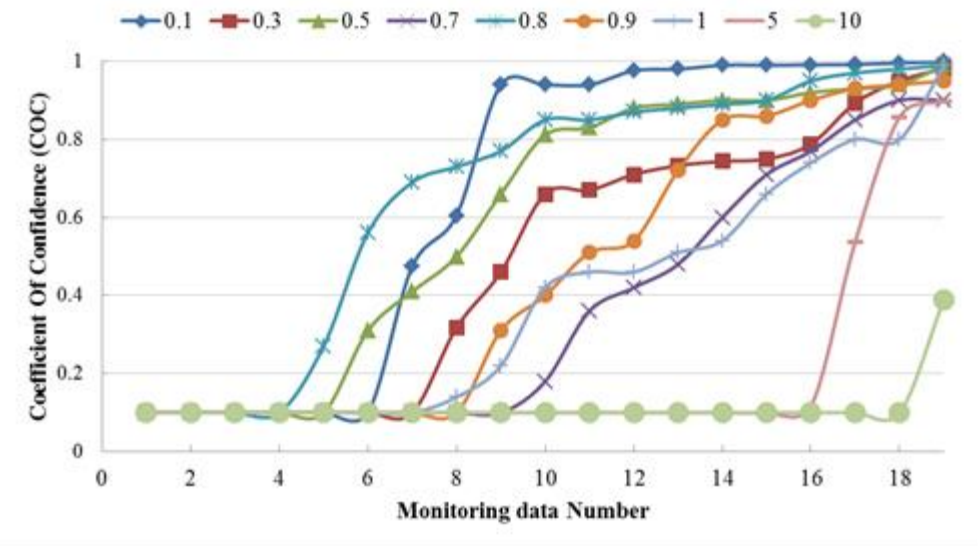


Figure 4.6 Estimated Coefficients Of Confidence (COC) for different U values

The Br injection histories in sources D and E were recovered with a high level of accuracy by the uncertainty-based (fuzzy) source identification model; however, a large error is associated with the Br release history estimated for source C. Moreover, the crisp method found a better estimate of the injection concentration at this source location, compared with the uncertainty-based method. The reason could be related to the HC uncertainty in the study area. In Fig. 4.7 the difference between the average HC values of the three realizations and the HC values shown in Fig. 4.4 is presented. Fig. 4.7 is a representative of the HC uncertainty in this field. The counters show that there is some level of uncertainty associated with the area around source C. The contaminants move along the natural gradient which is from east to west, thus the uncertainty on the left side of the sources affects the accuracy of the estimated contaminant concentrations. As the result of this uncertainty, the COCs estimated for the monitoring locations along line C, are lower than other monitoring locations. Therefore, matching the estimated and observed Br concentrations along lines D and E

has a higher importance in the uncertainty-based source identification objective function.

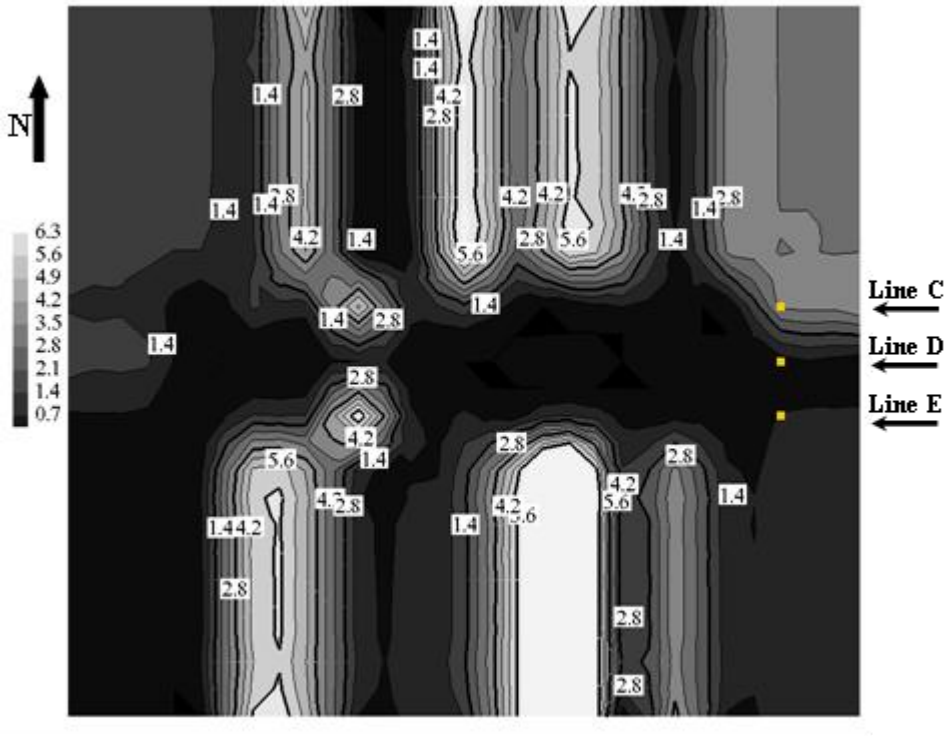


Figure 4.7 Hydraulic Conductivity uncertainty in Eastlakes Experimental site, unit is m/day

The area surrounding source C is a high permeability area with large HC values (Fig. 4.4) compared with the neighboring areas. In this area the number of closest sample data points for the interpolation purposes has a noticeable effect on the estimated HC field. As fuzziness or uncertainty can be reduced with the acquisition of additional information (Oberkampf et al., 2004; Ross, 2005), additional HC measurements in this area can reduce the associated uncertainty.

In this study, the uncertainty in HC values is characterized in the micro scale. Although large number of HC measurements are available (455), using the interpolated HC values when n=455 does not appropriately characterize the micro heterogeneity and

instead results in an averaged flow field. Considering the small scale of the study area, using all available HC measurements will result in a smooth HC field, in the scale of this study. By using only a small subset of the data available (n=3, 6, 9, and 12), the HC uncertainty is magnified in the micro scale. Table 4.3 shows the estimated contaminant source fluxes using crisp source identification model and HC field generated using n=455. The NAEE% value is 136% which demonstrates that IDW with n=455 does not truly characterize the actual HC field and associated uncertainty.

Table 4.3 Retrieved Br injection concentrations at potential source locations obtained using all available HC measurements

	Br Injection Concentration (mg/l)			
	Source C	**Source D**	**Source E**	**Source G**
Actual	186	186	186	0
Estimated	109	90	18	364

4. 6. Conclusions

This study presents a methodology to quantify hydrogeological parameter uncertainty to accurately estimate contamination release histories in polluted aquifers. The performance of the developed methodology is evaluated using actual measurement data from an experimental contaminated aquifer study area. The evaluation and remediation of contaminated aquifers requires accurate identification of the contaminant source locations and their release histories. In order to find the contaminant source characteristics, the groundwater flow and pollutant transport simulation models are utilized to optimally estimate the contaminant concentration in the study area, using potential source characteristics. Then the estimated concentrations are compared with the actual observed concentrations collected as monitoring data. The linked simulation-

optimization framework is utilized to search for the pollution source characteristics which exhibit the best match between the estimated and observed contaminant concentrations in the site. In order to obtain reliable solutions, the simulation models must be provided with reliable hydrogeological parameter values. In real-life scenarios, usually high levels of uncertainty and variability are associated with the available hydrogeological parameter values. In this study the effect of uncertainty in hydrogeological parameter values on the accuracy of flow and transport models estimates are quantified.

The proposed methodology also provides insight to the relation between the errors in flow and transport simulation models and variability in hydrogeological measurement data. Tracer test results at the ELE Site, located in Botany Sands Aquifer, Australia, are used to evaluate the performance of the proposed methodology. In this study area the hydrogeological heterogeneity in the micro scale, specifically the hydraulic conductivity, has substantial effect on the transport of pollutants. Ten tracers have been injected into the groundwater systems. Their movement under natural gradient was monitored by measuring concentrations in the groundwater at various locations and times after injection. Among available tracer information, Bromide is chosen as a typical conservative element.

The ASA-based linked simulation-optimization methodology was utilized to characterize the pollution sources using concentrations measured after tracer injection. The solution results demonstrate that the proposed methodology recovered pollution source characteristics more accurately compared with the methodologies which do not consider the effect of hydrogeological parameter uncertainty. In this ELE site, for the

Br contamination problem, utilizing the uncertainty quantification (fuzzy) based source identification methodology with appropriate *U* value results in about 33% improvement in accuracy compared with the crisp methodology. All methods were able to find the non-actual (dummy) source location.

The developed methodology provides the decision makers with a tool to incorporate the hydrogeological parameter value uncertainty for accurate identification of contaminant release histories in real-life study areas. Moreover, these solution results may be used to find the locations where available field data does not sufficiently characterize the flow field. Therefore, these solution results can help in identifying locations where additional hydrogeologic information may be collected to reduce uncertainty in the flow and transport simulation models.

The next chapter presents the development and evaluation of a new methodology potentially useful for increasing the accuracy of the contaminant source identification process utilizing uncertain contaminant concentration measurements.

5. CONTAMINANT SOURCE IDENTIFICATION USING UNCERTAIN CONCENTRATION MEASUREMENTS

5. 1. Introduction

The methodology presented in this chapter aims to use a two-objective approach to improve the accuracy of optimally estimating pollutant source characteristics in the presence of measurement error/ uncertainty.

In this chapter, a two-objective linked simulation-optimization approach is developed, which addresses the uncertainties associated with erroneous measurements utilized for groundwater contaminant source identification. The aim is to obtain a trade-off between two single objectives each of which may not be an ideal objective function separately, where uncertainties exist.

A three-dimensional transient flow and transport process in a contaminated site with point and distributed contaminant sources is modeled for the performance evaluation of the proposed methodology based on the two objectives of source characterization. The performance is evaluated for various degrees of uncertainties. The purpose is to compare the results with the performance obtained using single objective optimal source identification in the presence of concentration measurement error.

5. 2. Two-objective linked simulation-optimization methodology

Application of an effective and efficient optimization model requires careful definition of objective functions and possible constraints, and selection of an appropriate optimization algorithm. Therefore, in this section, the objective functions and constraints of a proposed two-objective optimization model for unknown groundwater

contaminant source characterization is presented. Then, the selected optimization algorithm and the linked-simulation optimization algorithm are discussed.

5.2 .1 . Two-objective contaminant source characterization methodology formulation

The optimal recovery of contaminant source release histories generally involves examining a set of candidate source characteristics to determine the set which results in the best pollutant plume with respect to the observed contaminant concentrations. Therefore, it can be defined as an optimization model, in which the objective functions are expressed as measures of the goodness of fit between the observed concentrations, and the simulated concentrations corresponding to the candidate solutions. The two objectives of the optimization model for optimal characterization of contaminant sources under uncertainty are defined by Eqs. 5-1 and 5-2, and the constraints are defined by the constraints set 5-3 and 5-4.

$$F_1 = Minimize \sum_{k=1}^{nk} \sum_{iob=1}^{nob} \frac{\left| Cest_{iob}^{k} - Cobs_{iob}^{k} \right|}{Cobs_{iob}^{k} + \eta} \times 100 \tag{5-1}$$

$$F_2 = Minimize \sum_{k=1}^{nk} \sum_{iob=1}^{nob} \left| Cest_{iob}^{k} - Cobs_{iob}^{k} \right| \tag{5-2}$$

Subject to

$$Cest_{iob}^{k} = f(D, HC, \theta, x_i, y_i, z_i, q_i^t) \quad i = 1,...,N; \quad t = 1,...,T \tag{5-3}$$

$$q_{\min} \le q_i^t \le q_{\max} \quad i = 1,...,N; \quad t = 1,...,T \tag{5-4}$$

where *nk*, *nob*, *N* and T are the total number of concentration observation time periods, available monitoring locations, candidate source locations and pollutant injection stress periods, respectively. $Cest_{iob}^{k}$ and $Cobs_{iob}^{k}$ are the concentrations estimated by the simulation model and observed concentrations at observation location *iob* and at the end of time period *k*, respectively. η is a constant which is sufficiently large to prevent any individual term in (5-1) becoming indeterminate due to the value of any observed concentration becoming very small. *D*, *HC* and θ are the dispersion coefficient, hydraulic conductivity and porosity, respectively. x_i, y_i, z_i and q_i^t are the Cartesian coordinates of candidate contaminant source *i*, and the contaminant release flux for candidate location *i* during stress period *t*, respectively. $q_{\max}$, $q_{\min}$ are the upper and lower bounds for the contaminant release fluxes, respectively.

Both objectives minimize the difference between the observed and estimated concentrations. The first objective minimizes the normalized difference between the estimated and measured concentrations at monitoring locations and times. The second objective does not use any normalizing factor. The estimated contaminant concentrations are calculated using flow and transport simulation models (first constraint, Eq. 5-3). The flow and transport simulation models are externally linked to the optimization algorithm, forming a linked simulation-optimization source identification model (Datta et al., 2009). In this way, for each set of candidate source characteristics, the simulation models estimate the contaminant concentration at monitoring locations (*iob*) and at time steps (*k*). Then these estimated values are transferred back to the optimization model to calculate the corresponding objective function value (Amirabdollahian & Datta, 2013). In this study, for the groundwater

flow simulation purpose, MODFLOW-2000 (Zheng et al. 2001) is utilized. MODFLOW is a computer program that numerically solves the three-dimensional transient groundwater flow equation. The contaminant transport process is simulated using MT3DMS (Zheng & Wang 1999). MT3DMS is a computer program that numerically solves the three-dimensional transient multi-spacious contaminant transport equation in ground water systems. Readers are referred to Chapter 3 for more information on the governing equations describing the flow and transport processes. Eq. 5-4 represents the second constraint which limits the source release fluxes to specify upper and lower bounds. In this study, the locations of sources are unknown and are decision variables. Therefore the source flux lower bound is set to 0. The non-actual or dummy and inactive sources must be correctly identified by zero release fluxes for all stress periods.

Generally, a contaminant source identification problem is considered as well posed, if the following conditions are satisfied (Tikhonov & Arsenin, 1978):

1. A solution exists,
2. The solution is stable,
3. The solution is unique.

In groundwater contamination problems, the propagation of contaminant starts from one or more sources, thus certainly a solution exists for the problem. The groundwater simulation models utilized in this study are proven to be stable and convergent (Jha & Datta, 2013). In order to ensure that the source identification problem has a unique solution, uncertainties in the modeling process (model, parameters, and measurement uncertainties) must not be present and the unknown

parameters should be piecewise constant (Sun, 1994). In this study, the following idealized assumptions are considered:

1. There is no model uncertainty; i. e. the governing equations and utilized numerical models used for simulation of groundwater flow and transport are able to simulate the exact natural process in forward runs.
2. The unknown variables are piecewise constant; e.g. the unknown fluxes are constant in every stress period.

However, the parameter values and/or observed concentrations are measured or/and recorded erroneously. Therefore, the uniqueness criterion is not satisfied and the contaminant source identification using uncertain measurement data is an ill-posed problem. The non-uniqueness in model identification means that the information available to define the model does not allow a single or unambiguous solution to the identification problem. This indicates that multiple models might give equally acceptable fits to the observation data (Beven, 2006). Therefore, it is vital to select a source identification model which will find multiple sets of source characteristics providing equally acceptable fits with the erroneous observation data. This is the reason a two-objective model was selected to obtain the Pareto-optimal solution instead of a single so called “optimum” solution.

In groundwater contaminant source characterization problems, even when there is no error associated with the inputs, there may not be a unique solution. Also, only if the global optimum solution is found, it represents the exact solution to the source identification problem. Non-uniqueness in the system response may even then, introduce alternate optimal solutions, although each globally optimal (Datta, 2002). In

the optimal contaminant source identification process, this global optimum represents the actual contaminant source characteristics. Therefore, by feeding actual source characteristics in the forward simulation run, the estimated concentrations ($Cest_{iob}^{k}$) equal observed concentrations ($Cobs_{iob}^{k}$) and both 5-1 and 5-2 result in zero value. Thus, in the case of no uncertainty, the proposed model reduces to a single objective one, and utilizing either objective functions should, ideally, return the global optimum; however, in an ill-posed problem, there is no unique answer to the problem. Despite this fact, many modelers have concentrated on the search for a single "optimum" but at the risk of avoiding important issues of model acceptability and uncertainty (Beven, 2006).

In this study, the aim is to find a set of possible solutions which shows acceptable performance in terms of two objective functions obtained as a solution. This set is called the Pareto-optimal or non-dominated front. In the two-objective source identification model, the search algorithm first searches for the source characteristics which result in better (non-dominated) performance, considering both objectives. Finally it reaches a point (in the decision space) where it is not possible to improve both objectives at the same time. For example, for the sake of simplicity, we assume there are two available observed concentrations $Cobs_1^1$=1 and $Cobs_2^1$=100. The best estimations of concentrations found so far are $Cest_1^1$=0.8 and $Cest_2^1$=80 (using set X_1 of source characteristics). The search algorithm is not able to further improve the result with respect to both objective functions. However, there is a set of source characteristics (X_2) which will result in $Cest_1^1$=0.7 and $Cest_2^1$=87. Substituting these values in Eqs. 5-1 and 5-2, the F_1 and F_2 values corresponding to X_1 values are 39.82

and 20.20, and corresponding to X_2 values are 42.7 and 13.3, respectively. Therefore, X_2 is superior to X_1 considering objective 2 (13.3< 20.2) which conflicts with objective 1 (42.7> 39.82). This demonstrates the tradeoff between the two objectives and the sacrifice necessary in one objective function value in order to achieve improvement in the other objective function value, for a Pareto-optimal or, non-dominated solution. It represents a vector optimization problem, for which a single optimal solution cannot be identified without further preference ordering of the Pareto-optimal solutions.

5.2 .2 . Two-objective contaminant source identification process

The two important components of the methodology are demonstrated in this section as, the selected optimization algorithm and the linked simulation-optimization process. The evolutionary optimization methods (such as GA, SA and ASA) are well suited for finding solutions to various types of optimization problems since they ideally do not make any assumption about the mathematical characteristics of the objective function. Moreover, due to the nature of these algorithms, it is much simpler to link the optimization procedure to the numerical simulation models such as physical flow and contaminant transport processes. In this study, a Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb, 2002) is linked to the numerical flow and transport simulation models to determine the contaminant source characteristics.

There are two goals in two-objective optimization: 1) fast convergence to the Pareto-optimal set, and 2) maintaining diversity in solutions of the Pareto-optimal set. NSGA-II (Deb, 2002) is a fast multi-objective optimization method based on the GA. In this method, elitism is observed since it is proven that elitism can significantly speedup the performance of GA, by preserving the good solutions from one generation

to the next, once they are found. Moreover, a crowding distance comparison approach is included in this algorithm, which preserves the diversity in the Pareto-optimal set. For details, readers are referred to Deb (2002).

In the proposed linked simulation-optimization based source identification methodology, the population of candidate source characteristics is generated by the algorithm contained in NSGA-II. Then the corresponding objective values are estimated using the concentrations estimated by MODFLOW and MT3DMS numerical simulation models, which are linked externally to the optimization model. Then the population is sorted based on the non-domination. Next, a new population is generated using selection, crossover and mutation operators. This process is continued until the specified stopping criteria are satisfied. Figure 5.1 demonstrates the schematic diagram of the NSGA-II linked simulation-optimization procedure.

5. 3. Performance evaluation

Performance of the proposed methodology is evaluated and compared with the performance of the single objective ASA-based linked simulation optimization contamination source identification (Jha & Datta, 2013). ASA is an improved variant of SA optimization search algorithm, by automating the adjustment of parameters controlling the temperature schedule and random step selection. ASA performance is less sensitive to user-defined parameters compared with SA. The efficiency of the ASA and GA-based contaminant source identification models were compared in Jha and Datta (2013). Their single objective models aimed to optimize the same objective function defined in Eq. 5-1. They tested both the ASA and GA based models using

error free and erroneous concentration observation data and hydrogeological parameter values. They concluded that the source characteristics identified by ASA are closer to actual source characteristics when compared to the results obtained by GA, even when measurement and parameter uncertainty exists. Moreover, ASA is computationally more efficient than GA. Based on these considerations, the authors selected ASA as the efficient optimization algorithms to solve a single objective source identification model. The ASA-based single objective model solutions are compared with the solution results obtained using the NSGA-II algorithm.

5.4 .1 . Study area

An illustrative study area is used to evaluate the performance of the proposed methodology. The synthetic concentration measurement data, obtained using simulation models with specified physical parameters, facilitates the evaluation of the methodology under known uncertainty conditions, without having to consider the unknown reliability of the field data. This is necessary and applicable only for performance evaluation purposes.

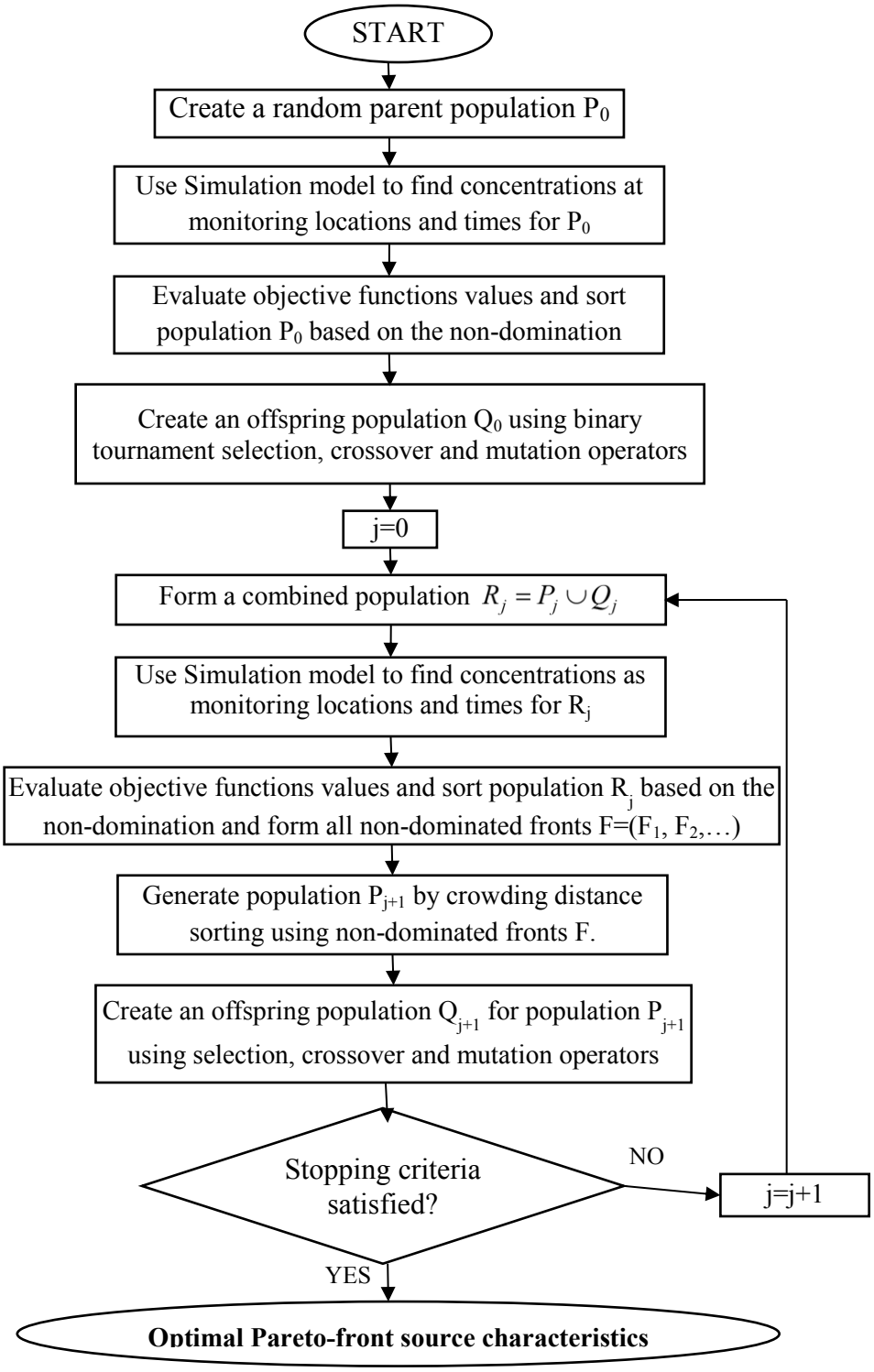


Figure 5.1 Schematic diagram of the NSGA-II based linked simulation-optimization contamination source identification procedure

Figure 5.2 shows the plan view of the illustrative three-dimensional study area measuring 3350m × 2490m × 60m and consisting of two unconfined layers. The boundary to the north is a no-flow boundary and the remaining ones are specified head boundaries corresponding with the average stage of the rivers located at these boundaries. It is assumed the primary influx to the system is through recharge due to rainfall. Also, there are creeks in the area, and two water extraction wells are assumed to be present (shown by filled circles in Figure 5.2). The flow condition is transient. There are two active point contaminant sources and one landfill (distributed contaminant source). A snap shot of the contamination plumes is shown in Figure 5.2. The cell marked with a star is a potential pollutant source location and numbers show the contaminant monitoring locations.

The monitoring locations are selected arbitrarily. Monitoring locations 1-10 are approximately 10 m below the ground surface (layer one) and locations 11-14 are about 30 m below the ground (layer two). The field hydro-geological parameters are given in Table 5.1.

The study period is eight years which is divided into four equal-length stress periods of two years each. All sources are considered as potentially active in all stress periods and the pollutant flux from each of the sources is assumed to be constant over a specified stress period. In the source identification model three potential sources are specified, only two of them being actual sources. The dummy (not actual) source is introduced to evaluate the accuracy of the proposed methodology in correctly identifying actual sources.

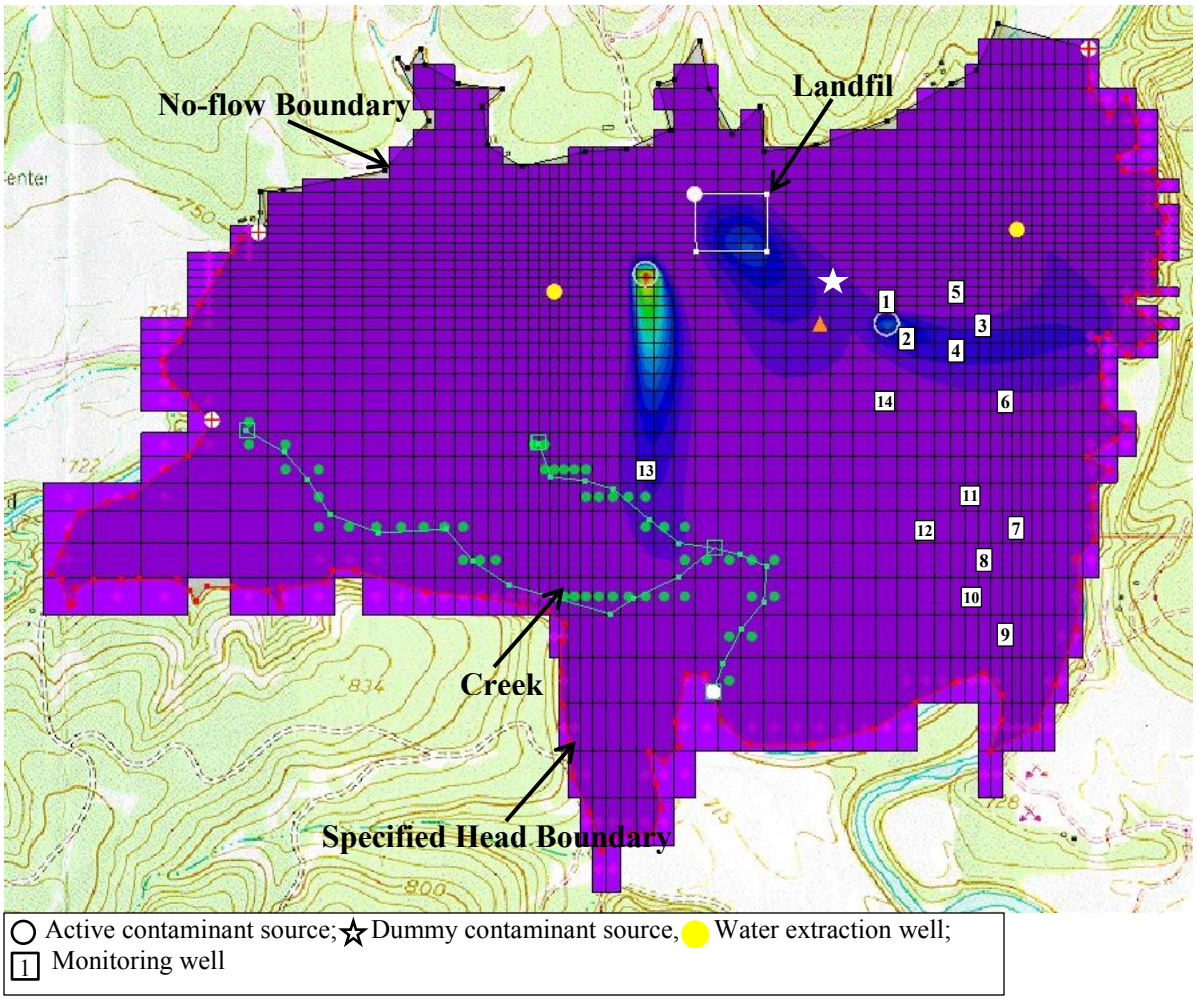


Figure 5.2 The study area.

The source identification decision variables include four variables corresponding to each stress period for all the potential point sources. There are 12 decision variables specifying the point sources and one decision variable which specifies the landfill leachate during the study period. The contaminant concentration monitoring data are collected every six months.

Since, this is an illustrative study area, the actual contaminant source fluxes are known for performance evaluation purposes. Initially, error-free concentration

measurements at the designated monitoring locations and time intervals are generated using the simulation models (MODFLOW and MT3DMS).

Table 5.1 Hydrogeologic parameters for the study area**.**

Parameter	Unit	Value
Number of Cells in x-direction	-	42
Number of Cells in y-direction	-	60
Number of Cells in z-direction	m	2
Horizontal Hydraulic Conductivity	m/d	8
Vertical Hydraulic Conductivity	m/d	3
Specific Storage	1/m	0.2
Porosity	-	0.2
Longitudinal Dispersivity	m	20
Horizontal Transverse Dispersivity	m	2
Vertical Transverse Dispersivity	m	1
Initial contaminant concentration	ppm	0
Diffusion Coefficient	-	0
Upper and Lower Bounds for Source Fluxes	Kg/d	0-100000

5.4 .2 . Error in concentration measurement data

In order to study the effect of contaminant concentration measurement error, the synthetically generated (simulated) concentration measurement values are perturbed using Eq. 5-5.

$$Cobs_{iob}^{k} = Cobs_{error\ free,iob}^{k}(1 + e_{iob}^{k}) \tag{5-5}$$

where $Cobs_{error\ free,iob}^{k}$ is the error-free concentration measurements at monitoring location *iob* and time *k*. e_{iob}^{k} is a measurement error factor for monitoring location *iob* and time *k*, which is randomly generated using a Gaussian probability distribution. To generate the concentration measurements with random measurement errors the mean and standard deviation of the erroneous concentration measurements are required

(Harmel & Smith, 2007). The error-free concentration measurement represents the mean, and the standard deviation, σ is estimated using Eq. 5-6.

$$\sigma = \phi \times Cobs^{k}_{error\ free,iob} \tag{5-6}$$

where ϕ is a fraction between 0 and 1.0 used to generate various degrees of measurement errors.

5.4 .3 . Error in field hydrogeological parameter values

Solution of the flow and transport processes requires the knowledge of various soil hydrogeologic parameters. Considering one of the field hydrogeologic parameters, spatial variations of Hydraulic Conductivity (HC) plays a critical role in transport of contaminants in groundwater systems. Considerable amounts of hydraulic conductivity data are required to obtain a reasonable degree of confidence in simulations of site behaviour. In real-life, generally very limited amounts of HC measurement data are available as inputs to the flow simulation model. Therefore, the flow model uses averaged values to generate the head distribution. This may lead to substantial uncertainty in the source identification process.

In order to test the performance of the proposed methodology under parameter uncertainty, the observed contaminant concentrations are generated using a heterogeneous non-uniform HC field; however, in the source identification program, the flow simulation model uses an average value of HC.

A realistic presentation of a porous medium can include a HC field with several heterogeneous layers. In these cases, the HC is measured at a limited number of locations. Then an interpolation method is required to find the entire HC field. Ordinary

Kiriging (OK) (Cressie, 1988) is a geostatistic interpolation method which depicts the spatial auto-correlation of the measured sample points using a variogram. It krigs the measured HC value to the entire field using a model variogram.

In this study, to generate the heterogonous non-uniform HC field, a limited number of values are generated as measured HC values at specific locations. Following the work of Freeze (1975), these values are generated randomly using a log-normal distribution. Then OK with a known model variogram is used to define the entire HC field.

5.4 .4 . Performance evaluation criteria

Evaluating the performance of the contamination source identification approaches in identifying accurate source flux histories, the errors in estimating source fluxes are calculated. Average Normalized Absolute Error of Estimation (*%ANAEE*) is the measure of errors in estimation of source fluxes, as presented in Eq. 5-7.

$$\%ANAEE = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left|(q_i^t)_{est} - (q_i^t)_{act}\right|}{\sum_{i=1}^{N}\sum_{t=1}^{T}(q_i^t)_{act}} \times 100 \qquad (5\text{-}7)$$

where $(q_i^t)_{est}$ and $(q_i^t)_{act}$ are the estimated and actual contaminant source fluxes at potential source location *i* and stress period *t*, respectively. Smaller *%ANAEE* values mean that the utilized source identification algorithm is able to recover source fluxes with less associated error. The results of both two-objective and single-objective approaches after 50,000 times of simulation model execution are reported, to have a fair

comparison between the performances of the two methods. The population size for NSGA-II is 50, thus it is executed for 100 generations. Note that ASA is not a population-based simulation algorithm.

The final solution results of NSGA-II are obtained as a non-dominated or, Pareto-front. Any of the points located on the Pareto-front can be chosen to be the optimum result; however, further considerations are necessary in order to select one solution to the problem, as estimates of contaminant source release histories. Therefore, the rule in Eq. 5-8 is utilized choosing a single solution from the Pareto-front. An averaging criterion is utilized to obtain the single solution based on the non-dominated set identified. The method is explained below.

$$IF \begin{cases} Q_1^{FOBJ1} < fobj1_i < Q_3^{FOBJ1} \\ AND \\ Q_1^{FOBJ2} < fobj2_i < Q_3^{FOBJ2} \end{cases} \quad THEN\ \eta_i = 1 \quad ELSE \quad \eta_i = 0 \quad i = 1,...,50 \tag{5-8}$$

where $fobj1_i$, $FOBJ1$ and $fobj2_i$, $FOBJ2$ are the i[th] final (non-dominated) objective function value and a vector containing all the final objective function values of the resulting Pareto-front associated with objective 1 and 2, respectively. Q_1^{FOBJ1}, Q_3^{FOBJ1} and Q_1^{FOBJ2}, Q_3^{FOBJ2} are the first and third quartiles of the non-dominated objective function values for objectives 1 and 2, respectively. η_i is a flag which shows if the decision variable values associated with the i^{th} point on the Pareto-front (i=1,…,50 as number of solutions in the non-dominated front is the same as population size) is used to estimate a single outcome for the NSGA-II. The average decision variable value corresponding to the Pareto-front points with flag 1, is the single outcome of the NSGA-II based method. Figure 5.3 is used to demonstrate the procedure for estimating

a single solution to the problem using the Pareto-front. The continuous and dashed lines show the first and third quartiles of the non-dominated objective function values 5-1 and 5-2, respectively. The flag values for points inside the square are 1 and for the rest of the points are 0.

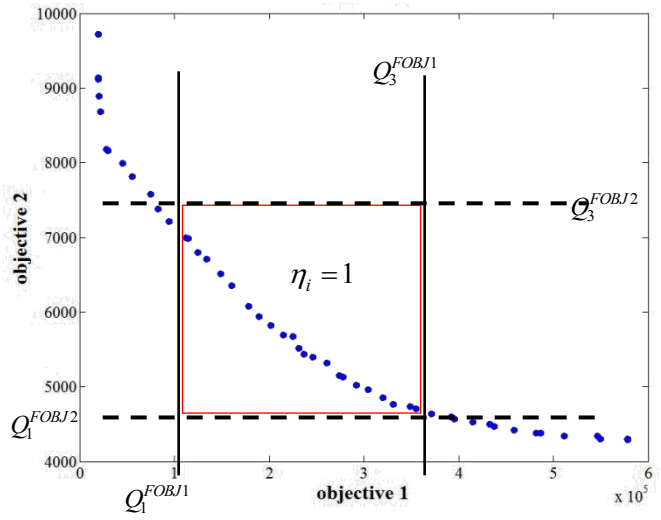


Figure 5.3 An example of Pareto-front and the region used to find a single solution of the two-objective NSGA-II based approach.

5. 4. Results

In this section, the results of utilizing the following three approaches are presented and compared.

Method 1 (M1): The two-objective NSGA-II-based source identification approach using objectives 1 and 2 (Eqs. 5-1 and 5-2).

Method 2 (M2): Single-objective ASA-based source identification approach using objective 1 (Eq. 5-1).

Method 3 (M3): Single-objective ASA-based source identification approach using objective 2 (Eq. 5-2).

5.4 .1 . Error-Free pollutant concentration measurements

M1, M2, and M3 are utilized to find source characteristics, initially using error-free contaminant measurement data. The results are recorded after 50,000 simulations, and are presented in Table 5.2.

Table 5.2 Contaminant source identification outcomes.

	Error Free			**Erroneous, ϕ =0.1**			**Erroneous, ϕ =0.2**		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
%ANAEE	38.1	36.8	37.7	48.2	68.1	45.7	60.1	107.7	53.9
Objective Function 1	9	8	-	1.9E+4	1.6E+4	-	2.1E+4	1.7E+4	-
Objective Function 2	1	-	1	756	-	440	1.1E+3	-	1.0E+3
	Erroneous, ϕ =0.3			**Erroneous, ϕ =0.4**			**Erroneous, ϕ =0.5**		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
%ANAEE	50.1	93.3	56.0	61.5	85.2	69.9	89.1	120.5	97.3
Objective Function 1	2.1E+4	1.7E+4	-	2.3E+4	1.8E+4	-	8.2E+5	2.0E+4	-
Objective Function 2	1.8E+3	-	1.5E+3	1.8E+3	-	1.6E+3	2.6E+3	-	2.4E+3

The *%ANAEE* associated to M1 is 38.1%, which is almost the same as the values obtained by M2 (36.8%) and M3 (37.7%). All approaches showed acceptable performance with respect to the landfill. Theoretically, all utilized methods should be able to find the actual source fluxes; however, the source 1's retrieved fluxes seem too far from the actual values. Monitoring point 13 is the only location which is able to capture release contaminants from source 1 and also is located too far from this source. Therefore, the inappropriate selection of monitoring locations with respect to source 1 is the main reason for inaccuracy in retrieving the release fluxes at source 1. Since the final objective function values related to all three methods are almost zero, 50,000 algorithm simulations seems to be an appropriate number to reach the near-optimal

result by these methods. The estimated values for objectives 1 and 2 associated with the 50th and 100th generation of M1 are presented in Figure 5.4. The scatter plot shows that the proposed approach does not result in a two-objective Pareto-front, and the search algorithm is converging toward a single optimum solution, in the case of error-free measurement data. This is expected, as the two objective functions considered here, may lose their conflicting nature when the idealistic scenario of error-free contaminant measurements is assumed to prevail.

Figure 5.4 The two-objective NSGA-II source identification results using error free measurement data.

5.4 .2 . Erroneous pollutant concentration measurements

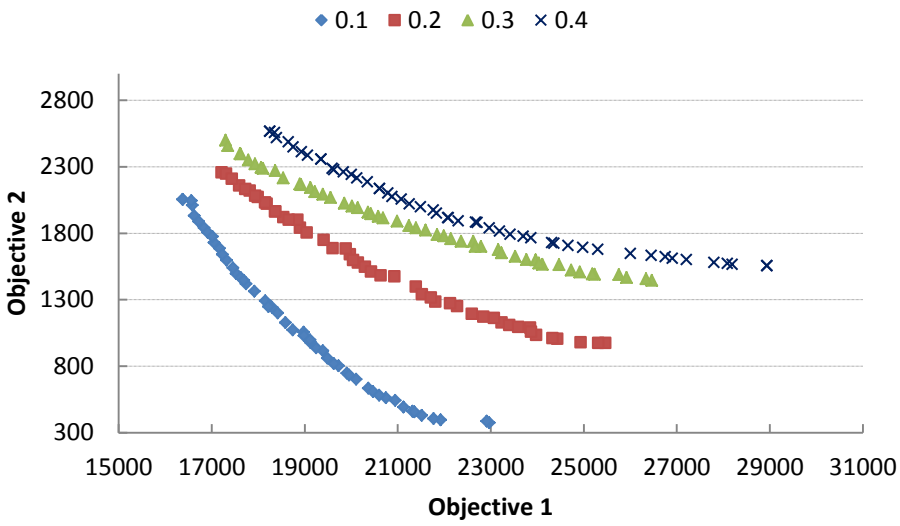
For the illustrative study area, the erroneous measured concentrations are generated using Eq. 5-5. Different levels of error intensity were added to the error-free measurement data. The error level intensity is controlled by the fraction ϕ in Eq. 5-7, which defines the standard deviation of the random normally distributed error values. The results obtained using M1, M2, and M3 associated with ϕ=0.1, ϕ=0.2, ϕ=0.3, ϕ

=0.4, and ϕ=0.5 error intensities, respectively, are presented in Table 5.2. Figure 5.5 shows the objective 1 and 2 values associated with the 100th generation of the solution results estimated by M1.

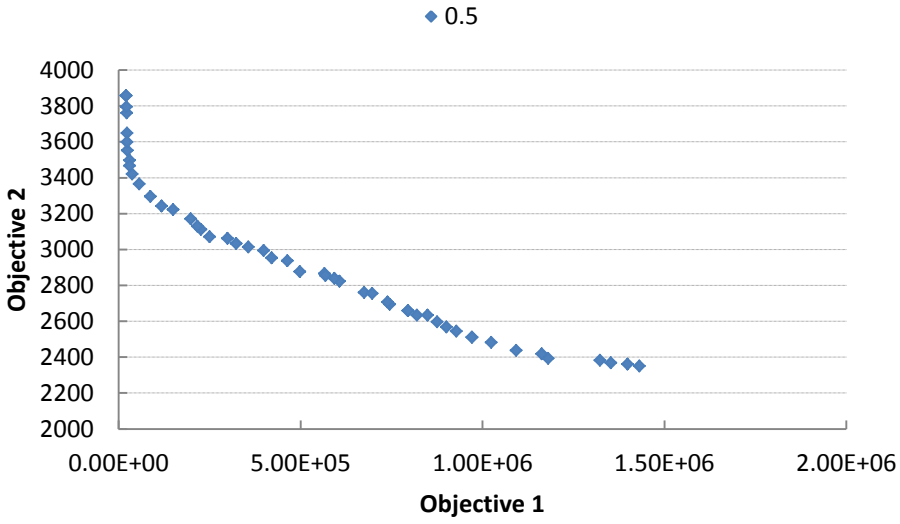
5.4 .3 . Parameter value uncertainty

For the performance evaluation purpose, the HC values are assumed to be known at three locations, denoted by stars as shown in Figure 5.6. At each location, HC values are known at two elevations, corresponding to two layers of the study area. HC values are generated randomly using a log-normal distribution with mean of 20 m/day and standard deviations of 0.2. In total 6 values (3 locations x 2 layers) are generated analogous to the measured HC values in the field. Using OK, the known values are interpolated throughout the study area. The model variogram is spherical with a range of 1000, sill of 4.5, and nugget of 0 (Cressie, 1988). Figure 5.6 illustrates the HC contours in the first layer of the study area.

The contaminant observation data are generated using the heterogeneous non-uniform field. Then measurement errors are added to the generated contamination observations using Eq. 5-5 and ϕ=0.3. M1, M2 and M3 are utilized to find the contaminant source characteristics using these erroneous contaminant observation data. Note that in the simulation model linked to the optimization algorithm, the study area is considered homogenous with HC equals 20 m/day. The solution results are presented in Table 5.3.



(a)



(b)

Figure 5.5 The two-objective NSGA-II source identification results using erroneous measurement data; a) ϕ=0.1, ϕ=0.2, ϕ=0.3, and ϕ=0.4, b) ϕ=0.5

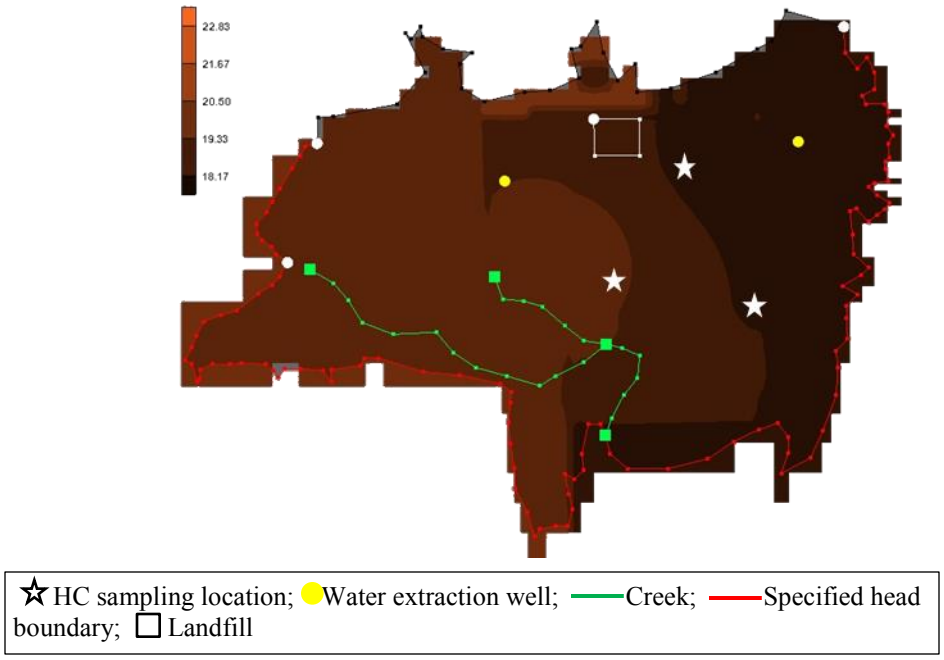


Figure 5.6 Hydraulic conductivity counters. Unit is m/day

5. 5. Discussion

Comparing the estimated *%ANAEE* values, M1 and M3 converged to source characteristics with similar associated *%ANAEE* values when using erroneous concentration measurement data with low error levels (ϕ=0.1 and ϕ=0.2); however, for the cases with moderate to high error levels (ϕ=0.3 and ϕ=0.4), M1 delivers relatively more accurate results (smaller *%ANAEE*) compared with M3.

M1 (using two-objective NSGA-II approach) performs better compared with M2 in all cases associated with erroneous measurements. Objective 1 (Eq. 5-1) uses the observed concentration value to normalize the objective function. In the case of uncertainty, often it is not possible to find a solution which satisfies all the observed values. Having the observed concentration in the denominator (Eq. 5-1), the optimization algorithm based on this equation tends to match smaller concentrations

(larger objective function improvement). This is more evident in the results associated when ϕ=0.5 (Table 5.4). Most of the fluxes estimated by M2 are zero or very low, since the algorithm tries to only match very low observed concentrations.

Table 5.3 The estimated contaminant source release fluxes in the presence of parameter and concentration measurement uncertainties

	%ANAEE	Objective function 1	Objective function 2	Source	Flux (kg/day) $\times 10^3$			
					Stress Period 1	Stress Period 2	Stress Period 3	Stress Period 4
M1	58.0	2.5E+4	1.3E+3	**1**	67	34	57	70
				2	58	27	16	0.3
				3	0	0	3	1
				Landfill	20			
M2	128.8	2.0E+4	-	**1**	62	12	67	99
				2	2	8	25	0.2
				3	0	0	3	35
				Landfill	7			
M3	69.5	-	1.4E+3	**1**	8	14	48	70
				2	75	22	15	0.2
				3	0.1	0	0	0.5
				Landfill	25			

Moreover, in the case of low measurement error levels (ϕ=0.1 and ϕ=0.2), the *%ANAEE* associated with M1 is 5% and 11% respectively larger than the corresponding value estimated by M3. By examining the *%ANAEE* values, it can be concluded that using M1 in the case of low measurement errors is not advantageous compared with the single-objective approach The estimated *%ANAEE* obtained using M1 is 41% and 79% i.e., smaller than the estimated values obtained by using M2. Moreover, the two-objective approach results are delivered as a Pareto-front. In order to compare the results with the single-objective methods, Eq. 5-8 is used to find a single

solution. The Pareto-front contains some solutions with individually larger and smaller *%ANAEE* values compared with the values reported in Table 5.2.

Figure 5.7 presents the *%ANAEE* values estimated for the 100th generation of the results obtained by M1 in the form of a box-whisker plot. Fifty individuals are in each generation. Figure 5.7 shows the maximum, 1st quartile, median, 3th quartile and minimum errors associated with these 50 possible contaminant source characteristics obtained from the last generation of M1. In this figure, the black dots show the values presented in Table 5.2 for the two-objective approach M1. As the results for M1 presented in Table 5.2 are averaged values, the Pareto-front actually contains solutions with smaller *%ANAEE* than the one presented in Table 5.2. For instance, in the case of ϕ=0.2 the smallest estimated *%ANAEE* obtained using M1 is 52.03% which is smaller than 53.93% reported for M3 in Table 5.2. Therefore, the actual performance of the two-objective approach may not be accurately presented in the results shown in Table 5.2.

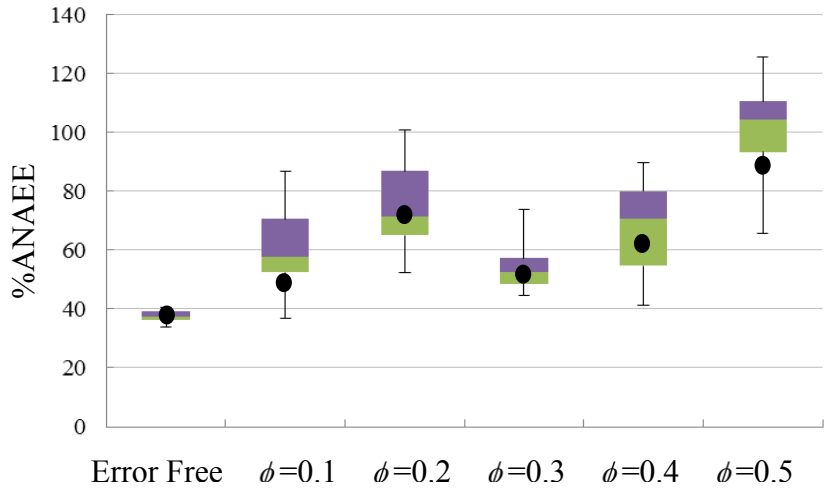


Figure 5.7 The box-whisker plot of the %ANAEE values associated with the 100th generation of M1

Table 5.4 The contamination source release fluxes

	Source	Flux (kg/day) $\times 10^3$			
		Stress Period 1	Stress Period 2	Stress Period 3	Stress Period 4
Actual	**1**	50.	60	15.	0.5
	2	70.	30	10	0
	3	0	0	0	0
	Landfill			20	
Erroneous ϕ=0.5 M1	**1**	59	51	79	44
	2	26	12	19	0.3
	3	6	2	3	23
	Landfill			8	
Erroneous ϕ=0.5 M2	**1**	3	63	70	54
	2	0	0	0.1	0.7
	3	0	0	29	2
	Landfill			0	
Erroneous ϕ=0.5 M3	**1**	4	46	97	95
	2	45	28	17	0
	3	5	0	0.9	4
	Landfill			13	

Actual aquifers are often heterogonous and non-uniform. Therefore, the proposed methodology was tested in a heterogeneous and non-uniform hydraulic conductivity field. Adding the parameter value uncertainty to the contaminant concentration measurements uncertainty, should result in more uncertainty in the source identification procedure. Applying three approaches (M1-M2-M3), the estimated *%ANAEE* values are 58.0%, 128.8%, and 69.5% for M1, M2, and M3, respectively. Therefore, in the case of parameter uncertainty the two-objective approach demonstrates better performance in retrieving contaminant source fluxes.

The box-whisker plots are representative of levels of non-uniqueness or/and possibly the ill-posed nature associated with this illustrative source characterization problem. The range between maximum and minimum *%ANAEE* can be used as a

measure of the non-uniqueness in the problem. These values can also represent the ill-posed nature of the inverse problem to a certain extent. The measured concentrations are perturbed using randomly distributed errors (Eq. 5-4). Therefore, the trend in level of uncertainty shown in Figure 5.5 does not correlate very well with the error levels shown along the horizontal axes of Figure 5.7.

Results with associated relatively small *%ANAEE* exist in the Pareto-front. In real life contamination problems, the actual contaminant source fluxes are unknown. Therefore, it is not possible to find the best solution (on the Pareto-front) using *%ANAEE*. In this study uncertainty is attributed to the lack of knowledge about the correct values for the measurements, and this uncertainty is reducible in contaminated groundwater sites by acquiring more data. In addition the problems due to non-uniqueness and ill-posed nature of the source characterization inverse problem can be addressed by designing efficient and effective concentration measurement networks in the field. This is an important issue that must be addressed, and this illustrative problem highlights the need for proper and efficient contaminant concentration measurement network designs to improve the accuracy and reliability of the source characterization. Therefore, as an extension to this study, a systematic and dedicated monitoring network design methodology is discussed in the next chapter. This methodology may lead to reduction in the uncertainty and help to find more accurate solutions on the Pareto-front.

5. 6. Conclusions

Accurate identification of contaminant source characteristics is required for effective and efficient pollution management and efficient design of contamination remediation strategies. The contaminant concentrations collected at monitoring locations are used to identify the contaminant sources in terms of location, activity duration, and source release fluxes; however, there are often various sources of error associated with the measured contaminant concentrations which are called uncertainty. The presence of such uncertainties which cannot be easily quantified using specified probability distribution functions, has an adverse effect on the accuracy of the contaminant source characterization process.

Generally, contaminant source fluxes are retrieved using a linked simulation-optimization approach which aims to minimize the difference between observed and simulated concentrations. The simulated concentration is estimated using candidate source characteristics and the linked flow and transport simulation models. In this study, a two-objective approach is proposed. The first objective minimizes the normalized difference between the estimated and measured concentrations at monitoring locations and times. The second objective does not use any normalizing factor. NSGA-II is utilized as an efficient multi-objective optimization algorithm.

Ideally, when there is no associated error, there is a unique solution to the source identification procedure. Therefore, the two-objective approach should converge to a single solution using error-free measurements. This solution is the same solution obtained by the single objectives being used one at a time. The solution to a source identification problem is non-unique and to a certain extent ill-posed. Therefore,

various solutions, each with a possibility of being the true source characteristics, may be achieved. When the contaminant concentrations are erroneous, it is not possible to match all the observed and simulated concentrations, and the non-uniqueness problem becomes more evident. In the two-objective approach, the first objective function is normalized using observed concentrations. This objective function emphasizes matching smaller observed concentrations (larger objective function improvement). The second objective function, which is not normalized, tries to match the high concentrations. The two-objective approach can find the possible solutions as a Pareto-front.

The performance of the proposed methodology was evaluated for an illustrative study area. The two-objective solution results are presented as an optimum Pareto-front. To be able to compare the performance of the two-objective and single-objective methods, the solutions which are located between the first and third quartile of objective functions 1 and 2 values, on the final Pareto-front, are averaged. It is demonstrated that the resulting single solution obtained by the two-objective approach has the capability to retrieve the source fluxes more accurately in the case of moderate to high measurement error levels.

In real-world problems, actual aquifers are often heterogeneous and non-uniform. In real life, generally very limited amounts of hydrogeological data are available as inputs to the simulation models. The models simulating the physical processes are often based on averaged values. To test the performance of the proposed methodology in real fields, the method has been applied to a contaminated aquifer with heterogeneous and non-uniform hydraulic conductivity values. It is demonstrated that in heterogeneous

fields, the two-objective approach is able to find more accurate source characteristics compared with single-objective approaches while using erroneous contamination measurement data. These evaluations also highlighted the need for appropriate design of contaminant concentration monitoring network which can improve the source characterization results.

In the next chapter, a new contaminant monitoring network design methodology is presented. This proposed methodology is designed for increasing the accuracy and reducing uncertainty in the contaminant source characterization process.

6. SEQUENTIAL MONITORING NETWORK DESIGN METHODOLOGY INCORPORATING UNCERTAINTY AND REDUNDANCY REDUCTION FOR IMPROVED CONTAMINANT SOURCE IDENTIFICATION

6.1. Introduction

The characteristics of contaminant sources are identified using available sparse and often erroneous contaminant concentration measurements collected in the field. These measurements are obtained using an arbitrary or predesigned monitoring network (Amirabdollahian & Datta, 2013). Due to the complexity of contamination movement in the groundwater and the large scale of the contaminated sites, data collecting can be very difficult and costly. Therefore, it is vital to design contaminant monitoring networks. In this chapter a new monitoring network design methodology is presented which has two aims: (1) minimize uncertainty in recovered contaminant source characteristics; and (2) reduce redundancy in collected contaminant observation data.

One of the difficulties in identification of contaminant sources is uncertainty. There are various sources of uncertainty associated with the available field hydrogeological data, flow boundary condition, site geological properties, etc. The measurement error in contaminant observation data is another source of uncertainty which makes the contaminant source characterization process more difficult. One of the objectives of the proposed monitoring network design approach is to reduce the effect of these uncertainties and improve the accuracy of recovered source characteristics.

Moreover, drilling observation wells and collecting and analyzing the contaminated groundwater involve costs. Also, maximum coverage of the contamination plume should be provided using the fewest possible monitoring locations. Therefore, the second objective of the proposed monitoring network design is to minimize the redundancy in captured information using a predefined number of monitoring locations. The number of monitoring locations is explicitly selected based on the budgetary constraints.

Since the pollution transport process is dynamic in time and space, a sequential monitoring network design and optimal contaminant source characterization approach is required. The initial contaminant source characteristics are estimated using contaminant observation data collected at initial arbitrary existing monitoring locations. The two-objective source characterization methodology presented in Chapter 5 is utilized. The set of source characteristics identified as the near optimal values are used to design a monitoring network. Sequentially, the new selected monitoring locations are used in addition to existing ones to collect additional contaminant observation data, once the new monitoring wells are implemented. A better estimation of contaminant source characteristics is obtained by using newly collected and available observation data from the designed monitoring network.

In this chapter, first the proposed objective functions related to the uncertainty and redundancy reduction are presented. Next, the Non-dominated Sorting Genetic Algorithm II (NSGAII, (Deb, 2002)) optimization algorithm is utilized to find optimal monitoring locations among potential locations for each sample collection time is briefly discussed. The sequential monitoring network and contaminant source

characterization methodology is then presented. It is followed by discussion of the performance evaluation results for an illustrative contaminated groundwater aquifer study area.

6.2. Methodology

The proposed sequential contaminant source identification and monitoring network design methodology has three important components. The first two components are related to the uncertainty and redundancy reduction objective functions. The third component constitutes of the sequential procedure of designing and implementing the monitoring network while utilizing the new measurement data in the form of feedback information.

6.2. 1. Uncertainty reduction objective function for monitoring network design

The first proposed objective function in designing the monitoring network is to reduce uncertainty in estimated contaminant source characteristics. The methodology presented in this chapter is integrated with the contaminant source identification methodology presented in Chapter 5. Therefore, the monitoring network design is performed using available preliminary information about the contaminant sources. The vector containing the contaminant source characteristics are presented in Eq. 6-1.

$$\begin{aligned} &Q = (\vec{q}_1, \vec{q}_2, ..., \vec{q}_p) \quad p = 1,...NP \\ &\vec{q}_p = (q^1_{1,p}, q^2_{1,p}, ..., q^t_{s,p}) \quad s = 1,..,NS, \quad t = 1,...,ST \end{aligned} \tag{6-1}$$

where $\vec{q}_p$ is p^{th} set of available estimate of source release fluxes. NP is the total number of available set of estimated contaminant release fluxes. $q^t_{s,p}$ is the p^{th} estimated release

flux at candidate source location s and stress period t. *NS* and *ST* are total number of candidate source locations and stress periods, respectively. *Q* is the vector containing all available estimates of contaminant release fluxes. In this chapter vector *Q* is the result of initial execution of the source identification model; however, if some initial or prior information is available about the contaminant sources, they can be incorporated as Q for the purpose of monitoring network design.

The first monitoring network design objective function implicitly aims to reduce the uncertainty in the identified contaminant source characteristics using Eqs. 6-2 to 6-4. The chosen locations are such that the summation of the deviations between the temporal average concentration value and the estimated concentration at a specific time and location is maximized.

$$Objective\ function\ 1: OBJ1 = Max\left(\sum_{i=1}^{N}\sum_{k=1}^{NT} X_i \frac{\sqrt{\sum_{p=1}^{NP}\left(Cest_{i,k}^{p} - Cest_{i,k}^{ave}\right)^2}}{NP-1}\right) \tag{6-2}$$

Subject to:

$$Cest_{i,k}^{p} = f(D, K, \theta, \vec{q}_p) \tag{6-3}$$

$$\sum_{i=1}^{N} X_i \le Max_{Monitoring} \tag{6-4}$$

where *N* and *NT* are the number of candidate monitoring locations and monitoring design time steps, respectively. $Cest_{i,k}^{p}$ is the estimated concentration at monitoring location *i* and at time *k* using $\vec{q}_p$ as candidate solutions for the source characteristics. $Cest_{i,k}^{ave}$ is the estimated concentration at monitoring location *i* and time *k*, averaged over all possible sets of source release histories. X_i is a flag which has a value of one when

the candidate location i is selected, otherwise it is zero. Eqs. 6-3 and 6-4 are constraints. Constraint 6-3 represents the groundwater flow and transport process simulation. Eq. 6-4 limits the total number of selected monitoring locations to $Max_{Monitoring}$. $Max_{Monitoring}$ is generally selected based on the budgetary constraints.

Objective function 1 focuses on locations where the standard deviation of estimated concentrations would be maximum. By selecting the monitoring locations based on this objective function, the source identification model would be able to differentiate between candidate source characteristics $\vec{q}_p$, in order to obtain an optimal solution. The resulting choice of the optimal monitoring network would reduce uncertainty in prediction of plume movement and improve the reliability of estimated contaminant source characteristics.

6.2. 2. Redundancy reduction objective function for monitoring network design

Redundancy in monitoring network design results in economic inefficiency of the monitoring process. Therefore, redundancy reduction is an important factor in selecting monitoring locations in contaminated aquifers (Dhar & Datta, 2010). The second objective function is aimed at reducing redundancy in the designed monitoring network.

The proposed objective function incorporates the Inverse Distance Weighting (IDW) method for spatial interpolation of concentration measurements. The spatial interpolation is required to estimate the concentration at all unmonitored locations. The choice of interpolation scheme is an important factor. Since the number of available

monitoring locations is limited and spatial measurement data are sparse, a geostatistical interpolation scheme is not a suitable option. Therefore the IDW method is selected in this study. In this interpolation method, the interpolated values are dependent on the distance between monitored and unmonitored locations. The details of IDW method are presented in Chapter 3. The redundancy reduction objective function is defined below:

$$Objective\ function\ 2: OBJ\ 2 = Min\left(\sum_{i=1}^{N}(1-X_i)\sum_{k=1}^{NT}\sum_{p=1}^{NP}\sqrt{\left(Cest_{i,k}^{p}-C\,int_{i,k}^{p}\right)^2}\right) \tag{6-5}$$

Subject to:

$$C\,\mathrm{int}_{i,k}^{p} = X_i\,Cest_{i,k}^{p} + (1-X_i)\frac{\sum_{l=1}^{TotalN} w_{l,i}\,Cest_{l,k}^{p}}{\sum_{l=1}^{TotalN} w_{l,i}} \tag{6-6}$$

$$w_{l,i} = 1/d(x_i,x_l)^2 \tag{6-7}$$

$$Cest_{i,k}^{p} = f(D,K,\theta,\vec{q}_p) \tag{6-8}$$

$$\sum_{i=1}^{N} X_i \le Max_{Monitoring} \tag{6-9}$$

where $C\,\mathrm{int}_{i,k}^{p}$ is the interpolated contaminant concentration at potential monitoring location i and time k using candidate source characteristics solutions $\vec{q}_p$, which is estimated by Eq. 6-6. Constraints 6-6 and 6-7 basically represent the IDW interpolation method. When a monitoring location i is selected ($X_i = 1$), the interpolated concentration is the same value as the estimated concentration; however, the spatial interpolation method is utilized for estimating the concentrations at potential monitoring locations which are not selected (X_i=0). The estimated concentrations at

newly selected and existing monitoring locations are interpolated to compute the concentration at those potential monitoring locations which are not selected as optimal locations. Total$N$ is the total number of monitoring locations including the newly selected and existing ones. $w_{l,i}$ is a coefficient estimated based on the inverse distance of potential monitoring locations i and l. x_i and x_l are the three-dimensional coordinates of potential monitoring locations i and l, respectively.

Objective function 2 aims to select monitoring locations that provide better interpolated values for unselected locations among all potential monitoring wells. In more detail, the estimated concentrations (using the transport simulation model) at all potential monitoring wells are compared with interpolated values at unselected well locations. Minimizing objective function 2 ensures that the selected monitoring wells can relatively provide (through interpolation) accurate information about unselected wells. The redundancy reduction objective function ensures that the selected contaminant observation information at monitored locations provides acceptable coverage for the unmonitored locations. This objective function aims to minimize the possibility of having redundant contaminant observation information from the designed monitoring network.

6.2. 3. Sequential two-objective monitoring network design methodology

The proposed monitoring network design methodology has a two-objective optimization framework. The objective functions are defined as Eqs. 6-2 and 6-5. The constraints are defined as Eqs. 6-3 and 6-4 and Eqs. 6-6 to 6-9. The flow and transport simulation models are linked to the monitoring network design model using constraints

6-3 and 6-8. In this chapter MODFLOW (Zheng et al., 2001) and MT3DMS are used as the three- dimensional flow and contaminant transport simulating models. The Pareto-front of the optimal monitoring network designs (non-dominated solutions) are found using NSGA-II (Deb, 2002) which is a fast multi-objective optimization method based on the GA. In this method, elitism is observed since it is proven that elitism can significantly speedup the performance of GA, by preserving the good solutions once they are found. Besides this, a crowding distance comparison approach is included in this algorithm, which preserves the diversity in the Pareto-optimal set.

In the sequential contaminant source identification and monitoring network design methodology, the estimated source fluxes using source identification model are used as feedback into the monitoring network design model. Figure 6.1 shows a schematic diagram of the proposed sequential methodology.

The contaminant source identification model first estimates the source fluxes using initial available concentration observation data. The forward contaminant transport simulation model is utilized to predict future contaminant concentration distribution using estimated source fluxes. Next the proposed monitoring network design methodology is utilized to find the suitable monitoring locations for the next monitoring time step. When the new monitoring network is designed, the contaminant concentration observation data are collected at the newly selected locations and the existing ones. The source identification model is executed again using all available observed concentration data, and if the stopping criteria are satisfied, the optimal contaminant source characteristics are achieved. Otherwise a monitoring network is designed for the next monitoring time step and the sequential process continues.

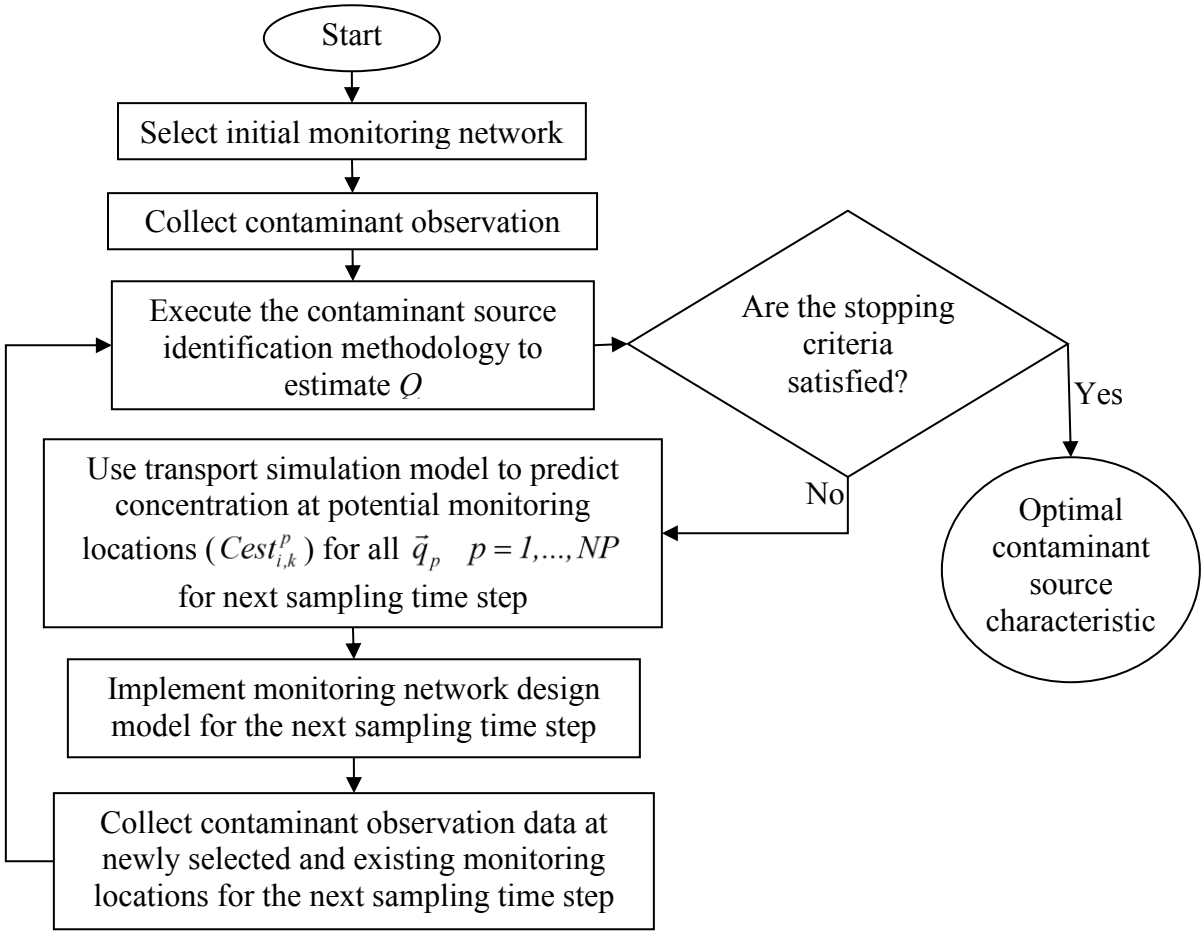


Figure 6.1 Schematic diagram of sequential contaminant source identification and monitoring network design methodology

The sequential process described earlier adds additional information to both source identification and monitoring network design models. The accuracy of estimated source fluxes improves by adding additional optimally selected monitoring location and thereby utilizing concentration measurement data from these selected locations. On the other hand, sequential improved estimates of the source fluxes result in better prediction of future contaminant plume distribution, and consequently enhance the

utility of the designed monitoring network. Therefore, in the sequential feedback methodology, both source identification and monitoring network design models benefit from additional concentration measurements obtained as feedback information from the designed monitoring networks, thus sequentially resulting in improved contaminant source characteristics estimates.

6.3. Performance evaluation of the proposed methodology

The proposed methodology is evaluated using an illustrative three-dimensional contaminated aquifers and synthetic hydrogeologic data. Using an illustrative study area helps to evaluate the performance of the developed methodology properly by eliminating the possibility of attributing the results to any subjective unknown such as quality of the input data, without quantification. In the illustrative study area, all properties including geology, boundary conditions, and hydrogeologic parameter values of the study area are assumed to be known for this performance evaluation purpose only. The performance evaluation is conducted using both error-free and erroneous synthetic contaminant concentration measurement data, in order to evaluate the potential applicability of the methodology in real contamination problems. In the next chapter, the proposed sequential contaminant source identification and monitoring network design methodology is evaluated for a real-life contaminated urban aquifer site.

6.3. 1. Study Area

Details of the illustrative study area are described in Chapter 5 and shown in Figure 5-2. The three-dimensional study area consists of two unconfined layers. There is a creek

and two extraction wells in the area. Two active point contaminant sources and one landfill as a distributed source are included in the model. The initial study period is eight years which is divided into four stress periods. All sources are active during these four stress periods and the pollutant flux from each of the sources is assumed to be constant over a specified stress period. In order to evaluate the model performance one dummy (not actual) source is also introduced as a potential source. Therefore, the source identification decision variables are the contaminant fluxes at three potential source locations for each stress period and the release flux at the landfill. In total there are 13 decision variables.

Fourteen arbitrary locations are selected to represent the initial arbitrary monitoring network. Monitoring locations 1-10 are approximately 10 m below the ground surface (layer one) and locations 11-14 are approximately 30 m below the ground surface (layer two). Estimates for the contaminant release fluxes are identified using the contaminant source identification methodology presented in Chapter 5.

6.3. 2. Optimal monitoring network

The main purpose of this section is to evaluate the performance of the proposed source identification and monitoring network design methodology. A more rigorous application to a real contaminated aquifer site is described in the next chapter. This performance evaluation with synthetic data is limited to one sequence only. The solution results are discussed to show the possibility of using this methodology for improving the source characterization, compared with an arbitrary network for concentration measurement. The monitoring network design model is solved in each design sequence to find optimal monitoring locations for the next sampling time steps.

The study period is extended for another one year (in total the study period is nine years). The contaminant observation data are assumed to be collected every six months. Thus, the monitoring network methodology is utilized to find the suitable sampling locations for the next six-month and 12-month sampling times.

There are 20 potential monitoring wells (n=20). Table 6.1 shows the locations of the potential monitoring wells. The potential monitoring locations are located in both the first and second layers. $Max_{Monitoring}$ is specified as 5. The proposed two-objective monitoring network design model is executed to find the five most suitable monitoring locations out of 20 potential ones. The optimum NSGA-II parameters including population size and maximum number of generations are 100 and 500, respectively. The best set of NSGA-II parameters is selected using sensitivity analysis. The stopping criterion is the maximum number of generations.

6.3. 3. Linked simulation-optimization model for identification of unknown source fluxes using designed monitoring network

The aim of designing a suitable monitoring network is to increase the efficiency of the contaminant source identification model in recovering the unknown source release histories. Introducing additional relevant contaminant concentration observation data to the source identification model will improve the accuracy of estimated source release fluxes; however, it is expected that when the observed concentrations are collected at optimal monitoring locations, the accuracy of estimated source fluxes is substantially improved compared to those obtained using an arbitrary monitoring network.

Table 6.1 Potential monitoring well locations.

Well ID	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Layer	1	1	1	1	1	1	1	1	1	1
Row	25	30	26	33	33	20	27	35	33	24
Column	21	25	14	25	30	51	29	29	34	35
Well ID	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Layer	1	1	1	1	2	2	2	2	2	2
Row	17	29	27	29	32	34	29	31	23	25
Column	35	35	49	43	30	31	29	38	42	36

To evaluate the performance of the proposed methodology, the unknown contaminant release fluxes are estimated using the two-objective source identification model presented in Chapter 5. The objective functions and constraints are defined by Eqs. 5-1 to 5-4. Concentration measurements from the optimal designed monitoring network are used to estimate the pollution sources flux histories. These evaluation results are compared with the recovered source release histories, estimated using five arbitrary monitoring networks. For each arbitrary monitoring network, five monitoring wells are selected randomly from the specified potential well locations. Because in this section the performance of the proposed methodology is evaluated using illustrative study area, the observed aquifer responses are simulated using MODFLOW and MT3DMS simulation models, along with appropriate initial and boundary conditions. In real-field applications the aquifer responses are monitored using collected contaminated water samples from the selected monitoring wells.

In the next stage of the evaluation, the performance of the proposed methodology is tested in the presence of concentration measurement errors. The simulated

concentration using MODFLOW and MT3DMS models are perturbed to represent the effect of random measurement error.

In order to study the effect of contaminant concentration measurement error, the synthetically generated (simulated) concentration measurement values are perturbed using Eq. (6-10).

$$Cobs_{iob}^{k} = Cobs_{error\ free,iob}^{k}(1 + e_{iob}^{k}) \tag{6-10}$$

where $Cobs_{error\ free,iob}^{k}$ is the error free concentration measurements at monitoring location *iob* and time *k*. e_{iob}^{k} the is measurement error for monitoring location *iob* and time *k*, which is randomly generated using a Gaussian probability distribution. The mean and standard deviation of the erroneous concentration measurements are required (Harmel & Smith, 2007). The error-free concentration measurement represents the mean, and the standard deviation. For erroneous data, σ, is estimated using Eq. 6-11.

$$\sigma = \phi \times Cobs_{error\ free,iob}^{k} \tag{6-11}$$

where ϕ is a fraction between 0 and 1.0, which is used to generate various degrees of measurement errors.

The improved efficiency of the source identification model is evaluated, when using sequential concentration measurements from designed monitoring network, and when measurement error exists. The linked simulation-optimization model is solved using perturbed concentration measurement data and the solution is compared with the results obtained using error-free measurement data.

6.4. Results and Discussion

The two-objective optimal monitoring network design model is solved and Figure 6.2 shows the Pareto-optimal front. The Pareto-optimal front includes the optimal networks with respect to both objective functions 1 and 2. The population size is 50. As shown in Figure 6.2, there are only 12 different Pareto-optimal results in the final generation of the NSGA-II optimization model. It is seen that as the value of OBJ2 decreases, the value of OBJ1 decreases and vice versa; however, the objective is to minimize OBJ2, while maximizing OBJ1. This essentially shows the conflicting nature of the two objective functions and their trade-offs. Based on the defined objective functions 1 and 2, better monitoring network design results in maximum values for objective function 1 and minimum values for objective function 2; however, any point on the Pareto-optimal front can be selected as a suitable design for monitoring network based on the decision maker's criteria.

In this study, both the objective functions Eqs. 6-2 and 6-5 are considered to be equally important. Although a compromise solution is not based on quantitative assignment of equal weights to the two conflicting objectives, it is obvious from the Pareto optimal solution shown in Figure 6.2, the point shown with the arrow in Figure 6.2 can be selected to represent a compromise optimal solution. This is based on the fact that at this location further small marginal reduction in objective function 2 (minimization) will result in comparatively large marginal sacrifice in objective function 1 (maximization). Therefore this particular solution is a good candidate for selection to represent the optimal monitoring network design. The objective function 1 and 2 values corresponding to this point are 139.5 mg/l and 720090 mg/l, respectively.

The resulting optimal monitoring locations are wells 9, 11, 12, 14 and 17. Figure 6.3 shows the contaminant concentration at these wells for the next six-month and 12-month sampling times. These concentration values shows an increasing trend with time, and appears to be relevant to source identification as well, as the measured concentrations are substantial.

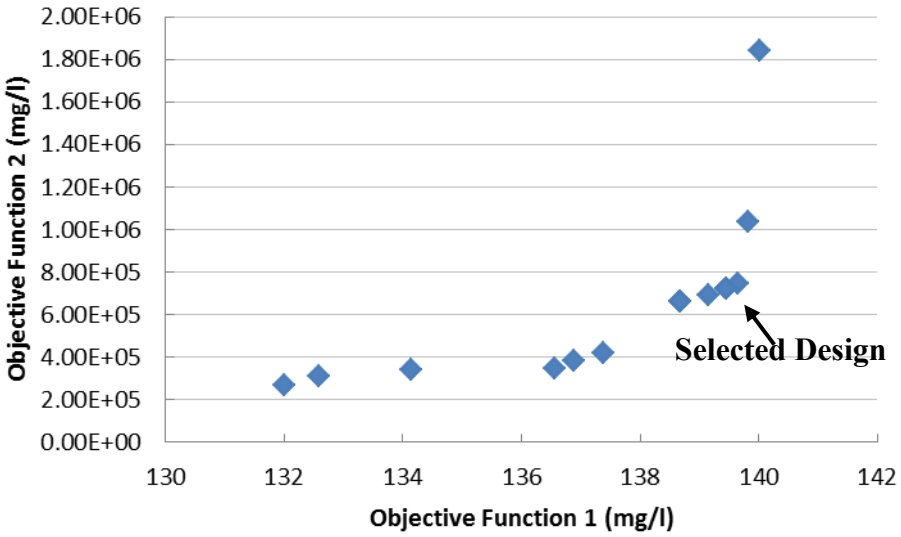


Figure 6.2 Optimal monitoring network design Pareto-front.

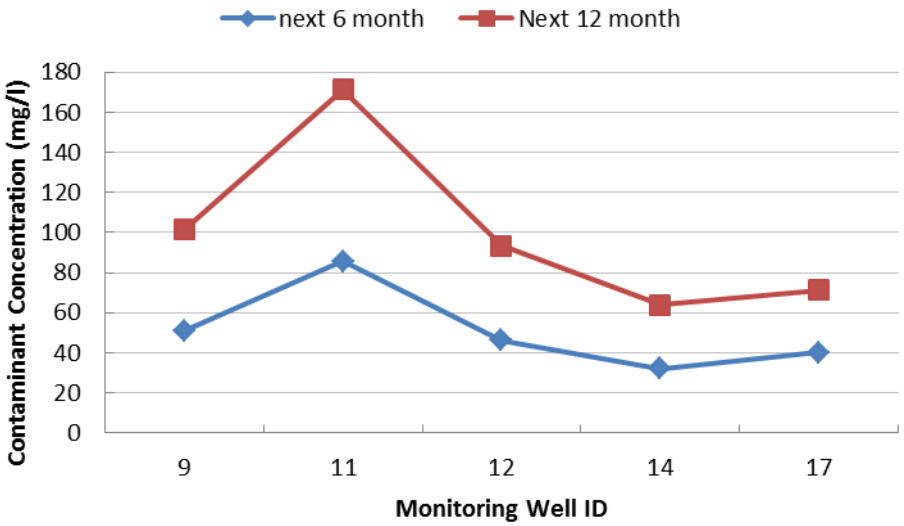


Figure 6.3 Error-free contaminant concentration (mg/l)

In order to evaluate the utility of the monitoring network design methodology for efficient contaminant source identification, the linked source identification model (presented in Chapter 5) is solved to find contaminant source characteristics, by utilizing nine years of contaminant concentration measurement data collected at 14 initial arbitrary monitoring locations, and additional one year of concentration data collected at five newly selected monitoring wells and existing arbitrary ones. The concentrations measured in the next 6 and 12 months after initial source identification are used as feedback information to improve accuracy in the next sequence of the source identification process.

The source identification model presented in Chapter 5 is also based on a two-objective optimization model. An NSGA-II optimization algorithm with population size 50 and 100 maximum number of generations is selected. The stopping criterion is the maximum number of generations. The decision variables are contaminant release flux

for each source location i and stress period j, $q_{i,j}$;i=1,2,3 and j=1,…,4, and $q_{distributed}$ which is the contaminant flux released from the distributed source. Therefore, there are 13 decision variables. The solution results are in the form of a Pareto-optimal front. To be able to compare the estimated source fluxes using designed monitoring network and five arbitrary selected monitoring networks, the estimated value for each decision variable is averaged over all members of the last generation (Pareto-front) of the NSGA-II source identification optimization model.

Figure 6.4 shows the estimated source release fluxes using the five selected monitoring locations in the designed monitoring network (shown as MN in Figure 6.4) and those obtained by using five arbitrary monitoring networks (RM1, RM2, RM3, RM4, and RM5), with error-free concentration measurement data.

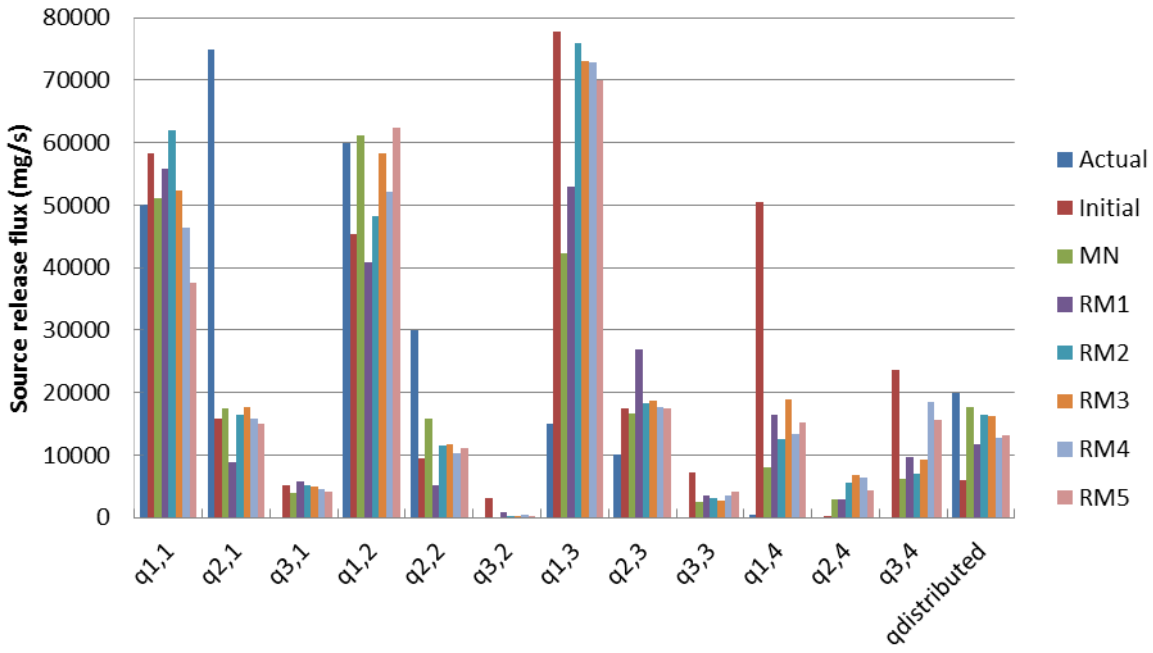


Figure 6.4 Recovered source histories using error-free measurement data and the designed and arbitrary monitoring networks.

In Figure 6.4, the actual release source fluxes are shown along with the initial and estimated ones. The initial source release fluxes are the estimated values in the

previous sequence of solving the source identification model (Q in Eq. 6-1). It is used as the initial population in the source identification models. As expected, almost all estimates of source fluxes show improvement when compared with the initial values utilizing each of the networks MN and RM1-RM5. This is a direct result of adding more information to the source identification model; however, the improvement is greater when using the designed monitoring network (MN). Normalized Absolute Error of Estimation (*NAEE%)* (Eq. 3-15) is estimated for each source and reported in Table 6.2. In (Eq. 3-15) the denominator is the summation of actual source release fluxes. Actual release fluxes for the dummy (not actual) source is zero for all stress periods. It is not possible to estimate *NAEE%* for the dummy source. In Table 6.2, the error associated with the dummy source (source 3) is the summation of the absolute difference between the actual release flux (zero) and the estimated ones over all four stress periods.

Table 6.2 Source release flux estimation error using different monitoring networks**.**

	Initial	**MN**	**RM1**	**RM2**	**RM3**	**RM4**	**RM5**	**Average increase**
NAEE% Source 1	108	29	63	77	64	65	67	36
NAEE% Source 2	76	70	96	78	79	81	78	12
Error (mg/s) Source 3	3.9E4	1.3E4	1.9E4	1.5E4	1.7E4	2.7E4	2.4E4	1.2E4
NAEE% Distributed Source	70	11	41	17	18	36	33	18

As shown in Table 6.2, estimation errors associated with sources one, three and the distributed source improved compared to the initial estimates, by adding more

contaminant observation data to the source identification model using both designed or arbitrary monitoring networks; however, for all sources the improvement is more substantial when using the optimally designed monitoring network.

Therefore, using a designed monitoring network instead of arbitrary selected monitoring locations can improve the efficiency of the source identification model. Also, the designed monitoring network has a direct positive effect on the costs associated with the contaminated site investigation. Moreover, utilizing a designed monitoring network can shorten the long process of contaminated site investigation plan. For instance, in this illustrative study area, in the current sequence of source identification and monitoring network design models, the *NAEE%* associated with the distributed source is 11% which is an acceptable identification tolerance. Therefore, any remediation or management plan related to this source can be taken after this sequence of model executions; although, by using an arbitrary monitoring network design, more sequences of model execution are required to achieve acceptable accuracy. Just as a reminder, each sequence of model solution requires another 12 months (every six-month) contaminant observation data in addition to installing five more monitoring locations, which is time consuming and costly. For source 2, utilizing the arbitrary monitoring networks resulted in increased source flux estimation error compared with the initial solution estimate. This can be related to the location of newly selected monitoring locations with respect to source 2, and also the superimposition of contamination plumes released from other sources. For source 2, utilizing the designed monitoring network reduced the source flux estimation error compared with the initial ones.

In order to evaluate the applicability of the proposed methodologies when measurement error exists, three different sets of erroneous measurement data are randomly generated at the newly selected monitoring locations and existing ones (monitoring locations 1-14) using Eqs. 6-11 and 6-12. The observed concentrations associated with the next 6 and 12 months monitoring times are perturbed and the observation data collected in the first eight years are unaffected. For all three sets of erroneous measurement data ϕ is 0.5. The linked simulation-optimization source identification model is solved using the erroneous measurements. The estimated source release fluxes and the associated estimation errors are shown in Figure 6.5 and Table 6.3, respectively.

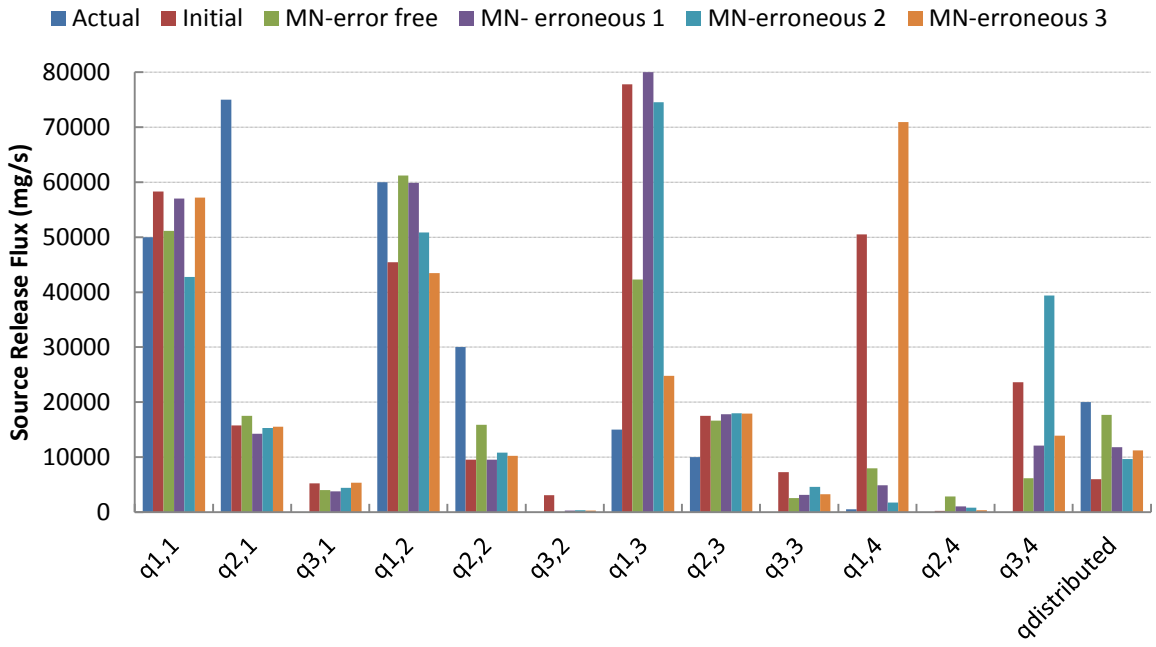


Figure 6.5 Recovered source histories using error-free and erroneous measurement data collected at the designed monitoring networks

Table 6.3 Source release flux estimation error using error-free and erroneous measurement data.

	Initial	Error Free	Erroneous 1	Erroneous 2	Erroneous 3	Average increase
NAEE% Source 1	108	29	61	61	82	39
NAEE% Source 2	76	70	78	76	76	7
Error (mg/s) Source 3	3.9E4	1.3E4	1.9E4	4.8E4	2.2E4	1.7E4
NAEE% Distributed Source	70	11	41	51	43	34

As expected and also based on the results presented in Table 6.3 and Figure 6.5, the source release histories are recovered more accurately when there is no measurement error; however, in real contaminated aquifer cases, the measurement error is one of the most common sources of error in characterization of contaminant sources. On average, using erroneous concentration data results in 39%, 7%, and 34% more error (in terms of *NAEE%*) in estimated source fluxes for sources 1, 3 and distributed respectively, compared with the case where the error-free measurement data are utilized to characterize the sources. With respect to source 2 which is the non-actual or dummy one, the average estimate of flux utilizing the erroneous measurement data increased the source flux estimation error by 1.7E4 mg/s (1.5E3 kg/day) compared to the release flux estimated using the error-free measurement data.

In order to examine the advantage of utilizing the designed monitoring network compared with the arbitrary selected ones, the average source flux estimation error using error-free measurement data collected at arbitrary monitoring locations is compared with the estimated source fluxes using error-free measurement data, collected at the designed monitoring network. The comparison is shown in the last column of Table 6.2. Comparing the corresponding average increase in errors presented in the last

column of Table 6.2 and 6.3, even in the case of measurement errors, the designed monitoring network is more efficient than the arbitrary ones (using error-free measurement) for sources 2. The estimation errors for sources 1 and 3 are relatively similar in both cases. It is expected that adding measurement error to the data collected at arbitrary monitoring locations will increase the corresponding source flux estimation error. For the distributed source estimation, measurement error has a significant effect on the accuracy of estimated source flux. This may also stem from the number of monitoring locations capturing the contaminant released from this source.

6.5. Conclusion

In this chapter a two-objective monitoring network design methodology is proposed which is used sequentially with the linked simulation-optimization source identification model. The monitoring network design objectives are to reduce uncertainty in recovered source histories and to reduce redundancy in selecting the location of contaminant observation wells. By selecting the monitoring locations based on the first objective function, the source identification model would be able to differentiate between possible source releases histories. The second objective function ensures that the selected monitoring well locations are able to provide acceptable coverage of information for the unmonitored locations. An NSGA-II optimization model is used to find the optimal Pareto-front of the monitoring network designs.

The performance evaluations show that the proposed methodology is capable of efficiently identifying the unknown contaminant source characteristics including

locations and release fluxes. A well-designed monitoring network can substantially increase the efficiency of the contaminant source identification model

Using the sequential source identification and monitoring network design methodology, the initial estimates of contaminant source release histories are used as feedback into the monitoring network design methodology. The sequential process adds additional information to both source identification and monitoring network design models in the form of feedback information. The accuracy of estimated source fluxes improves by adding additional optimally selected monitoring location and observation data. On the other hand, improved estimates of the source fluxes result in better prediction of future contaminant plume distribution, and consequently improve the utility of the designed monitoring network.

The performance of the proposed methodology is evaluated for an illustrative contaminated study area utilizing synthetic data. Five monitoring locations are optimally selected out of 20 available potential locations using the proposed two-objective monitoring network design methodology. To show the utility of the proposed methodology, the linked simulation-optimization source identification model is solved using the new optimally designed monitoring network and five randomly generated arbitrary monitoring networks. The solution results show that utilizing the designed monitoring network increases the accuracy and efficiency of the source identification model solutions compared to those obtained with the arbitrary selected locations. To evaluate the performance of the designed monitoring network in the presence of measurement error, the estimated concentrations are perturbed using generated random normally distributed error terms. The linked simulation-optimization source

identification model is solved using erroneous observation data and the results are compared with the estimated source fluxes using error-free observation data. The proposed methodology also exhibits acceptable performance in the presence of measurement error which shows the utility of this method in real-field conditions.

In the next chapter the proposed sequential source identification and monitoring network design methodology is evaluated for a real-life contaminated aquifer.

7. APPLICATION OF THE DEVELOPED SEQUENTIAL SOURCE IDENTIFICATION AND MONITORING NETWORK DESIGN METHODOLOGIES TO FIELD DATA

7.1. Introduction

In this chapter, the proposed sequential contaminant source identification and monitoring network design methodologies presented in the previous chapter are tested for a real urban contaminated study area located in New South Wales, Australia. Due to confidentiality requirements, some information related to the location of this site is not included.

A brief review of the background of the problem and the calibrated flow and transport simulation models are presented. The sequential source identification and monitoring network design methodology are then applied for this site. Finally the solution results are presented and the applicability of the proposed methodology to a real-life contaminated aquifer is discussed.

7.2. Background of the problem

The selected polluted aquifer is part of a suburban town located at the Upper Macquarie Groundwater management area (Jha & Datta, 2015a; Prakash & Datta, 2015) in New South Wales. The detected pollution in this area is BTEX which was first detected as vapour pollution in the basement of buildings. BTEX is an acronym for benzene, toluene, ethylbenzene, and xylenes.

The investigation regarding the extent and intensity of pollution started in October 2006 by installing extensive monitoring wells in the polluted aquifer region. The wells have been used to monitor and record groundwater contaminant concentrations and the groundwater hydraulic heads. In the course of time, more monitoring wells were added. BTEX components consists of some of the Volatile Organic Components (VOC) found in petroleum derivatives such as petrol (gasoline) ("BTEX Definitions," 2015). Although there was no systematic procedure in selecting the location of these monitoring wells, they are mainly located close to potential sources of contamination such as petrol stations and fuel storage tanks.

The concentration monitoring data were collected at different times during the investigation period from October 2006 to July 2011. Using the collected concentration data, the preliminary identified source of pollution was a leaking underground storage tank located at a petrol station; however, more contaminant sources may exist. Moreover, the activity duration of this source in addition to the release flux at different times is unknown.

The aim of this study is to find the location of any other potential source of contaminants and release history (including activity duration and release fluxes) associated with each source. The contaminant observation data were collected at 75 wells, however the aim of this study is to find optimal source histories using relatively limited concentration measurement data. Therefore the aim is to establish that it is possible to characterize the sources with a planned monitoring network design without utilizing a large set of monitoring wells, some of which may be redundant. This fewer number of monitoring wells compared with the number of wells implemented earlier

for monitoring can establish the fact that a much less costly monitoring plan can be sufficient. It is postulated that the large number of monitoring wells results in redundancy in monitoring. Moreover the concentration measurement data were collected approximately every three months during the investigation period (October 2006 to July 2011); however, one of the other aims of this study is to shorten the investigation period for the scenario present in the site, utilizing the proposed methodologies. It may be noted that with some concentration measurements missing and also some wells installed later during the investigation period, this problem becomes more challenging.

The aim of this study is to show that a more efficient and less costly monitoring network can be implemented, while utilizing the proposed sequential contaminant source identification and monitoring network design methodology (Chapter 6). A limited number of monitoring locations (from the available monitoring wells) were selected using the two-objective monitoring network design. The contaminant source identification model utilizes the contaminant concentration data at these selected locations to identify contaminant source characteristics. The identified source characteristics were used to design the monitoring network for the following sequences of model executions. The process was repeated until acceptable convergence in identified source characteristics was achieved.

7.3. The simulation model of the aquifer study area

The utilized simulation models for groundwater flow and contaminant transport process were MODFLOW (Zheng et al., 2001) and MT3DMS (Zheng & Wang, 1999). The

calibrated flow and transport models developed by Prakash (2014); Prakash and Datta (2015) were used in this study. For completeness, the characteristics and properties of the utilized flow and transport simulation models are briefly explained in the next two sections. For more information, readers may refer to Prakash and Datta (2015).

7.3 .1 . Groundwater flow modeling of the study area

Groundwater Modeling System (GMS) 7.0 software was utilized to develop the calibrated MODFLOW flow model. Figure 7.1 shows the "investigation area" which measures 608 m by 864 m. The impermeable bed of the aquifer is approximately 164 m below the ground surface.

Figure 7.1 Plan view of the study area. (Investigation area is marked)

7.3 .1. 1 . Model Area

Since there were no distinct geological formations to be used as the natural boundary conditions of the investigation area, a much larger area (measuring 2.187 km by 2.426 km) was considered for calibration of the flow model. The larger area is called the "model area". Figure 7.2 shows the model area. The Macquarie River forms the constant head boundary condition at the western side of the "model area". The other boundaries were modeled as constant head boundary conditions where the hydraulic head at other boundaries was estimated by Puech (2010).

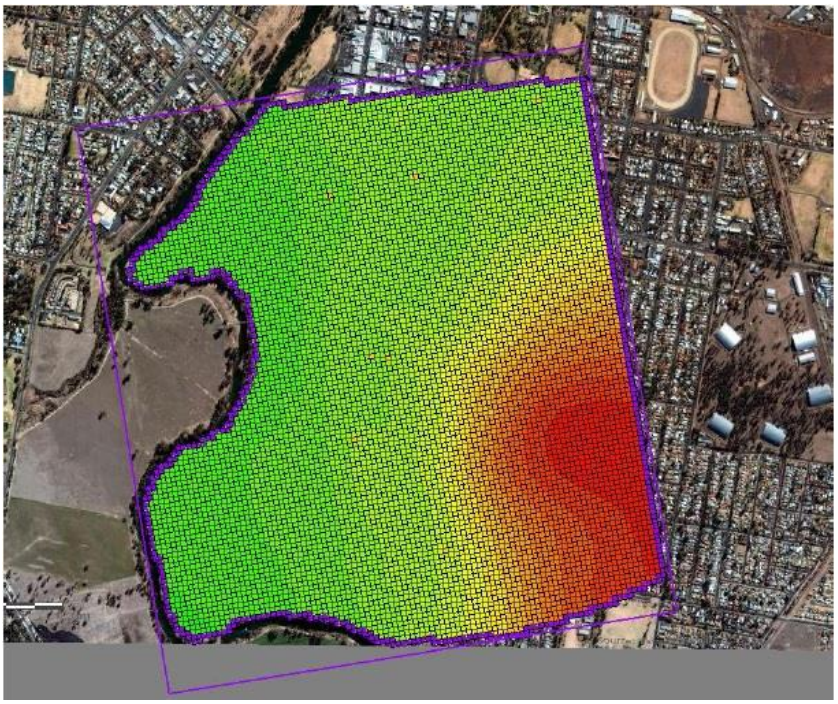


Figure 7.2 Plan view of the "model area" (Prakash, 2014).

The ground topography is from south-east towards the river in the west. The ground elevation ranges from 292 mAHD to 251 mAHD. The limited number of available bore-hole logs divides the geology of the model area into three distinct layers.

The top, middle and bottom layers are comprised of tertiary alluvium, quaternary alluvium and bedrock, respectively. The thickness of layers varies from one point to another and is estimated by interpolating available point data from bore-hole logs.

Rainfall and river are the two main sources of recharge in the model area. The average catchment rainfall is 583 mm/year in the wet season. Since the model area is largely suburban area, only 10% of rainfall was assumed to contribute as recharge and the rest was considered as surface runoff. Recharge was considered to be uniform throughout the model area.

The study period was from 1 January 1995 until 31 December 2012 and was divided into 18 stress periods of one year each. For the groundwater flow model MODFLOW transient simulation model was used. The flow model physical and hydrogeological properties are listed in Table 7.1 (Prakash, 2014).

There are eight extraction wells in the model area. The extraction rates were variable for each stress period and they were estimated based on the total annual extraction rate for all wells reported by the City Council and the long-term extraction rates were derived from Puech (2010).

The flow model was calibrated using available hydraulic head measurements obtained at 31 observation locations. The calibration period was from October 2006 to July 2011. The calibration targets were set to be within 1m from observed hydraulic heads with a confidence level of 90%. The boundary conditions were adjusted to achieve satisfactory calibration results.

Table 7.1 The physical and hydrogeological properties of the model area

Parameter	Unit	Value
Maximum length of study area	m	2187.1
Maximum width of study area	m	2425.6
Grid spacing in x-direction, Δx	m	21.87
Grid spacing in y-direction, Δy	m	21.08
Number of layers in z-direction		3 (layers' thickness are adjusted based on the geology)
K_{xx} (Layer 1, Layer 2, Layer 3)	m/day	12.37, 16.24, 0.001
K_{yy} (All layers)	m/day	0.2
θ (All layers)	dimensionless	0.27
Longitudinal Dispersivity (D_x)	m	12
Traverse Dispersivity (D_t)	m	6
Horizontal Anisotropy	dimensionless	1.5
Specific Yield S_y (All layers)	dimensionless	0.1
Specific Storage S_s (All layers)	m^{-1}	0.000006

7.3 .1. 2 . Investigation Area

The hydraulic head measurements are available within the investigation area. Therefore, the calibrated simulation model for the "model area" is a well-calibrated simulation model for the "investigation area". The model for the investigation area was derived from the calibrated model using GMS 7.0 regional to local feature. Refined grid size was selected for the investigation area and it was discretized into 75 rows, 50 columns and three layers. Figure 7.3 shows the plan view of the investigation area. All

geological, hydraulic and hydrogeological information were transferred to the investigation area flow model. The boundary conditions were considered to be time varying specified head boundary conditions. The values of time varying specific heads at boundaries of the investigation area were imported for the calibrated model of the model area.

7.3 .2 . Pollutant transport simulation model in the investigation area

The three-dimensional transient pollutant transport model was developed using MT3DMS module within GMS 7.0. The transport of the petrochemical pollutant BTEX was simulated within the investigation area. The pollutant sources were modeled as point sources using the well feature in MODFLOW and MT3DMS. The pollutant sources were considered as injection wells in the flow model and a relatively small flow rate (1 m^3/day) was assigned to them. The small flow rate ensured that the pollutant sources did not have any significant effect on the flow regime of the study area. In the transport simulation model, the injection concentration for each source must be determined and assigned. The initial concentration of BTEX in the aquifer was zero. All the transport parameters are shown in Table 7.1. The transport simulation model used the flow field generated by the flow model to estimate contaminant concentrations in the investigation area over time.

7.4. Performance evaluation of the proposed sequential source identification and monitoring network design methodologies

In this section the source identification methodology (Chapter 5) and the monitoring network design methodology (Chapter 6) are sequentially used to find the contaminant

source characteristics. Based on the preliminary available information, there are two potential sources of contaminants in the area. In total, 74 contaminant observation wells are available and 63 wells are located within the investigation area. These wells have been installed at different investigation stages and the contaminant concentration measurements have been carried out at various time stages. In general, the monitoring times are at every three months interval; however, many concentration measurements are missing during a number of time periods or, the collection of samples has been discontinued after few monitoring stages at some monitoring locations.

In Figure 7.4, the red marks show the potential source locations, and the available monitoring wells are shown by green dots. The blue line shows the boundary of the investigation area. The contaminant observation period was between October 2006 and July 2011. The study period was between January 1995 and January 2013 and was divided into 18 stress periods of one year each. The performance evaluation had two parts. First the source identification methodology was applied to find an estimate of contaminant source fluxes using an arbitrarily selected monitoring network. Next, the monitoring network design methodology selected the optimal contaminant monitoring network. This process was continued sequentially until satisfactory optimal source characteristics were obtained.

Figure 7.3 Plan view of the discretized investigation area.

7.4. 1. Contaminant Source Identification Model

The two-objective contaminant source identification model (Chapter 5) was used to estimate the contaminant source histories. The potential contaminant points source 1 and 2 are located at cell (1,17,29), and (1,16,24), respectively. In this chapter the location of points are reported as layer number, row number and column number. The study period is divided into 18 stress periods each of 1 year (1 January 1995 – 1 January 2013); however, based on the available contaminant observation data, the

contaminant sources may be active in 10 stress periods (1 January 1999- 1 January 2009). Therefore, the source identification decision variables are the BTEX release from each source at ten different stress periods. There are 20 decision variables and these correspond to two potential source locations and 10 stress periods ($S_{i,j}$ i=1,..2 and j=1,…,10).

Figure 7.4 Plan view of the investigation area and the monitoring and potential source locations.

The aim of this study is to evaluate the sequential source identification and monitoring network methodologies. Therefore, at each sequence of model execution, the monitoring network for the next monitoring time step was designed. The monitoring locations for the first sequence were selected arbitrarily as the initial monitoring locations.

The two-objective source identification model was executed using available BTEX concentration measurements collected at the monitoring locations. An NSGA-II optimization algorithm was utilized to solve the source characterization problem. The population size and maximum number of generations were 50 and 100, respectively. The stopping criterion for optimization was the maximum number of generations. The maximum and minimum possible source fluxes specified were 0 and 100 g/s, respectively.

7.4. 2. Monitoring network design model

The estimated source fluxes were used as feedback to design the monitoring network for the next sequence of model execution. At each sequence of model execution, three new monitoring locations were selected for the next monitoring time step. The potential monitoring locations were the available contaminant observation wells installed in the investigation area; however, at some monitoring wells, the observed concentrations were not reported at some of the monitoring times. Therefore, for each monitoring design sequence, those locations with no observed concentration were taken out of the potential locations. Moreover, few potential monitoring locations could not capture any

information regarding the identification of contaminant sources. They were either very far from the potential source locations or, are installed where the contaminants cannot reach them regardless of the source characteristics. Therefore, to make the evaluation process more efficient, these locations are also eliminated from the list of potential source locations. For the elimination process, it was assumed that both sources were active with the highest possible release flux (100 g/s). Then the transport model was used to predict BTEX concentration at all potential monitoring wells. The locations which exhibited zero concentration were located out of possible contaminant plumes and did not have any substantial effect on the source identification process.

The two-objective monitoring network design methodology (Chapter 6) was utilized. An NSGA-II optimization algorithm was applied. The population size and maximum number of generations were 100 and 100, respectively. The stopping criterion was the maximum number of generations.

7.5. Results

The three arbitrarily selected initial monitoring locations were MW02, MW05, and MW18, located at (1,20,26), (1,22,26), and (1,21,28), respectively. The monitoring time was January 2009. Figure 7.5 shows the concentration breakthrough of the selected monitoring locations

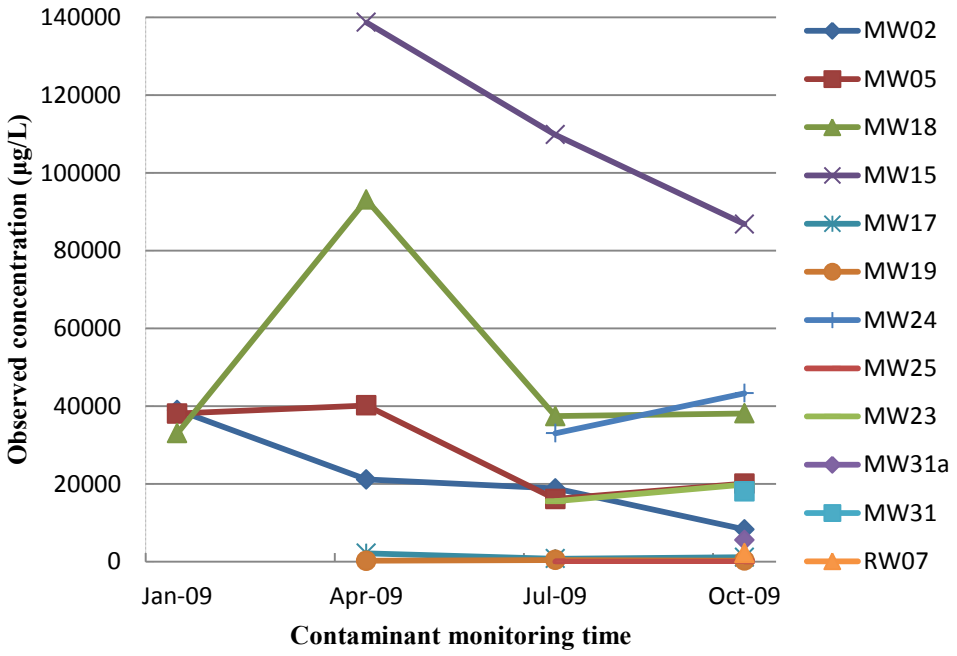


Figure 7.5 Breakthrough curves at monitoring locations utilized in the source identification procedure

7.5 .1 . First sequence

The source identification model was executed to find the contaminant release fluxes at potential sources using contaminant observation data collected at the three initial arbitrary locations. Three observed concentrations collected in January 2009 were utilized. The NSGA-II optimization model is utilized. Initially, the optimization stopping criterion was set at 100 generations; however, due to the very limited number of observed concentrations, the optimization algorithm converged after nine generations. Figure 7.6 shows the Pareto-front related to the final generation after the first sequence of source identification process.

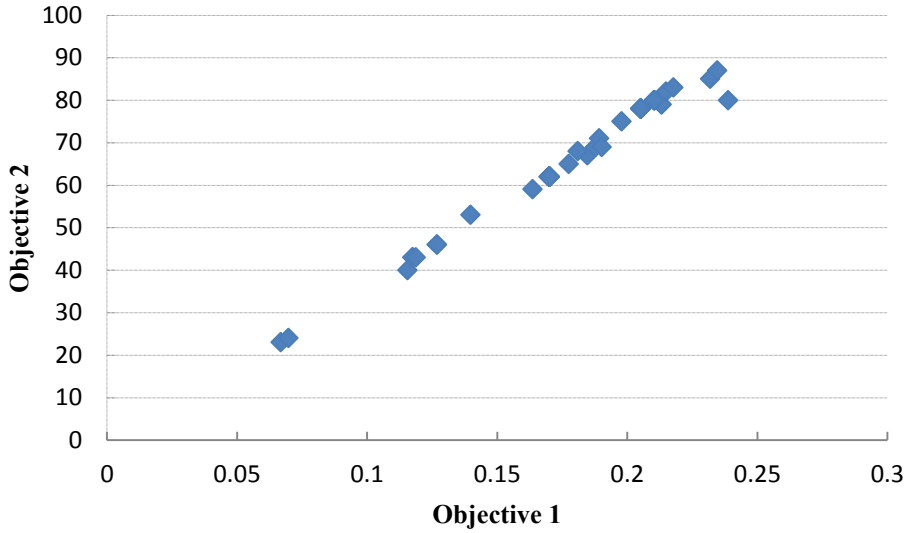


Figure 7.6 Objective function values related to the final generation of the source identification model for the first sequence

In figure 7.6, the source identification linked simulation-optimization algorithm converges to one set of source fluxes (not a Pareto-front). Having one optimal solution for the two-objective source identification model is due to the very limited number of observed concentrations (three) in the first sequence. The estimated optimal source fluxes are presented in Table 7.2.

7.5 .2 . Second sequence

The monitoring network for the next monitoring time step (April 2009) had to be designed. Three new monitoring locations were selected using the two-objective monitoring network design model. The vector containing all available estimates for the release histories (Q), corresponds to figure 7.6, was estimated in the first sequence. The monitoring network was selected based on the uncertainty and redundancy reduction objective functions (objective 1 and 2, respectively). Likewise using the method explained earlier, the monitoring locations which could not capture any information (in

April 2009) in addition to already selected monitoring locations (in sequence one) were eliminated from the list of potential monitoring locations. In the monitoring network design procedure, in total there were 40 potential monitoring locations from which three were selected as observation wells for April 2009. Monitoring wells M15, M17 and M19 located at cells (1,24,28), (1,18,24), and (1,17,26) were selected.

The contaminant observation concentrations data were recorded in April 2009 at monitoring locations M02, M05, M18 (already existing) and M15, M17, and M19 (newly selected) wells. The observed concentrations in July 2009 in addition to the concentrations collected in January 2009 (already existing) were utilized in the source identification model to improve the accuracy of the source flux estimation. In the second sequence, the initial population of the linked simulation-optimization algorithm was the final generation estimated in the first sequence. Figure 7.7 shows the objective function values related to the final generation of the source identification model.

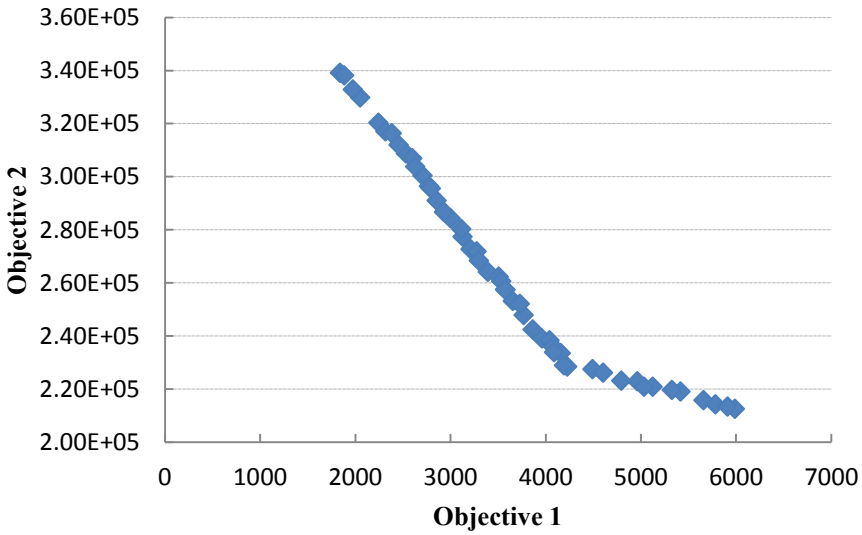


Figure 7.7 Objective function values related to the final generation of the source identification model. Second sequence

At this stage, since the final solution is presented in the form of a Pareto-front, the release fluxes for each decision variable are averaged over all members of the final generation and presented in Table 7.2.

7.5 .3 . Third sequence

The new monitoring locations were again selected for the next monitoring time (July 2009). The two-objective monitoring network design methodology was utilized to select three new monitoring well locations. The already-selected locations and the ones which were not able to capture any information at this monitoring time stage were eliminated from the list of all potential monitoring locations at this stage. Therefore, 31 potential monitoring locations were available. Monitoring wells M24, M25 and M23 located at cells (1,30,25), (1,24,20), and (1,25,32) were selected.

The contaminant observed concentrations data were recorded in July 2009 at monitoring locations M02, M05, M18, M15, M17, and M19 (already existing) and M24, M25, and M23 (newly selected) wells. The observed concentrations in July 2009, in addition to the concentrations collected in January 2009 and April 2009 (already existing) were utilized in the source identification model to improve the accuracy of the source flux estimation. In the third sequence, the initial population of the linked simulation-optimization model was the final generation estimated in the second sequence. Figure 7.8 shows the objective function values related to the final generation of the source identification model.

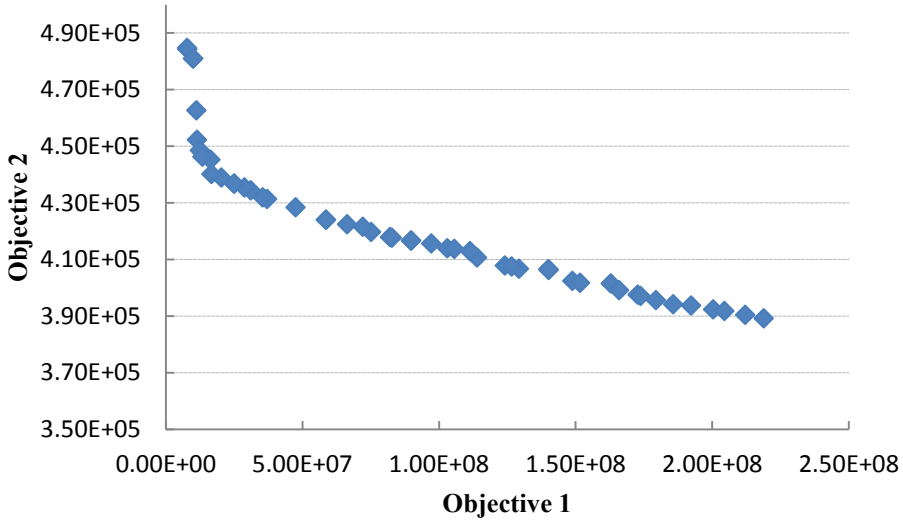


Figure 7.8 Objective function values related to the final generation of the source identification model. Third sequence.

The release fluxes estimated for each decision variable, in the last generation, are averaged over all members of the final generation and presented in Table 7.2.

7.5 .4 . Fourth sequence

The same process was continued in the fourth sequence. First the monitoring network for the contaminant sample collection in October 2009 was designed. The two-objective monitoring network design methodology was utilized to select three optimal locations out of 27 available potential locations. Monitoring wells M31a, M31, and RW7 located at cells (1,37,23), (1,38,23), and (1,39,21), respectively, were selected.

The contaminant observation concentrations data were measured in October 2009 at monitoring locations M02, M05, M18, M15, M17, M19, M24,M25, and M23 (already existing) and at M31a, M31, and RW7 (newly selected) wells. The observed concentrations in October 2009 in addition to the concentrations collected in January 2009, April and July 2009 (already existing) were utilized in the source identification

model to improve the accuracy of the source flux estimation. In the fourth sequence, the initial population of the linked simulation-optimization algorithm was the final generation estimated in the third sequence. Figure 7.9 shows the objective function values related to the final generation of the source identification model.

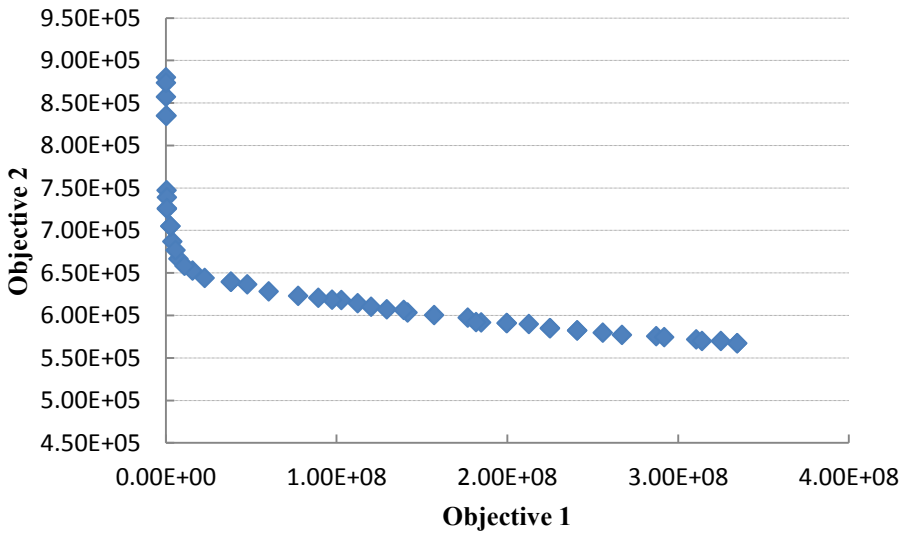


Figure 7.9 Objective function values related to the final generation of the source identification model. Fourth sequence.

The release fluxes estimated for each decision variable, in the last generation, are averaged over all members of the final generation and presented in Table 7.2.

7.5 .5 . Evaluation

In order to evaluate the effectiveness of the proposed sequential contaminant source identification and monitoring network design methodologies in recovering contaminant source release histories, a comparatively accurate estimate of source release fluxes was required. Therefore, the contaminant source identification methodology was executed using all available contaminant observation data (63 monitoring wells) for January,

April, July, and October 2009. In total 252 observed concentrations were utilized to estimate the contaminant release fluxes using the two-objective source identification model. In the NSGA-II optimization algorithm, the population size and maximum number of generations were 50 and 400, respectively. The maximum and minimum possible source fluxes were 0 and 100 g/s, respectively. The stopping criterion was the maximum number of generations, selected as 400 generations. The maximum number of generations in each sequence of the sequential method was 100 and the model was executed for four sequences. Therefore, having set the maximum number of generations for evaluation purposes at 400, helped to provide a more systematic comparison between the effectiveness of methods without considering the calculation costs. In Table 7.2, the optimal source fluxes estimated using all available measured concentrations from the site are presented. These source flux estimates are designated as evaluation source fluxes, for comparison purposes. Figure 7.10 shows the objective function values related to the final generation of the source identification model.

7.7. Discussion

The effectiveness of the proposed sequential source identification and monitoring network design methodology is evaluated using Normalized Absolute Error of Estimation (*%NAEE*). This measure is estimated using Eq. 7-1 (Jha & Datta, 2013).

$$NAEE(\%) = \frac{\sum_{i=1}^{N} \left| q_{est}^{i} - q_{evaluate}^{i} \right|}{\sum_{i=1}^{N} q_{evaluate}^{i}} \times 100 \qquad (7\text{-}1)$$

where N is the number of stress periods and q^i_{est} and $q^i_{evaluate}$ are the estimated and the evaluation source fluxes for stress period i, respectively. *NAEE%* is calculated for potential source location one and presented in Figure 7.11. Since the second potential source location is dummy (non-actual), the evaluation source fluxes are zero for all stress periods which make the %NAEE indeterminate. Therefore, the Absolute Error of Estimation (AEE) is estimated for potential source location two (Eq. 7-2) and presented in Figure (7.12).

Table 7.2 Estimated contaminant source release fluxes (g/s)

	$S_{1,1}$	$S_{1,2}$	$S_{1,3}$	$S_{1,4}$	$S_{1,5}$	$S_{1,6}$	$S_{1,7}$	$S_{1,8}$	$S_{1,9}$	$S_{1,10}$
1st sequence	33	94	26	26	25	0	0	0	10	9
2nd sequence	36	88	24	34	31	5	0	2	0	0
3rd sequence	21	48	23	14	3	3	7	22	0	1
4th sequence	17	37	13	3	2	20	12	6	2	1
Evaluation	22	33	14	1	1	26	15	4	4	10
	$S_{2,1}$	$S_{2,2}$	$S_{2,3}$	$S_{2,4}$	$S_{2,5}$	$S_{2,6}$	$S_{2,7}$	$S_{2,8}$	$S_{2,9}$	$S_{2,10}$
1st sequence	48	61	13	62	50	0	32	2	0	29
2nd sequence	8	34	1	1	0	0	0	0	0	0
3rd sequence	0	0	0	0	0	0	0	0	0	0
4th sequence	0	0	0	0	0	0	0	0	0	0
Evaluation	0	0	0	0	0	0	0	0	0	0

$$AEE = \sum_{i=1}^{N} \left| q^i_{est} - q^i_{evaluate} \right| \tag{7-2}$$

where N is the number of stress periods and q^i_{est} and $q^i_{evaluate}$ are the estimated and the evaluation source fluxes (obtained with all available concentration measurements) for stress period i, respectively.

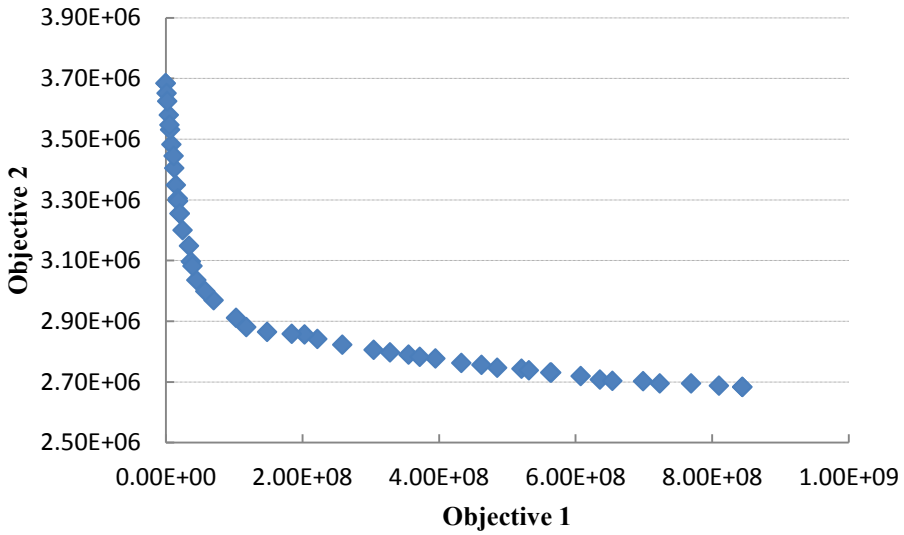


Figure 7.10 Objective function values related to the final generation of the source identification model. Evaluation

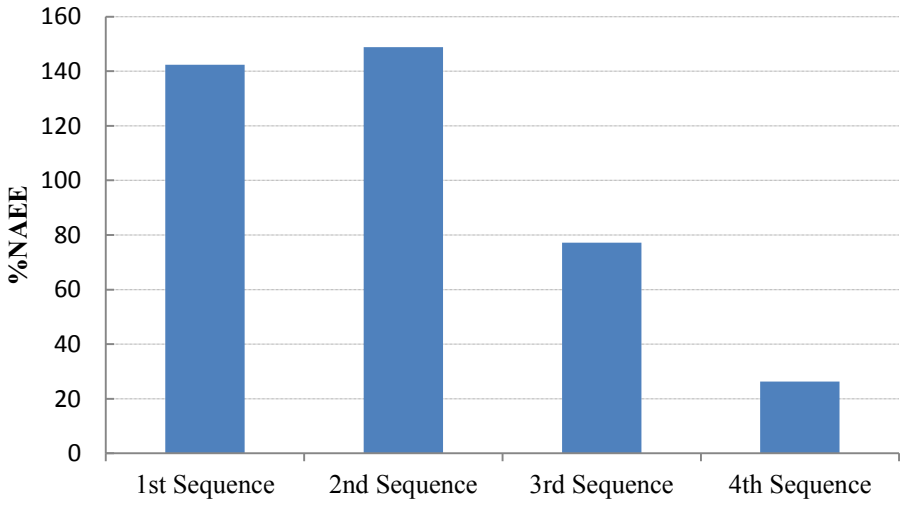


Figure 7.11 *NAEE%* for potential source location one

As Figure 7.11 shows, the *NAEE%* values estimated for source 1 show a decreasing trend by proceeding with the sequences. The 2nd sequence does not show any improvement in estimating source fluxes (in terms of *NAEE%*), however, it exhibits about 7% more estimation error. Proceeding from the 1st sequence to the 2nd one, adding more concentration observation data, generally should reduce the source flux estimation error. Therefore, this 7% increase in the estimation error may be attributed to the level of accuracy of observed concentrations collected in April 2009 and the measurement error at this monitoring time step. The *NAEE%* for the source fluxes estimated at the end of fourth sequence is 24%. Having a closer look on Table 7.2, all the release fluxes estimated for source one are relatively accurate except $S_{1,10}$. It can be noted that $S_{1,10}$ is the contaminant flux release at source location one, in 2009. The concentration observation data utilized in the source identification model were collected between January and October 2009. Therefore, the large flux estimation error associated to $S_{1,10}$ (900%) can be tracked back to the short monitoring period.

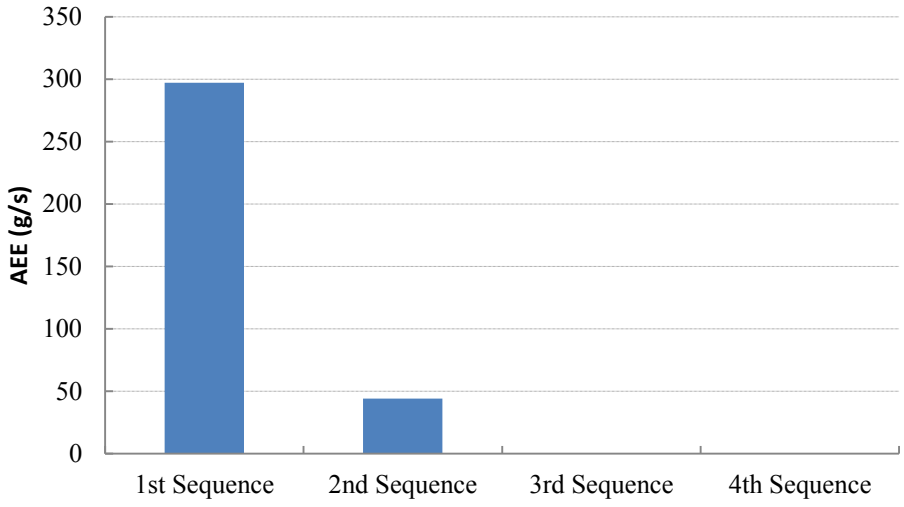


Figure 7.12 AEE for potential source location two.

The evaluation source fluxes for source two (Table 7.2) correctly show this potential source location is dummy or non-actual. In Figure 7.12, the AEE of the estimated source fluxes shows that the sequential source identification and monitoring network design methodology is able to find the non-active source locations after two sequences. Therefore, the sequential method successfully identifies the actual source locations using a relatively limited number of observation data (three and nine concentration measurements in sequences 1 and 2, respectively), compared to the benchmark source estimates based on an extensive monitoring network consisting of 63 monitoring wells and 252 observed concentrations.

7.8. Conclusions

In this chapter the proposed sequential contaminant source identification and monitoring network design methodologies are evaluated for an urban contaminated study area. The flow and transport simulation models are calibrated for the study area using available hydraulic head measurements. The sequential methodologies are applied to the area in four sequences. In the first sequence, three arbitrary monitoring wells are selected and the source identification model is applied to find the optimal source release fluxes. Then these estimates of contaminant release fluxes are used to select three new monitoring wells for the next monitoring time step using the two-objective monitoring network design methodology. The source identification model is executed again to improve the accuracy of estimated release fluxes. Then this sequential process is repeated for four sequences.

To evaluate the performance of the developed methodologies, the source identification model is executed using all available observed concentrations from an extensive monitoring wells network. These estimates provide the benchmark source fluxes for evaluation purpose. Comparing the recovered source histories with the benchmark source release fluxes shows that the proposed sequential method is able to find relatively accurate source characteristics with limited observed concentrations in a real life contaminated aquifer study area. Moreover, the sequential methodology is capable of finding the location of non-active potential sources, using a very limited number of observed concentrations and a short time span of observation. Therefore, utilizing the proposed methodology to characterize contaminant sources can save the decision makers substantial money and time.

8. SUMMARY AND CONCLUSION

In this chapter a summary of the new methodologies developed in this study is presented. The limitations associated with these methodologies and possible future research directions are highlighted. The conclusions from the performance evaluation results are also highlighted.

In contrast with the traditional view of science where uncertainty represents an undesirable state, a state that must be avoided at all costs, in this study new methodologies are presented to accept and incorporate some level of uncertainty in the contaminant source identification and monitoring network design procedure. Many previous researchers have tested their methodologies using erroneous/ uncertain hydrogeologic parameter values, such as hydraulic conductivity and contaminant measurement error; however, this study focuses on explicitly incorporating these sources of uncertainty in the contaminant source identification and monitoring network design methodologies.

Lack of information about field hydrogeologic parameters results in non-random uncertainty in estimated contaminant concentrations using flow and transport simulation models. Inaccurate simulation results decrease the accuracy of the linked simulation-optimization based contaminant source identification. In this study, the uncertainty in hydraulic conductivity parameter values is quantified using uncertainty-based linked simulation-optimization contaminant source identification models. Multiple realizations of hydraulic conductivity fields were generated using the IDW method with different sets of interpolation parameters.

Utilizing simulated concentrations in these realizations, the COC is estimated for each available spatial and temporal contaminant concentration monitoring data. Incorporating the COC in the source identification process increases the accuracy of recovered contaminant source characterization. The efficiency of the uncertainty-based source identification methodology is compared with the crisp model in an illustrative study area. Results show that the uncertainty based methodology is able to characterize the contaminant sources more accurately compared with the previous crisp source identification methodologies. In crisp methodology the hydrogeologic parameter uncertainty is not considered in the methodology.

The accuracy of solution results and the computational efficiency of the proposed methodology are compared with SGS, which is a probabilistic approach. The evaluation results demonstrate that the SGS method adds more uncertainty to the already uncertain source identification model. Performance evaluation results provide insight into the inter-relation between errors in the source identification process and available spatial hydraulic conductivity values. The proposed methodology is potentially useful in quantifying parametric uncertainty, when precise information is not available.

The performance of the proposed uncertainty-based source identification model is evaluated for a real experimental aquifer study area. The utilized experimental site is located in a polluted industrial area in NSW, Australia. A tracer test has been conducted in the area and the contaminant concentrations are recorded in different locations. The performance evaluation results show the applicability of the proposed methodology in real contaminated sites.

In order to incorporate the contaminant observation measurement error/ uncertainty in the source identification methodology, in this study, a two-objective approach is proposed. The first objective minimizes the normalized difference between the estimated and measured concentrations at monitoring locations and times. The second objective does not use any normalizing factor. The NSGA-II is utilized as an efficient multi-objective optimization algorithm.

When the measurements are erroneous, the source identification problem is non-unique. Therefore, various solutions with a possibility of being the true source characteristics may be achieved. When the contaminant concentration measurements are erroneous, it is not possible to match all the observed and simulated concentrations. In the two-objective approach, the first objective function is normalized using observed concentrations. Therefore, this objective function emphasizes matching smaller observed concentrations (larger objective function improvement); however, the second objective function, which is not normalized, tries to match the high concentrations. Therefore, the two-objective approach finds the set of possible solutions as a Pareto-front.

The performance of the proposed methodology is evaluated in an illustrative study area and the results are presented as an optimum Pareto-front. It is demonstrated that the retrieved source flux estimates obtained using the proposed two-objective approach are more accurate compared with the solution results obtained using each objective function separately.

Furthermore, a new two-objective monitoring network design methodology is developed which is sequentially utilized with the linked simulation-optimization source

identification model. The monitoring network design objectives are to reduce uncertainty in recovered source histories, and reduce redundancy in selecting the location of contaminant observation wells. By selecting the monitoring locations based on the first objective function, the source identification model would be able to differentiate between possible source releases histories. The second objective function ensures that the selected monitoring well locations are able to provide acceptable coverage for the unmonitored locations. An NSGA-II optimization model is used to find the optimal Pareto-front of the monitoring network designs.

The performance of the proposed methodology is evaluated in an illustrative contaminated study area. Results show that utilizing the designed monitoring network increases the accuracy and efficiency of the source identification model compared with the arbitrary selected locations. Moreover, the proposed methodology exhibits acceptable performance in the presence of measurement error which shows the utility of this method in real-field conditions. Moreover, the proposed sequential source identification and monitoring network design methodology is evaluated for a real urban contaminated aquifer located in NSW, Australia. Comparing the recovered source histories with the benchmarked source release fluxes shows that the proposed sequential method is able to find relatively accurate source characteristics with a limited number of observed concentrations in a real contaminated field. Moreover, the sequential methodology is capable of finding the location of non-active potential sources, using a very limited number of observed concentrations and short pollutant investigation periods. Utilizing the proposed methodology to characterize contaminant

sources can save decision makers substantial money and time, and require a very limited set of monitoring locations, essentially reducing redundancies in the network.

The performance evaluation results for both illustrative and real contaminated aquifers demonstrate the applicability of the proposed methodologies in groundwater contamination problems. There are some limitations to the methodologies developed in this study which can be used as research directions by future researchers. The main limitations are:

i. The uncertainty-based linked simulation optimization methodology considers quantification of hydraulic conductivity parameter uncertainty. The methodology is tested using a number of hydraulic conductivity fields; however, the methodology needs to be evaluated for more scenarios and more rigorous performance evaluation is needed to establish the applicability of the methodology.

ii. The other hydrogeologic parameters' uncertainty including dispersivity, porosity etc., must be considered as sources of uncertainty in the contaminant source identification methodology.

iii. The geological sources of uncertainty such as fractures and boundary conditions must be considered as possible sources of uncertainty in a contaminant source identification model.

iv. The evaluation results presented in this study are limited in scope. More rigorous evaluation processes must be completed to clearly establish wide applicability of the proposed methodologies and to extend these methodologies.

REFERENCES

Abebe, A., Guinot, V., & Solomatine, D. P. (2000). *Fuzzy Alpha-Cut vs. Monte Carlo Techniques in Assessing Uncertainty in Model Parameters*. Paper presented at the Proc. 4- The International Conferences on Hydrodynamics, Iowa City, USA.

Abrishamchi, A., Tajrishy, M., & Shafieian, P. (2005). Uncertainty Analysis in QUAL2E Model of Zayandeh-Rood River. *Water Environment Research, 77*(3), 279-286. doi:10.2175/106143005X41861

Amirabdollahian, M., & Datta, B. (2013). Identification of contaminant source characteristics and monitoring network design in groundwater aquifers: an overview. *Journal of Environmental Protection, 4*(5A), 26-41.

Amirabdollahian, M., & Datta, B. (2014). Identification of Pollutant Source Characteristics under Uncertainty in Contaminated Water Resources systems using Adaptive Simulated Anealing and Fuzzy Logic. *International Journal of GEOMATE, 6*(1), 757-762.

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

ASCE Task Committee. (2003). Long-Term Groundwater Monitoring Design; The State of The Art. Reston, VA.

Ayvaz, M. T. (2007). Simultaneous Determination of Aquifer Parameters and Zone Structures with Fuzzy C-Means Clustering and Meta-Heuristic Harmony Search Algorithm. *Advances in Water Resources, 30*(11), 2326-2338. doi:10.1016/j.advwatres.2007.05.009

Bagtzoglou, A. C., Tompson, A. F. B., & Dougherty, D. E. (1991). Probabilistic Simulation for Reliable Solute Source Identification in Heterogeneous Porous Media. *Water Resources Engineering Risk Assessment, 29*(NATO ASI Series), 189-201.

Bagtzoglou, A. C., Tompson, A. F. B., & Dougherty, D. E. (1992). Application of Particle Methods to Reliable Identification of Groundwater Pollution Sources. *Water Resources Management, 6*, 15-23.

Bashi-Azghadi, S. N., & Kerachian, R. (2010). Locating monitoring wells in groundwater systems using embedded optimization and simulation models. *Science of the Total Environment, 408*(10), 2189-2198. doi:10.1016/j.scitotenv.2010.02.004

Beck, P. H. (2000). *Transport of conservative and reactive inorganic elements in the saturated part of a hetrogeneous sand aquifer, Botany Basin, Sydney, Australia.* (PhD), University of New South Wales, Sydney, Australia.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology, 320*(1–2), 18-36. doi:10.1016/j.jhydrol.2005.07.007

Beven, K., Buytaert, W., & Smith, L. A. (2012). On virtual observatories and modelled realities (or why discharge must be treated as a virtual variable). *Hydrological Processes, 26*(12), 1905-1908. doi:10.1002/hyp.9261

Bogardi, I., Bardossy, A., & Duckstein, L. (1985). Effect of Parameter Uncertainty on Calculated Sediment Yield. *Advances in Water Resources, 8*(2), 96-101. doi:10.1016/0309-1708(85)90006-5

BTEX Definitions. (2015). Retrieved from http://toxics.usgs.gov/definitions/btex.html

Cameron, E., & Peloso, G. F. (2001). An Application of Fuzzy Logic to the Assessment of Aquifers' Pollution Potential. *Environmental Geology, 40*(11-12), 1305-1315.

Carroll, R. W. H., & Warwick, J. J. (2001). Uncertainty Analysis of the Carson River Mercury Transport Model. *Ecological Modelling, 137*(2-3), 211-224.

Chadalavada, S., & Datta, B. (2008). Dynamic Optimal Monitoring Network Design for Transient Transport of Pollutants in Groundwater Aquifers. *Water Resources Management, 22*(6), 651-670. doi:10.1007/s11269-007-9184-x

Chadalavada, S., Datta, B., & Naidu, R. (2011a). Optimisation Approach for Pollution Source Identification in Groundwater: An Overview. *International Journal of Environment and Waste management, 8*(1-2), 40-61.

Chadalavada, S., Datta, B., & Naidu, R. (2011b). Uncertainty Based Optimal Monitoring Network Design for a Chlorinated Hydrocarbon Contaminated Site. *Environmental Monitoring and Assessment, 173*(1-4), 929-940. doi:10.1007/s10661-010-1435-2

Cieniawski, S. E., Eheart, J. W., & Ranjithan, S. (1995). Using Genetic Algorithms to Solve a Multiobjective Groundwater Monitoring Problem. *Water Resources Research, 31*(2), 399-409. doi:10.1029/94wr02039

Clemo, T. (2003, 11-14 September). *Improved Water Table Dynamics in Block-Centered Finite-Difference Flow Models.* Paper presented at the MODFLOW and More 2003: Understanding through Modeling, Golden, Colorado, USA.

Copty, N. K., & Findikakis, A. N. (2000). Quantitative Estimates of the Uncertainty in the Evaluation of Ground Water Remediation Schemes. *Ground Water, 38*(1), 29-37. doi:10.1111/j.1745-6584.2000.tb00199.x

Cressie, N. (2015). *Statistics for Spatial Data* (Vol. 2nd). Hoboken: Wiley.

Datta, B. (2002). Discussion of "Identification of Contaminant Source Location and Release History in Aquifers" by Mustafa M. Aral, Jiabao Guan, and Morris L. Maslia. *Journal of Hydrologic Engineering, 7*(5), 399-400. doi:10.1061/(ASCE)1084-0699(2002)7:5(399)

Datta, B., Beegle, J. E., Kavvas, M. L., & Orlob, G. T. (1989). *Development of an Expert System Embedding Pattern Recognition Techniques for Pollution Source Identification*. Project Completion Technical Report submitted to United States Geological Survey (U.S.G.S.). University of California, Davis. Retrieved from NTIS, Springfield, Virginia, USA.

Datta, B., Chakrabarty, D., & Dhar, A. (2008). Optimal Dynamic Monitoring Network Design and Identification of Unknown Groundwater Pollution Sources. *Water Resources Management, 23*(10), 2031-2049. doi:10.1007/s11269-008-9368-z

Datta, B., Chakrabarty, D., & Dhar, A. (2009). Simultaneous Identification of Unknown Groundwater Pollution Sources and Estimation of Aquifer Parameters. *Journal of Hydrology, 376*(1-2), 48-57.

Datta, B., Chakrabarty, D., & Dhar, A. (2011). Identification of Unknown Groundwater Pollution Sources using Classical Optimization with Linked Simulation. *Journal of Hydro-Environment Research, 5*(1), 25-36.

Datta, B., & Dhiman, S. D. (1996). Chance-Constrained Optimal Monitoring Network Design for Pollutants in Ground Water. *Journal of Water Resources Planning and Management, 122*(3), 180-188.

Datta, B., & Kourakos, G. (2015). Preface: Optimization for groundwater characterization and management. *Hydrogeology Journal, 23*(6), 1043-1049. doi:10.1007/s10040-015-1297-3

Datta, B., Prakash, O., Campbell, S., & Escalada, G. (2013). Efficient Identification of Unknown Groundwater Pollution Sources Using Linked Simulation-Optimization Incorporating Monitoring Location Impact Factor and Frequency Factor. *Water Resources Management, 27*(14), 4959-4976. doi:10.1007/s11269-013-0451-8

Datta, B., Prakash, O., & Sreekanth, J. (2014). Application of genetic programming models incorporated in optimization models for contaminated groundwater systems management: Springer.

Deb, K. (2002). A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Transaction of Evolutionat Computation, 6*(2), 182-197.

Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: geostatistical software library and user's guide*. Oxford: Oxford University Press.

Dhar, A., & Datta, B. (2007). Multiobjective Design of Dynamic Monitoring Networks for Detection of Groundwater Pollution. *Water Resources Planning and Management, 133*(4), 329-338.

Dhar, A., & Datta, B. (2009). Global Optimal Design of Ground Water Monitoring Network using Embedded Kriging. *Ground Water, 47*(9), 806-815.

Dhar, A., & Datta, B. (2010). Logic-Based Design of Groundwater Monitoring Network for Redundancy Reduction. *Water Resources Planning and Management, 136*(1), 88-93.

Dixon, B. (2005). Groundwater Vulnerability Mapping: A GIS and Fuzzy Rule Based Integrated Tool. *Applied Geography, 25*(4), 327-347. doi:10.1016/j.apgeog.2005.07.002

Dixon, B., Scott, H. D., Dixon, J. C., & Steele, K. F. (2002). Prediction of Aquifer Vulnerability to Pesticides using Fuzzy Rule-Based Models at the Regional Scale. *Physical Geography, 23*(2), 130-153.

Dokou, Z., & Pinder, G. F. (2009). Optimal Search Strategy for the Definition of a DNAPL Source. *Journal of Hydrology, 376*(3), 542-556. doi:10.1016/j.jhydrol.2009.07.062

Dokou, Z., & Pinder, G. F. (2011). Extension and Field Application of an Integrated DNAPL Source Identification Algorithm that Utilizes Stochastic Modeling and a Kalman Filter. *Journal of Hydrology, 398*(3), 277-291. doi:10.1016/j.jhydrol.2010.12.029

Dou, C., Woldt, W., & Bogardi, I. (1999). Fuzzy Rule-Based Approach to Describe Solute transport in the Unsaturated Zone. *Journal of Hydrology, 220*(1-2), 74-85.

Dou, C., Woldt, W., Bogardi, I., & Dahab, M. (1997). Numerical Solute Transport Simulation using Fuzzy Sets Approach. *Journal of Contaminant Hydrology, 27*(1–2), 107-126. doi:10.1016/s0169-7722(96)00047-2

Evans, D. J. (1993). *A physical and hydrochemical characterisation of a sand aquifer in Sydney, Australia.* (MSc), University of New South Wales, Sydney, Australia (unpubl.).

Freeze, R. A. (1975). A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resources Research, 11*(5), 725-741. doi:10.1029/WR011i005p00725

Freeze, R. A., James, B., Massmann, J., Sperling, T., & Smith, L. (1992). Hydrogeological Decision Analysis: 4. The Concept of Data Worth and Its Use in the Development of Site Investigation Strategies. *Ground Water, 30*(4), 574-588. doi:10.1111/j.1745-6584.1992.tb01534.x

Fu, T., Chen, H., Zhang, W., Nie, Y., & Wang, K. (2015). Vertical distribution of soil saturated hydraulic conductivity and its influencing factors in a small karst catchment in Southwest China. *Environmental Monitoring and Assessment, 187*(3), 1-13. doi:10.1007/s10661-015-4320-1

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation.* New York: Oxford University Press.

Gorelick, S. M., Evans, B., & Remson, I. (1983). Identifying Sources of Groundwater Pollution: An Optimization Approach. *Water Resources Research, 19*(3), 779-790. doi:10.1029/WR019i003p00779

Gurarslan, G., & Karahan, H. (2015). Solving inverse problems of groundwater-pollution-source identification using a differential evolution algorithm. *Hydrogeology Journal, 23*(6), 1109-1119. doi:10.1007/s10040-015-1256-z

Harmel, R. D., Cooper, R. J., Slade, R. M., Haney, R. L., & Arnold, J. G. (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Transactions of the ASABE, 49*(3), 689-701.

Harmel, R. D., & Smith, P. K. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *Journal of Hydrology, 337*(3–4), 326-336. doi:http://dx.doi.org/10.1016/j.jhydrol.2007.01.043

Herrera, G. S., & Pinder, G. F. (2005). Space-Time Optimization of Groundwater Quality Sampling Networks. *WATER RESOURCES RESEARCH, 41*(12), W12407. doi:10.1029/2004wr003626

Hudak, P. F. (1998). A Method for Designing Configurations of Nested Monitoring Wells Near Landfills. *Hydrogeology Journal, 6*(3), 341-348. doi:10.1007/s100400050157

Hudak, P. F. (2001). Effective Contaminant Detection Networks in Uncertain Groundwater Flow Fields. *Waste Management, 21*(4), 309-312. doi:10.1016/s0956-053x(00)00079-9

Hudak, P. F., Loaiciga, H. A., & Marino, M. A. (1995). Regional-Scale Ground Water Quality Monitoring via Integer Programming. *Journal of Hydrology, 164*(1–4), 153-170. doi:10.1016/0022-1694(94)02559-t

Ingber, L. (1996). Adaptive simulated annealing (ASA): Lessons learned. *Control Cybern., 25*(1), 33-54.

Jankowski, J., & Beck, P. (2000). Aquifer heterogeneity: Hydrogeological and hydrochemical properties of the Botany Sands aquifer and their impact on contaminant transport. *Australian Journal of Earth Sciences; copyright © Geological Society of Australia, reprinted by permission of Taylor & Francis Ltd, www.tandfonline.com on behalf of Geological Society of Australia, 47*(1), 45-64. Retrieved from http://dx.doi.org/10.1046/j.1440-0952.2000.00768.x

Jarsjö, J., Bayer-Raich, M., & Ptak, T. (2005). Monitoring groundwater contamination and delineating source zones at industrial sites: Uncertainty analyses using integral pumping tests. *Journal of Contaminant Hydrology, 79*(3), 107-134. doi:10.1016/j.jconhyd.2005.05.011

Jha, M., & Datta, B. (2011). Simulated annealing based simulation-optimization approach for identification of unknown contamination sources in groundwater aquifers. *Desalination and Water treatment, 32*(1-3), 79-85.

Jha, M., & Datta, B. (2012). Application of Simulated Annealing in Water Resources Management: Optional Solution of Groundwater Contamination Source Characterization Problem and Monitoring Network Design Problems. In M. De Sales Guerra Tsuzuki (Ed.), *Simulated Annealing Single and Multiple Objective Problems* (pp. 157-174). Rijeka, Croatia: InTech.

Jha, M., & Datta, B. (2013). Three-Dimensional Groundwater Contamination Source Identification Using Adaptive Simulated Annealing. *Journal of Hydrologic Engineering, 18*(3), 307-317. doi:doi:10.1061/(ASCE)HE.1943-5584.0000624

Jha, M., & Datta, B. (2014). Linked Simulation-Optimization based Dedicated Monitoring Network Design for Unknown Pollutant Source Identification using Dynamic Time Warping Distance. *Water Resources Management, 28*(12), 4161-4182. doi:10.1007/s11269-014-0737-5

Jha, M., & Datta, B. (2015a). Application of Dedicated Monitoring–Network Design for Unknown Pollutant-Source Identification Based on Dynamic Time Warping. *Journal of Water Resources Planning and Management, 141*(11), 4015022. doi:10.1061/(ASCE)WR.1943-5452.0000513

Jha, M., & Datta, B. (2015b). Application of Unknown Groundwater Pollution Source Release History Estimation Methodology to Distributed Sources Incorporating Surface-Groundwater Interactions. *Environmental Forensics, 16*(2), 143. doi:10.1080/15275922.2015.1023385

Kollat, J. B., & Reed, P. M. (2007a). A Computational Scaling Analysis of Multiobjective Evolutionary Algorithms in Long-term Groundwater Monitoring Applications. *Advances in Water Resources, 30*(3), 408-419. doi:10.1016/j.advwatres.2006.05.009

Kollat, J. B., & Reed, P. M. (2007b). A Framework for Visually Interactive Decision-making and Design using Evolutionary Multi-objective Optimization (VI DEO). *Environmental Modelling and Software, 22*(12), 1691-1704. doi:10.1016/j.envsoft.2007.02.001

Koohpayehzadeh Esfahani, H., & Datta, B. (2015). Simulation of reactive geochemical transport processes in contaminated aquifers using surrogate models. *International Journal of GEOMATE, 8*(15), 1190-1196.

Li, J., Huang, G. H., Zeng, G., Maqsood, I., & Huang, Y. (2007). An Integrated Fuzzy-Stochastic Modeling Approach for Risk Assessment of Groundwater Contamination. *Journal of Environmental Management, 82*(2), 173-188. doi:10.1016/j.jenvman.2005.12.018

Li, Y., & Chan Hilton, A. B. (2007). Optimal Groundwater Monitoring Design using an Ant Colony Optimization Paradigm. *Environmental Modelling and Software, 22*(1), 110-116. doi:10.1016/j.envsoft.2006.05.023

Liu, C. X., & Ball, W. P. (1999). Application of Inverse Methods to Contaminant Source Identification from Aquitard Diffusion Profiles at Dover AFB, Delaware. *Water Resources Research, 35*(7), 1975-1985. doi:10.1029/1999wr900092

Loaiciga, H. A. (1989). An Optimization Approach for Groundwater Quality Monitoring Network Design. *Water Resources Research, 25*(8), 1771-1782. doi:10.1029/WR025i008p01771

Loaiciga, H. A., Charbeneau, R. J., Everett, L. G., Fogg, G. E., Hobbs, B. F., & Rouhani, S. (1992). Review of Ground-Water Quality Monitoring Network Design. *Journal of Hydraulic Engineering, 118*(1), 11-37.

Mahar, P. S., & Datta, B. (1997). Optimal Monitoring Network and Ground-Water-Pollution Source Identification. *Water Resources Planning and Management, 123*(4), 199-207.

Mahar, P. S., & Datta, B. (2001). Optimal Identification of Ground-Water pollution Sources and Parameter Estimation. *Water Resources Planning and Management-ASCE, 127*(1), 20-29.

Mahinthakumar, G. K., & Sayeed, M. (2005). Hybrid Genetic Algorithm - Local Search Methods for Solving Groundwater Source Identification Inverse Problems. *Water Resources Planning and Management, 131*(1), 45-57.

Mahinthakumar, G. K., & Sayeed, M. (2006). Reconstructing Groundwater Source Release Histories using Hybrid Optimization Approaches. *Environmental Forensics, 7*(1), 45-54. doi:10.1080/15275920500506774

Mathon, B., Ozbek, M., & Pinder, G. (2010). Dempster–Shafer Theory Applied to Uncertainty Surrounding Permeability. *Mathematical Geosciences, 42*(3), 293-307. doi:10.1007/s11004-009-9246-0

McKinney, D. C., & Loucks, D. P. (1992). Network Design for Predicting Groundwater Contamination. *Water Resources Research, 28*(1), 133-147. doi:10.1029/91wr02397

Meyer, P. D., & Brill, E. D., Jr. (1988). A Method for Locating Wells in a Groundwater Monitoring Network under Conditions of Uncertainty. *Water Resources Research, 24*(8), 1277-1282. doi:10.1029/WR024i008p01277

Meyer, P. D., Valocchi, A. J., & Eheart, J. W. (1994). Monitoring Network Design to Provide Initial Detection of Groundwater Contamination. *Water Resources Research, 30*(9), 2647-2659. doi:10.1029/94wr00872

mGstat V 0.99. (2004). mGstat V 0.99 Retrieved from http://sourceforge.net/projects/mgstat/files/

Mirghani, B. Y., Mahinthakumar, K. G., Tryby, M. E., Ranjithan, R. S., & Zechman, E. M. (2009). A parallel evolutionary strategy based simulation–optimization approach for solving groundwater source identification problems. *Advances in Water Resources, 32*(9), 1373-1385. doi:10.1016/j.advwatres.2009.06.001

Mirzaei, R., & Sakizadeh, M. (2016). Comparison of interpolation methods for the estimation of groundwater contamination in Andimeshk-Shush Plain, Southwest of Iran. *Environmental Science and Pollution Research, 23*(3), 2758-2769. doi:10.1007/s11356-015-5507-2

Montas, H. J., Mohtar, R. H., Hassan, A. E., & AlKhal, F. A. (2000). Heuristic Space–Time Design of Monitoring Wells for Contaminant Plume Characterization in Stochastic Flow Fields. *Journal of Contaminant Hydrology, 43*(3–4), 271-301. doi:10.1016/s0169-7722(99)00108-4

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika, 37*(1/2), 17.

Mugunthan, P., & Shoemaker, C. A. (2004). Time Varying Optimization for Monitoring Multiple Contaminants under Uncertain Hydrogeology. *Bioremediation Journal, 8*(3-4), 129-146. doi:10.1080/10889860490887509

Neupauer, R. M., Borchers, B., & Wilson, J. L. (2000). Comparison of Inverse Methods for Reconstructing the Release History of a Groundwater Contamination Source. *Water Resources Research, 36*(9), 2469-2475. doi:10.1029/2000wr900176

Nunes, L. M., Paralta, E., Cunha, M. C., & Ribeiro, L. (2004). Groundwater Nitrate Monitoring Network Optimization with Missing Data. *Water Resources Research, 40*(2), 1-18. doi:10.1029/2003wr002469

Oberkampf, W. L., Helton, J. C., Joslyn, C. A., Wojtkiewicz, S. F., & Ferson, S. (2004). Challenge problems: uncertainty in system response given uncertain parameters. *Journal Reliability Engineering and System safety, 85*, 11-19.

OHagan, A., & Oakley, J. E. (2004). Probability is perfect, but we can't elicit it perfectly. *Reliability Engineering & System Safety, 85*(1–3), 239-248. doi:10.1016/j.ress.2004.03.014

Passarella, G., Vurro, M., D'Agostino, V., & Barcelona, M. J. (2003). Cokriging Optimization of Monitoring Network Configuration Based on Fuzzy and Non-Fuzzy Variogram Evaluation. *Environmental Monitoring and Assessment, 82*(1), 1-21. doi:10.1023/a:1021662705554

Prakash, O. (2014). *Optimal monitoring network design and identification of unknwon pollutant sources in polluted aquifers.* (PhD), James Cook University, Townsville.

Prakash, O., & Datta, B. (2014). Multiobjective Monitoring Network Design for Efficient Identification of Unknown Groundwater Pollution Sources Incorporating Genetic Programming–Based Monitoring. *Journal of Hydrologic Engineering, 19*(11), 4014025. doi:10.1061/(asce)he.1943-5584.0000952

Prakash, O., & Datta, B. (2015). Optimal characterization of pollutant sources in contaminated aquifers by integrating sequential-monitoring-network design and source identification: methodology and an application in Australia. *Hydrogeology Journal, 23*, 1089-1107.

Puech, V. (2010). *Upper macquarie groundwater model*. Technical report VW04680. Office of water, NSW government and national water commission. Retrieved from Australia.

Reed, P., Minsker, B., & Valocchi, A. J. (2000). Cost-Effective Long-Term Groundwater Monitoring Design using a Genetic Algorithm and Global Mass Interpolation. *Water Resources Research, 36*(12), 3731-3741. doi:10.1029/2000wr900232

Reed, P., & Minsker, B. S. (2004). Striking the Balance: Long-Term Groundwater Monitoring Design for Conflicting Objectives. *Water Resources Planning and Management, 130*(2), 140-149.

Ross, T. J. (2005). *Fuzzy Logic with Engineering Applications*. Chichester: John Wiley & Sons, Limited.

Russo, D., & Bouton, M. (1992). Statistical-Analysis of Spatial Variability in Unsaturated Flow Parameters. *Water Resources Research, 28*(7), 1911-1925.

Sax, T., & Isakov, V. (2003). A Case Study for Assessing Uncertainty in Local-Scale Regulatory Air Quality Modeling Applications. *Atmospheric Environment, 37*(25), 3481-3489. doi:10.1016/s1352-2310(03)00411-4

Schlumberger Water Services. (2011). Visual MODFLOW help. Retrieved from http://www.swstechnology.com/help/vmod/index.html?vm_ch5_run5.htm

Schulz, K., & Huwe, B. (1997). Water Flow Modeling in the Unsaturated Zone with Imprecise Parameters using a Fuzzy Approach. *Journal of Hydrology, 201*(1–4), 211-229. doi:10.1016/s0022-1694(97)00038-3

Schulz, K., & Huwe, B. (1999). Uncertainty and Sensitivity Analysis of Water Transport Modelling in a Layered Soil Profile using Fuzzy Set Theory. *Journal of Hydroinformatics, 1*, 127-138.

Shojaei, M., Nazif, S., & Kerachian, R. (2015). Joint uncertainty analysis in river water quality simulation: a case study of the Karoon River in Iran. *Environmental Earth Sciences, 73*(7), 3819-3831. doi:10.1007/s12665-014-3667-x

Singh, D., & Datta, B. (2010). *Sequential Characterization of Contaminant Plumes using Designed Monitoring Network*. Paper presented at the IAHR-APD Congress, 7th IUMWC, Auchland.

Singh, R. M., & Datta, B. (2006). Identification of Groundwater Pollution Sources using GA-Based Linked Simulation Optimization Model. *Journal of Hydrologic Engineering, 11*(2), 101-109.

Singh, R. M., & Datta, B. (2007). Artificial Neural Network Modeling for Identification of Unknown Pollution Sources in Groundwater with Partially Missing Concentration Observation Data. *Water Resources Management, 21*(3), 557-572. doi:10.1007/s11269-006-9029-z

Singh, R. M., Datta, B., & Jain, A. (2004). Identification of Unknown Groundwater Pollution Sources Using Artificial Neural Networks. *Water Resources Planning and Management, 130*(6), 506-514.

Skaggs, T. H., & Kabala, Z. J. (1994). Recovering the Release History of a Groundwater Contaminant. *Water Resources Research, 30*(1), 71-79. doi:10.1029/93wr02656

Skaggs, T. H., & Kabala, Z. J. (1995). Recovering the Release History of a groundwater Contaminant Plume - Method of Quasi-Reversibility. *Water Resources Research, 31*(11), 2669-2673.

Smith, L., & Schwartz, F. W. (1980). Mass transport: 1. A stochastic analysis of macroscopic dispersion. *Water Resources Research, 16*(2), 303. doi:10.1029/WR016i002p00303

Smith, L., & Schwartz, F. W. (1981). Mass transport: 3. Role of hydraulic conductivity data in prediction. *Water Resources Research, 17*(5), 1463. doi:10.1029/WR017i005p01463

Smith, R. I., Fowler, D., Sutton, M. A., Flechard, C., & Coyle, M. (2000). Regional Estimation of Pollutant Gas Dry Deposition in the UK: Model Description, Sensitivity Analyses and Outputs. *Atmospheric Environment, 34*(22), 3757-3777. doi:10.1016/s1352-2310(99)00517-8

Snodgrass, M. F., & Kitanidis, P. K. (1997). A Geostatistical Approach to Contaminant Source Identification. *Water Resources Research, 33*(4), 537-546.

Sreekanth, J., & Datta, B. (2011). Comparative Evaluation of Genetic Programming and Neural Network as Potential Surrogate Models for Coastal Aquifer Management. *Water Resources Management, 25*(13), 3201-3218. doi:10.1007/s11269-011-9852-8

Sun, N.-Z. (1994). *Inverse problems in groundwater modeling*. Boston: Kluwer adademic.

Tiedeman, C., & Gorelick, S. M. (1993). Analysis of Uncertainty in Optimal Groundwater Contaminant Capture Design. *Water Resources Research, 29*(7), 2139-2153. doi:10.1029/93wr00546

Tikhonov, A. N., & Arsenin, V. Y. (1978). *Solutions of Ill-Posed Problems*: American Mathematical Society.

U.S. EPA. (2005). *Roadmap to Long-Term Monitoring Optimization* (EPA 542-R-05-003). Retrieved from http://www.epa.gov/tio/download/char/542-r-05-003.pdf

Wagner, B. J. (1992). Simultaneous Parameter Estimation and Contaminant Source Characterization for Coupled Groundwater Flow and Contaminant Transport Modelling. *Journal of Hydrology, 135*(1), 275-303. doi:10.1016/0022-1694(92)90092-a

Warwick, J. J., & Cale, W. G. (1986). Effects of Parameter Uncertainty in Stream modeling. *Journal of Environmental Engineering, 112*(3), 479-489.

Woldt, W., Bogardi, I., Kelly, W. E., & Bardossy, A. (1992). Evaluation of Uncertainty in a 3-Dimensional Groundwater Contamination Plume. *Journal of Contaminant Hydrology, 9*(3), 271-288.

Woodbury, A. D., & Ulrych, T. J. (1996). Minimum Relative Entropy Inversion: Theory and Application to Recovering the Release History of a Groundwater Contaminant. *Water Resources Research, 32*(9), 2671-2681.

Wu, J., Zheng, C., & Chien, C. C. (2005). Cost-Effective Sampling Network Design for Contaminant Plume Monitoring under General Hydrogeological Conditions. *Journal of Contaminant Hydrology, 77*(1–2), 41-65. doi:10.1016/j.jconhyd.2004.11.006

Wu, J., Zheng, C., Chien, C. C., & Zheng, L. (2006). A Comparative Study of Monte Carlo Simple Genetic Algorithm and Noisy Genetic Algorithm for Cost-Effective Sampling Network Design under Uncertainty. *Advances in Water Resources, 29*(6), 899-911. doi:10.1016/j.advwatres.2005.08.005

Yang, A. L., Huang, G. H., & Qin, X. S. (2010). An Integrated Simulation-Assessment Approach for Evaluating Health Risks of Groundwater Contamination Under Multiple Uncertainties. *Water Resources Management, 24*(13), 3349-3369. doi:10.1007/s11269-010-9610-3

Yegnan, A., Williamson, D. G., & Graettinger, A. J. (2002). Uncertainty Analysis in Air Dispersion Modeling. *Environmental Modelling and Software, 17*(7), 639-649. doi:10.1016/s1364-8152(02)00026-9

Yeh, H.-D., Chang, T.-H., & Lin, Y.-C. (2007). Groundwater Contaminant Source Identification by a Hybrid Heuristic Approach. *Water Resources Research, 43*(9), 1-16. doi:10.1029/2005wr004731

Yeh, W. W. G. (2015). Review: Optimization methods for groundwater modeling and management. *Hydrogeology Journal, 23*(6), 1051-1065. doi:10.1007/s10040-015-1260-3

Yenigül, N. B., Elfeki, A. M. M., Gehrels, J. C., van den Akker, C., Hensbergen, A. T., & Dekking, F. M. (2005). Reliability Assessment of Groundwater Monitoring Networks at Landfill Sites. *Journal of Hydrology, 308*(1), 1-17. doi:10.1016/j.jhydrol.2004.10.017

Yu, X. W. (1994). *Study of physical and chemical properties of groundwater and surface water in the northern part of the Botany Basin, Sydney.* (PhD), University of New South wales, Sydney, Australia (unpubl).

Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control, 8*(3), 338-353. doi:10.1016/s0019-9958(65)90241-x

Zhang, Y., Pinder, G. F., & Herrera, G. S. (2005). Least Cost Design of Groundwater Quality Monitoring Networks. *WATER RESOURCES RESEARCH, 41*(8), W08412. doi:10.1029/2005wr003936

Zheng, C., Hill, M. C., & Hsieh, P. A. (2001). *MODFLOW-2000, The U.S. Geological Survey Modular Ground-Water Model- User Guid to the LMT6 Package, The Linkage with MT3DMS for Multi-Species Mass Transport Modeling.* Open File Report 01-82. Retrieved from Denver, Colorado.

Zheng, C., & Wang, P. P. (1999). *A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems*. Retrieved from Washington, DC. http://inside.mines.edu/~epoeter/583/14/discussion/MT3DMS_manual.pdf

APPENDIX I

Notation

AEE	Absolute Error of Estimation
$ANAEE$	Average Normalized Absolute Error of Estimation
C	Solute concentration in groundwater (ML^{-3})
C_{obs}	Error free concentration at monitoring locations
C_{pert}	Perturbed concentration values
$Cest_{iob}^{k}$	Concentrations estimated by the simulation model at observation location *iob* and at the end of time period *k*
$Cest_{iob,r}^{k}$	Estimated concentration at monitoring location *iob* and time *k* using *r* hydraulic conductivity realization
$Cest_{iob,nr}^{k}$	Estimated concentration at monitoring location *iob* and time period *k* using nr^{th} SGS generated hydraulic conductivity realization
$Cest_{i,k}^{p}$	Estimated concentration at monitoring location *i* and at time *k* using $\vec{q}_p$ as candidate solutions for the source characteristics
$Cest_{i,k}^{ave}$	Estimated concentration at monitoring location *i* and time *k*, averaged over all possible sets of source release histories
$Cestn_{iob}^{k}$	Average concentration estimated at the monitoring location *iob* and time period *k*
$Cestnr_{iob}^{k}$	Average concentration estimated at the monitoring location *iob* and time period *k* using all SGS generated realizations

$C int_{i,k}^{p}$	Interpolated contaminant concentration at potential monitoring location i and time k using candidate source characteristics solutions
$Cobs_{error\ free,iob}^{k}\ k$	Error-free concentration measurements at monitoring location *iob* and time k
$Cobs_{iob}^{k}$	Observed concentrations at observation location *iob* and at the end of time period k
D	Dispersion coefficient
$D_{j,k}$	Dispersion coefficient tensor (L^2T^{-1})
d_i	Distance between the estimated point and the sample point
d_{ij}	Distance between points i and j
e_{iob}^{k}	A measurement error factor for monitoring location *iob* and time k
F_{iob}^{k} , $F_{i,j}$	Coefficient of uncertainty
$FOBJ1, FOBJ2$	Vector containing all the final objective function values of the resulting Pareto-front
$fobj1_i$, $fobj2_i$	i^{th} final (non-dominated) objective function value of the resulting Pareto-front
H	Potentiometric head (L)
HC	Hydraulic conductivity
I	Moran's I
K_r	Hydraulic conductivity field realization
K_{xx}, K_{yy}, and K_{zz}	Hydraulic conductivities (L/T) along the x, y, and z coordinate axes
$Max_{Monitoring}$	Maximum number of selected monitoring locations

N	Total number of candidate source locations
n, n_r	Number of closest sampled data points used for interpolation purposes
$NAEE\%$	Normalized Absolute Error of Estimation
nk	Total number of concentration observation time periods
nn	A constant
nob	Total number of available monitoring locations
NP	Total number of available set of estimated contaminant release fluxes
NR	Number of hydraulic conductivity realizations
NS	Total number of candidate source locations
NT	Total number of monitoring design time steps
p	The exponent parameter.
$p(COC \le \eta_1)$, $p(COC \le \eta_2)$	The probabilities that the COC value is less than or equal to η_1 and η_2
Q	Vector containing all available estimates of contaminant release fluxes
Q_1^{FOBJ1}, Q_3^{FOBJ1}, Q_1^{FOBJ2}, Q_3^{FOBJ2}	First and third quartiles of the non-dominated objective function values
q_{est}^i and q_{act}^i	Estimated and actual source fluxes for stress period i
$q_{evaluate}^i$	Evaluation source fluxes for stress period i
$q_{\max}$ and $q_{\min}$	Upper and lower bound for contaminant release fluxes
$\vec{q}_p$	p^{th} set of available estimate of source release fluxes

$q^t_{s,p}$	p^{th} estimated release flux at candidate source location *s* and stress period *t*
R	Number of hydraulic conductivity realizations
S_S	Specific storage of the porous material (L^{-1})
S_{ud}	A uniform random number between -1 and +1
ST	Total number of stress periods
t	Time
$TotalN$	Total number of monitoring locations including the newly selected and existing ones
U	Normalizing factor
u_j	Groundwater velocity in three dimensions (LT^{-1})
x_i, y_i, z_i	Cartesian coordinates of candidate contaminant source *i*
x_i , x_l	Three-dimensional coordinates of potential monitoring locations *i* and *l*
X_i	A flag
W	Volumetric flux per unit (T^{-1})
w_i	Weights related to each $Z(x_i)$ value
$w_{i,j}$	Weight applied based on distance between locations *i* and *j*
$Z(x_i)$	IDW interpolated value
α	A constant
β	A fraction between 0 and 1.0 and illustrates the level of measurement error
η_i	A flag
θ	Porosity

$\mu_{\text{iob}}^{\text{k}}$	Coefficient of Confidence (COC) assigned to monitoring location *iob* at the end of time period *k*
μ_Y	The mean value
σ	Standard deviation
ϕ	A fraction between 0 and 1.0
$\sum_{p=1}^{NR} R_p$	Chemical reaction terms ($ML^{-3}T^{-1}$)