Development of Integrated Methodologies for Optimal Characterization of Reactive Contaminant Sources and Monitoring Network Design in Polluted Aquifer Sites

Thesis submitted by

Hamed K. Esfahani, B.Sc. Eng., M.Tech

for the degree of Doctor of Philosophy (Ph.D.)

in the School of Science and Engineering

James Cook University, Australia

August 2016

STATEMENT OF ACCESS

I, the undersigned, author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Theses network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and I wish this work to be embargoed until August 2017, after which date I do not wish to place any further restriction on access to this work.

Signature Date

STATEMENT OF SOURCES DECLARATION

I herewith declare that I have produced this thesis without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This thesis has not previously been presented in identical or similar form to any other Australian or foreign examination board.

Signature Date

ELECTRONIC COPY

I, the undersigned, the author of this work, declare that the electronic copy of this thesis provided to James Cook University Library is an accurate copy of the print thesis submitted.

Signature Date

STATEMENT OF CONTRIBUTION OF OTHERS

Financial contribution towards this PhD project was received from:

1. James Cook University, Graduate Research School Scholarship (JCUPRS) in 2013.

Apart from the financial assistance, the following have contributed to this PhD project as specified hereunder:

Dr. Bithin Datta: Dr. Datta supervised the entire PhD project and helped conceptualise the problem, suggested several ways of solving it and provided insights on the tools and technologies used in this thesis.

Lynne A Nathan: Ms Nathan copy edited and professionally proof-read the thesis.

Acknowledgements

Foremost, I am deeply indebted to my supervisor, Dr. Bithin Datta, whose patience, kindness and vast academic experience, have been invaluable to me. Without his continuous and unconditional support, this thesis would not have been possible.

I am also thankful to Graduate Research School at James Cook University for generously funding my research and providing substantial support for attending several conferences during the course of my study.

I must also thank the School of Science and Engineering and the High Performance Computing Unit at James Cook University for providing excellent infrastructure and a collaborative work environment. My colleagues at James Cook University, especially the PhD candidates in Engineering, have been my support network throughout the tenure of my PhD candidature. They made my stay at JCU a lot more enjoyable and fun-filled.

Last, but not least, I would like to thank my wife, Sara, for her understanding and love during the past few years. Her support and encouragement was in the end what made this dissertation possible. She has been my inspiration and motivation for continuing to improve my knowledge and move my career forward. I also thank my wonderful daughter, Adrina, for always making me smile.

Publications produced during PhD candidature

Peer-reviewed:

1. Esfahani, H. K. and B. Datta, 2016. Fractal Singularity Index Based Multi-objective Monitoring Network Design for Unknown Contaminant Source Characterization in an Aquifer with Reactive Transport Processes. Journal of Water Resource Research. Under Review.

2. **Esfahani, H. K.** and B. Datta, 2016. Genetic Programming Based Simulation of Complex Reactive Geochemical Transport Processes with Application to an Abandoned Mine Site Aquifer in Queensland Australia. Environmental Technology and Innovation (ETI) journal. Under Review.

3. **Esfahani, H. K.** and B. Datta, Linked Optimal Reactive Contaminant Source Characterization in Contaminated Mine Sites: A Case Study. Journal of Water Resource Planning and Management (ASCE). Accepted.

4. Datta, B., **Esfahani, K. H.** et al, 2016. Preliminary Hydrogeologic Modeling and Optimal Monitoring Network Design for a Contaminated Abandoned Mine Site Area: Application of a Developed Monitoring Network Design Software. Journal of Water Resource and Protection. Vol. 8, pp. 46-64.

5. **Esfahani, K. H.**, Datta, B., 2015, 'Development of Methodologies for Optimal Characterization of Reactive Contaminant Sources in Polluted Aquifer Sites'. Int. J. GEOMATE, Vol.8 No.1, pp. 1190-1196, and also Presented in GEOMATE 2014.

6. **Esfahani, K.H.** and B. Datta, Use of Genetic Programming Based Surrogate Models to Simulate Complex Geochemical Transport Processes in Contaminated Mine Sites, in: A.H. Gandomi, Alavi, A. H., Ryan, C., Editor, Handbook of Genetic Programming Application, Springer International Publishing, Switzerland. 2015b, pp. 359-379.

Conference presentations:

1. **Esfahani, K.H.** and B. Datta, 2016**.** Simulation of Complex Geochemical Processes in Contaminated Aquifer Sites Using Surrogate Models As Approximate Simulators. International Conference of Civil Engineering, Tehran, Iran.

2. **Esfahani, K.H.** and B. Datta, 2015**.** Simulation of Complex Geochemical Processes in Contaminated Aquifer Sites Using Surrogate Models As Approximate Simulators. 6th International Contaminated Site Remediation Conference (Clean-up 2015), Melbourne, Australia.

3. Datta, B., **Esfahani, K. H.** et al, 2015**.** Recent Developments in Identification of Unknown Contamination Sources and Monitoring Network Design for Contaminated Groundwater Systems. 6th International Contaminated Site Remediation Conference (Clean-up 2015), Melbourne, Australia.

4. **Esfahani, K.H.**, Datta, B., 2015, 'Approximate Simulation of Complex Geochemical Transport Processes by Genetic Programming Based Surrogate Models with Application to a Closed Down Mine Site in Queensland Australia'. Presented in 2nd IWA Malaysia Young Water Professional Conference (YWP15), Kuala Lumpur, Malaysia.

Abstract

Groundwater is one of the most important natural resources in many parts of the world; however it is widely polluted due to human activities. Reliable groundwater management and remediation strategies are generally developed following the identification of groundwater pollution sources, where the measured data from monitoring locations are utilized to estimate the unknown pollutant source location, magnitude and duration of activity. However, accurately identifying characteristics of unknown contaminant sources is a challenging task due to uncertainties in terms of predicting source flux injection, hydro-geological and geo-chemical parameters, and the concentration observation field measurement. Although sufficient concentration measurement data are essential to accurately identify source characteristics, available data are often sparse and limited in quantity. Therefore, this inverse problem of characterizing unknown groundwater pollution sources is often considered ill-posed, complex and non-unique.

Different methods have been utilized to identify pollution source characteristics; however, the linked simulation-optimization approach is one effective method to obtain acceptable results under uncertainties in complex real life scenarios. With this approach, the numerical flow and contaminant transport simulation models are externally linked to an optimization algorithm, with the objective of minimizing the difference between measured concentration and estimated pollutant concentration at observation locations. Concentration measurement data are very important to accurately estimate pollution source properties; therefore, optimal design of the monitoring network is essential to gather adequate measured data at desired times and locations.

A simulation model should be utilized to accurately describe the aquifer processes properties in terms of hydro-geochemical parameters and boundary conditions. However, the simulation of the transport processes becomes complex when the pollutants are chemically reactive. An additional difficulty with linked simulation-optimization models is that an optimal solution generally requires huge computation time, due to iterative repeated solution of the numerical flow and transport simulation models. To address this, Genetic Programming based surrogate models may be used to approximate the numerical simulation model in the linked simulation-optimization model for source characterization. Therefore, the aim of the present study is to demonstrate the feasibility and efficiency of a developed methodology to optimally identify or characterize the unknown distributed pollution sources with chemically reactive species in complex contaminated aquifers.

Because the accuracy and reliability of the source characterization process depends on the quality and extent of the spatial and temporal concentration measurement data, a relevant issue is the design and implementation of a suitable and efficient monitoring network under conditions of various uncertainties. This is especially true, where the initial measurement data available are sparse and obtained from arbitrary monitoring locations. Therefore a new two objective Pareto optimal monitoring network design methodology is developed. This design methodology utilizes Fractal Singularity Mapping Technique to determine plume boundaries, information used to select potential monitoring locations for contaminant concentration monitoring. This approach substantially improves the source characterization efficiency as demonstrated for illustrative study areas. In order to improve the efficiency and accuracy of the source characterization methodology in real life sites where

contamination is evident, but the monitoring data are very sparse and arbitrary, the monitoring network design model is integrated with the source characterization process by sequentially utilizing the source characterization model to estimate the sources. This information then, is utilized to design and implement a cost effective monitoring network. This sequential and iterative methodology is shown to improve the source characterization efficiency and accuracy, even when dealing with a hydrogeochemically complex aquifer system with multiple reactive species.

The performance of the linked source characterization model is also evaluated by limited application to real life sites, which included a complex, abandoned mine site in Queensland, Australia. The sequential source characterization and monitoring network design methodology is applied to a contaminated aquifer in an urban area in Australia. Several techniques are utilized in the proposed methodology to increase the efficiency of the source characterization including trained and tested Genetic Programming based surrogate models, Adaptive Simulated Annealing optimization algorithm, Fractal Singularity Mapping Technique, and Statistical Kriging interpolation.

The study includes the following steps: 1. The flow and reactive contaminant transport simulation model is utilized to simulate the aquifer processes; 2. Trained Genetic Programming (GP) based meta-models are developed using the simulated response of the aquifer to randomly generated source fluxes. The selected GP models replace the numerical simulation model in the linked simulation-optimization model for source characterization. 3. Two objectives Pareto-optimal design of a monitoring network for sequential characterization of pollutant sources uses a linked simulation-optimization model incorporating Adaptive Simulated Annealing as the optimization algorithm. 4. Integrated source identification and monitoring network design is carried out to obtain

sufficient accuracy in characterization of source properties. 5. The performance of the developed methodologies is evaluated by limited application of the developed methodologies to real life sites.

Table of Contents

List of Figures

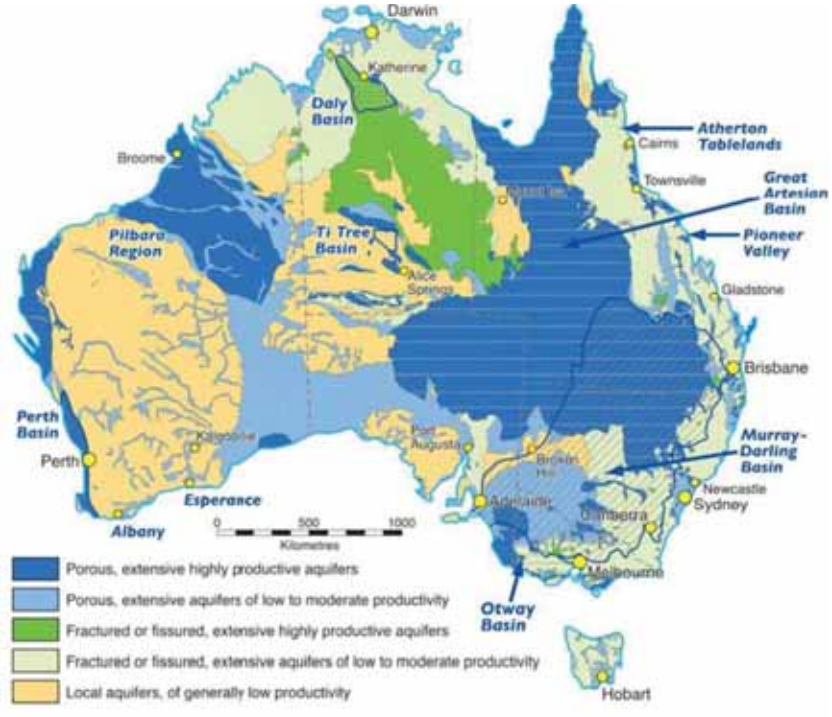
List of Tables

1. Introduction

An important step in effective groundwater conservation and remediation is accurate identification of groundwater pollution sources. Contaminant source characteristics identification is widely considered as the first necessary step in developing groundwater management strategies. Here, the pollutant concentration measured at monitoring wells is utilized to estimate the unknown pollutant source location, magnitude and duration of activity. Another step is the optimal design of a monitoring network based on initial arbitrary well locations to accurately estimate pollution source characteristics. This study presents linked simulation-optimization, and sequential monitoring network design based methodologies for characterization of unknown groundwater pollution source characteristics. The optimization models are solved using an Adaptive Simulated Annealing (ASA) optimization algorithm. Genetic Programming (GP) and Fractal singularity Mapping Technique are also utilized to identify unknown contaminant source characteristics and optimal design of monitoring network, respectively in complex contaminated aquifers such as mine sites. The performances of the developed methodologies are evaluated for different complex scenarios of groundwater pollution such as distributed chemical reaction-based mineral deposit waste. These methodologies are also applied to real life contaminated aquifers to demonstrate the potential applicability of these methodologies.

Groundwater, currently accounting for 30 per cent of Australia's total water consumption, is one of the most important natural resources in Australia, especially in areas where evaporation rates exceed rainfall (Magee 2009). Average annual groundwater usage is approximately 3,500 GL. Water used for agricultural irrigation is the largest

single user group, comprising on average approximately 60% of Australia's annual groundwater use. Other user groups include mining (12%), manufacturing and other industries (17%), household water supply (5%), and as an input into potable water supply networks (9%) (NCGRT 2013). Figures 1.1 and 1.2 show the Australia's groundwater resources and Australia's reliance on groundwater, respectively. Widespread pollution of groundwater occurs due to human activities without considerable environmental management. Some of the main pollution sources for groundwater are agricultural waste, waste stream of the water and wastewater treatment, landfills, septic and storage tanks, industrial waste and mine waste drainage (Zhang et al. 2007). In this research, it is proposed to focus mainly on groundwater pollution resulting from mining wastes especially in abandoned mine sites. This category of sources is one of the most complex causes of



groundwater pollution in many locations throughout the world.

Figure 1.1 Australia's groundwater resources (LAU et al. 1987).

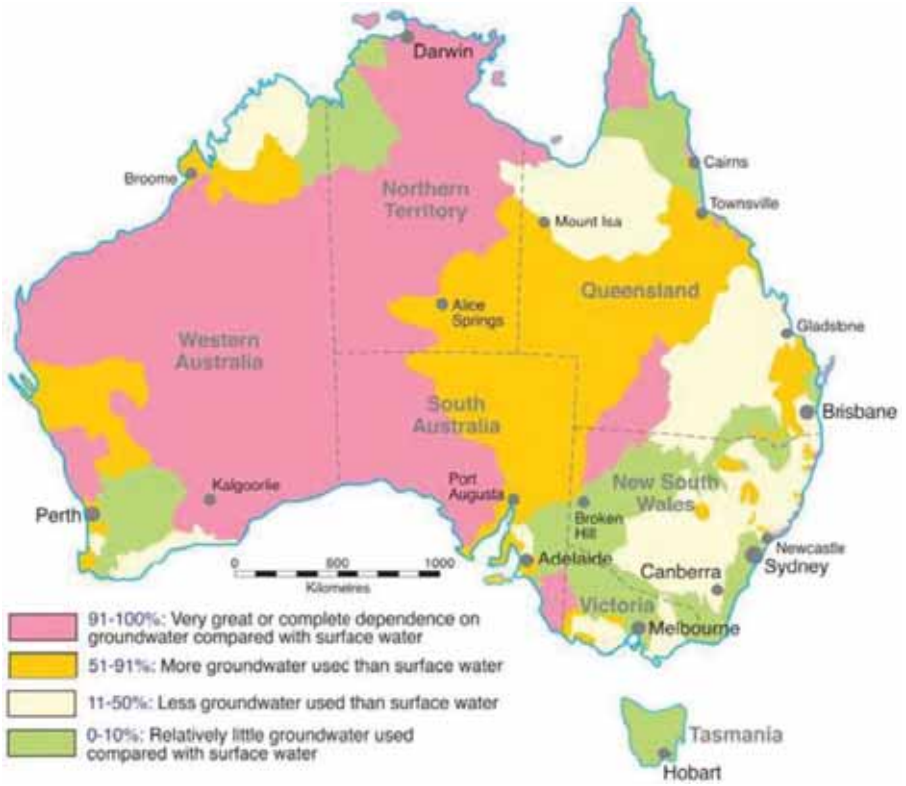


Figure 1.2 Australia's reliance on groundwater (NCGRT 2013).

Acid mine drainage (AMD) or acid rock drainage (ARD) is metal-rich, acidic water that constitutes the wastewater from metal mines or coal mines containing sulphur compounds (Kalin et al. 2006). This highly toxic fluid is involved in a series of chemical weathering reactions which result in the formation and movement of highly acidic water rich in heavy metals. Therefore, considerable amounts of acidity, metals and other pollutions are released into the environment (Smith and Huyck 1999). Thus, Acid Mine

Drainage has a stable hazardous impact on water resources and may cause long-term harmful effects on humans, animals and plants. These contaminants often leak to the underground system without any control. Figure 1.3 shows the primary AMD chemical reactions and Figures 1.4 and 1.5 illustrate the harmful impact of AMD on the environment.

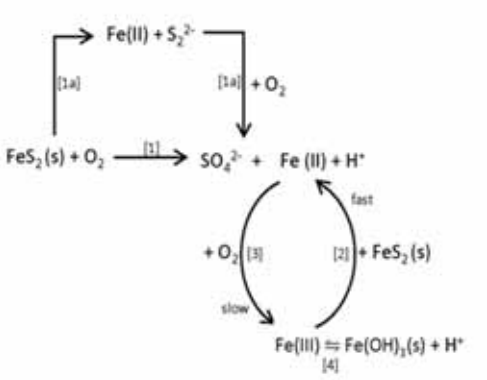


Figure 1.3 Primary AMD chemical reactions.(Kalin et al. 2006)

Figure 1.4 The effect of significant amounts of AMD in surface water.

The primary step required for protection and management of polluted aquifers is identifying the pollution source characteristics, and the ability to predict future pathways of groundwater pollution. Identification of unknown pollutant sources can be improved by using data from an optimally designed monitoring network. Reliable information about the location and release history of pollution sources is considered necessary for planning effective remediation.

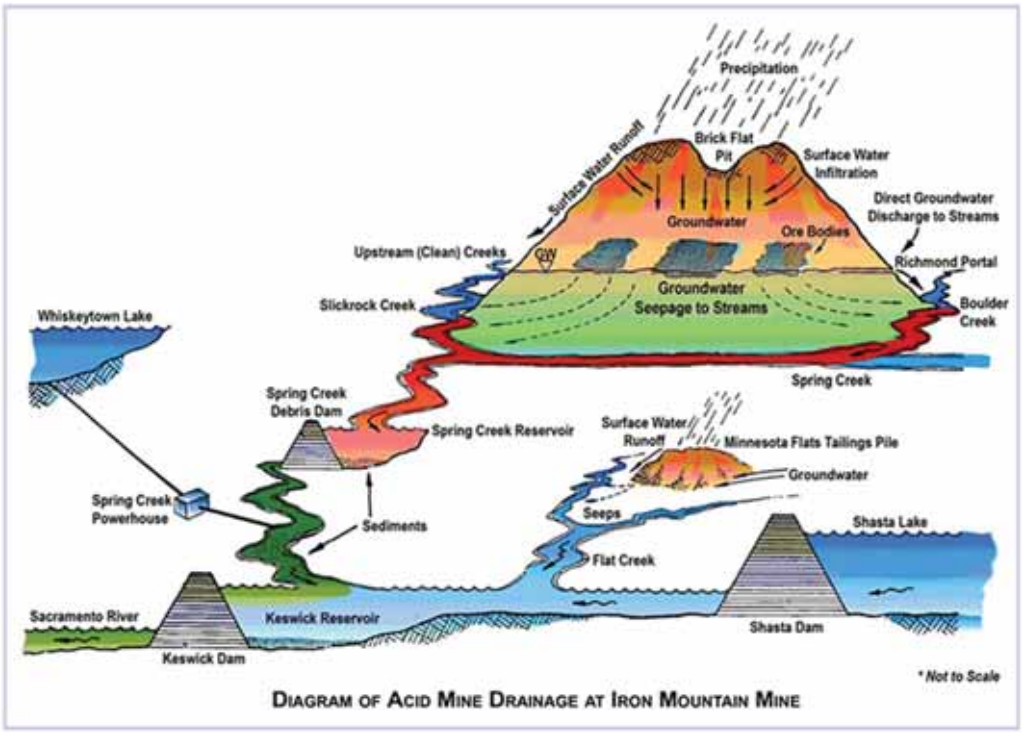


Figure 1.5 AMD movement in surface and subsurface water.

1.1. Unknown Groundwater Pollutant Source Characterisation

There are several serious obstacles to accurately identifying characteristics of the contaminant sources. Some difficulties are the uncertainties in accurately predicting source flux injection in the aquifer, generally encountered sparsity of concentration

observation data in the field, and the non-uniqueness (Datta 2002) in the response to subjected hydraulic and chemical stresses in the aquifer (Datta et al. 2013). Available information are often limited to the pollutant concentration measured from one or more arbitrary wells, guesses about the potential location of the pollutant sources and also hydrogeochemical parameter measurements (porosity, hydraulic conductivity, etc.) (Singh and Datta 2006). Based on these inputs, the following results are expected:

- Location of pollutant sources.
- Magnitude of pollution released into the groundwater.
- The pattern of the contaminant distribution in the aquifer.
- Starting time and period of sources activities.

Although chemical fingerprinting and non-reactive groundwater flow and transport processes have been widely used for solving this problem, unknown source characterization cannot be effectively addressed without incorporation of the chemical-reaction transport simulation (Yeh et al. 1994). Various factors influence chemical behaviours of contaminants in groundwater, including water temperature, pH, and presence of other ions which must be incorporated in the flow and transport processes (Yeh et al. 2004). Finite-difference or finite-element methods have been used to solve the governing equations numerically. Inverse model solution methods have been widely used as effective methods to find the source characteristics identification from a set of measured data. Based on this concept, the unknowns (location, magnitude, and period of activity of the contaminant sources) are determined by solving for fitness in the field concentration measured data and simulated concentration values (Datta et al. 1989). Some challenges exit when solving an inverse problem due to non-uniqueness in solution, ill-

posed nature of the problem, and difficulties in identifying a unique optimum solution (Datta 2002). The main reasons are insufficient observation data, measurement error or uncertainties in model definition (Mahar and Datta 1997).

A more efficient approach is to use a linked simulation-optimization approach to minimize a least square error criterion (Singh and Datta 2004). In this method, the simulation of the physical and chemical processes is incorporated in an optimization model to solve the inverse problem. The previous linked simulation-optimization approach utilized a linear programming and response matrix along with forward simulation. Recently, evolutionary algorithms such as Genetic Algorithm and Adaptive Simulated Annealing have been used as optimization algorithm in linked simulation-optimization methodology for solving the inverse problem. The evolutionary algorithms are more likely to find a global optimal solution to the source characterization problem and also make it easier to use linked simulation-optimization approach.

1.2. Sequential Monitoring Network Dedicated to Source Characterization

Source characterization in highly complex aquifers is the first step of site remediation and water resource management that requires a large amount of concentration measurement data. However, in real-world scenarios several constraints generally limit the number of contamination monitoring wells and hence the availability of field data. Therefore, new techniques are required to optimally design effective monitoring networks for accurate unknown contaminant source characterization (Datta and Dhiman 1996). The sequential monitoring network design (Prakash and Datta 2015) develops the

identification of unknown source characteristics. An important factor related to the optimal design of an effective optimal monitoring network is the initial specification of a suitable set of potential monitoring well locations. In this study, Local Singularity Mapping Technique is utilized to obtain potential well locations used as input to the optimal design model. Also multi-objective optimization is used for solving the optimal monitoring network design model. In the multifractal based singularity mapping technique, a local feature is similar to the whole, in terms of shape and structure. Based on the singularity indices obtained from the singularity mapping technique, the study area is classified into subsets of contaminated and clean areas. The boundaries between these two subset areas can be identified as the contamination plume boundary. This information can be used to identify potential monitoring wells locations for designing the optimal network.

The proposed methodology utilizes a multi objective optimization algorithm for solving a two objective optimal monitoring network design model. The optimization model is linked to a calibrated numerical simulation model simulating flow and transport processes in the aquifer. Although constraining the maximum number of permissible designed monitoring locations, the designed monitoring network improves the source identification results. The optimal locations reduce the possibility of missing the impact of potential contaminant sources on spatiotemporal concentrations at the monitoring locations. At the same time, the designed monitoring network decreases the degree of non-uniqueness in the measured set of possible aquifer responses to geochemical stresses. Sequential optimal monitoring networks can provide vital feedback information in identifying unknown source locations, and in efficient characterization of unknown groundwater pollution sources.

1.3. Research Objectives

This study is aimed at developing optimization based methodologies for characterization of unknown groundwater sources in highly complex contaminated aquifers. In the absence of any useful information about source locations and magnitude of pollution, source characterization is even more difficult. Availability of only limited amounts of observed pollutant concentration data from randomly located monitoring wells further increases the difficulty. Real life scenarios may have multiple point/non-point sources, each having different patterns of pollution releases, and missing measurements from the observation data.

The main goal of the present study is to develop a set of methodologies to optimally identify or characterize unknown distributed pollution sources with reactive species as well as designing an effective monitoring network for pollution source identification in hydrogeologically and geochemically complex polluted aquifers such as an abandoned mine site with contaminated aquifer. The performance of the developed methodology is evaluated by applying the methodology to illustrative polluted aquifer sites, i.e., a previously operational gold and copper mine sites in Queensland, Australia.

Specific objectives of this research are as follows:

I. Implement a numerical three dimensional transient contaminant transport simulation model incorporating multiple species reactive contaminant transport in a contaminated mine site study area.

II. Develop trained and tested surrogate models based on Genetic Programing (GP) to substitute numerical simulation models as approximate simulator of the physical

processes in linked simulation-optimization models for optimal source characterization.

III. Develop a linked simulation-optimization based methodology for optimal unknown groundwater pollution source characterization for aquifer polluted by point and distributed reactive pollutant sources.

IV. Develop an optimal monitoring network design methodology using fractal singularity technique incorporating uncertainty for complex polluted aquifer sites, incorporating multiple objectives of design.

V. Develop and evaluate an integrated source identification and sequential monitoring network design methodology applicable to aquifers polluted by non-point sources and reactive pollutant species.

VI. Evaluate the performance and potential applicability of developed methodologies for a real contaminated aquifer site including a contaminated mine site.

1.4. Organization of the Thesis

This thesis comprises seven chapters including the first chapter: Introduction. The second chapter of the thesis reviews the state-of-art of various topics discussed in this study.

The third chapter presents the development and performance evaluation of Genetic Programming (GP) based surrogate models as an approximate simulator of flow and transport processes in aquifers. These surrogate models can be incorporated within the linked simulation optimization models to drastically reduce the computational time by replacing the numerical simulation model within the linked simulation optimization

model. The reduction in computation time involving repeated simulations ensures the computational feasibility of applying the source characterization methodology to a regional scale complex contaminated aquifer with reactive chemical species under aquifer parameters uncertainties. The process of training and testing the surrogate models computational advantages and the feasibility of using the GP based surrogate models as approximate simulators are demonstrated for complex transport processes with chemically reactive multiple species.

The fourth chapter discusses the source characterization methodology application to a complex mine site area, which covers three main components. The first component consists of the calibration and implementation of the numerical simulation model for multiple species reactive contaminant transport process for the aquifer site. The second component consists of the development of GP models as an approximate simulator within the linked simulation-optimization methodology for estimation of unknown pollutant source characteristics in terms of their location and magnitude. The third component consists of development and application of Adaptive Simulated Annealing (ASA) optimization algorithm based linked simulation-optimization model for unknown source characterization. The application of this methodology to a contaminated mine site in Queensland, Australia is discussed.

The fifth chapter presents the development of an optimal monitoring network design model potentially useful for improved characterization of contaminant sources in terms of flux magnitude, activity time and source locations. The optimal monitoring design is based on multi-objective linked optimization algorithm, and the fractal singularity index. Local Singularity Mapping Technique is utilized to obtain potential well locations. Also multi-objective optimization is utilized for Pareto-optimal design of the monitoring

network to improve the accuracy of groundwater pollution source characterization. Efficiency of the developed methodology is demonstrated by evaluating the performance of the source identification model, when utilizing concentration measurements obtained by using the designed network, with those obtained using measurements from arbitrary monitoring networks.

The sixth chapter presents an application of the methodologies developed in previous chapters to a real life polluted urban groundwater aquifer site in New South Wales, Australia. Feedback based methodology for efficient identification of unknown pollutant source characteristics, integrating sequential monitoring network design with a source identification model is utilized.

Chapter 7 presents a brief discussion of salient features of this study and conclusions. Some of the limitations of the methodologies developed are also mentioned.

Important existing literature relevant to this study is reviewed in the following chapter.

2. Review of Literature

This chapter provides a brief overview of the literature relevant to solving the unknown groundwater pollutant source characterisation problem. Variations of the source characterization problem, in terms of the different unknown source characteristics that were addressed and the solution methodologies that were adopted, are discussed. Various approaches for monitoring network design for more efficient solution of the unknown groundwater pollutant source characterisation problem are also reviewed. An overview of the tools and techniques utilized for simulation, optimization and data interpolation, are discussed in this chapter.

2.1. Unknown Groundwater Pollution Source Characterisation

Contamination in groundwater aquifers can remain undetected for long periods of time. Often, pollutants are detected long after the activity at the sources may have started, or may have ceased to exist. Pollution in groundwater is generally detected by a water supply well or a group of wells in arbitrary locations. Identification of contaminant source characteristics in terms of flux magnitude, activity time and source locations from sparse pollutant concentration measurement, belongs to a category of inverse problems which are generally ill-posed (Yeh 1986). Ill-posed problems exhibit discontinuous dependence on data and high sensitivity to measurement errors. This problem is considered ill-posed since its solution does not satisfy following conditions: existence, uniqueness and stability. For instance, a similar contaminant plume can be obtained from different

combinations of source characteristics and small errors in input can result in large estimation errors.

Various methodologies have been suggested for detecting the groundwater contaminant sources. These methods can be classified as the follows:

- Heat transport inversion based approach
- Direct approach
- Analytical and regression based approach
- Probabilistic and geostatistical simulation approach
- Optimization based approach

A detailed review of Source identification methodologies can be found in Neupauer et al. (2000); Atmadja and Bagtzoglou (2001); and Bagtzoglou and Atmadja (2005).

One of the initial attempts of solving the source characterization problem was a solution approach based on using the inversion of heat transfer (Cannon 1966). Due to similarity between heat and mass transport models and flow and contaminant transport ones, hydro-dynamics can be used to overcome ill-posed based problems in source identification models. However, such approaches require that the parameters used in mathematical models are precisely known. This may be possible in the case of heat transfer as the medium is mostly homogeneous and parameter values can be accurately measured. The same assumptions do not hold true in groundwater systems as they are highly heterogeneous and the aquifer properties such as hydraulic conductivity, and porosity, are average approximations for the entire study area. Therefore, the use of heat transport inversion methods has been limited.

Tikhonov regulation technique was utilized as a direct approach for unknown source identification (Skaggs and Z.J. 1994, Liu and Ball 1999). This technique transformed the ill-posed algebraically indeterminate inverse problem into a minimization problem to get a unique solution. However, even a small error in the input data significantly reduced the solution accuracy. Other direct approaches, namely Marching-Jury Backward Beam Equation (MJBBE) (Atmadja and Bagtzoglou 2001, 2003), particle based censored random walk (Ababou et al. 2010), and reverse time particle tracking (Bagtzoglou 2003), have been applied for identifying the source characteristics.

An analytical and regression based approach has been applied in characterization of unknown groundwater pollution sources. Some of the significant studies using these methods are Sidauruk et al. (1998); Alpati and Kabala (2000) ; and Ala and Domenico (1992). However, these analytical solutions work only for a limited number of cases where the aquifer is homogeneous with simple geometry and flow conditions.

Bagtzoglou et al. (1991, 1992) presented a probabilistic frame work for solving the source characterization problem. They use random walk method to solve the transport equation backwards in time. Neupaur and Wilson (1999, 2005) used an adjoint method to find backward in time location and travel time probabilities. Though these methods can address the problem of non-homogeneity in the porous media parameters, they required solving the inverse governing stochastic equations. The requirement for extensive computational resources limits the applicability of these methods to simplified one-dimensional or simple two-dimensional problems.

Due to the ill-posed nature of inverted transport equation, a different approach for identification of source characteristics as simulation-optimization has been utilized. It

couples the forward time contaminant simulation model with the optimization techniques. This approach avoids the problem of non-uniqueness and stability associated with formally solving the inverse problem. However, the iterative nature of simulation model usually requires increased computational effort. Various techniques are proposed in literature based on a coupled simulation-optimization methodology to solve the source characterization problem. Some of the representatives are discussed below.

An optimization based methodology for the characteristics of unknown groundwater pollution sources has been presented by many researchers. Gorelick et al. (1983) presented the response matrix approach in the identification of unknown pollution source models. In their study, they combined forward-time simulations with an optimization model using linear programming and multiple regressions. Datta et al. (1989) developed and evaluated the performance of a new methodology utilizing optimal with the pattern-matching technique for unknown source characteristics identification. This methodology was based on stochastic dynamic programming for minimizing the loss of misrecognition. Uncertainties in the identification process were addressed using an Expert System.

Wagner (1992) developed a methodology to estimate hydrogeologic parameters as well as source characterization. In this attempt, Wagner used a non-linear maximum likelihood estimation problem. Unknown parameters of hydrogeological and source parameters were estimated based on hydraulic head and pollutant concentration data. Mahar and Datta (1997, 2001) used non-linear optimization algorithm to estimate the source characteristics (location, magnitude, and duration of activity). Mahar and Datta (1997; 2001) were the first to couple groundwater sources' identification and monitoring network design to increase the accuracy of characterization process.

Aral et al. (2001) used progressive genetic algorithm (PGA) as a nonlinear optimization model to identify pollutant source locations and release histories. The proposed methodology by Singh and Datta (2004) estimated both the unknown pollution sources and hydrogeological parameters using an artificial neural network. Singh and Datta (2006) evaluated the performance of the developed methodology by using an external flow and transport simulation model to the genetic algorithm (GA)-based optimization model. They linked an external simulator to the optimization program properly to increase the efficiency and flexibility of the model.

Yoon et al. (2007) proposed a three-dimensional groundwater flow and solute transport model (MODFLOW-GWT) along with simulated annealing (SA) optimization algorithm and tabu search (TS). Ayvaz (2010) developed a simulation-optimization based model using a heuristic harmony search (HS) optimization technique to solve the unknown groundwater pollution source characteristics. The study illustrated the feasibility of identifying numbers of unknown pollution sources along with their locations and release histories in aquifer systems. This study used MODFLOW and MT3DMS for the flow and transport concentration model simulation, respectively.

Chandalavada et al. (2012) developed agroundwater source characterization methodology using feedback monitoring measurements obtained from a sequentially designed monitoring network. Their study presented a GA-based simulation-optimization link to obtain the optimal source location from the identified potential locations. They demonstrated the efficiency of the developed methodology for a real life polluted aquifer site. Prakash and Datta (2013) developed a methodology to identify unknown source characteristics using feedback from optimal monitoring network. This source identification methodology applied simulated annealing (SA) as the optimization

algorithm. Concentration measurement data obtained from the implemented sequentially designed monitoring networks were utilized for source characterization. Jha and Datta (2013) extended the inverse problem methodology for optimization of a groundwater pollution source identification model using the adaptive simulated annealing algorithm (ASA) in linked simulation-optimization model. They utilized a dedicated optimal monitoring network to improve the accuracy of source identification. Jha and Datta (2015b) utilized ASA as optimization algorithm for distributed source characterization in complex study areas. Optimal solutions obtained in their study show that the linked simulation-optimization based methodology is potentially applicable for the characterization of spatially distributed pollutant sources, typically present at abandoned mine site areas.

2.2. Monitoring Network design for Unknown Groundwater Pollution Source Characterisation

Monitoring networks are integral to groundwater management. Design of monitoring network may have different objectives, and vary as per aquifer properties and budgetary constraints. Monitoring networks are essentially installed for extracting information which would assist in achieving the underlying objectives for which the monitoring network was installed in the first place. A large body of literature exists, dealing with the design of monitoring networks for different groundwater quality management objectives.

Designing an optimal monitoring network is dependent on different factors such as aquifer parameters, the selected objectives, specific constraints and limitations. Some of the design methodologies for an optimal monitoring network design proposed during the

last two decades include: detection of contamination (Loaiciga et al. 1992, Loaiciga and Hudak 1993); reducing the cost of monitoring for groundwater quality monitoring (Loaiciga 1989, Fethi et al. 1994); multiple objectives groundwater monitoring network design (Reed and Minsker 2004); using monitoring network design for source identification and redundancy reduction with feedback information (Dhar and Datta 2007, Datta et al. 2009); sampling strategy in space and time using Kalman filter (Kollat et al. 2011, Reed and Kollat 2013); long term monitoring network using multi-objective simulation–optimization model under uncertainties (Luoa et al. 2016); and Optimal contaminant source characterization by integrating sequential-monitoring-network design and source identification (Prakash and Datta 2015).

Meyer and Brill Jr (1988) presented a methodology to design an optimal monitoring network incorporating contaminant movement simulation under uncertain conditions. The Monte Carlo technique was used to apply transport parameters of uncertainty in the contaminant concentration distribution. In addition, Meyer et al. (1994) developed a method that incorporates conditions of uncertainty in monitoring network design. They illustrated three objectives: (i) minimizing the number of observation wells for monitoring network, (ii) maximizing the probability of detecting a pollutant leak, (iii) minimizing the expected area of pollution at the time of detection.

Chandalavada and Datta (2007) extended optimal monitoring network design methodology to include uncertainty in both source and aquifer parameters. A genetic algorithm and a geostatistical tool, kriging, are used for solving the optimization model and computing the variances of estimated concentrations at potential monitoring locations, respectively. Their study applied the results of two following objective functions to obtain optimal design of monitoring locations: i) minimization of the

summation of unmonitored pollutant concentrations at potential locations where the probable concentration value is large and ii) minimization of estimated concentration variances at potential locations.

Dhar and Datta (2007) also developed an optimal design methodology of well locations using optimization models. Based on a transient pollutant transport process, optimal design of monitoring networks was obtained using dynamic wells locations in different time stages. This time varying network design minimized costs in comparison with designing all well locations in a single stage. Dhar and Datta (2010) presented a method for optimally minimizing the number of monitoring wells using an inverse distance weighting method and linear optimization model. Chandalavada et al. (2011b) developed a methodology to track the pollutant plume with minimum integrate monitoring wells. This optimized design of monitoring network incorporated with the application of uncertainty in estimating the pollutant concentration. The developed methodology was applied to a polluted defence site in Australia.

Chandalavada et al. (2011) designed an optimal monitoring network to delineate the pollutant plume using a minimum number of monitoring wells. Monitoring wells were installed at locations having minimum measurement uncertainty. The uncertainty in the study area was quantified by using concentration estimation variances at all the potential monitoring locations. Jha and Datta (2012) used a Dynamic Time Warping system to estimate the starting time of the source activity. The estimated starting time of the source activity was further utilized in comparing the observed pollutant concentration measurements with the simulated pollutant concentration measurements correctly in time in a linked simulation optimization model. ASA was used to solve the optimization problem. Prakash and Datta (2015) proposed a new feedback-based methodology for

efficient identification of unknown pollutant-source characteristics, integrating sequential-monitoring-network design and source identification processes. They evaluated the performance of the proposed methodology to a real polluted aquifer in New South Wales, Australia.

2.3. Relevant Tools and Techniques

Relevant literature on tools for modelling groundwater flow and transport processes, data interpolation technique, optimization algorithm, and regression modelling used at different stages throughout this study are discussed in this section.

2.3.1. Flow and Transport Modelling with Reactive Contaminant Transport

Mathematical models are abstractions that represent the physical processes describing the cause and effect relationship. In groundwater systems, these models are used to simulate the processes of flow and solute transport to compute the concentration of a dissolved chemical species in an aquifer at any specified time and place.

Groundwater flow is generally governed by Darcy's law and conservation of mass. The theoretical basis for the equation describing solute transport has been well documented in (Bear 1979). Analytical solutions are also available Javandel et al. (1984), but are limited to unrealistically idealised scenarios. Numerical methods, namely finite-difference and finite-element methods are commonly used for the solution of mathematical equations used for flow and transport simulation. A number of numerical simulation models using finite difference and finite element methods have been developed, some of which have been discussed by Wang and Anderson (1982); Anderson and Woessner (1992); Zheng

and Bennett (1995); and Domenico and Schwartz (1998). However, only those numerical simulation models with a reactive equation transport in aquifers are discussed here.

In many cases, an accurate description of the contaminant transport in aquifers is obtained if both chemical and physical behaviours of contaminant species are incorporated. Algebraic equilibrium equations are often used to describe fast physicochemical processes (involving adsorption, ion exchange, complexation, dissolution, etc.). As identified by Rubin (1983), a system of partial differential and algebraic equations can be used to simulate the transport processes with chemical reactive contaminants.

To solve the geochemical transport problem, two different approaches: the one-step (the physical and chemical equations combined in one procedure) and two-step (two separate procedures for Physical transport equations and chemical equations) procedures can be used. The benefit of the two-step procedure is its reasonable flexibility in switching between chemical systems for replacing different chemical reaction subroutines. For instance, PHREEQE (Parkhurst et al. 1982) is a widely used computer code using the two-step procedure for solution. Herzer and Kinzelbach (1989) combined an iterative two-step procedure both for transport and chemical reaction into one single two-dimensional finite difference transport simulation code. Their study demonstrated that the maximum error occurs in heterogeneous equilibrium reaction (between a solid and liquid phase). This error is acceptable as long as it remains smaller than the physical dispersion process.

Tebes-Stevensa et al. (1998) developed a reactive transport simulation code (FEREACT) to investigate the chemical reaction-based transport for cobalt and EDTA

(Ethylene Diamine Tetra-acetic Acid) through an iron oxide-coated sand column. The code included both geochemical and microbiological reaction processes. Their study evaluated equilibrium and three types of kinetic (adsorption, surface dissolution and biodegradation) reactions in aquatic environment with two-dimensional steady-state groundwater flow.

Zhang et al. (2007) developed the chemical reaction-based simulation by M (species numbers) partial differential equations (PDEs). They divided total reactive transport equations into two sets of equations: a set of thermodynamic equilibrium equations and a set of reactive transport equations with kinetic-variables without equilibrium reactions. Using kinetic-variables rather than individual species simplifies the reaction terms in equations and also reduces the number of reactive transport equations. They suggested a methodology with robust and efficient numerical integration by expulsion of equilibrium reactions from reactive transport equations.

Prommer et al. (2002a) illustrated Numerical experiments of both non-reactive and reactive transport at a hydrocarbon-contaminated field site. The study used the PHT3D (Prommer 2002b) model coupling the non-reactive model MT3DMS (Zheng and Wang 1999) with the reactive geochemical model PHREEQC-2 (Parkhurst and Appelo 1999). Moreover, MT3DMS is widely used in various chemical reactive transport simulators such as RT3D (Clement 1997), SEAM3D (Waddill and Widdowson 1998), BioRedox (Carey et al. 1999), DART (Freedman and Ibaraki 2002), PHAST (Parkhurst et al. 2004) and PHWAT (Mao et al. 2006). These models have been coupled with saturated media flow codes. On the other hand, one-dimensional movement simulators such as HP1 (Simunek et al. 2009) and RICH-PHREEQ, (Wissmeier and Barry 2010) have been used

in recent years as unsaturated porous media flow models incorporating detailed geochemistry.

Wissmeier and Barry (2011) presented a software tool for groundwater flow and transport movement in combination with geochemistry in two- and three-dimensional systems. Their study coupled the reaction-based transport equations in the computer source code, IPHREEQC (Charlton and Parkhurst 2011), to the finite element solver COMSOL (COMSOL 2010) for flow and transport simulations.

HYDROGEOCHEM (Yeh and Tripathi 1991) was the first comprehensive simulator of hydrologic transport and geochemical reaction in saturated-unsaturated media. It iteratively solves the two-dimensional transport and geochemical equilibrium equations. Yeh et al. (1995) developed HYDROGEOCHEM as a Hybrid Lagrangian-Eulerian Finite Element Model (LEHGC) to simulate transport through Saturated-Unsaturated Media. This modification increases the efficiency of solution and allows the use of larger time steps.

Yeh and Salvage (1997) introduced HYDROGEOCHEM 2.0 with an expanded LEHGC 1.1 (Yeh et al. 1995) with mixed equilibrium and kinetic geochemical reactions. The modified algorithm of the LEHGC developed a more effective LEHGC 1.1. This version of the simulator is more robust for using more grid nodes than LEHGC. HYDROGEOCHEM 4.0 (Li 2003) was the last version of this code in two dimensions. This simulator combined heat transfer with reactive geochemical and biochemical transport to incorporate the effect of thermal transport.

HYDROGEOCHEM 5.0 (HGCH) (Yeh et al. 2004) (Sun 2004) was developed as the latest version of the code in three dimensions. Based on HYDROGEOCHEM 4.0, this

code was developed to solve comprehensive heat, reactive geochemical and biochemical transport. In addition, this simulator provides three options for numerical approximations for reactive geochemical and biochemical transport and heat transfer equations: conventional finite element methods (FEM), hybrid Lagrangian-Eulerian FEM, or hybrid Lagrangian-Eulerian FEM for interior nodes plus FEM for boundary nodes. Nevertheless, HYDROGEOCHEM 5.0 has some limitations. These include (1) single fluid phase can be applied and (2) dual-porosity media cannot be used effectively (Yeh et al. 2004). This proposed study uses HYDROGEOCHEM 5.0 as the numerical model for simulating transport of reactive chemical species in an aquifer.

AMD and ARD are hazardous contaminant sources for groundwater, therefore relevant literature describing its production/movement procedure in aquifers are reviewed here. Detailed causes and effects of acid mine drainage has been reviewed in recent years for instance Johnson and Hallberg (2005); Akcil and Koldas (2006) and Kalin et al. (2006). Various sulphide minerals constitute AMD based on their chemical weathering reactions such as pyrite (FeS2), pyrrhotite (Fe1-xS), chalcopyrite (CuFeS2), arsenopyrite (FeAsS), etc. Nordstrom and Alpers (1999). Beside mining activities, rocks' surface weathering in presence of water, oxygen and microorganisms produces AMD. These contaminants are considerable threats for water resources.

The low pH in the AMD environment provides a good base for metal and heavy metal ions solution such as, iron, aluminium and manganese, and also the low concentration of zinc, cadmium, copper and lead (Smith and Huyck 1999). Various earlier studies discuss AMD with a focus on oxygen consumption for weathering reactions with simplified geochemistry consequences (Jaynes et al. 1984, Elberling et al. 1994, Wunderly et al. 1996, Werner and Berglund 1999, Bain et al. 2000). These studies attempt to model

AMD behaviour in groundwater incorporating different physical processes and oxidation of iron sulphide.

Except for iron sulphide oxidation by O_2, geochemical processes are slow and limited by (bio)geochemical kinetics, which can be neglected or assumed as mass-balance or local equilibrium. Scharer et al. (1994), Lichtner (1996) and Mayer et al. (2003) mention aluminosilicate weathering and/or Fe(II) oxidation in tailings' deposits with geochemistry limitation. As they found, these limitations have an impact on the composition of AMD; however, a wide range of (bio)geochemical reactions was not assessed. Salmon and Malmstrom (2004) focused on geochemical processes that contribute to concentrations of major ions in the groundwater and that determine the proton balance within pyritic tailings deposits with negligible carbonate mineral content.

Zhao et al. (2012) evaluated reaction transport movements of the sulfide minerals' components and heavy metals in groundwater in the Dabaoshan Mine area in China, using the PHREEQC geochemical software. They simulated chemical reactions based on Fe^{2+}/Fe^{3+} which is generated by acid mine drainage. Demers et al. (2013) developed a methodology for numerical analysis of contaminated drainage under near-neutral conditions. Their study produced valuable results using flow and geochemical reaction-based transport simulation code, MIN3P (Mayer et al. 2002), along with laboratory measured data. Also Molson et al. (2008) used MIN3P to acid mine drainage (AMD) flow and transport simulation. Their study showed that covered waste deposits with the capillary barrier effects (CCBEs) reduce AMD generation. This covered soil layer remains approximately saturated, which limited oxygen availability and AMD production.

2.3.2. Techniques for Geostatistical Data Interpolation

Data interpolation requires estimating the value of a variable at an unmeasured location from observed values at surrounding locations. Matheron (1963) developed a method of geostatistical data interpolation called “kriging”. Kriging is a collection of generalised linear regression techniques for minimizing an estimation variance defined from a prior model for covariance (Deutsch and Journel 1998). An overview of kriging based geostatistical data interpolation techniques can be found in Journel and Huijbregts (1978); and Cressie (1990).

Geostatistical data interpolation was initially developed with an emphasis on solving mining related problems but has found wide application in all major engineering fields. Geostatistical kriging has been extensively applied to solve various groundwater management problems. Reed et al. (2000) developed a long term cost-effective monitoring network design for polluted aquifer sites. The method combined a transport simulation model, plume interpolation, and a genetic algorithm to identify cost-effective sampling plans. Inverse Distance Weighting (IDW), Ordinary Kriging (OK) and a hybrid method that combines the two approaches were used for plume interpolation.

Wu et al. (2005) extended the work by Reed et al. (2000) by introducing the first and second moments of a three-dimensional pollutant plume as new constraints in the optimization formulation. Application of geostatistical kriging in groundwater monitoring network design can also be found in Yeh et al. (2006); and Feng-guang et al. (2008). A MATLAB open source code, mGstat version 0.99 (Hansen 2004) and kriging packages from the Geostatistical Software Library (GSLIB) (Deutsch and Journel 1998) are used for data interpolation in this study.

2.3.3. Evolutionary Optimization Algorithm

Characterization of unknown groundwater pollution sources is often formulated as an optimization problem and can be solved using different optimization algorithms. Choice of optimization algorithm largely depends on the type of problem to be solved. In this study, Adaptive Simulated Annealing (ASA) is used as the solution algorithm to solve the optimization problem.

Objective functions for solving unknown groundwater pollution source characterization are complex multi-variate optimization problems. Such formulations are highly non-linear, containing several local and global optima. Simulated annealing is a meta-heuristic search algorithm capable of escaping from local optima. Its use of hill-climbing moves to escape local optima makes SA efficient in solving non-convex optimization problems. Its ease of implementation of complex objective functions, and convergence to a global optimal solution, enhances its suitability for solving ill-posed inverse problems, as is the case with unknown groundwater pollution source characterization. A number of methodologies have been proposed using different optimization algorithms for improving the source identification results as reported in Datta and Kourakos (2015), Yeh (2015) and Ketabchi and Ataie-Ashtiani (2015).

Simulated Annealing (SA) first introduced by Kirkpatrick et al. (1983), is an extension of the metropolis Algorithm (Metropolis et al. 1953). The basic concept of SA is derived from thermodynamics. Each step of the SA algorithm replaces the current solution by a random nearby solution, chosen with a probability that depends on the difference between the corresponding function values and algorithm control parameters that is gradually decreased during the process. SIMANN is a FORTRAN code for SA developed by Goffe

is publicly available since 1996. Ingber (1993) developed SA with c-language program as Adaptive Simulated Annealing (ASA). ASA is a version of the SA algorithm in which the algorithm parameters are automatically adjusted as the algorithm progresses. These parameters control temperature, schedule, and random step selection. Thus, the main advantage of the ASA algorithm is that it is more efficient and less sensitive to user defined parameters. Goldberg (1989) presented a single objective optimization tool, real coded genetic algorithm (RGA). In addition, Deb (2001) modified RGA to make it more efficient, and simple to use. The majority of utilized optimization algorithm include: genetic algorithm (GA) (Wu et al. 2005, Chandalavada et al. 2011b), simulated annealing (SA) (Jha and Datta 2011, Prakash and Datta 2015), adaptive simulated annealing (ASA) (Esfahani and Datta 2016), and Genetic Programming (GP) (Datta et al. 2014, Esfahani and Datta 2015a, 2015b).

2.3.4. Surrogate Models

Linked simulation-optimization model solutions incorporating simulation of very complex physical processes generally require huge computation time, due to iterative repeated solution of the flow and transport simulation part within the optimization algorithm. To address this, surrogate models approximating the simulation model can be used together with the optimization algorithm to increase the feasibility and efficiency of the source characterization methodology. These surrogate models are generally computer programs describing the relationship between output values (e. g., pollutant concentration at well locations and times) and input values (e. g., pollutant flux at potential pollutant source locations). A response matrix approach as an initial linear surrogate model was earlier used to simulate the aquifer responses (Zhou et al. 2003,

Abarca et al. 2006). More recently, proposed non-linear surrogate models include Artificial Neural Network (ANN) and Genetic Programming (GP) (Koza 1994) based surrogate models.

Artificial Neural Networks (ANN) has been widely used as approximate surrogate models for groundwater simulation (Aly and Peralta 1999). Rogers et al. (1995) presented one of the earliest attempts using ANN as a surrogate for a coastal groundwater flow model. They demonstrated the substantial saving in terms of computation time by using ANN and Genetic Algorithmic (GA) based meta-model (surrogate model) within a linked simulation-optimization model for evolving optimal groundwater management strategies. Replacing groundwater simulation models with ANN-base surrogate models were developed by Bhattacharjya and Datta (2005), Bhattacharjya and Datta (2009), Bhattacharjya et al. (2007) and Dhar and Datta (2009). McPhee and Yeh (2006) used ordinary differential equation surrogates to approximately simulate groundwater flow and transport processes. Even the most popularly used trained ANN-based surrogate modelling approach obtains the optimal model formulation by trial and error (Bhattacharjya and Datta 2005).

Bhattacharjya et al. (2007) used ANN as an approximate simulation for substituting the three dimensional flow and transport simulation model to simulate the complex flow and transport processes in a coastal aquifer. Bhattacharjya and Datta (2009) used the trained ANN-based surrogate models for approximating density dependent saltwater intrusion process in coastal aquifer to predict the complex flow and transport processes. Dhar and Datta (2009) used ANN as a surrogate model for simulation of flow and transport in the multiple objective non-dominated front search process resulting in saving a huge amount of computational time.

Genetic programming is an evolutionary optimization algorithm based on the concepts of genetics and natural selection. It bears strong resemblance to GA. A GP model is essentially a highly fit computer program describing the relationship between output values and inputs, evolved using genetic programming (Koza 1994). GP is often used to perform symbolic regression. Most conventional regression algorithms optimize the parameters for a pre-specified model structure. However, with GP, the model structure and parameters are determined simultaneously. GP optimises the parameter values of a given model structure within predefined parameter space to find a highly fit computer program that produces desired output for a particular set of inputs. Compared to other regression techniques, the most important advantage of GP is its ability to optimize both the variables and constants of the candidate models without initial model structure definition. This approach makes GP a strong surrogate model to characterize the model structure uncertainty.

GP typically codes solutions as tree structured variable length chromosomes. The first step towards development of GP was performed by Cramer (1985), in which he developed the first tree structured GAs for basic symbolic regression. Classification rules using structured GA were developed by Forsyth and Roy (1986). However, it was Koza (1994) who coded the GP algorithm in LISP. This was applied to a wide range of problems, including symbolic regression and classification.

Recently genetic programming has been utilized in hydrological applications in several researches (Dorado et al. 2002, Makkeasorn et al. 2008, Wang et al. 2009). Trained GP-based surrogate models have been used to substitute the simulation models for runoff prediction, river stage and real-time wave forecasting (Savic et al. 1999, Khu et al. 2001, Sheta and Mahmoud 2001, Whigham and Craper 2001, Babovic and Keijzer 2002, Gaur

and Deo 2008). In addition, GP has been applied to approximate modelling of different geophysical processes including flow over a flexible bed (Babovic and Abbott 1997); urban fractured-rock aquifer dynamics (Hong and Rosen 2002); temperature downscaling (Coulibaly 2004); rainfall-recharge process (Hong et al. 2005); soil moisture (Makkeasorn et al. 2006); evapotranspiration (Parasuraman et al. 2007b); and saturated hydraulic conductivity (Parasuraman et al. 2007a). Zechman et al. (2005) developed trained GP-based surrogate models as an approximate simulation of groundwater flow and transport processes in a groundwater pollutant source identification problem.

GP is domain independent, and this flexibility renders it the capability to be used for structural optimisation of various engineering problems. However, GP has not been widely applied in groundwater resource management problems. Sreekanth and Datta (2010) implemented GP as meta-model to replace the flow and transport simulation of density dependent saltwater intrusion in coastal aquifers for ultimate development of optimal saltwater intrusion management strategies. Sreekanth and Datta (2011b, 2012) compared two non-linear surrogate models based on GP and ANN models, respectively, and showed that the GP based models perform better in some aspects. These advantages include: simpler surrogate models, optimizing the model structure more efficiently, and parsimony of parameters. Datta et al. (2013) developed a monitoring network design feedback methodology to identify unknown source characteristics using GP surrogate model. They proposed a combination of GP with linked simulation-optimization to identify the unknown pollutant sources with limited number of concentration measured data. Replacing the simulation groundwater model by GP-based ensemble surrogate models in linked simulation-optimization developed methodology was addressed by Datta et al. (2014) and Sreekanth and Datta (2011) which improve the computational

efficiency and obtains reasonably accurate results under aquifer hydrogeologic uncertainties. In this study, this feature has been exploited for the design of an optimal monitoring network and is explained in Chapter 3. A professional software package, Discipulus^{TM} 5.1 (RML Technologies Inc.) is used in this study for GP modelling.

2.3.5. Fractal Singularity Mapping Technique

Optimal monitoring network design methodologies have been obtained from initial potential monitoring well locations. Therefore, the initial potential well locations affect the optimal network design. The potential well locations depend on the selected objectives for monitoring network design; for example, the boundary of the contaminant plume affects the potential locations for unknown source characterization using optimal monitoring network design. A detailed review of this technique can be found in Zuo and Wang (2016).

The main attraction of fractal/multifractal theory lies in its ability to quantify irregular and complex phenomena or processes that exhibit similarity over a wide range of scales, which is termed self-similarity (Mandelbrot 1989). Since the concept of fractal was introduced by Mandelbrot in the 1960s, a number of studies were applied to geological processes and phenomena to characterize the spatial distributions of concentrations and the relationship between tonnage and grade of deposits (Cheng et al. 1994). Some of recent studies which apply fractal and multifractal based methodologies include: earthquake (Turcotte 2002), flooding (Cheng 2008), geoscience (Chen et al. 2015) and groundwater contamination plume delineation (Datta et al. 2016). Fractal models such as Number-Size model (N-S), Concentration-Area model (C-A) (Cheng et al. 1994), Spectrum-Area model (S-A) (Cheng et al. 1999, 2000), simulated size–number (SS–N)

model (Sadeghi et al. 2015), singularity index (Chen et al. 2015), and Concentration-Volume model (C-V) (Afzal et al. 2011) have been developed for geochemical data analysis.

Considering the fact that different points within a small vicinity may have different singularity exponents, Chen et al. (2007) proposed the concept of the local singularity iteration algorithm as an improvement of parameter estimation involved in the local singularity analysis. Cheng (2012) proposed a density–area power-law model, attempting to systematically confirm that singularity analysis is useful for recognition of weak geochemical anomalies. Agterberg (2012a, 2012b) noticed that singularity exponents are linearly related to logarithmically transformed element concentration values, which can be used to measure the small-scale nugget effect caused by measurement error and microscopic randomness. Zuo et al. (2015) found that the local singularity index calculated by the original algorithm is influenced by background values and proposed a modified algorithm to overcome these shortcomings.

Considering the difference of elemental mobility in areas with overburden, Xiao et al. (2014) investigated the possibility of mobile elements by means of accumulation coefficient analysis to determine a set of suitable indicator elements for further analysis. This procedure is necessary for identification of weak anomalies from buried mineralization. Utilizing the advantage that the singularity exponent can identify heterogeneity, Wang et al. (2013) proposed a new model, the fault trace-oriented singularity mapping technique, to characterize anisotropic mineralization-associated geochemical signatures. Furthermore, a tectonic-geochemical exploration model that focuses on fault controlled and geochemical halo-associated mineralization was also constructed by Wang et al. (2012). Note that a singularity distribution map derived from a

singularity mapping technique requires further analysis for modelling mineral prospectively, which usually involves setting a series of thresholds to divide the singularity exponents into binary patterns and then calculating the spatial correlation between these patterns and known deposits and occurrences. These correlation indices, termed Student's t-values, can be estimated by the weights of evidence method. Further details on this method can be found in related literature (Cheng et al. 1994).

2.4. Motivation for this Study

Based on review of existing methodologies for water resource management, there is scope to further develop computationally efficient methodologies for unknown groundwater contaminant characterization in highly complex aquifer study areas such as mine site areas. Reactive transport of chemical species, in contaminated groundwater systems, especially with multiple species, is a complex and highly non-linear process. In order to increase the model reliability for real field data, uncertainties in hydrogeological parameters and boundary conditions are needed to be considered as well. In this study, flow and chemically reactive transport processes are numerically simulated incorporating uncertainties in complex contaminant aquifers.

Simulation of such complex geochemical processes using efficient numerical models is generally computationally intensive. Also, the determination of an optimal contaminated aquifer management and remediation strategy often requires repeated solutions of complex and nonlinear numerical flow and contamination process simulation models. To address this combination of issues, and uncertainties in predicting the response of the aquifer, trained ensemble Genetic Programming (GP) surrogate models are utilized as

approximate simulators of these complex physical processes in the contaminated aquifer. In this study, the proposed source characterization methodology uses GP based surrogate models as approximate simulation model in the linked simulation-optimization approach. The efficiency and computational feasibility of source characterization using GP models are evaluated for complex study areas.

Source characterization in highly complex aquifers requires a large amount of concentration measurement data. However, in real-world scenarios several constrains generally limit the number of contamination monitoring wells and hence the availability of field data. Therefore, new techniques are required to optimally design effective monitoring networks for accurate unknown contaminant source characterization. An important factor related to the optimal design of an effective optimal monitoring network is the initial specification of a suitable set of potential monitoring well locations. In this study, Local Singularity Mapping Technique is utilized to obtain potential well locations used as input to the optimal design model. Also the multi-objective optimization model is used for solving the optimal monitoring network design model.

Systematic and planned monitoring at optimal monitoring locations can provide vital feedback information for efficient source characterization. This aspect has received only limited attention in the past. This study, therefore, incorporates a sequential monitoring network design and gathered information feedback based methodology for efficient identification of unknown groundwater contaminant source characteristics. The linked simulation-optimization model is solved using observation data from sequentially designed monitoring networks to optimize the source characterization results.

The trained and tested GP based surrogate models as approximate simulator is discussed in Chapter 3.

3. Use of Genetic Programming Based Surrogate Models to Simulate Complex Geochemical Transport Processes in Contaminated Mine Sites

Similar versions of this chapter have been published and copyrighted to appear in the following journals:

- Esfahani, K.H., Datta, B., 2015, 'Development of Methodologies for Optimal Characterization of Reactive Contaminant Sources in Polluted Aquifer Sites'. Int. J. GEOMATE, Vol.8 No.1, pp. 1190-1196, and also Presented in GEOMATE 2014.
- Esfahani, K.H. and B. Datta, 2015, Use of Genetic Programming Based Surrogate Models to Simulate Complex Geochemical Transport Processes in Contaminated Mine Sites, in: A.H. Gandomi, Alavi, A. H., Ryan, C., Editor, Handbook of Genetic Programming Application, Springer International Publishing, Switzerland. 2015b, pp. 359-379.

This chapter presents the development of Genetic Programming (GP) based ensemble surrogate models to serve as computationally efficient approximate simulators of complex groundwater contaminant transport process with reactive chemical species under

aquifer parameters uncertainties. Performance evaluation of the developed methodology for a contaminated aquifer site resembling an actual abandoned mine site is also discussed.

3.1. Background of the Problem

Reactive transport of chemical species, in contaminated groundwater systems, especially with multiple species, is a complex and highly non-linear process. Simulation of such complex geochemical processes using efficient numerical models is generally computationally intensive. In order to increase the model reliability for real field data, uncertainties in hydrogeological parameters and boundary conditions are needed to be considered as well. Also, often the development of an optimal contaminated aquifer management and remediation strategy requires repeated solutions of complex and nonlinear numerical flow and contaminant transport process simulation models. To address these combination of issues, trained ensemble Genetic Programming (GP) surrogate models can be utilized as approximate simulators of these complex physical processes in the contaminated aquifer. For example, use of trained GP surrogate models can reduce the computational burden in solving linked simulation based groundwater aquifer management models (Sreekanth and Datta 2011) by orders of magnitude. Ensemble GP models trained as surrogate models can also incorporate various uncertainties in modelling the flow and transport processes. The development and performance evaluation of ensemble GP models to serve as computationally efficient approximate simulators of the complex groundwater contaminant transport process with reactive chemical species under aquifer parameters uncertainties are presented. Performance evaluation of the ensemble GP models as surrogate models for the reactive

species transport in groundwater demonstrates the feasibility of its use and the associated computational advantages. In order to evolve any strategy for management and control of contamination in a groundwater aquifer system, a simulation model needs to be utilized to accurately describe the aquifer processes in terms of hydro-geochemical parameters and boundary conditions. However, the simulation of the transport processes becomes complex and extremely non-linear when the pollutants are chemically reactive. In many contaminated groundwater aquifer management scenarios, an efficient strategy is necessary for effective and reliable remediation and control of the contaminated aquifer. Also, a hydrogeologically complex aquifer site, for example, mining site, acid mine drainage (AMD) and the reactive chemical species together with very complex geology, complicates the characterization of contamination source location and pathways.

In such contamination scenarios, it becomes necessary to develop optimal source characterization models, and strategies for future remediation. Solution of optimization models either for source characterization, or optimal management strategy development requires the incorporation of the complex physical processes in the aquifer. Also, most of the developed optimization models for source characterization or remediation strategy development require repeated solution of the numerical simulation models within the optimization algorithm. This process is enormously time consuming and often restricts the computational feasibility of such optimization approaches.

In order to overcome these computational restrictions, and to ensure computational feasibility of characterizing sources and pathways of contamination it is computationally advantageous to develop surrogate models which can be trained using solutions obtained from rigorous numerical simulation models. A number of attempts have been reported by researchers to develop surrogate models for approximately simulating the physical

processes. Especially the use of trained Artificial Neural Network (ANN) models has been reported by a number of researchers (Ranjithan et al. 1993). However, the architecture of an ANN model needs to be determined by extensive trial and error solutions, and may not be suitable to deal with the simulation of very complex geochemical processes in a contaminated aquifer site such as mine sites. Genetic Programming (GP) based surrogate models may overcome some of the limitations of earlier reported surrogate models. Therefore, this chapter develops GP models to approximately simulate 3-dimensional, reactive, multiple chemical species transport in contaminated aquifers.

Trained and tested GP based surrogate models are developed using the simulated response of a complex contaminated aquifer to randomly generated source fluxes. An ensemble GP model is an extension of the GP modelling technique capable of incorporating various uncertainties in a contaminated aquifer system data. These ensemble GP models are trained and tested utilizing transient, three dimensional groundwater flow and transport simulation models for an illustrative study area, hydrogeologically representing an abandoned mine site in Australia. Performance of the developed surrogate models is also evaluated by comparing GP model solutions with solution results obtained by using a rigorous numerical simulation of the aquifer processes. The three dimensional finite element based transient flow and contaminant transport process simulator, HYDROGEOCHEM 5.0 (HGCH) is used for this purpose. Reactive transport processes incorporating acid mine drainage in a typical mine site is simulated. Comparison of the solutions obtained with the surrogate models and the numerical simulation model solution results show that the ensemble GP surrogate models

can provide acceptable approximations of the complex transport process in contaminated groundwater aquifers, with a complex geochemical scenario.

The performance of the developed surrogate models is evaluated for an illustrative study area to establish the suitability of GP models as surrogate models for such complex geological processes. These surrogate models if suitable will ensure the computational feasibility of developing optimization based models for source characterization, and help in the development of optimum strategies for remediation of large contaminated aquifer study areas. This chapter will demonstrate the utility and feasibility of using trained and tested ensemble GP models as a tool for approximate simulation of the complex geochemical processes in contaminated mine sites.

In this chapter, the main objectives are to develop ensemble genetic programming based surrogate models to approximately simulate the complex transport process in a complex hydrogeologic system with reactive chemical species, and to illustrate its efficiency and reliability for a contaminated aquifer resembling an abandoned mine site. The numerical model formulations as well as use of ensemble genetic programming based surrogate models are described in the methodology section and the results are presented and discussed in the results and discussion section.

3.2. Methodology

The methodology developed includes two main components. In the first step, the numerical simulation model for the flow and transport processes is described. The complex chemical reactive transport process in the contaminated aquifer is simulated by using the HGCH, a 3-dimensional coupled physical and chemical transport simulator.

The hydrogeochemical data and boundary conditions for the illustrative study site are similar to an abandoned mine site in Queensland, Australia. Trained ensemble GP based surrogate models are then developed to approximately obtain concentrations of the chemical contaminants at different times in specified locations while incorporating uncertainties in hydrogeological aquifer parameters like hydraulic conductivity. Comparison of the spatiotemporal concentrations obtained as solution by solving the implemented numerical three dimensional reactive contaminant transport simulation model (HGCH), and those obtained by using ensemble GP models are then presented to show the potential applicability and the efficiency of using GP ensemble surrogate models under aquifer uncertainties.

3.2.1. Simulation Model of Groundwater Flow and Geochemical Transport Processes

Numerical mathematical models utilize sets of finite element or finite difference equations that represent the discretised physical process of flow and chemically reactive transport in the groundwater system. These models are solved to compute the concentration of a dissolved chemical species in an aquifer at any specified time and place. A finite element based numerical method is used to solve the governing equations of groundwater flow and transport processes in this study.

The fundamental principle of conservation of mass of fluid or of solute is applied for deriving mathematical equations that describe groundwater flow and transport processes. The governing principle of conservation of mass, often known as the “continuity equation”, is used in conjunction with mathematical equations for the relevant processes to obtain equations describing flow or transport.

The general equations for flow through saturated-unsaturated media are obtained based on following components: i) fluid continuity, ii) solid continuity, iii) Fluid movement (Darcy's law), iv) stabilization of media, and v) water compressibility (Yeh et al. 1994). Following governing equations are used:

$$\frac{\rho}{\rho_0} F \frac{\partial h}{\partial t} = \nabla . \left[\boldsymbol{K} . \left(\nabla h + \frac{\rho}{\rho_0} \nabla z \right) \right] + \frac{\rho^*}{\rho_0} q \tag{3.1}$$

F is the generalized storage coefficient (1/L) defined as:

$$F = \alpha' \frac{\theta}{n_e} + \beta' \theta + n_e \frac{ds}{dh} \tag{3.2}$$

K is the Hydraulic conductivity tensor (L/T) is:

$$\boldsymbol{K} = \frac{\rho g}{\mu} \boldsymbol{k} = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} \frac{\rho_0 g}{\mu_0} \boldsymbol{k}_s \, k_r = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} \boldsymbol{K}_{so} \, k_r \tag{3.3}$$

V is the Darcy's Velocity (L/T) described as:

$$\boldsymbol{V} = -K \left[\frac{\rho}{\rho_0} \nabla h + \nabla z \right] \tag{3.4}$$

Where:

θ: effective moisture content (L3/ L3);

h: pressure head (L);

t: time (T);

z: potential head (L);

q: source or sink of fluid [(L3/ L3)/T];

ρ_0: fluid density without biochemical concentration (M/ L3);

ρ: fluid density with dissolved biochemical concentration (M/ L3);

ρ*: fluid density of either injection (= ρ*) or withdraw (= ρ) (M/ L3);

μ_0: fluid dynamic viscosity at zero biogeochemical concentration [(M/L)/T];

μ: the fluid dynamic viscosity with dissolved biogeochemical concentrations [(M/L)/T];

α′: modified compressibility of the soil matrix (1/L);

ß: modified compressibility of the liquid (1/L);

ne: effective porosity (L3/L3);

S: degree of effective saturation of water;

G: is the gravity (L/T2);

k: permeability tensor (L2);

ks: saturated permeability tensor (L2);

Kso: referenced saturated hydraulic conductivity tensor (L/T);

kr: relative permeability or relative hydraulic conductivity (dimensionless);

When combined with appropriate boundary and initial conditions, the above equations are used to simulate the temporal-spatial distributions of the hydrological variables, including pressure head, total head, effective moisture content, and Darcy's velocity in a specified study area.

The contaminant transport equations in groundwater systems can be derived based on:

i. advective transport, in which dissolved chemicals are moving with the flowing groundwater;

ii. hydrodynamic dispersion, in which molecular and ionic diffusion and small-scale variations in the flow velocity through the porous media cause the paths of dissolved molecules and ions to diverge or spread from the average direction of groundwater flow;

iii. fluid sources, where water of one composition is introduced into and mixed with water of a different composition;

iv. chemical reactions, in which some amount of a particular dissolved chemical species may be added to or removed from the groundwater as a result of chemical, physical, biological, dispersion/diffusion, source/sink, and biogeochemical reactions in the water or between the water and the solid aquifer materials or other separate liquid phases as the major transport processes can be written as follows (Yeh, 2000):

$$\frac{D}{Dt}\int_{v}\theta C_i dv = -\int_{\Gamma} n.(\theta C_i)V_i d\Gamma - \int_{\Gamma} n.J_i d\Gamma + \int_{v}\theta r_i dv + \int_{v} M_i dv, i \in M \qquad (3.5)$$

Where

C_i: the concentration of the i-th species in mole per unit fluid volume (M/L3);

ν: the material volume containing constant amount of media (L3);

Γ: the surface enclosing the material volume ν (L2);

n: the outward unit vector normal to the surface Γ;

Ji: the surface flux of the i-th species due to dispersion and diffusion with respect to relative fluid velocity [(M/T)/L2];

θri: the production rate of the i-th species per unit medium volume due to all biogeochemical reactions [(M/L3)/T];

Mi: the external source/sink rate of the i-th species per unit medium volume [(M/L3)/T];

M: the number of biogeochemical species;

Vi: the transporting velocity relative to the solid of the i-th biogeochemical species (L/T).

The transport equation is solved to compute the concentration of a dissolved chemical species in an aquifer at any specified time and place in the study area. In order to solve the transport equation, linear pore water velocity needs to be known for the study area. Hence, it becomes necessary to first calculate the hydraulic head distribution using a groundwater flow simulation model.

3.2.1.1. Numerical Models for Simulation of Groundwater Flow and Transport Processes

Analytical methods for solution of the finite element equations describing groundwater flow and solute transport require that the properties and boundaries of the flow system be highly idealized. Analytical methods provide exact solutions, but come at the cost of over simplifying assumptions of the complex field environment. Alternatively, for problems where analytical methods are inadequate, finite element equations are approximated

numerically. Mathematical models of groundwater flow or transport assume the variables to be continuous. In order to apply numerical methods of solution, the study area is discretised into nodes and elements. Continuous variables are replaced with discrete variables defined at these elements. In order to solve these governing equations of groundwater flow and solute transport, it is necessary to specify the boundary condition. Boundary condition is generally specified as values of head or solute concentration around a boundary (Dirichlet condition), or the flux or concentration gradient around a boundary (Neumann condition). In some cases it is also possible to specify a combination of these two boundary conditions. In order to solve for transient conditions, it is also essential to know the initial conditions. Since groundwater flow and chemically reactive transport are inherently transient, it is necessary to know the initial condition in order to solve the governing equations. The initial condition is essentially the starting head or pollutant concentration in groundwater aquifer.

HYDROGEOCHEM 5.0 (HGCH), consisting of the numerical flow simulator and physio-chemical transport simulator is a computer program that numerically solves the three-dimensional groundwater flow and transport equations for a porous medium. The finite-element method is used in this simulation model.

3.2.2. Genetic Programming Based ensembles Surrogate Model

GP models are used in this study to evolve surrogate models for approximately simulating flow and transport processes in a contaminated mine site. Trained GP models are developed using the simulated response of the aquifer to randomly generated source fluxes. GP, a branch of genetic algorithms (Koza 1994), is an evolutionary algorithm-based methodology inspired by biological evolution to find computer programs that

perform a user-defined task (Sreekanth and Datta 2011b). Essentially, GP is a set of instructions and a fitness function to measure how well a computer model has performed a task. The main difference between GP and genetic algorithms is the representation of the solution. GP creates computer programs in the Lisp or scheme computer languages, as the solution. Genetic algorithms create a string of numbers that represent the solution.

The main operators applied in genetic programming as in evolutionary algorithms are crossover and mutation. Crossover is applied on an individual by simply replacing one of the nodes with another node from another individual in the population. With a tree-based representation, replacing a node means replacing the whole branch (Figure 3.1). This adds greater effectiveness to the crossover operator. The expressions resulting from crossover are very different from their initial parents. Mutation affects an individual in the population. It can replace a whole node in the selected individual, or it can replace just the node's information. To maintain integrity, operations must be fail-safe or the type of information the node holds must be taken into account. For example, mutation must be aware of binary operation nodes, or the operator must be able to handle missing values.

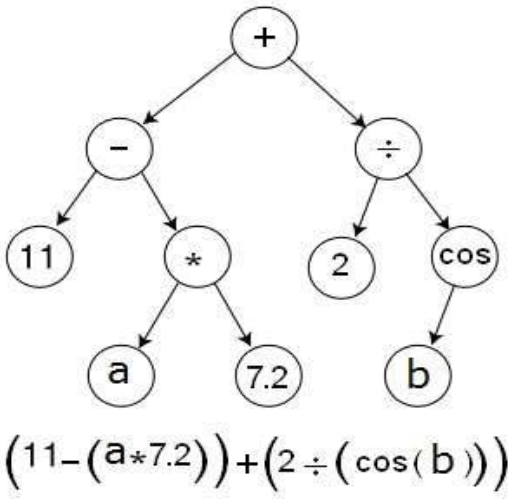


Figure 3.1 Function represented as a tree structure

GP utilizes a set of input-output data which are generated randomly by using the flow and contaminant transport simulation models. The numerical Simulation model creates M number of out-put sets from M number of input sets, which is generated by using random Latin Hypercube sampling in defined ranges. The performance of each GP program is an evaluated formulation in terms of training, testing the validation using the set of input-output patterns. The testing data evaluate the model performance for new data using the fitness function obtained in the training phase. Non-tree representations have been proposed and successfully implemented, such as linear genetic programming which suits the more traditional imperative languages (Banzhaf et al. 1998). The commercial GP software Discipulus (Francone 1998) performs better by using automatic induction of binary machine code. In the proposed methodology, GP software (Discipulus 5.2) is used to solve and generate GP models. Discipulus uses Linear Genetic Programming (LGP) which utilizes input variables in line-by-line approach. This objective of this program is minimizing difference in value between the output estimated by GP program on each

pattern and the actual outcome. The fitness objective functions are often absolute error or minimum squared error. Almost two-thirds of the input-output data sets obtained from the numerical simulation model are utilized for training and testing the GP model. The remaining data sets are used to validate the GP models. The r-square value shows the fitness efficiency to the GP models (Sreekanth and Datta 2010).

The trained ensemble GP surrogate models are evaluated to verify the performance of the surrogate models approximating flow and transport processes simulation with reactive chemical species, under hydrogeological uncertainties. Input data sets for training the GP models are generated randomly by Latin Hypercube sampling within defined ranges. The aquifer hydrogeological uncertainties considered include uncertainties in estimating hydraulic conductivity; water content, and specified groundwater hydraulic head boundary conditions.

3.3. Performance Evaluation of Developed Methodology

In order to evaluate the performance of the proposed methodology, ensemble GP based surrogate models are utilized for an illustrative study area (Figure 3.2). The specified hydrogeologic conditions resemble a homogeneous and isotropic aquifer. In order to evaluate the methodology, the ensemble GP surrogate models are first trained using the sets of solution results obtained using the 3-D finite element based flow and reactive transport simulation model. Then the surrogate model solution results are compared with the actual numerical simulation solution results. Once trained and tested, the GP models are utilized for simulating the transport process in the study site.

The areal extent of the specified study area for performance evaluation purpose is 10000 m^2, with complex pollutant sources including a point source and a distributed source. The spatial concentrations are assumed to be measured at different times at ten arbitrary observation well locations. The thickness of the aquifer is specified as 50m with anisotropic hydraulic conductivity in the three directions. The boundaries of the study area are specified as no-flow boundary for the top and bottom sides while left and right sides of the aquifer have constant head boundaries with specified hydraulic head values. The total head decreases from top to bottom and left to right gradually. The illustrative aquifer is shown in Figure 3.2.

As shown in Figure 3.2, the dark blue areas represent the contaminant sources S(i) which include distributed and point sources. Concentration data from monitoring well locations, shown as black rectangular points, are used to train, test and validate the GP model formulations.

Table 3.1 shows dimensions, hydrogeological properties, and boundary conditions of the study area which are utilized in the numerical model to simulate groundwater flow and chemical reactive transport processes. The synthetic concentration measurement data used for the specified polluted aquifer facilitates evaluation of the developed methodology. These synthetic concentration measurement data at specified observation locations are obtained by solving the numerical simulation model with known pollution sources, boundary conditions, initial conditions, and hydrogeologic as well as geochemical parameter values. In the incorporated scenario, copper (Cu^{2+}), Iron (Fe^{2+}) and sulphate (SO_4^{2-}) are specified as the chemical species in the pollutant sources. The associated chemical reactions are listed in Table 3.2.

Incorporating deviations given by ±5 percent of the originally specified or mean value of the hydrolgeologic parameters to represent uncertainties with hydrogeological properties of the synthetic study area results in nine different scenarios, based on different hydraulic conductivity and boundary conditions values ($N_{ij;}$ i=1, 2, 3; j=1, 2, 3). The first digit (i) indicates an index for the hydraulic conductivity error measure, and the second one (j) represents an index for the hydraulic head as boundary condition. For example, the scenarios with the specified hydraulic conductivity data given in Table 3.1 is indicated with index i = 2. Similarly, the 5 percent lower and higher values for hydraulic conductivity in comparison with the originally specified hydrogeologic data (Table 3.1) are indicated with number 1 and 3, respectively. The same pattern is applied for the j values, for utilizing different values of boundary conditions.

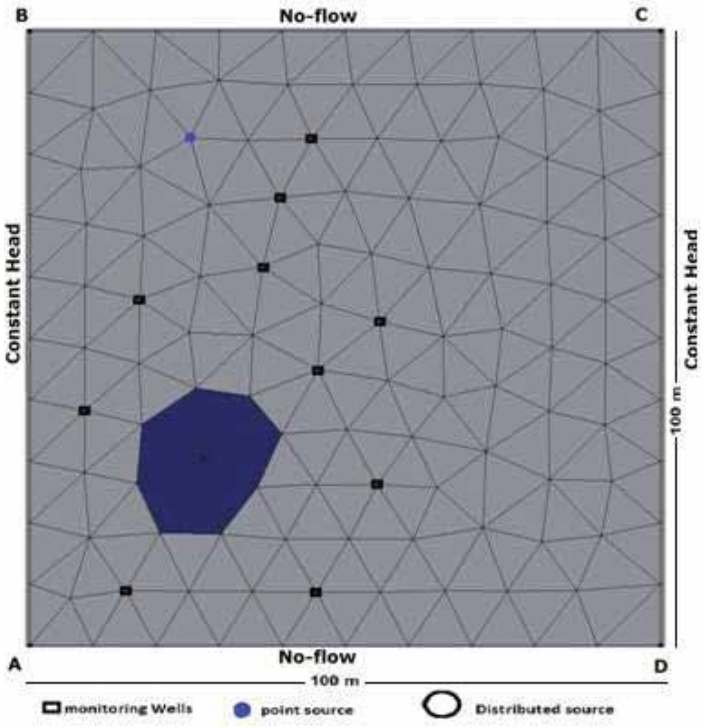


Figure 3.2 Illustrative study area (Total Head: A= 37 m, B= 40 m, C= 33m, D = 30 m)

Table 3.1 Simulation model parameters for i,j = 2,2

Aquifer Parameter	Unit	Value
Dimensions (Length* Width* Thickness) study area	m*m*m	100*100*50
Number of Nodes		387
Number of elements		1432
Hydraulic conductivity, Kx,Ky,Kz	m/d	10.0,5.0,3.0
Effective porosity, Ɵ		0.3
Longitudinal dispersivity, αL	m/d	10.0
Transverse dispersivity, αT	m/d	6.0
Horizontal anisotropy		1
Initial contaminant concentration	Mole/lit	0-5
Diffusion coefficient		0

Table 3.2 Typical Chemical reactions in the contaminant transport process

Chemical Reaction Equations	Constant Rate (Log k)*
Equilibrium Reactions	
(1) $Cu^{2+} + H_2O \leftrightarrow Cu(OH)^{+} + H^{+}$	-9.19

(4) Cu2+ + SO42- ↔ CuSO4	2.36
(7) Fe2+ + SO42- ↔ FeSO4	2.39
(9) 4Fe2+ + 4H+ ↔ 4Fe3+ +2H2O	8.5
(14) Fe3+ + SO42- ↔ FeSO4+	4.05
(15) Fe3+ + SO42- + H+ ↔ FeHSO42+	2.77
Kinetic Reactions	
(17) FeOOH(s) +3H+ ↔ Fe3+ + 3H2O	Kf=0.07

* Constant rates are taken from (Ball and Nordstrom 1992)

3.3.1. Generation of training and testing patterns for the ensemble GP models

The total time of source activities is specified as 800 days, subdivided into eight similar time intervals of 100 days each. The actual pollutant concentration from each of the sources is presumed to be constant over each stress period. The pollutant concentration of copper, iron as well as sulphate in the pit is represented as Cpit(i), Fepit(i) and Spit(i) respectively, where i indicates the time step value, and also C(i), Fe(i) and S(i) represent copper, iron and sulphate concentrations in the point sources, respectively at different time steps.

An overall of sixteen concentration values for each contaminant are considered as explicit variables in the simulation model. The concentration measurements are simulated for a time horizon of 800 days from the start of the simulation. The pollutant concentrations are assumed to be the resulting concentrations at the observation wells at every 100 days interval and this process is continued at all the observation locations till t=800 days. Only

for this methodology evaluation purpose, these concentration measurements are not obtained from field data, but are synthetically obtained by solving the numerical simulation model for specified initial conditions, boundary conditions and parameter values. In actual application these measurement data need to be simulated using a calibrated flow and contaminant simulation model. However, using field observations for calibration, and then for evaluation of a proposed methodology results in uncertain evaluation results as the quality of the available measurement data cannot be quantified most of the time. Therefore as often practiced, synthetic aquifer data is used for this evaluation of the proposed methodology.

The comprehensive three-dimensional numerical simulation model was used to simulate the aquifer flow and chemical reactive transport processes due to complex pollutant sources in this study area. Different random contaminant source fluxes as well as different realizations of boundary conditions and hydraulic conductivities were generated using Latin hypercube sampling. For random generation purpose, 10 percent of mean aquifer parameter values are considered as Maximum error for the uncertainties of aquifer parameters. HGCH was utilized to obtain the concentrations resulting from each of these concentration patterns. The simulated concentration measurement data at monitoring network and the corresponding concentration of contaminants at sources form the input and output, respectively, in input-output pattern. Totally, 8000 concentration patterns for all the ten concentration observation locations were used in this evaluation. Eight input-output patterns were defined based on different time steps. Genetic programming models were obtained using each of these data sets to create ensemble GP based surrogate models. Each data set was split into halves for training and testing the genetic programming-based surrogate models.

Surrogate models were developed for simulating pollutant concentrations at the observation locations at different times resulting from the specified pollutant sources at different times under hydrogeological uncertainties. All the GP models used a population size of 1000, and mutation frequency of 95. The Discipulus, commercial Genetic Programming software, was used to develop the surrogate models. The model was developed using default parameters values of Discipulus. The GP fitness function was the summation squared deviation between GP model generated and actually simulated concentration values at the observation locations and times.

3.4. Evaluation of Solution Results

The flow and concentration simulation results for the study area obtained using the numerical HGCH simulation model are shown in Figures 3.3 to 3.6. The flow movement, represented by total head contours in top layer and also velocity vectors are shown in Figures 3.3, 3.4 and 3.5 respectively. Figures 3.3, 3.4 and 3.5 show the hydraulic heads for flow. The contours show a gentle slope from point B towards D. Figure 3.6 shows the copper concentration distribution in the study area which shows the complex transport processes with reactive chemical species. The concentration of sulphate remains almost the same, while iron concentration is lower in groundwater. Based on pH changes the iron can react and cease to be in solute phase, thus removed from groundwater.

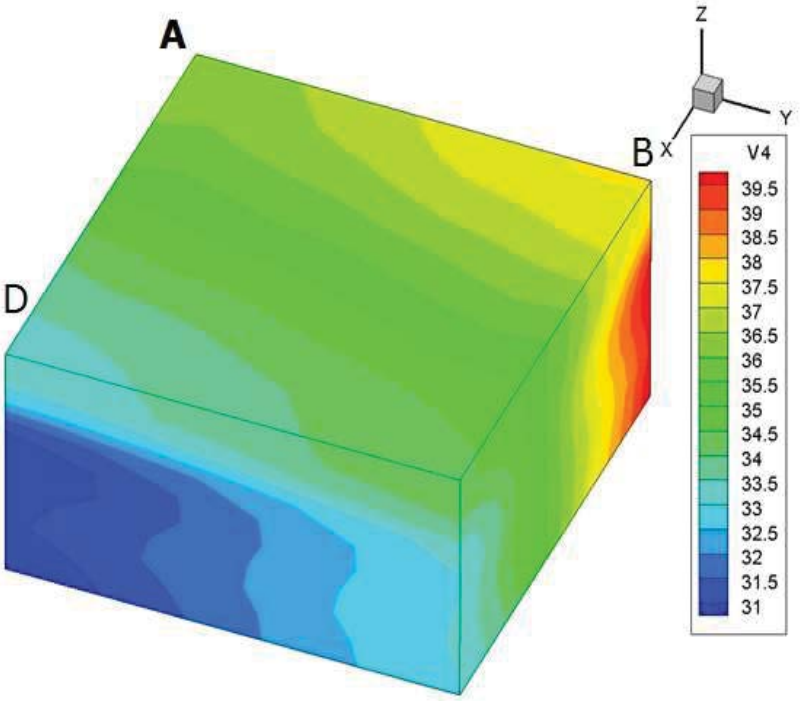


Figure 3.3 3-Dimensional view of hydraulic head distribution

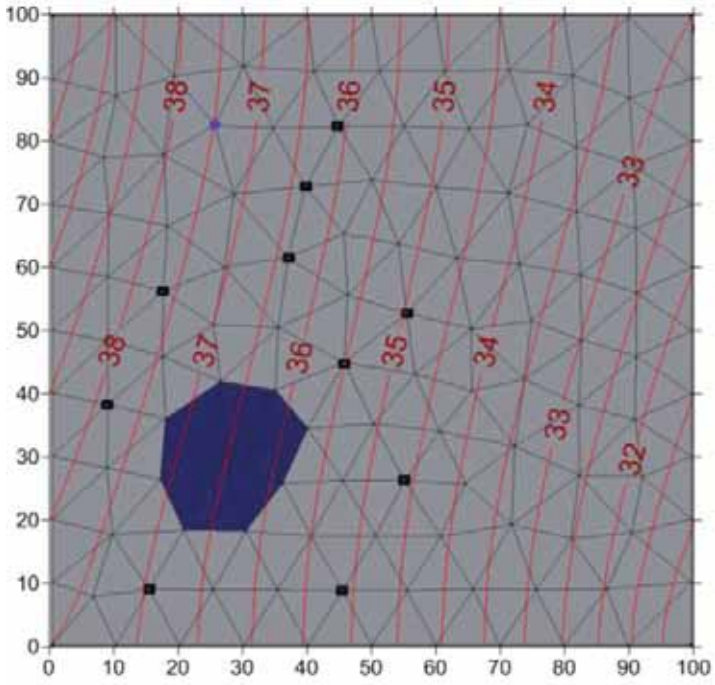


Figure 3.4 Hydraulic head contours (m)

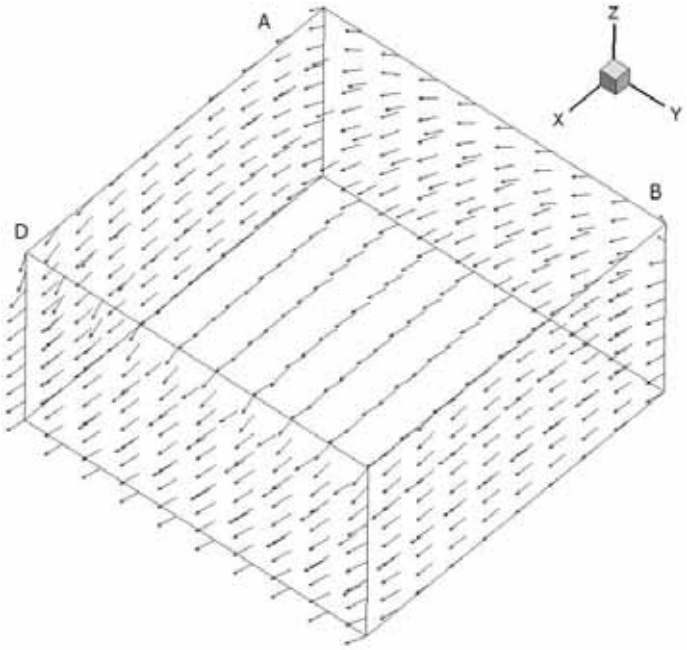


Figure 3.5 Velocity vectors of groundwater movement

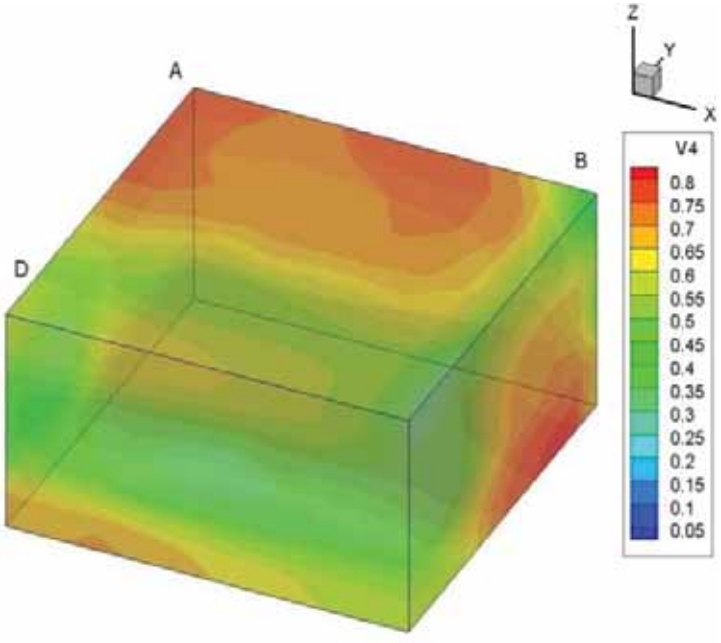
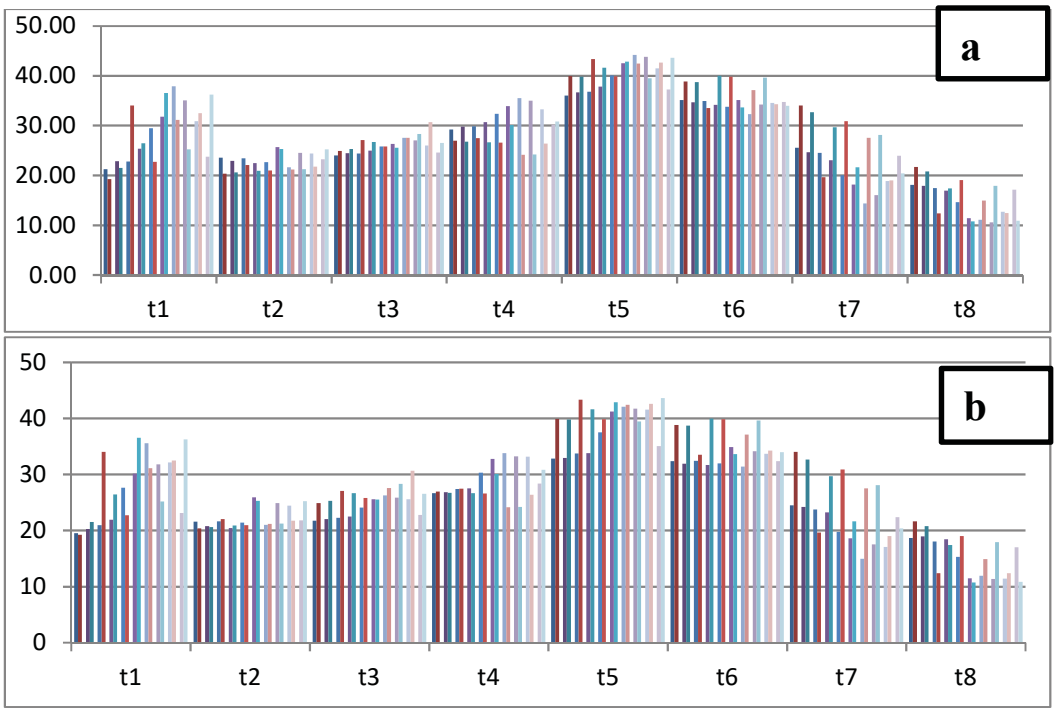


Figure 3.6 Copper concentration (Mole/lit) distribution in the study area

The results obtained using the developed ensemble genetic programming based surrogate models for approximate simulation of pollutant concentrations are compared with the numerical simulation results obtained using the HGCH. Nine different scenarios are considered. These nine scenarios are characterized by different hydraulic conductivity value realizations and hydraulic head boundary conditions. Each of these variables are randomized within 10 (± 5) percent errors in the mean values (assumed same as the actual values) for hydrogeological parameters and boundary conditions. The ensemble GP based surrogate models are obtained using the scenario incorporating uncertainties along with random source concentration data. The uncertainties in the parameter values of the scenarios are within the range of input data which are used to create the ensemble GP models. Figure 3.7 (a, b and c) illustrate these comparison results in which one scenario for one particular hydraulic conductivity is selected for obtaining simulated output data from the HGCH model at each monitoring location. Each time step is marked on the x-axis. Each of the bars corresponds to contaminant concentration in each well, obtained by HGCH and ensemble GP models.

Figure 3.7 shows that the results obtained from the ensemble GP based surrogate models are very close to the simulated results obtained using the numerical simulation model, and also incorporates parameter uncertainties. Figure 3.8 shows the summation of normalized error at each of the observation locations for each monitoring network averaged over the 8 time periods. It is noted that, the ensemble GP models provide relatively accurate results for concentrations at observation locations. Although the boundary conditions are different, the normalized errors for all the three scenarios with same hydraulic conductivity are almost the same. The most important advantage of using the developed GP models is that the numerical simulation model requires long computational time

usually several hours for a typical study area, while ensemble genetic programming surrogate models deliver the solution results typically in a fraction of a second. Also the ensemble GP models directly incorporate hydrogeologic uncertainties in the modelled system. Therefore, the computational advantage of using the ensemble GP for approximate simulation of complex reactive transport processes in aquifers is enormous, if the errors in simulation are within acceptable range. Especially, this computational time saving could be critical in development and solution of linked simulation-optimization models (Datta et al. 2014) for management of contaminated aquifers.



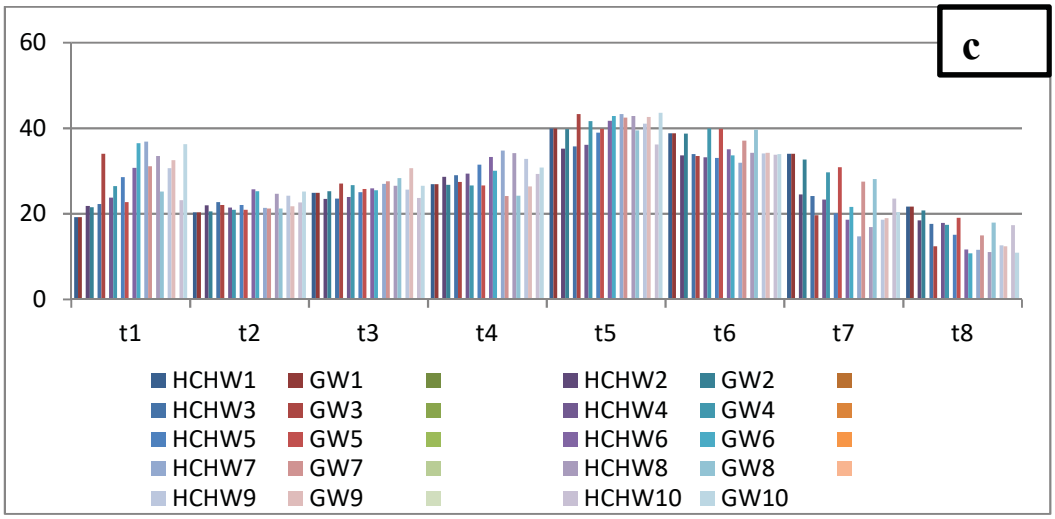


Figure 3.7 Comparison of ensemble GP model solutions with HGCH simulation results for specified parameter values defined by a Lower bound on uncertain aquifer parameter values, b Actual or mean parameters values and c Upper bound on aquifer parameter values. (GW1: Concentration data at well number 1 based on GP formulation, HCHW1: Concentration data at well number 1 based on HGCH simulation)

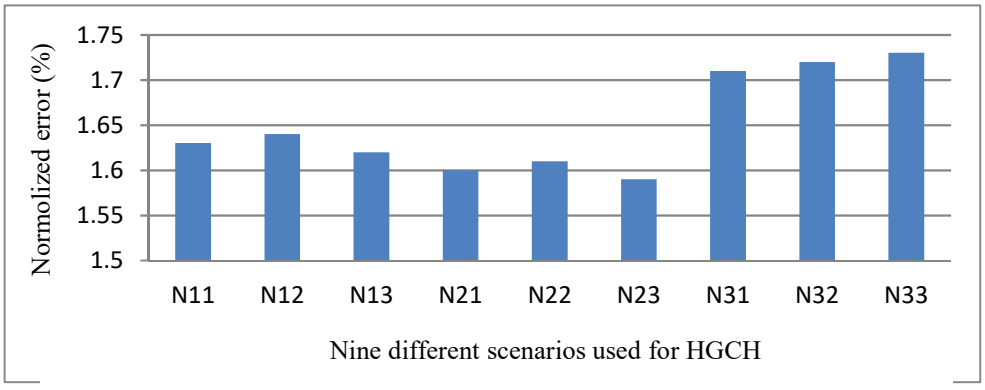


Figure 3.8 Normalized error for all scenarios

3.5. Summary and Conclusions

Although surrogate models are widely used in solving groundwater management problems replacing the actual complex numerical models, often the main issue is the

accuracy and reliability of surrogate model predictions under input data uncertainties. This chapter describes the development of a methodology based on ensemble GP surrogate models to substitute numerical simulation for approximate simulation of the chemically reactive multiple species transport process in a contaminated aquifer resembling the geochemical characteristics of an abandoned mine site. The evaluation results show the applicability of this methodology to approximating the complex reactive transport process in an aquifer. The developed ensemble GP models result in increasing the computation efficiency and computational feasibility, while providing acceptable results.

The linked simulation-optimization approach is an effective method to identify source characterization and monitoring network design under uncertainties in complex real life scenarios. The unknown contaminant source characterization is important for developing robust remediation strategies for groundwater management. The main difficulty with linked simulation-optimization models generally is the required huge computation time, due to iterative repeated solution of the numerical flow and transport simulation models. To address this, ensemble GP based surrogate models may be used to approximate the numerical simulation model under uncertainties, in the linked simulation-optimization model. Ensemble GP based surrogate models can increase efficiency and feasibility of developing optimal management strategies for groundwater management in geochemically complex contaminated aquifers such as mine sites, while at the same time incorporating uncertainties in defining the hydrogeologic system. The evaluations results show that it is feasible to use ensemble GP models as approximate simulators of complex hydrogeologic and geochemical processes in a contaminated groundwater aquifer incorporating uncertainties in describing the physical system.

Next chapter presents the potential applicability of using GP based surrogate models as approximate simulator in the linked simulation-optimization methodology to increase the feasibility of source characterization in complex contaminant aquifers.

4. Linked Optimization–Simulation for Optimal Reactive Contaminant Source Characterization in Contaminated Mine Sites

Partial contents of this chapter have been published, or are under review for publication, as detailed below:

Esfahani, H. K. and B. Datta, 2016. Genetic Programming Based Simulation of Complex Reactive Geochemical Transport Processes with Application to an Abandoned Mine Site Aquifer in Queensland Australia. Environmental Technology and Innovation (ETI) journal. Under Review.

- Esfahani, H. K. and B. Datta, Linked Optimal Reactive Contaminant Source Characterization in Contaminated Mine Sites: A Case Study. Journal of Water Resource Planning and Management (ASCE). 142 (12).

In this chapter, application of a developed methodology for characterization of distributed contaminant sources in a contaminated mine site aquifer involving complex geochemical processes is presented.

4.1. Background of the Problem

Linked simulation-optimization models are increasingly being used for identification of unknown groundwater pollution source characteristics in contaminated aquifers, which is an essential step for determining effective and reliable groundwater management and remediation strategies for a polluted aquifer (Chandalavada et al. 2012, Datta and Kourakos 2015). Reactive transport of chemical species, in contaminated groundwater systems, especially with multiple species, is a complex and highly non-linear process. Simulation of such complex geochemical processes using numerical models is generally computationally intensive. In order to increase the model reliability for real life field conditions, uncertainties in hydrogeological parameters and boundary conditions are needed to be considered as well. Also, often the development of an optimal contaminated aquifer management and remediation strategy requires an enormous number of repeated solutions using complex and nonlinear numerical flow and contamination process simulation models. Therefore using robust numerical simulation models in linked simulation-optimization methodology increases the computational burden extensively, and may affect the computational feasibility and efficiency of the methodology. To address the combination of these issues, Genetic Programming (GP) based surrogate models trained using numerical simulation models can be utilized as approximate simulators of the complex physical and geochemical processes in the contaminated aquifer. Ideally, these surrogate models also incorporate uncertainties in hydrogeologic parameter values. Trained and tested surrogate models for approximately simulating the transport processes can decrease the computational time substantially and improve the

efficiency, feasibility, and reliability of the linked simulation-optimization based source characterization methodology.

Initial attempts in the identification of unknown groundwater pollution sources proposed the use of linear optimization model based on linear response matrix approach (Gorelick et al. 1983) and statistical pattern recognition (Datta et al. 1989). Other methodologies proposed include: determining optimal estimates of the unknown model parameters and source characteristics using non-linear maximum likelihood estimation based inverse models (Wagner 1992); source identification using embedded nonlinear optimization technique (Mahar and Datta 1997, Mahar and Datta 2000, Mahar and Datta 2001); Genetic Algorithm (GA) based optimization algorithm (Aral et al. 2001, Singh and Datta 2006); Artificial Neural Network (ANN) approach; classical optimization based approach (Datta et al. 2009); heuristic harmony for source identification (Ayvaz 2010); Simulated Annealing (SA) based linked optimization algorithm for source identification (Jha and Datta 2011, Chandalavada et al. 2012, Prakash and Datta 2014a, 2015); and differential evolution algorithm for groundwater source identification (Elçi and Ayvaz 2014, Gurarslan and Karahan 2015). Review of different optimization techniques for solving source identification problems are presented in Chandalavada et al. (2011), Ketabchi and Ataie-Ashtiani (2015), Sreekanth and Datta (2015) and Amirabdollahian and Datta (2013).

The present chapter focuses on the feasibility and applicability of using GP based surrogate models for a contaminated aquifer site with very complex, multiple species reactive transport processes, and at the same time presents an application to mining sites for characterizing sources of contamination in terms of location, magnitude and duration of activity. Also, Esfahani and Datta (2015b) did look at the feasibility of using the GP

based surrogate model to approximate the flow and transport processes only for a hypothetical study area with complex geochemical processes. The feasibility of using linked simulation-optimization models incorporating GP based surrogate models as an approximate simulator of the complex hydrogeologic processes in a real life mine site aquifer, and limited demonstration of the feasibility of the proposed methodology are addressed in this chapter. This chapter also provides a typical example of implementing flow and transport simulation models for contaminated aquifer sites with complex hydrogeology and geochemistry, where actual field measurement data are very limited and sparse. Use of the surrogate models improve the computational efficiency in a linked simulation-optimization setting and produces reasonably accurate results under limited aquifer hydrogeologic uncertainties.

Some of the earlier studies (Sreekanth and Datta 2011b, Datta et al. 2013, Datta et al. 2014, Prakash and Datta 2014b) did propose the use of GP based surrogate models for approximating the aquifer flow and transport process simulation. The present chapter focuses on the feasibility and applicability of using GP based surrogate models for a contaminated aquifer site with very complex, multiple species reactive transport processes. The goal is to incorporate the GP based surrogate model for characterizing sources of contamination in terms of location, magnitude and duration of activity in a typical abandoned mine site. This study is an attempted application to a contaminated aquifer site with very complex geochemical and hydrogeologic characteristics.

In this chapter, the developed methodology for unknown contaminant sources characterization is applied to a previously operational mine site in Queensland, Australia and the related contaminated groundwater system. The developed methodology uses trained genetic programming (GP) models as surrogate models within a linked

simulation-optimization methodology. The performance of the optimal source characterization methodology using the surrogate models is evaluated to establish the potential applicability of this approach. One of the objectives is to demonstrate the feasibility of using genetic programming based surrogate models to approximately simulate the complex transport processes in a geologically and geochemically complex hydrogeological system with reactive chemical species. The surrogate models can be then linked to an optimization algorithm to develop the unknown pollution sources identification methodology, based on surrogate models as approximate simulators of complex geochemical processes. Uncertainties in the estimation of hydrogeologic parameters such as hydraulic conductivity are incorporated in the trained GP models by considering training patterns that are generated by randomized parameter values.

4.2. Methodology

The proposed linked simulation-optimization based methodology for characterization of unknown pollution sources utilizes GP-based surrogate models as approximate simulators of the flow and transport processes in the aquifer study area contaminated by multiple reactive chemical species. The GP models were trained and validated by using the randomly generated aquifer stresses (source fluxes, recharge, etc.) as inputs, and the corresponding numerically simulated concentrations at specified locations as outputs. Therefore, the hydraulic heads were not modelled explicitly using the GP models. However, the contaminant transport model requires the flow model for obtaining the concentrations as solutions. Therefore, the flow model is implicitly incorporated in the GP models, although explicit values of the hydraulic heads are not obtained as solutions from the GP models. Also, field measurement data in general, are not adequate for

training surrogate models. Training of surrogate models require many sets of input (stresses) and outputs (responses) for many plausible scenarios. These plausible scenarios in terms of stresses can be generated statistically, while the actual contaminant source fluxes are not known.

The development of the proposed methodology and its application to an illustrative real life contaminated mine site in Queensland, Australia includes three main steps. In the first step, the flow and transport simulation model based on the estimated hydrological and geochemical properties of the mine site area is implemented. HYDROGEOCHEM 5.0 (HGCH) (Sun 2004, Yeh et al. 2004), a numerical three-dimensional coupled physical and chemical transport process simulator is utilized to simulate the flow and the reactive chemical transport processes, also incorporating the reactive contaminants' behaviours in the contaminated aquifers. The simulation results are compared with field measurement data (Wels et al. 2006) to obtain acceptable calibration results based on measured flow and concentration data. In the next step, GP models are trained utilizing numerical simulation model solutions to approximately compute the concentrations of the chemical contaminants at specified locations resulting from specific contaminant sources. The trained GP based surrogate models incorporate uncertainties in hydrogeological aquifer parameters like hydraulic conductivity, porosity and boundary conditions. Finally, the trained GP based surrogate models are linked to Adaptive Simulated Annealing (ASA) optimization algorithm within a linked simulation-optimization model to obtain the optimal characterization in terms of location, magnitude and time of activity of the unknown contaminant sources.

The concentration measurement data for this mine site are based on Wels et al. (2006). These data are very limited in number and cover only two years of monitoring. Our

concentration calibration results are compared with these limited measurement data. Additional concentration measurement data were not available. This study should also provide an illustration of the limitations in modelling flow and transport processes in such a site with very limited field measurements. The limitations in concentration data availability, as well as very limited information on aquifer parameter values posed a typical challenge in our attempt to develop a well-calibrated simulation model. This is also one of the reasons that a large number of iterations with various plausible assumptions regarding the recharges, boundary conditions and initial head and concentration values were necessary to achieve an acceptable calibrated model. The flow and transport simulation models implementation should be considered, keeping in mind these limitations and challenges that may be typical of a contaminated aquifer site resembling such an abandoned mine.

4.2.1. Source Characterization Model Using Linked Simulation-Optimization

In this methodology, Genetic Programming (GP) based surrogate models are trained and tested as approximate simulators of flow and transport processes under aquifer uncertainties (Sreekanth and Datta 2012). Therefore, a robust simulation model is required to generate GP models. A three-dimensional numerical model HYDROGEOCHEM 5.0 (HGCH), incorporating the governing equations (5.1) to (5.5), is used for simulation. The flow and chemically reactive transport simulation model HGCH is explained in detail in Chapter 3, section 3.2.1.

The general equations for flow through saturated-unsaturated media are obtained based on following components: 1) fluid continuity, 2) solid continuity, 3) Fluid movement (Darcy's law), 4) stabilization of media, and 5) water compressibility (Yeh et al., 1994)

(equation 4.1-4.4). Contaminant transport equations used in the HGCH model can be derived based on mass balance and biogeochemical reactions (Yeh et al. 2000). The general transport equation using advection, dispersion/diffusion, source/sink, and biogeochemical reaction as the major transport processes can be written as equation (4.5) (Sun 2004):

$$\frac{\rho}{\rho_0} F \frac{\partial h}{\partial t} = \nabla . \left[\boldsymbol{K} . \left(\nabla h + \frac{\rho}{\rho_0} \nabla z \right) \right] + \frac{\rho^*}{\rho_0} q \tag{4.1}$$

F is the generalized storage coefficient (1/L) defined as:

$$F = \alpha' \frac{\theta}{n_e} + \beta' \theta + n_e \frac{ds}{dh} \tag{4.2}$$

K is the Hydraulic conductivity tensor (L/T) is:

$$\boldsymbol{K} = \frac{\rho g}{\mu} \boldsymbol{k} = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} \frac{\rho_0 g}{\mu_0} \boldsymbol{k}_s \, k_r = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} \boldsymbol{K}_{so} \, k_r \tag{4.3}$$

V is the Darcy's Velocity (L/T) described as:

$$\boldsymbol{V} = -K \left[\frac{\rho}{\rho_0} \nabla h + \nabla z \right] \tag{4.4}$$

$$\frac{D}{Dt} \int_v \theta C_i dv = - \int_\Gamma n . (\theta C_i) V_i d\Gamma - \int_\Gamma n . J_i d\Gamma + \int_v \theta r_i dv + \int_v M_i dv, i \in M \tag{4.5}$$

When combined with appropriate boundary and initial conditions, equations (4.1) – (4.4) are used to simulate the temporal-spatial distributions of the hydrological variables, including pressure head, total head, effective moisture content, and Darcy's velocity.

Where:

θ: effective moisture content (L3/ L3);

h: pressure head (L);

t: time (T);

z: potential head (L);

q: source or sink of fluid [(L3/ L3)/T];

ρ_0: fluid density without biochemical concentration (M/ L3);

ρ: fluid density with dissolved biochemical concentration (M/ L3);

ρ*: fluid density of either injection (= ρ*) or withdraw (= ρ) (M/ L3);

μ_0: fluid dynamic viscosity at zero biogeochemical concentration [(M/L)/T];

μ: the fluid dynamic viscosity with dissolved biogeochemical concentrations [(M/L)/T];

α′: modified compressibility of the soil matrix (1/L);

ß: modified compressibility of the liquid (1/L);

ne: effective porosity (L3/L3);

S: degree of effective saturation of water;

G: is the gravity (L/T2);

k: permeability tensor (L2);

ks: saturated permeability tensor (L2);

Kso: referenced saturated hydraulic conductivity tensor (L/T);

kr: relative permeability or relative hydraulic conductivity (dimensionless);

C_i: the concentration of the i-th species in mole per unit fluid volume (M/L3);

ν: the material volume containing constant amount of media (L3);

Γ: the surface enclosing the material volume ν (L2);

n: the outward unit vector normal to the surface Γ;

Ji: the surface flux of the i-th species due to dispersion and diffusion with respect to relative fluid velocity [(M/T)/L2];

θri: the production rate of the i-th species per unit medium volume due to all biogeochemical reactions [(M/L3)/T];

Mi: the external source/sink rate of the i-th species per unit medium volume [(M/L3)/T];

M: the number of biogeochemical species;

Vi: the transporting velocity relative to the solid of the i-th biogeochemical species (L/T).

GP models utilize a set of input-output data for training purposes, which are generated randomly by the simulation model. The numerical Simulation model creates M number of output sets from M number of input sets. The input data are generated by using random Latin hypercube sampling (LHS) in a defined range (Esfahani and Datta 2015a, 2015b). The trained and tested GP based surrogate models are explained in detail in Chapter 3, sections 3.2.2. The performance of each GP program formulation is evaluated in terms of

training, testing, and validation using the set of input-output patterns. The testing process evaluates the model performance for new data sets without developing a new fitness function. To compare the GP and HGCH results at the same location, the normalized error criterion (f) is used as defined by the following equation:

$$\boldsymbol{f} = \sum_{k=1}^{n_t} \sum_{job=1}^{n_{ob}} \frac{ABS(C_{hgch}{}_{iob}^{k} - C_{obs}{}_{iob}^{k})}{C_{hgch}{}_{iob}^{k}} \tag{4.6}$$

Where:

$C_{hgch}{}_{iob}^{k}$: the concentration simulated by the HGCH model at observation monitoring location i_{ob} and at the end of time period k (ML^{-3});

n_t: the total number of monitoring time steps;

n_{ob}: the total number of observation wells;

n_k: Total number of concentration observation time periods.

This study utilizes multiple realization optimization incorporating a surrogate modelling approach to solve for the unknown groundwater sources characterization in the mine site area. Here, the multiple realizations pertain to the concentration of the different chemical species obtained from different surrogate models in the ensemble. The surrogates developed using genetic programming are coupled to the ASA algorithm. The methodology is schematically represented in Figure 4.1. For each observation well, one individual GP model is trained, tested and validated using the input-output patterns provided by numerical simulation model. All GP based models are linked to the optimization algorithm. These approximate simulation models utilize the unknown source

concentration candidate solutions generated by the optimization algorithm to obtain the estimated contaminant concentration at monitoring locations, where concentration measurement field data are available. Then the optimization algorithm evaluates the objective function. The objective function value is defined as function of the differences between observed and simulated concentration values. The optimal source characterization is obtained by solving the optimization model to minimize the objective function. Mahar and Datta (2001) defined the objective function of a simulation-optimization model for unknown concentration source characterization as follows:

$$Minimize\ F1 = \sum_{k=1}^{n_k} \sum_{job=1}^{n_{ob}} (C_{gp}{}_{iob}^{k} - C_{obs}{}_{iob}^{k})^2 . w_{iob}^{k} \tag{4.7}$$

$$C_{gp}{}_{iob}^{k} = f(x, y, z, C_{si}) \tag{4.8}$$

The weight w_{iob}^{k} can be described as follows: (4.9)

$$w_{iob}^{k} = \frac{1}{(C_{obs}{}_{iob}^{k} + n)}$$

Where:

$C_{gp}{}_{iob}^{k}$: the concentration estimated by the GP models at observation monitoring location i_{ob} and at the end of time period k (ML^{-3}); n_{ob}: Total number of observation wells;

$Cobs_{iob}^{k}$: Observed concentration at well iob and at the end of time period k;

f(x, y, z, C_{si}): the simulated concentration obtained from the transport simulation model at an observation location and source concentrations C_{si}.

w_{iob}^{k} : Weight corresponding to observation location iob, and the time period k.

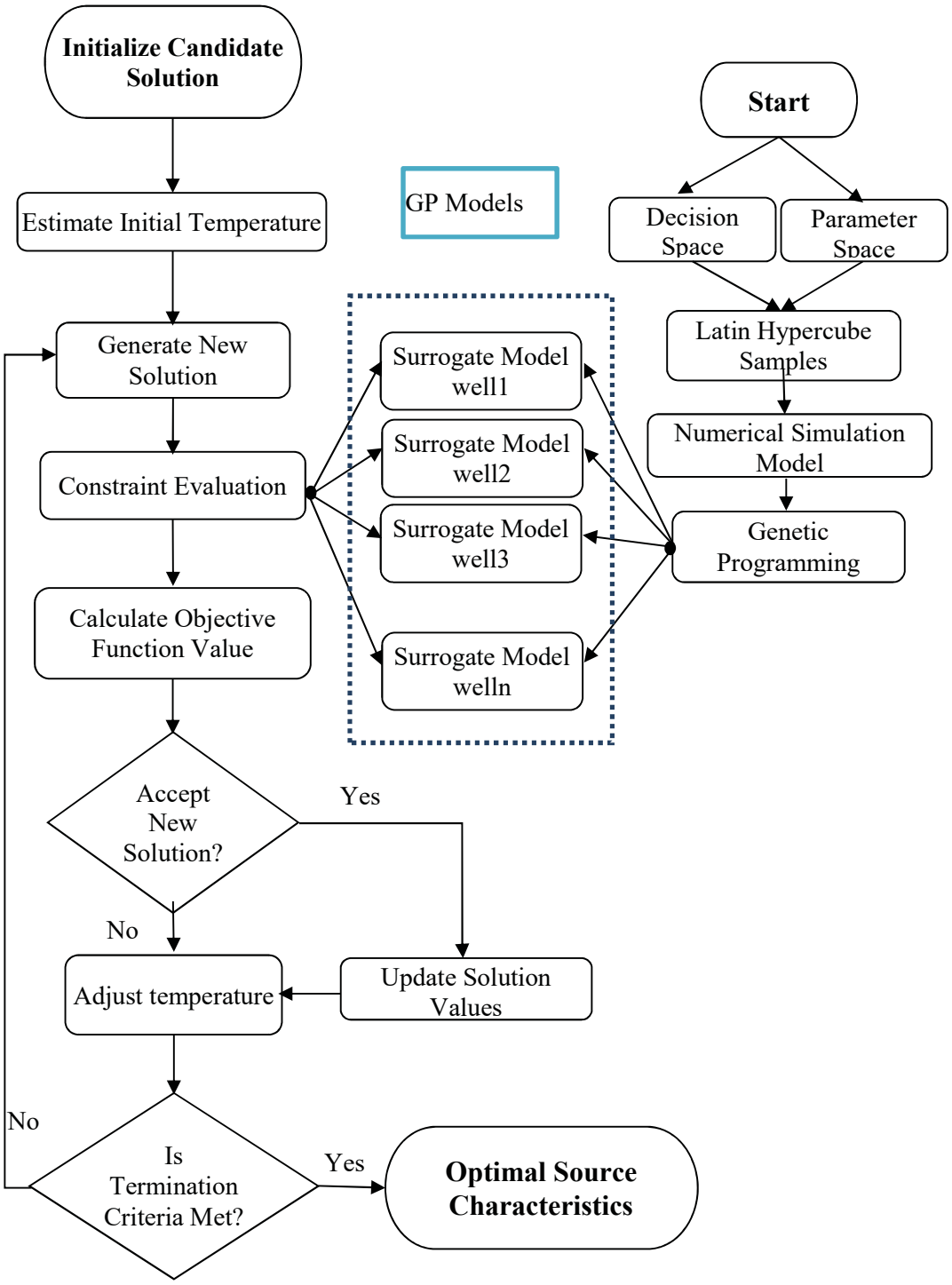


Figure 4.1 Schematic linked simulation-optimization model using GP-based surrogate models

n (in Eq 4.9) is an appropriate constant where errors at low concentrations do not dominate the solution. The present study utilizes the Adaptive Simulated Annealing as the optimization algorithm for optimal source characterization model. This algorithm is chosen based on its comparative efficiency in reaching a global optimal solution.

4.3. Performance Evaluation

To evaluate the performance of the proposed optimal contaminant source characterization methodology linked to surrogate models, available concentration measurement data at existing monitoring locations in the selected study area are utilized as a part of the real life illustrative scenario. These data are also utilized to calibrate the numerical simulation model developed using HGCH.

4.3.1. Description of the selected mine site

The historic mine site utilized as an illustrative contaminated aquifer in this study is located in central Queensland, Australia (Figure 4.2). This copper, gold and silver mine was operational for nearly 100 years between 1889 and 1990. Dee River forms a natural boundary on the southern side of the study area which inflows into Don and Dawson River and finally into the Fitzroy River. Acid Rock Drainage (ARD) and Acid Mine Drainage (AMD) generated from the flooded open cut and waste deposits and their tailings, are likely contamination sources at the mine site and the Dee River downstream of the mine. These distributed pollution sources affect groundwater quality as well as surface water quality. The environmental issues still remain and the potential sources of contamination (the pit, waste deposits and waste dumps) as well as possible pathways of

contamination have not been characterized adequately. This study is an attempt to develop a suitable methodology with a potential to accomplish this characterization task. Several remediation measures were undertaken for this area. However, systematic and proper identification of contaminated sources, including the characterization of ARD and AMD chemical sources and pathways are yet to be accomplished. This is required to develop effective and efficient remediation strategies for optimal groundwater management, based on the estimation of magnitude, location and activity time of the contaminant sources.

Figure 4.2 Plan view of the historic mine site (map images: Map data © 2016 GBRMPA, Google; satellite image: Imagery © 2016 CNES / Astrium, F2:2 DigitalGlobe, Map data © Google)

4.3.2. Topography and Hydrology

The mine site is located in a hilly region with highly varying topography. The range of surface level elevations is between 200m and 350m above the sea level. The natural topography has changed to some extent due to mining activities with generation of open pit, and deposition of waste rock dumps and tailings. The study area is defined as sub-tropical, sub-humid climatic region and the long-term average annual rainfall at the Mount Morgan mine site is 680 mm. The long-term evaporation is estimated to be 1507 mm/year according to the Australian Commonwealth Bureau of Meteorology reports(MLA 2008).

4.3.3. Geology

The geology of the site has been described in (Taube 1986, 2000, Ulrich et al. 2003). Devonian Mount Morgan Tonalite which has almost no permeability is the major ore in this site. Devonian rhyolitic Mine corridor volcanic rock has provided Au-Cu mineralization which is widely surrounded and intruded by the tonalite. Moreover, few dykes compartmented the area and it causes groundwater to discharge deeper from the mine site (Taube 2000). It is also suspected that long-term mining operations may have caused some geological changes in the subsurface, and are not documented. Also, the hydrogeologic and geochemical measurement data from the site are very sparse and very limited, adding to the difficulties in implementing a suitably calibrated numerical simulation model.

Pyrite is the main ore mineral in this site, which is followed by chalcopyrite and trace pyrrhotite, sphalerite, wurtzite, magnetite and gold (Edraki et al. 2005). The majority of

waste deposits are comprised of quartz–pyrite waste rock extracted from the mining activities. The absence of carbonate rocks to neutralize groundwater pH has exacerbated this situation (Edraki et al. 2005).

4.3.4. Groundwater flow and transport modelling and calibration

Groundwater flow and reactive chemical contaminant transport processes are simulated with specified hydrogeological parameters as given in Table 4.1. Sparse and limited hydraulic head measurement data, and very limited concentration measurement data at existing concentration monitoring locations are used for calibrating the numerical flow and reactive transport numerical simulation model. In actual application not specifically directed towards performance evaluation of the developed methodology, the actual concentration measurement data available at specified observation locations of the aquifer study area are to be utilized as measured concentration inputs to the optimal source characterization decision model.

A conceptual model for flow and transport processes for the mine site is solved using the HGCH flow and reactive chemical transport simulator. Site properties in terms of hydrogeological and geochemical features and boundary conditions are specified based on field investigation data reported in Wels et al. (2006), and based on earlier preliminary calibration efforts reported in Jha and Datta (2015a) and Datta et al. (2016). The flow model simulates the groundwater flow for average annual rainfall assuming a steady state condition. Other simplifying assumptions in this model included: negligible change in groundwater storage, constant recharge from unsaturated zone into groundwater flow, and no groundwater flow across the geographical boundaries of this model.

Regarding the hydro-geological system information (Taube 1986), the aquifer system at the mine site can be sub-divided into four layers for numerical modelling description. These are:

- Waste rock dumps and tailings,
- Highly weathered bedrock,
- Partially weathered bedrock,
- Tight bedrock.

Based on field data, each of these layers has variable thicknesses and they are defined as convertible layers. Potential sources of groundwater at this mine site are the Pit Lake, waste dumps and tailings (Figure 4.3). The open pit acts as a significant source/sink of groundwater. The pit and the Dee River water levels are assumed almost invariant with time in terms of the water level. There are 11 observation wells between the pit and river which monitor the groundwater level and pollution concentration. In this study, this information was obtained from published related papers (Edraki et al. 2005, Wels et al. 2006, 2009) and is used for model calibration.

There is a relatively high hydraulic conductance between pit and the Dee River (Zone 5 in Figure 4.3) due to alluvial deposits in the Dee River. Therefore, this area is the main path for transmitting significant quantities of groundwater relative to Dee River base flow. In the simulation model, larger hydraulic conductivity is defined for zone 5 in layer 2. The contaminants along with groundwater are transported in bedrock due to the presence of fractured bedrock (Wels et al. 2006, Jha and Datta 2015b).

Generally the Mount Morgan site, waste rocks and tailings contain majority amounts of pyrite [FeS2] which is followed by variable amounts of quartz [SiO2], jarosite [KFe3

(SO4)2(OH)6], pyrrhotite [Fe(1-x)S] and chalcopyrite[CuFeS2], with minor amounts of kaolinite [Al2Si2O5(OH)4], and albite [NaAlSi3O8]. The open pit and seepage sumps at the site contain acid–sulphate waters with a narrow range of low pH values (2.6–3.8) and high conductivities (8090–26,600 lS/ cm) (Edraki 2005). Metal concentrations in the open pit and sumps generally fall in a range from: 76–3073 mg/L Al; 13– 1487 mg/L Fe; 22–230 mg/L Mn; 3–138 mg/L Cu; 2– 81 mg/L Zn; 0.1–8 mg/L Pb; 0.44–6.55 mg/L Co; and 0.01–0.59 mg/L Cd in samples of August 2002 (Edraki 2005). The bulk of the materials (94 Mt; (Jones 2001)) currently exposed at the site are quartz–pyrite waste rock extracted from the Main Pipe mineralization. Oxidation of sulphides within these waste rocks and also tailings on the site (8 M tons; (Jones 2001)) produces acidic and metal-rich drainage. The scarcity of carbonate rocks that could have neutralized acid mine drainage has exacerbated this situation (Edraki 2005). The contaminant concentration in aqueous sources are reported in few published studies (Wels et al. 2006, 2009). Also, there are waste deposits which are likely to act as distributed sources producing AMD. As the AMD production is directly related to the amount of water and oxygen, the pollution concentration is dependent on water availability. Therefore, the unit of concentration is indirectly defined in terms of moles per litre.

Table 4.1 Aquifer's physical and hydrogeological properties

Aquifer Parameter	Unit	Value
Length of mine site	m	4000
Width of mine site	m	4000

Thickness of mine site	m	variable
Node numbers at each layer		2621
Element numbers at each layer		5010
Horizontal Hydraulic conductivity, K	m/d	
Layer 1		0.49683
Layer 2		0.05320
Layer 3		0.01437
Layer 4		0.00936
Zone 5		0. 79867
Vertical Hydraulic conductivity, K	m/d	
Layer 1		0.200
Layer 2		0.002
Layer 3		0.0005
Layer 4		0.0005
Zone 5		0.200
Effective porosity, Θ		0.28
Longitudinal dispersivity, αL	m/d	18

Transverse dispersivity, αT	m/d	6
Horizontal anisotropy		1
Initial contaminant concentration	g/lit	0-100
Diffusion coefficient		0
Average rainfall	mm/year	700

The major contaminant source of this site is the pit, in which initial slow generation of the ions from the solid phase have already taken place, and as a result a huge amount of contaminant ions are already present. The equation representing the initial slow generation of the ions are, therefore, not included in the modelling process for this case study. The initial slow AMD or ARD generation from the solid phase was not incorporated. The associated kinetic chemical equations and their production have been also discussed in previous studies (Taube 2000, Ulrich et al. 2003, Edraki et al. 2005), and therefore not repeated here. In this study, the transport processes of these chemical reactive ions in subsurface water is simulated and the linked simulation- optimization methodology developed for contaminant source characterization including chemical reactive spices. These reactive species include sulphate, iron and copper ions. The reaction equations incorporated in this study are clearly defined in Table 4.2. Also, the reaction coefficients etc., are given in the same table.

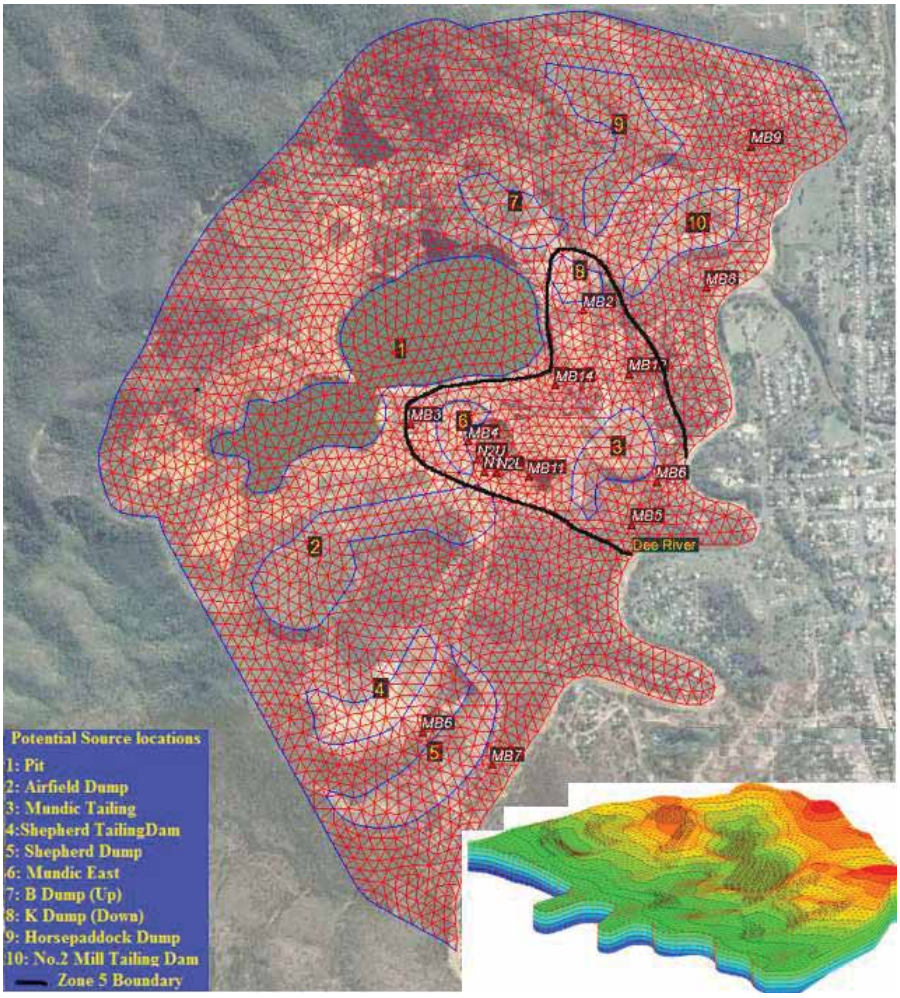


Figure 4.3 Model definition based on HGCH (MBX are Observation locations) for the mine site aquifer

Table 4.2 Typical chemical reactions during the contaminants' transport in unsaturated-saturated zone

Chemical Reaction Equations	Constant Rate (Log k)
Equilibrium Reactions	
$Cu^{2+} + H_2O \leftrightarrow Cu(OH)^{+} + H^{+}$	-9.19
$Cu^{2+} + 2H_2O \leftrightarrow Cu(OH)_2 + 2H^{+}$	-16.19
$Cu^{2+} + 3H_2O \leftrightarrow Cu(OH)^{3-} + 3H^{+}$	-26.9

$Cu^{2+} + SO_4^{2-} \leftrightarrow CuSO_4$	2.36
$Fe^{2+} + H_2O \leftrightarrow Fe(OH)^+ + H^+$	-9.397
$Fe^{2+} + 2H_2O \leftrightarrow Fe(OH)_2 + 2H^+$	-20.49
$Fe^{2+} + SO_4^{2-} \leftrightarrow FeSO_4$	2.39
$Fe^{2+} + SO_4^{2-} + H^+ \leftrightarrow FeHSO_4^+$	1.66
$4Fe^{2+} + 4H^+ \leftrightarrow 4Fe^{3+} + 2H_2O$	8.5
$Fe^{3+} + H_2O \leftrightarrow Fe(OH)^{2+} + H^+$	-2.187
$Fe^{3+} + 2H_2O \leftrightarrow Fe(OH)_2^+ + 2H^+$	-4.59
$Fe^{3+} + 3H_2O \leftrightarrow Fe(OH)_3 + 3H^+$	-12.56
$Fe^{3+} + 4H_2O \leftrightarrow Fe(OH)_4^- + 4H^+$	-21.58
$Fe^{3+} + SO_4^{2-} \leftrightarrow FeSO_4^+$	4.05
$Fe^{3+} + SO_4^{2-} + H^+ \leftrightarrow FeHSO_4^{2+}$	2.77
Kinetic Reactions	
$Fe(OH)_{3(s)} + 3H^+ \leftrightarrow Fe^{3+} + 3H_2O$	K_f=0.05
$FeOOH_{(s)} + 3H^+ \leftrightarrow Fe^{3+} + 3H_2O$	K_f=0.07
$FeOOH_{(s)} \leftrightarrow$ =FeOH	K_f=0.05

Copper (Cu++), sulphate (SO4--) and iron (Fe++) are introduced as initial pollutants in the sources, which involved chemical reactions and are listed in Table 4.2. The source activities are considered for the duration of up to 800 days. 800 days are used as total time horizon for this scenario as the observation concentration data are available for

around two years duration. The assumed initial condition for contaminant concentration is specified in the simulation model based on the available data and extrapolation within the area. Kriging is used for different subareas based on the information available in Edraki et al. (2005). Therefore, 800 days only captured the source impacts since time equal to zero. The actual pollutant concentration from each of the sources is presumed to be evenly distributed over the duration of the source activity. The pollutant concentration of copper as well as sulphate in the sources is represented as C(i) and S(i) respectively, where i represents the specified potential source identification number. It is assumed that there are maximum 10 sources (Figure 4.3). F(i) also shows the iron concentration in different sources for the same pattern. A total of 10 source concentration values for each contaminant are considered as explicit variables in the simulation model. These sources include the waste dumps and the pit. For the performance evaluation purpose here, the concentration measurements are simulated for 800 days since the start of the simulation.

The main aim of calibration was to obtain the optimal estimation of hydraulic conductivity and groundwater levels for the aquifer based on limited field measurement data. Hydraulic head measurement data from 11 observation locations spread across the impacted area were used for the simulation model calibration. The hydraulic head data used for model calibration and validation were recorded in 2002 and 2003 (Wels et al. 2006). A portion of the average rainfall intensity per year was specified as recharge for calibration of the simulation model also using measured head data from observation locations.

Calibration targets for the developed model were set to be within two metre intervals of the observed hydraulic head value in observation locations with a confidence level of 90 percent. True boundary condition is the most difficult to determine in the site with very

limited observed measurement data. Therefore, realistic boundary condition needs to be specified reflecting the site conditions. In the calibration process, one of the most difficult issues is to address properly the correct boundary conditions. Therefore, the boundary condition needs to be determined based on preliminarily calibration results as well. The model boundary conditions were manually adjusted to achieve the calibration targets. The measured and simulated heads were compared at selected location. In Figure 4.4(a), the white bars signify that calibration target is achieved for the shown well numbers, whereas black bars represent intermediate errors.

The transport simulation model was calibrated to a reasonable extent by comparing the available measured concentration values around the pit with those values obtained by simulation. The reaction coefficients and identification of the ions are based on the information available in Edraki et al. (2005). By incorporating the specified concentration initial conditions, the 800 days of simulation captured the impact of the past contaminant source activities, and the activities since the initial time of simulation. This calibration process was continued until acceptable calibration results were obtained. For the transport process calibration purpose, the concentration in the pit was specified. The contaminant source representing the pit was specified by specifying constant head boundary condition along the pit parameters. Recharge from the pit, therefore, mimics the contaminant flow. The recharge from the other areas was based on calibrated vertical recharge mainly based on average annual rainfall. The concentration at the other potential locations (e.g. waste dumps, waste deposits) was estimated based on the available concentration measured in the vicinity of these source locations. In the absence of any other data to characterize the sources, this approach was considered reasonable to provide ballpark figures. Therefore, the simulated concentrations were also compared with the available measured

concentration data at few monitoring points to ascertain if the flow and transport simulation models perform within expected range. This cannot be strictly described as calibration of the transport model; however, these comparisons provide some degree of confidence in the calibration process where available field measurement data are very limited. Figures 4.4(b) and 4.4(c) show the total head contours and concentration contours respectively after calibration. More rigorous calibration can be undertaken if a large number of spatiotemporal field measurement data are available.

Calibration and validation of a flow and transport model for such a complex study area, and typically with very limited amount of observation/measurement data available is very challenging. The calibration process is most challenging as in many other such abandoned mine sites in Australia, monitoring data even for the hydraulic head are sparse and covers may be few measurements over a couple of years, as in this study area. The contaminant transport model cannot be calibrated, as that requires information on sources (magnitude, locations and duration). That is the role of the source characterization part. Also, the hydrogeology is very complex. It is plausible that mining operations have altered the geology in the vicinity that needs to be addressed. Even before a monitoring program can be implemented it needs the basic idea regarding the possible sources and pathways. The hydraulic calibration is a result of continuous efforts over the years, with earlier attempts reported in Jha and Datta (2015b) and Datta et al. (2016). This calibration process reported in this study is more rigorous, comparatively more robust. The challenge is not only to simulate but then characterize the sources and pathways.

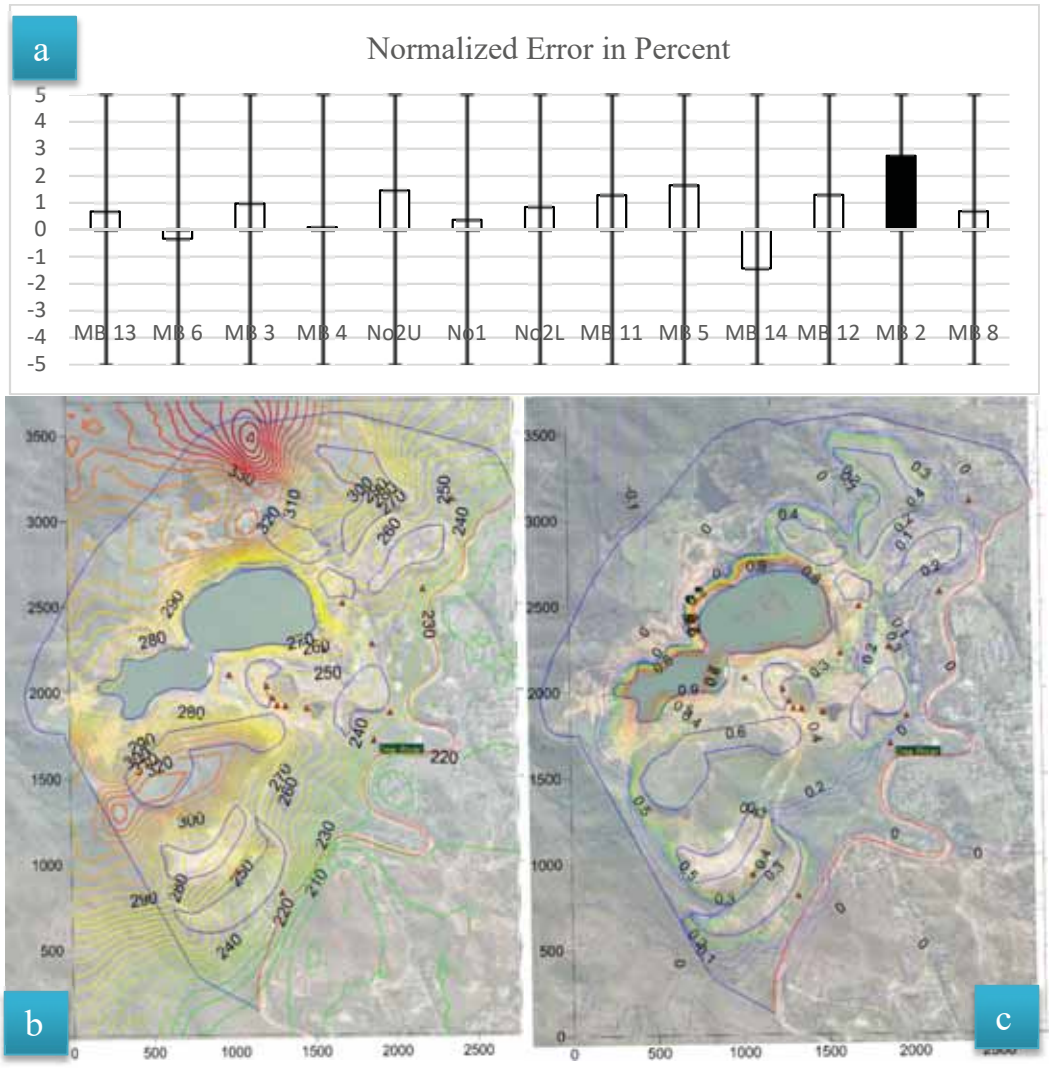


Figure 4.4 Model calibration results; a) Normalized error at observation wells; b) Total head contours; c) Concentration contours

The calibrated simulation model reflects the hydrogeologic conditions to the extent possible, and the source locations magnitudes, duration of activities are unknown. These uncertain parameters are randomly generated to simulate plausible responses using the numerical simulation model (HYDROGEOCHEM 5.0). The hydraulic calibrated parameters, and the available information on the chemical species and their associated geochemical parameters are incorporated. At least 400 simulated samples of the

simulated responses are then utilized in training the AI/pattern recognition based algorithm: Genetic Programming (GP). The main idea is to make the solution of the linked simulation-optimization possible by using the GP based approximate response. Of course, the GP performance is evaluated before accepting a GP model.

The geochemical processes on such sites are no doubt very complex, and even uncertain to some extent. This study is one attempt to address such complex cases while combining simulation of the flow and transport processes within an optimal decision model for source characterization. Therefore, there are two distinct parts of this study. The first part demonstrates the feasibility of developing GP based surrogate models to approximately simulate the flow and transport processes, in order to make the solution of the decision model possible within reasonable computational time and with reasonable accuracy. The second part deals with the solution of the contaminant source characterization model. No doubt the calibration of the flow and transport models are very important in this context. The adequacy of the simulation models in terms of the capabilities is also important. The 3-D multiple species HYDROGEOCHEM 5.0 model has been utilized as it satisfied the requirements of capabilities in terms of simulating the processes.

Heterogeneous hydrogeologic conditions are incorporated. There are different geological layers and also the top layer, which has the major contribution in the flow and transport processes, has different zones with different hydraulic conductivity, and other layers have smaller hydraulic conductivity.

The flow was considered steady-state averaged over one-year cycles to eliminate effects of some seasonal fluctuations. Because there has been no pumping from the aquifer site for the last twenty years or so, and the time interval for modelling was one year

(averaged) and based on data availability constraints, it was considered reasonable to assume flow process to have attained some form of steady state, approximately and averaged over the time interval. Limited head observation data also suggested that this assumption is reasonable. The mine site was operated from 1890 to 1990 (nearly a hundred years). Also, the mine site is abandoned since 1990. The time horizon considered for calibration was 800 days as observation data are available for two years. Porosity was one of the parameters calibrated.

Boundary condition is the most difficult to determine in the site with very limited observed measurement data. The subsurface geology is also very complex and has been modified to a certain extent due to a long mining operation. Therefore, realistic boundary condition needs to be specified reflecting the site conditions. In the calibration process, one of the most difficult issues is to assess and assign properly the correct boundary conditions in the absence of proper geophysical monitoring/measurements. Therefore, the boundary condition needs to be determined based on preliminarily calibration results as well.

A considerable amount of time and effort was spent for calibration of the model, due to very limited site information and measurement available. There are limitations of calibration which is dictated by data limitations. In well-monitored aquifer sites, the situation may be different and calibration process may be much simpler. The calibration reflects a plausible approach to obtain a calibrated and simulation model under less than ideal data availability scenario.

Various variations in boundary conditions included some modifications in assumed constant head boundary conditions, and also constant head boundary near the pit. All

these variations were based on limited perturbation of some of the head measurements available.

4.3.5. Typical Limitations in Implementation of a Calibrated Simulation Model

This case study demonstrates that implementation of the flow and transport simulation models is very challenging for such highly contaminated large number of abandoned mine sites in Australia. The limitations in the calibration of the flow and transport processes include: 1- very limited availability of hydrogeologic data; 2- complex geology of the site area; 3- sparse and limited hydrologic field measurement data over a couple of years; 4- complex geochemical processes including acid mine drainage, and 5- reactive multiple species. Adequate boundary condition and other physical and geochemical properties are determined based on sparse and limited observation data.

The contaminant transport model essentially cannot be calibrated, as that requires accurate information on contaminant sources (magnitude, locations, and duration). The characterization of the contaminant sources is the outcome of the source characterization model solution. Also, the hydrogeology of the site is very complex. This is also one of the reasons that a number of iterations with various plausible assumptions regarding the recharges, boundary conditions and initial head and concentration values had to be used iteratively to obtain an acceptable calibrated model. Therefore, the calibration process needs to be evaluated in view of these limitations. In this regional scale modelling of a study area with complex

hydrogeology and reactive contaminant transport processes, it is difficult to look at the processes in minute scale. Therefore, even though the transport may occur through preferential flow paths, the calibrated model is not suitable for identifying such low resistivity paths. The availability of field measurement data and their resolution are critical in determining the capabilities of the calibrated model. The simulation results show approximate match between the simulated concentrations and the average concentration measurement data between 2003 and 2006 at monitoring well locations. These results can be interpreted as a justification for adopting the numerical transport simulation model, in the absence of more extensive concentration measurements.

4.3.6. Approximate simulation using Genetic Programming

Sulphate is the main contaminant of AMD in the mine contamination sources. It is specified as ten input data sets S(i) (i=1 to 10), which consist of concentration values for each of the ten source locations. The GP based models have a limitation that it cannot formulate the relationship between multiple inputs and multiple outputs, in a single GP model. Therefore, for each observation well location, one separate GP formulation is required. In this case study, there are 11 observation locations specified, and two different scenarios were incorporated for each of these locations i.e., with and without incorporating the uncertainties or errors in estimating aquifer properties.

A single GP model corresponds to a specific observation location and an uncertainty scenario. Therefore, the total number of GP models generated are 22 for 11 observation locations. Each of these GP models is trained and tested utilizing simulated concentration

data based on randomized inputs for specified locations of concentration measurement. For random generation purpose, 10 percent maximum deviation from previously estimated values of the hydrogeologic parameters is considered for incorporating errors and uncertainties in the measured values. Hydrogeologic parameters which are considered uncertain include: hydraulic conductivity, soil moisture in the saturated and unsaturated zones, and water level at boundary areas (pit and Dee River). The HGCH is utilized as the simulation model to obtain the concentrations resulting from each of these input source flux patterns and incorporating aquifer properties. These constitute the simulated concentration measurement data at monitoring locations. The output data consists of the resulting contaminant concentration measurements due to these source fluxes at all the 11 monitoring wells at time t=800 days. 1000 data patterns comprised of inputs and the corresponding outputs are used in the GP models. Out of total data patterns, 40% is used for training, 30% for validation, and the rest of patterns for testing. A Latin Hypercube distribution (using MATLAB R2012b) was implemented for generating the random contaminant flux values ranging between 0 g/lit and 100 g/lit. The corresponding output data was simulated using HGCH code. DiscipulusTM (RML Technologies, Inc.) is used for training, validation, and testing of the GP models.

The field data, in general, are not suitable for training surrogate models. First, the surrogate model training required many sets of input (stresses) and outputs (responses) for many plausible scenarios. These plausible scenarios in terms of stresses can be generated statistically. Second, the actual contaminant source fluxes are not known. Also, for transient transport conditions the outputs specified are the concentrations varying with time for different contaminant flux input scenarios. There are 400 different input and corresponding output data sets, each representing different input source

concentrations and other hydrogeologic variables. It is no doubt necessary to ensure that the input source characteristics cover a sufficient range of source strength. Therefore, these were chosen carefully and certainly reflect variations in output in the form of concentrations. The source concentrations were randomized in the range equivalent of 0 to 100000 mg/l. This was considered sufficient and representative of the site conditions.

Uncertainty was incorporated in training of GP models by using randomly generated hydraulic conductivity, and porosity parameter values for the different layers and different zones within the study area as described in the manuscript. Totally random variations in these properties may result in problems with the calibrated model. To address this, the random variations were restricted to about 10% of the calibrated values.

There was no attempt to emphasize the statistical correlation and covariance as the geology is very complex and due to sparsity of data. It was not considered useful to make elaborate statistical assumptions. The source fluxes specified were based on LHS sampling and varied for each of the 8 different time periods at every potential source location.

4.3.7. Optimal Sources Characterization

Trained GP surrogate models are used as approximate simulators of the flow and transport processes in the linked simulation-optimization model for optimal source characterization of different distributed sources in the contaminated mine site area. Only for this evaluation purpose, concentration measurements are simulated using the calibrated numerical flow and transport simulation model to generate synthetic concentration measurement data at specified monitoring locations. These concentration

measurement data are generated for the study area for specified source characteristics. Then these synthetic concentration measurement data are utilized together with the linked simulation-optimization models incorporating the trained GP models, to evaluate the potential applicability, accuracy and feasibility of the developed methodology. Following the performance evaluation with synthetic data, the developed methodology was also applied to characterize groundwater sources based on contamination measurement in the contaminated abandoned mine site area. However, the source flux magnitudes or source activity starting time cannot be validated. There are two contaminant concentration observation data sets available for 2002 and 2003 (Wels et al. 2006, 2009).

4.4. Results and Discussion

The numerical simulation output obtained using HGCH are compared with the results of developed genetic programming based surrogate models at the observation wells. Figure 4.5 (a) illustrates these comparison results. Each well number is marked on the x-axis and each of the bars corresponds to contaminant concentration in each well, obtained by HGCH and GP models, and GP models incorporating parameter uncertainties including maximum 10 % error, respectively. However, these parameter values as given in Table 4.1 represent the values incorporated in further simulation of the flow and transport processes. Instead of specifying the exact values as obtained by a limited calibration process, it was considered reasonable to assume possible deviations between exact values and the calibrated values. Therefore, a random sampling of these parameters within the specified range (+/- 10%) was accomplished using Latin Hypercube Sampling and these values were then specified in the simulation process to obtain the aquifer responses in terms of flow and transport. Figure 4.5 (a) shows the simulated concentration values

obtained using the HGCH, GP models, and the GP models with incorporated parameter uncertainties, respectively. Figure 4.5 (a) illustrates that the results obtained using the GP models are very close to the numerical simulation results obtained using a numerical simulation model. The GP model results also incorporate limited amount of uncertainties not incorporated in the numerical simulation results shown. Figure 4.5 (b) shows the summation of normalized error for uncertainty incorporated GP models and actual simulation model at each of the observation locations. It can be observed that the uncertainty incorporated GP models provide relatively accurate results for concentration simulation at observation locations, even though incorporating complex multiple species reactive contaminant transport processes under uncertainties.

The benefit of using developed GP models is that the solution results can be obtained in typically fraction of a second and do not require long computational time unlike the complex numerical simulation models. Although, time required for training and validating the GP models is to be considered. Each numerical simulation of the flow and transport processes for this mine site aquifer using HGCH simulation model typically required several hours of CPU time. Especially, this computational time saving by using GP models could be critical in development and solution of linked simulation-optimization models (Datta et al. 2014) for evolving remediation strategies for contaminant aquifer management.

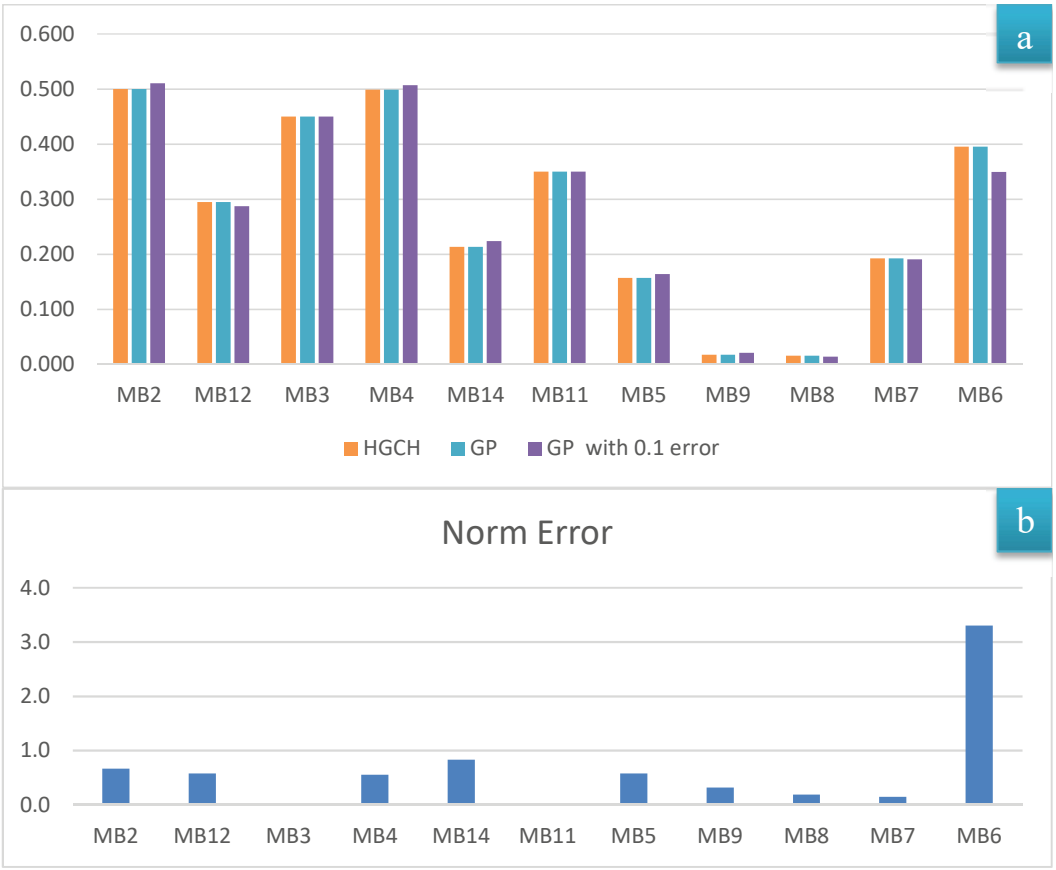


Figure 4.5 a) Comparison concentration (Mole/L) of solution results between simulation model, GP surrogate models and uncertainty incorporated GP surrogate models with 10% error in aquifer properties; b) Normalized errors for all wells sets using uncertainty incorporated GP models

The performance evaluation results of source identification using GP surrogate models linked ASA optimization algorithm are presented in Figure 4.6. The contaminant source concentration magnitude in mole per litre is shown on the y axis. In Figure 4.6, the linked optimization model solution results with GP models with no parameter uncertainty, as well as uncertainty incorporated GP surrogate models results are compared with simulation results for observation wells for specified pollution sources. The errors associated with source characterization using the GP models linked optimization model are very small, and the uncertainty incorporated GP models when linked to the

optimization model results in comparatively larger source estimation errors at few observation wells due to increasing site complexity with associated uncertainties. These performance evaluation results demonstrate that the developed methodology with GP based surrogate models are capable of optimally identifying the actual sources magnitude and location in the mine site area with small estimation errors. Figure 4.7 illustrates the results of optimal sources identification using GP models and modified GP models incorporating parameter uncertainties, linked to an optimization algorithm, using real field data from the contaminated mine site. The actual concentration measurement data from the mine site, used for this evaluation is for the year 2003 and reported in (Wels et al. 2006). The pollution concentrations were measured at observation wells and at aquatic sources. In this study, the sources considered are: aquatic source and waste deposit sources which release contaminants in the presence of water and oxygen.

Figure 4.6 shows the results for a synthetic case where the study area is the same but the sources were considered synthetic, i.e., known for simulation of simulated concentration observation values. These simulated concentration measurements are used to test the performance of the methodology. Figure 4.7 shows the results of characterization utilizing the concentration field measurements.

In the performance evaluation process, the aim is to evaluate if the source characterization model can recover the actual source characteristics based on the synthetic concentration measurements. Hence, in the linked simulation-optimization model for source characterization the calibrated simulation models were utilized. Therefore, the source characteristics can be recovered accurately using the synthetic concentration measurement data, which can be verified from the specified sources used to generate the

synthetics data. If the performance is satisfactory, the methodology should be useful to recover the actual source characteristics in the site.

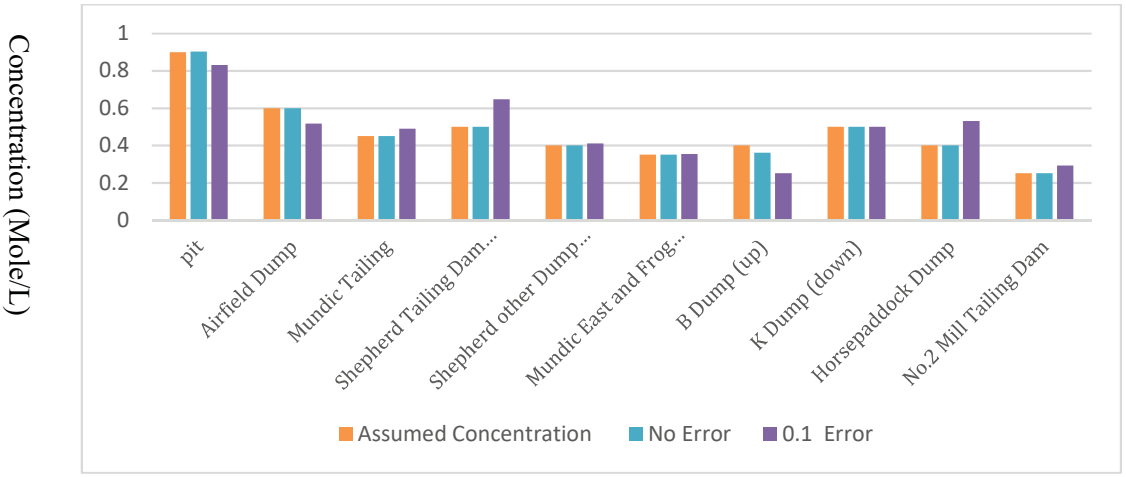


Figure 4.6 Comparison of actual concentrations (Mole/L) and concentrations obtained as ASA linked optimization model solution with synthetic data

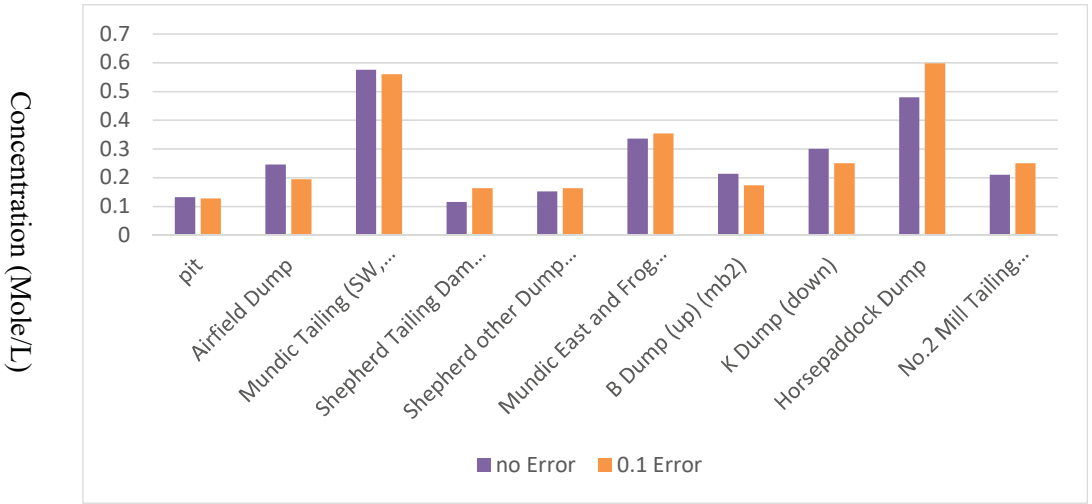


Figure 4.7 Best estimation of concentration (Mole/L) in distributed sources with GP based linked optimization model with actual field data

4.5. Conclusions

This study presents an application of a developed source characterization methodology incorporating trained GP based surrogate models for approximate simulation of the chemically reactive multiple species transport processes in a contaminated former mine site in Australia. The approximate simulators of the flow and transport processes were developed incorporating uncertainties in few of the hydrogeologic parameters. The performance evaluation results show the potential applicability of using GP based surrogate models as approximate simulators of complex hydrogeologic and geochemical processes in a complex contaminated aquifer site, while incorporating limited degree of uncertainties. The evaluation results illustrate the accuracy and reliability of surrogate model predictions under uncertainties applicable to real life sites. The developed GP models reduce the computational time and complexity, and appear to provide acceptable results for simulating the complex reactive contaminant transport processes in aquifers. Although the linked simulation-optimization approach is an effective method for identification of source characterization and monitoring network design under uncertainties in complex real life scenarios for groundwater management, this approach requires huge computation time, due to iterative repeated solution of the numerical flow and transport simulation models. The performance evaluation results show the efficiency and feasibility of using GP based surrogate models trained with an ensemble of contamination scenarios incorporating some degree of parameter uncertainties as well.

These preliminary evaluation results are encouraging and point towards the feasibility of using the proposed methodology to optimally characterize the sources and pathways of contamination in complex contaminated aquifer sites. This methodology should address

some of the existing limitations in optimal characterization of unknown contaminant sources in contaminated sites like abandoned mine sites. Such characterization is the essential first step in tackling effectively a very critical environmental problem, and designing effective remediation strategies. The limitations in the process of implementing adequately calibrated flow and transport simulation models for such a hydrogeologically and geochemically complex contaminated aquifer site with very limited availability of measurement data are also highlighted.

The next chapter (Chapter 5) presents the development and application of fractal singularity based design of optimal monitoring network design methodology, and its performance evaluation for a contaminated aquifer site in terms of unknown pollution source characterization efficiency.

5. Fractal Singularity Index Based Multi-Objective Monitoring Network Design for Unknown Contaminant Source Characterization in an Aquifer with Reactive Transport Process

Similar versions of this chapter have been published and copyrighted in the following journals:

- Esfahani, H. K. and B. Datta, 2016. Fractal Singularity Index Based Multi-objective Monitoring Network Design for Unknown Contaminant Source Characterization in an Aquifer with Reactive Transport Processes. Journal of Water Resource Research. Under Review.

In this chapter, Local Singularity Mapping Technique is utilized to obtain potential well locations used as input to the optimal design model. Also multi-objective optimization model is used for solving the optimal monitoring network design model to identify unknown groundwater contaminant source characteristics. The potential application of the developed methodology is demonstrated by evaluating the performance for an

illustrative mine site area. These performance evaluation results show the improved efficiency in source characterization when concentration measurements from the designed monitoring network are utilized.

5.1. Introduction

Groundwater is one of the valuable natural resources in many parts of the world; however it is widely contaminated due to human activities. Undoubtedly, characterization of unknown groundwater pollution sources is the first step to obtain reliable groundwater management and remediation strategies. The concentration measurement data at monitoring locations are utilized to estimate the unknown contaminant source locations, magnitudes and corresponding durations of activity. However, accurate characterization of contaminant sources is a challenging task due to uncertainties in: predicting source flux distribution, hydrogeological and geochemical parameters, and the concentration field measurement. Although a large number of observed pollutant concentration data spread over time and space is essential for accurate source characterization, budgetary constraints also restrict long-term and widespread spatiotemporal monitoring. Moreover, any arbitrary groundwater monitoring network may not characterize unknown groundwater pollution sources accurately. The complexity of unknown source characterization with limited arbitrary wells increases with incorporating reactive transport of chemical species with plume overlapping from several distributed sources such as mine site area aquifers. Therefore, more accurate estimation of pollution source characterization is obtained using the optimal design of a monitoring network in order to accurately estimate pollution source properties. A suitable monitoring network can be utilized to increase the reliability of source characterization in terms of location,

magnitude and duration of activity. Potential well locations to obtain the optimal monitoring network design should be selected carefully after considering likely scenarios of contamination. One of the effective methods to select the potential well locations can be to first determine the contaminant plume boundary and then select potential monitoring well locations within the boundary. This study proposes the fractal singularity mapping index methodology to obtain the contaminant transport plume boundary in a contaminated aquifer to increase the utility and efficiency of the optimal monitoring network design to characterize unknown distributed pollution sources accurately in a complex contaminated aquifer, such as an abandoned mine site.

The local fractal singularity mapping technique (FSMT) is used in this study to characterize the groundwater contamination plume, and to select potential observation locations to increase the efficiency and accuracy of the monitoring network design. The multi-objective design is based on a Pareto-optimal solution for two objectives: (1) maximize the summation of the product of estimated concentration gradients, and the simulated concentration at that location, and (2) minimize the maximum normalized error between actual concentration and those estimated by Kriging using the designed monitoring wells. This is achieved by selecting a mix of monitoring locations where the concentration gradient is large, and also those locations where the best estimate is obtained for the aquifer concentrations using kriging interpolation technique. Performance of the proposed source identification methodology is evaluated by solving an illustrative problem. Variation in the accuracy of source identification is demonstrated by making a trade-off between the previously stated conflicting objectives while designing an optimal monitoring network. To illustrate the FSMT efficiency, the source identification results obtained by using pollutant measurement data from the optimally

designed monitoring network with and without using singularity mapping guideline are compared. In addition, these results are compared with the results of unknown source characterization using measurement data from arbitrary monitoring networks with, and without, using singularity mapping guidelines.

5.2. Methodology

The proposed methodology consists of two steps. In the first step, FSMT is used to compute the singularity indices based on the assumption of source fluxes at sources and corresponding aquifer response at all nodes. The source fluxes are specified for this performance evaluation purpose. In a real life situation, information regarding the source fluxes will be necessary. The concentration values are required at every node to estimate the singularity indices. These values can be simulated using numerical simulation models, which is a necessity for the performance evaluation purpose. In field situations, it is possible to use measured concentrations only at monitoring locations, and interpolate them to each node using interpolation techniques e.g., Kriging. It is not essential in that case to estimate and specify the source fluxes, if the simulation model is not utilized to obtain concentration values. However, for performance evaluation purposes, it is essential to specify the sources and to simulate concentration measurements synthetically using numerical simulation models. FSMT is applied to estimate the likely contamination plume boundary which is then used as one of the guidelines for locating potential monitoring wells. In the second step, a multi-objective linked simulation-optimization formulation is used for optimal monitoring network design, with constraints on the maximum number of monitoring wells to be implemented. The Pareto-optimal solutions obtained from the two-objective model are used to design a set of Pareto-optimal

monitoring networks. The objectives are: (1) maximize the summation of the product of estimated concentration gradients, and the concentration at that location, and the simulated concentration at that location, and (2) minimize the maximum normalized error between actual concentration and those estimated with the interpolation models, based on monitoring data from designed monitoring locations.

The linked simulation-optimization model for optimal source identification is solved using the ASA algorithm. The ASA is used as the optimization algorithm for solving the optimization problem that minimizes the difference between the simulated and measured pollutant concentrations at the optimally chosen monitoring locations obtained as solution. The numerical simulation model for flow and transport processes in the linked optimization algorithm is substituted by Genetic Programming based surrogate models to increase the efficiency and decrease the computing time in the iterative search for optimal solution. The source-identification model is solved using concentration measurements from a chosen Pareto-optimal monitoring networks.

5.2.1. The Singularity Mapping Technique Methodology

Generalized fractal self-similarity is often characterized by a power-law relationship in the spatial or frequency domain (Chen et al. 2015). In the singularity mapping technique, the C-A (Concentration-Area) model is used. In this study, the FSMT in 2D map data is described as a power-law relationship between area A in a sampled region, and the total amount of a certain physical quantity μ(A) as Eq. (5.1) (Datta et al. 2016).

$$\langle \mu(A) \rangle = cA^{\alpha/2} \tag{5.1}$$

Here ⟨ ⟩ denotes the statistical expectation, α is the Holder exponent or singularity index, and c is a constant. The areal density value of μ(A) in the area A is defined by concentration ρ(A) as Eq. (5.2).

$$\langle \rho(A) \rangle = \frac{\mu(A)}{A} = cA^{\frac{\alpha}{2}-1} \tag{5.2}$$

As for the concept of singularity, Cheng (2007) considered mineralization to be a singular process due to an efficient process of element enrichment. A singularity mapping technique was proposed to characterize the degree of uniqueness of geological features, and it can detect potential targets that are often smoothed by traditional contouring methods. From the multifractal point of view, the singularity index characterizes how statistical behaviors change as the scale of geochemical values changes. Different points within close vicinity may have different singularity exponents (Cheng, 2007). In the singularity mapping technique, the indices are estimated using the window-based procedure. The improved window-based procedure (Chen et al. 2015) is conducted as per the following steps.

Define a set of square windows A(ri) with variable window sizes rmin=r1<r2<…<ri=rmax for a given sampling point on the map. Calculate concentration ρ[A(ri)]. In the improved method, the minimum vale (ρmin) is subtracted from all concentration values in each window.

Eq. (5.3) is defined by taking logarithm of Eq. (5.2).

$$\rho[A(r_i)] = (\alpha - 2)\log(r_i) + C \tag{5.3}$$

Therefore the slope of the log-log plot of Eq. (5.3) can be used for the singularity index (α) calculation. Based on the distribution of α, the 2D mapped area is classified into subsets of fractals and can be divided into following three cases. If an anomaly is convex, then $\rho(A)$ is decreasing function of A, and $\alpha<2$ it indicates high density and positive singularity. If an anomaly is concave, then $\rho(A)$ is increasing function of A and $\alpha>2$ indicates high density and negative singularity (Cheng 2007). If an anomaly is constant, then $\rho(A)$ is a constant and $\alpha=2$, it indicates a non-singular behaviour.

This approach has been developed in this study to detect the boundary of the contamination plume and determine the effective potential well locations relevant to source characterization. The contamination plume boundary is estimated utilizing the characteristics that almost horizontal gradient generally prevail near the plume boundary which corresponds to the inflection point of the anomaly. Around the inflection point, the value of ρ (A) estimated by the window-based method, does not change with change of Area A. Therefore, near the plume boundary, the third scenario is valid. This indicates that the singularity index $\alpha=2$ detects the boundary, while $\alpha<2$ and $\alpha>2$ specify the inside and outside of the boundary, respectively.

5.2.2. Multi-objective Optimization Algorithm for Monitoring Network Design

In this study, the purpose of the monitoring network design is to maximize the estimated concentration gradients while preserving the accuracy of estimated contaminant plume mass. Thus, there are two objectives to be optimized: (i) maximize the summation of estimated concentration gradients weighted by the concentrations at that location, and (ii) minimize the maximum normalized error of contaminant concentration estimated by utilizing the monitoring network and the simulated contaminant concentration data at the

potential well locations. The objective of finding well locations with maximum concentration gradient is conflicting in nature with the other objective of finding well locations which minimize the normalized error between the simulated and estimated concentrations at potential well locations. A multi-objective optimization model is formulated for the design of an optimal monitoring network with the previously stated conflicting objectives. One of the objectives needs to be sacrificed to improve the other objective and vice versa. The two-objective optimization model is solved by optimizing one of the objectives subject to the other objective defined as an implicit constraint. The number of monitoring wells to be selected is essentially governed by budgetary constraints. The two objectives of the multi-objective optimization model for optimal monitoring network design for accurate identification of unknown pollution sources are defined by Equations (5.4) and (5.6), respectively (Prakash and Datta 2014).

The multi-objective optimization model can be mathematically expressed as:

$$F1 = Maximize(\sum f_{i,j} C^*_{i,j} \left\{ \frac{\left|C^*_{i-1,j} - C^*_{i,j}\right| + \left|C^*_{i,j+1} - C^*_{i,j}\right|}{dx} + \frac{\left|C^*_{i,j-1} - C^*_{i,j}\right| + \left|C^*_{i,j+1} - C^*_{i,j}\right|}{dy} \right\}) \quad (5.4)$$

$$\varepsilon_{max} \geq C^*_{i,j} \geq \varepsilon_{min} \; \forall i, j \quad (5.5)$$

$$F2 = Minimize(Maximum \; (\frac{C_{sim} - C_{int}}{C_{sim} + N})) \quad (5.6)$$

Where $C^*_{i,j}$is simulated concentration in cell i, j. $f_{i,j}$ represents the binary decision variable to place or not to place a monitoring well at grid location i, j. fi,j≡ such that when fi,j value equal to 1 representing monitoring well to be placed at grid i, j, and zero otherwise. ε_{max} and ε_{min} are the high and low value of concentration, respectively. C_{int}

is interpolated (kriged) concentration in cell i, j. C_{sim} is simulated concentration in cell i, j.

The two-objective optimization model is solved using the constrained method (Datta and Peralta 1986, Prakash and Datta 2014). In the constrained method, one of the objective functions (F1) is maximized, constraining the minimum level of satisfaction of the second objective function (F2) as shown in Eq. (5.7)

$$\sum f_{i,j} C^*_{i,j} \left\{ \frac{\left|C^*_{i-1,j} - C^*_{i,j}\right| + \left|C^*_{i,j+1} - C^*_{i,j}\right|}{dx} + \frac{\left|C^*_{i,j-1} - C^*_{i,j}\right| + \left|C^*_{i,j+1} - C^*_{i,j}\right|}{dy} \right\}) - \gamma \geq 0 \quad (5.7)$$

Where γ is the minimum level of satisfaction of the second objective function F2, also termed as the trade-off constant.

Therefore, the resulting model can be solved iteratively as a single objective optimization model for different satisfaction levels of γ, thus a Pareto-optimal solution set is generated. The second objective function can be specified as a new implicit constraint. The upper limit and lower limit of γ are defined by equations 5.8 and 5.9 respectively. The upper limit of γ is equal to the maximum value of the second objective function F2 when solved as a single objective optimization [Eq. (5.9)]. To obtain the lower limit of γ, first the single objective optimization is solved to obtain the maximum value of the first objective function F1. Then, the minimum of γ is the value of the second objective function F2 corresponding to the optimal F1 [Eq. (5.9)].

$$Max F_2 \geq \gamma \quad (5.8)$$

$$F_{2\,Max\,F_1} \leq \gamma \quad (5.9)$$

Where F_{2MaxF1} is the value of the objective function F2 corresponding to the maximum value of the first objective function F1 when solved as a single objective model. All solutions obtained on a Pareto-optimal front correspond to a different Pareto optimal monitoring network.

5.2.3. Linked Simulation-Optimization Model for Optimal Contaminant Source Identification

Source identification in terms of magnitude of an unknown pollution source is often solved using a linked simulation-optimization approach. The linked simulation-optimization model simulates the physical and chemical processes of flow and solute transport within the optimization algorithm. The flow and solute transport simulation models are treated as an important binding constraint for the optimization model. Therefore, any feasible solution of the optimization model needs to satisfy the flow and the transport simulation model. The advantage of this approach is that it is possible to link any complex numerical model to the optimization model. However, running the simulation model for several thousand times to obtain the optimal source characterization is time consuming and may affect computational feasibility and efficiency of the methodology. To address these issues, trained and tested Genetic Programming (GP) based surrogate models are utilized as approximate simulators of the physical processes in the linked optimization algorithm to obtain the reasonable and acceptable results with enormous saving of CPU time. In this study, linked simulation-optimization based methodology for characterization of unknown pollution sources utilizes GP-based surrogate models as approximate simulators of the flow and transport processes in the

aquifer study area contaminated by multiple reactive chemical species (Esfahani and Datta 2016).

The methodology for source identification includes three main steps which are as follows:

1. Implementation of a numerical flow and transport simulation model based on the estimated hydrological and geochemical properties of the contaminated aquifer study area. In the performance evaluation reported here, the numerical simulation model is the three dimensional finite element based flow and transport model HYDROGEOCHEM 5.0 (HGCH) (Sun 2004, Yeh et al. 2004).
2. GP models are trained and tested using the numerical simulation model to be utilized as approximate simulation models to compute the concentrations of the chemical contaminants at specified locations resulting from specific contaminant sources.
3. The trained GP based surrogate models are linked to the Adaptive Simulated Annealing (ASA) optimization algorithm within a linked simulation-optimization model to obtain the optimal characterization of the unknown contaminant sources in terms of location, magnitude and time of activity.

5.2.4. Groundwater Flow and Transport Simulation Model

HYDROGEOCHEM 5.0 (HGCH) is a three-dimensional simulator of hydrogeological and geochemical reactive transport processes in saturated-unsaturated media. This robust simulator consisting of the flow simulation model and physio-chemical transport simulation model is used in this study as the numerical simulation model.

The general equations for flow through saturated-unsaturated media are obtained based on following compounds: 1) fluid continuity, 2) solid continuity, 3) fluid movement (Darcy's law), 4) stabilization of media, and 5) water compressibility (Yeh et al., 1994):

$$\frac{\rho}{\rho_0} F \frac{\partial h}{\partial t} = \nabla . \left[\boldsymbol{K} . \left(\nabla h + \frac{\rho}{\rho_0} \nabla z \right) \right] + \frac{\rho^*}{\rho_0} q \tag{5.10}$$

$$F = \alpha' \frac{\theta}{n_e} + \beta' \theta + n_e \frac{ds}{dh} \tag{5.11}$$

$$\boldsymbol{K} = \frac{\rho g}{\mu} \boldsymbol{k} = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} \frac{\rho_0 g}{\mu_0} \boldsymbol{k}_s \, k_r = \frac{(\rho/\rho_0)}{(\mu/\mu_0)} \boldsymbol{K}_{so} \, k_r \tag{5.12}$$

$$\boldsymbol{V} = -K \left[\frac{\rho}{\rho_0} \nabla h + \nabla z \right] \tag{5.13}$$

Where θ is effective moisture content (L3/ L3); h is pressure head (L); t is time (T); z is potential head (L); q is source or sink of fluid [(L3/ L3)/T]; ρ_0 is fluid density without biochemical concentration (M/ L3); ρ is fluid density with dissolved biochemical concentration (M/ L3); ρ* is fluid density of either injection (= ρ*) or withdraw (= ρ) (M/ L3); μ_0 is fluid dynamic viscosity at zero biogeochemical concentration [(M/L)/T]; μ is the fluid dynamic viscosity with dissolved biogeochemical concentrations [(M/L)/T]; α′ is modified compressibility of the soil matrix (1/L); ß is modified compressibility of the liquid (1/L); ne: effective porosity (L3/L3); S is degree of effective saturation of water; G is the gravity (L/T2); k is permeability tensor (L2); ks is saturated permeability tensor (L2); Kso is referenced saturated hydraulic conductivity tensor (L/T); kr is relative permeability or relative hydraulic conductivity (dimensionless); F is the generalized storage coefficient (1/L); K is the Hydraulic conductivity tensor (L/T); V is the Darcy's Velocity (L/T). When combined with appropriate boundary and initial conditions, the

above equations are used to simulate the temporal-spatial distributions of the hydrological variables, including pressure head, total head, effective moisture content, and Darcy's velocity.

Contaminant transport equations used in the HGCH model can be derived based on mass balance and biogeochemical reactions (Yeh et al. 2000). The general transport equation using advection, dispersion/diffusion, source/sink, and biogeochemical reaction as the major transport processes can be written as follows (Sun 2004):

$$\frac{D}{Dt}\int_{v}\theta C_i dv = -\int_{\Gamma} n.(\theta C_i)V_i d\Gamma - \int_{\Gamma} n.J_i d\Gamma + \int_{v}\theta r_i dv + \int_{v} M_i dv, i \in M \qquad (5.14)$$

Where C_i is the concentration of the i-th species in mole per unit fluid volume (M/L3); ν is the material volume containing constant amount of media (L3); Γ is the surface enclosing the material volume ν (L2); n is the outward unit vector normal to the surface Γ; Ji is the surface flux of the i-th species due to dispersion and diffusion with respect to relative fluid velocity [(M/T)/L2]; θri is the production rate of the i-th species per unit medium volume due to all biogeochemical reactions [(M/L3)/T]; Mi is the external source/sink rate of the i-th species per unit medium volume [(M/L3)/T]; M is the number of biogeochemical species; Vi is the transporting velocity relative to the solid of the i-th biogeochemical species (L/T).

5.2.5. Genetic Programing Based Surrogate Model Development

In this study, trained and tested Genetic Programming (GP) (Koza 1994) models are used to develop surrogate models as approximate simulators of transport process (Sreekanth and Datta 2012). GP models utilize sets of input-output data for training purpose, which

are generated randomly by the simulation model. The numerical Simulation model creates M number of output sets from M number of input sets, which is generated by using random Latin hypercube sampling (LHS) in a defined range (Esfahani and Datta 2015a, 2015b). The performance of each of the GP models is evaluated in terms of training, testing, and validation using the set of input-output patterns. The testing data evaluates the model performance for new data without developing a new fitness function. Therefore, the following steps are utilized to develop the GP based surrogate models:

1. The M number of input sets is generated using LHS in a specific range as assumed source concentrations.
2. The M output set are obtained at observation well location using the actual simulation model based on the hydrogeological properties of the aquifer.
3. The set of input-output patterns are fed into the GP based surrogate models to obtain GP formulation. For each observation well location, one separate GP formulation is required.
4. These data patterns are split into three categories for training, testing and validating the GP formulations.
5. The obtained GP formulations are utilized to approximately simulate the concentrations at specified locations resulting from specific contaminant sources.

5.2.6. Developing GP Surrogate Models Linked to Source Identification Model

This study utilizes multiple realization optimization incorporating surrogate modelling approach to solve the unknown groundwater sources characteristics in mine site area. Here, the multiple realizations pertain to the concentration of the different chemical species values obtained from different surrogate models in the ensemble. The surrogate

models are developed using genetic programming and coupled to the ASA algorithm. The approximate simulation models utilize the unknown source concentration candidate solutions generated by the optimization algorithm to obtain the estimated contaminant concentration at monitoring locations. Then the optimization algorithm evaluates the objective function. The objective function value is a function of the differences between surrogate models estimated and measured concentration values. The optimal source characterization is obtained by solving the optimization model to minimize the objective function. Mahar and Datta (2001) defined the objective function of a simulation-optimization model for unknown concentration source characterization as follows:

$$Minimize\ F1 = \sum_{k=1}^{n_k} \sum_{job=}^{n_{ob}} (C_{est}{}_{iob}^{k} - C_{obs}{}_{iob}^{k})^2 \,.\, w_{iob}^{k} \tag{5.15}$$

$$C_{est}{}_{iob}^{k} = f(x, y, z, C_{si}) \tag{5.16}$$

$$w_{iob}^{k} = \frac{1}{(C_{obs}{}_{iob}^{k} + n)} \tag{5.17}$$

Where n is an appropriate constant where errors at low concentrations do not dominate the solution, $C_{obs}{}_{iob}^{k}$ is the concentration measured data at observation monitoring location iob and at the end of time period k (ML-3), $C_{est}{}_{iob}^{k}$ is the concentration estimated by the simulation models at observation monitoring location iob and at the end of time period k (ML-3), nt is the total number of monitoring time steps, nob is the total number of observation wells, nk is total number of concentration observation time periods, nob is total number of observation wells, obs_{iob}^{k} is observed concentration at well iob and at the end of time period k, f(x, y, z, Csi) is the simulated concentration obtained from the

transport simulation model at an observation location and source concentrations Csi. w_{iob}^{k} is the weight corresponding to observation location iob, and the time period k.

The constraint set equation (5.16) represents the linked simulation model for flow and transport process simulation. This simulation model can be a numerical model, or, it can be replaced by an approximate simulator such as a trained and tested surrogate model.

The present study incorporates the Adaptive Simulated Annealing as the optimization algorithm for optimal source characterization model. This algorithm is chosen based on its comparative efficiency in reaching a global optimal solution. The methodology is schematically represented in Figure 5.1.

5.3. Performance Evaluation of Developed Methodology

To evaluate the performance of the proposed monitoring network design for the characterization of the distributed pollution sources using linked simulation-optimization, an illustrative heterogeneous, anisotropic aquifer is used, which resembles an abandoned (no longer operating) contaminated gold and copper mine site in Queensland, Australia as shown in Figure 5.2. S(i) represents the locations of potential pollutant sources. There are ten (i=1,2..,10) potential sources identified which are shown in figure 5.2. Groundwater flow and solute transport model are simulated with hydrogeological parameters as given in Table 5.1. Also, the reaction and reaction coefficients, etc. incorporated in this study are clearly defined in Table 5.2.

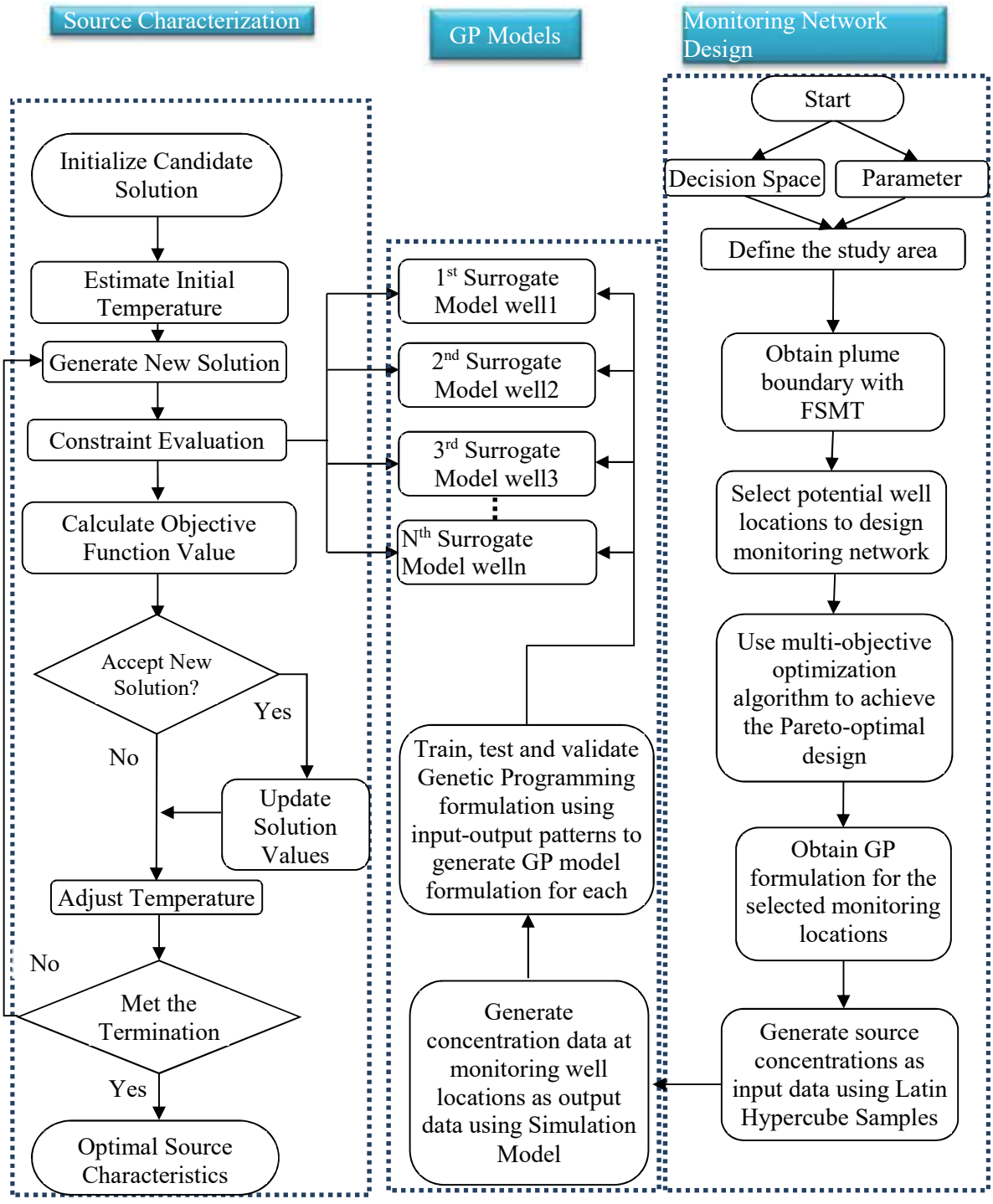


Figure 5.1. Schematic representation of optimal monitoring network design for source identification

Table 5.1 Aquifer's physical and hydrogeological properties

Aquifer Parameter	Unit	Value
Length of mine site	m	4000
Width of mine site	m	4000
Thickness of mine site	m	variable
Node numbers at each layer		2621
Element numbers at each layer		5010
Horizontal Hydraulic conductivity, K	m/d	
Layer 1		0.49683
Layer 2		0.05320
Layer 3		0.01437
Layer 4		0.00936
Vertical Hydraulic conductivity, K	m/d	
Layer 1		0.200
Layer 2		0.002
Layer 3		0.0005
Layer 4		0.0005
Effective porosity, Θ		0.28
Longitudinal dispersivity, αL	m/d	18
Transverse dispersivity, αT	m/d	6
Horizontal anisotropy		1
Diffusion coefficient		0
Average rainfall	mm/year	700

The synthetic concentration measurement data used for the performance evaluation purpose only, facilitates evaluation of the methodology without having to account for the unknown reliability of any field data or model calibration issues. The potential duration

of each source is 800 days, and the pollutant flux from each of the sources is assumed to be constant over a stress period. The pollutant flux from each of the sources is represented as S(i), where i represents the source number. A total of ten potential source fluxes (S1, S2, S3, S4, S5, S6, S7, S8, S9 and S10) are considered as explicit unknown variables in the source characterization model. The chemically reactive multiple species, and the large number of distributed contaminant sources increase the complexity of this scenario. The typical chemical reaction equations are illustrated in Table 5.2. Therefore, an effective monitoring network should be designed for effective and efficient unknown distributed source characterization. The study area and the hydraulic head contours are shown in Figures 5.2 and 5.3 respectively. The river side and the pit is specified by the constant head conditions.

Table 5.2 Typical chemical reactions during the contaminants' transport in mine site area

Chemical Reaction Equations	**Constant Rate (Log k)**
Equilibrium Reactions	
$Cu^{2+} + H_2O \leftrightarrow Cu(OH)^{+} + H^{+}$	-9.19
$Cu^{2+} + SO_4^{2-} \leftrightarrow CuSO_4$	2.36
$Fe^{2+} + H_2O \leftrightarrow Fe(OH)^{+} + H^{+}$	-9.397
$Fe^{2+} + 2H_2O \leftrightarrow Fe(OH)_2 + 2H^{+}$	-20.49
$Fe^{2+} + SO_4^{2-} \leftrightarrow FeSO_4$	2.39
$Fe^{2+} + SO_4^{2-} + H^{+} \leftrightarrow FeHSO_4^{+}$	1.66
$4Fe^{2+} + 4H^{+} \leftrightarrow 4Fe^{3+} + 2H_2O$	8.5
$Fe^{3+} + H_2O \leftrightarrow Fe(OH)^{2+} + H^{+}$	-2.187
$Fe^{3+} + 2H_2O \leftrightarrow Fe(OH)_2^{+} + 2H^{+}$	-4.59

$Fe^{3+} + SO_4^{2-} \leftrightarrow FeSO_4^{+}$	4.05
$Fe^{3+} + SO_4^{2-} + H^{+} \leftrightarrow FeHSO_4^{2+}$	2.77

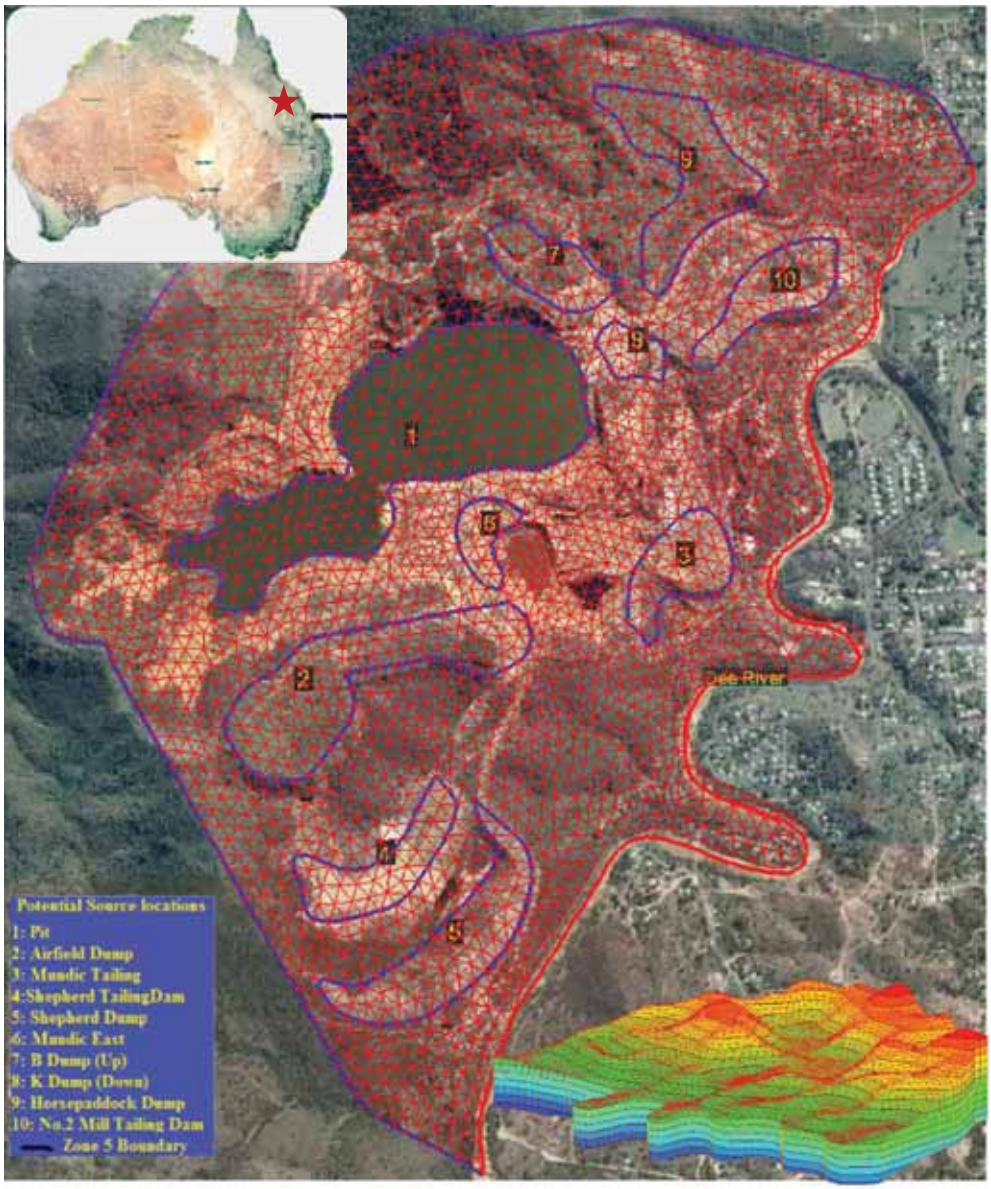


Figure 5.2 Plan and 3D Views of illustrative study area

The pit is represented as a contaminant source by considering unknown constant concentration boundary condition along the pit perimeter. The waste dumps are

considered as contaminant sources with unknown concentration in the recharge at its locations. Because of the complex nature of the study area considered, source concentrations were considered as constant over time at a location.

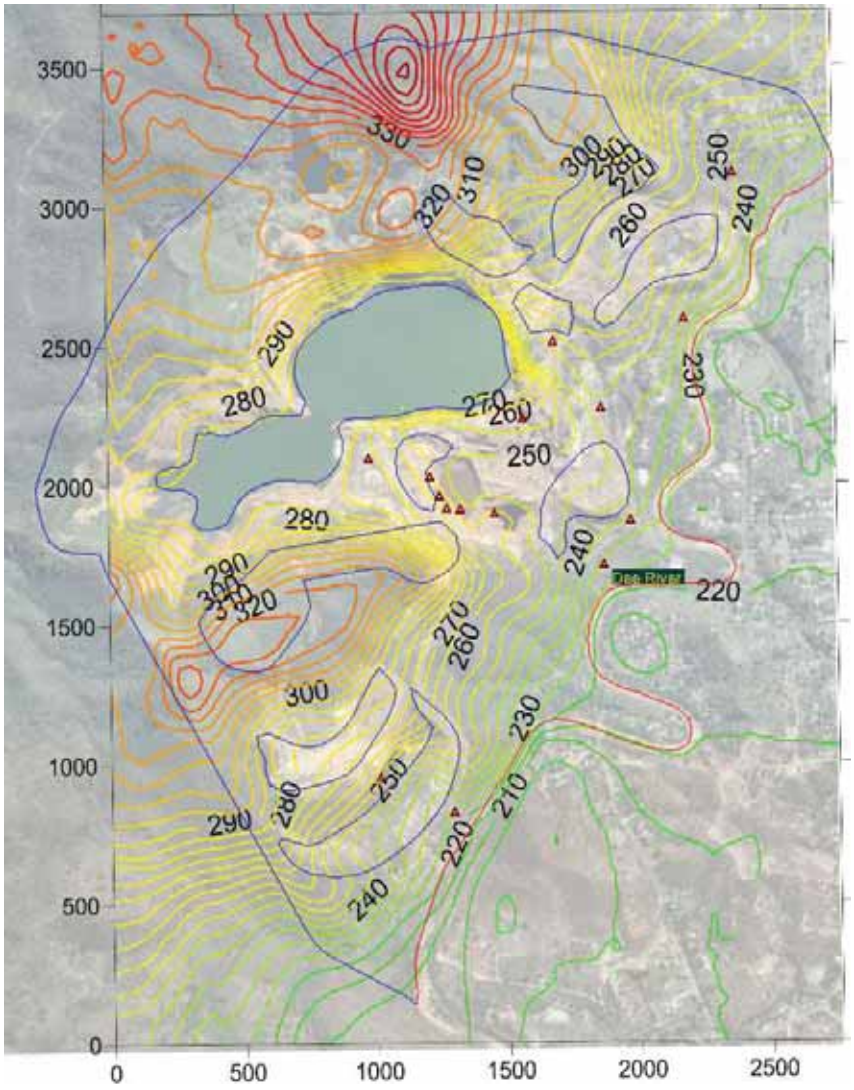


Figure 5.3 Total head contours (Esfahani and Datta 2016)

5.3.1. Potential monitoring well locations using Fractal Singularity Mapping Technique

In this section, the local singularity mapping method is utilized to delineate the contaminant plume in the polluted groundwater aquifer. The assumed contaminant plume for this complex scenario is obtained from flow and transport processes model solution with chemically reactive species, and the accuracy of the proposed methodology is evaluated for this complex mine site area aquifer.

The fractal singularity mapping method was used to delineate the contamination plume in the illustrative abandoned mine site aquifer area. In this scenario, all assumed pollutant sources are active and since they are situated relatively close to each other, overlapping of the pollutant plumes released from individual sources is expected. Figure 5.4 shows the singularity index contours for 800 days after the start of contaminant release at sources. Values more than 2, show the regions out of the plumes. The plume's inside regions are specified by values less than 2. Contour related to singularity index two is the representative of the plume boundaries. Based on available boundary plume obtained from FSMT, the potential of monitoring locations can be achieved effectively and accurately. Figure 5.5 shows the significant indicator t-value of the spatial correlation between the singularity indices and outcropping contaminated areas, wherein the higher the t-value, the stronger their spatial relationship. In our case, the t-value reaches a maximum when $\alpha = 2.004$. Figure 5.6 shows the potential well locations using FSMT boundary conditions and also the arbitrary potential well locations to compare the FSMT efficiency.

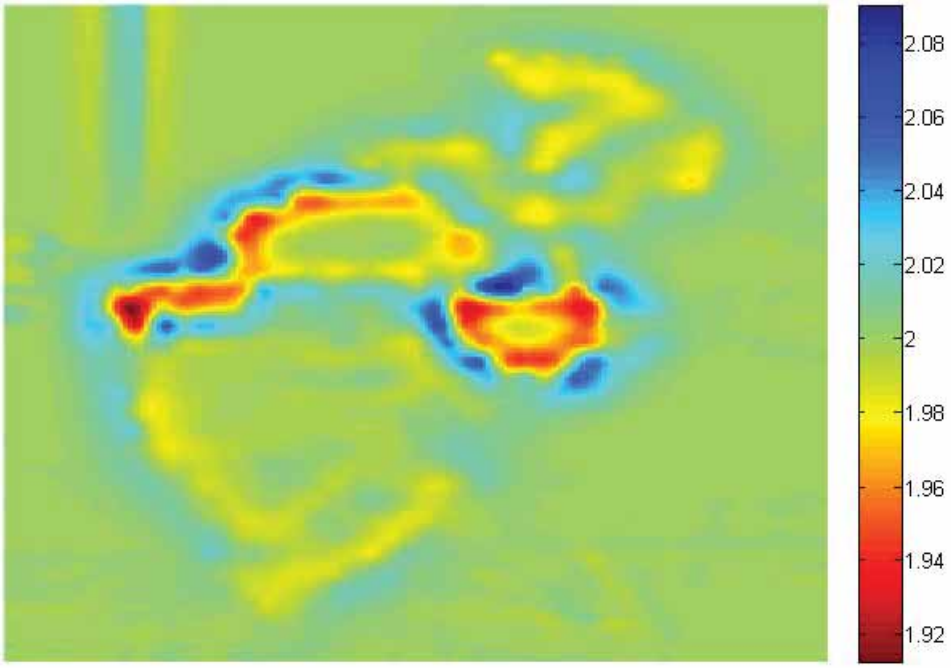


Figure 5.4 Singularity map in the study area

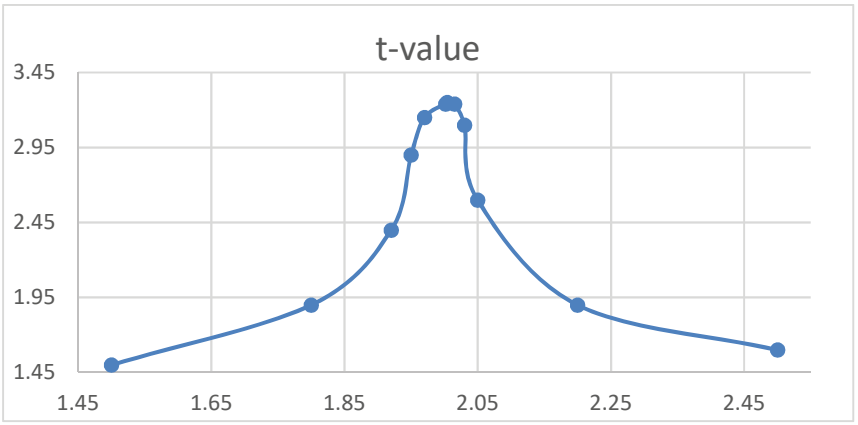


Figure 5.5 The relationship between the singularity index and t-value

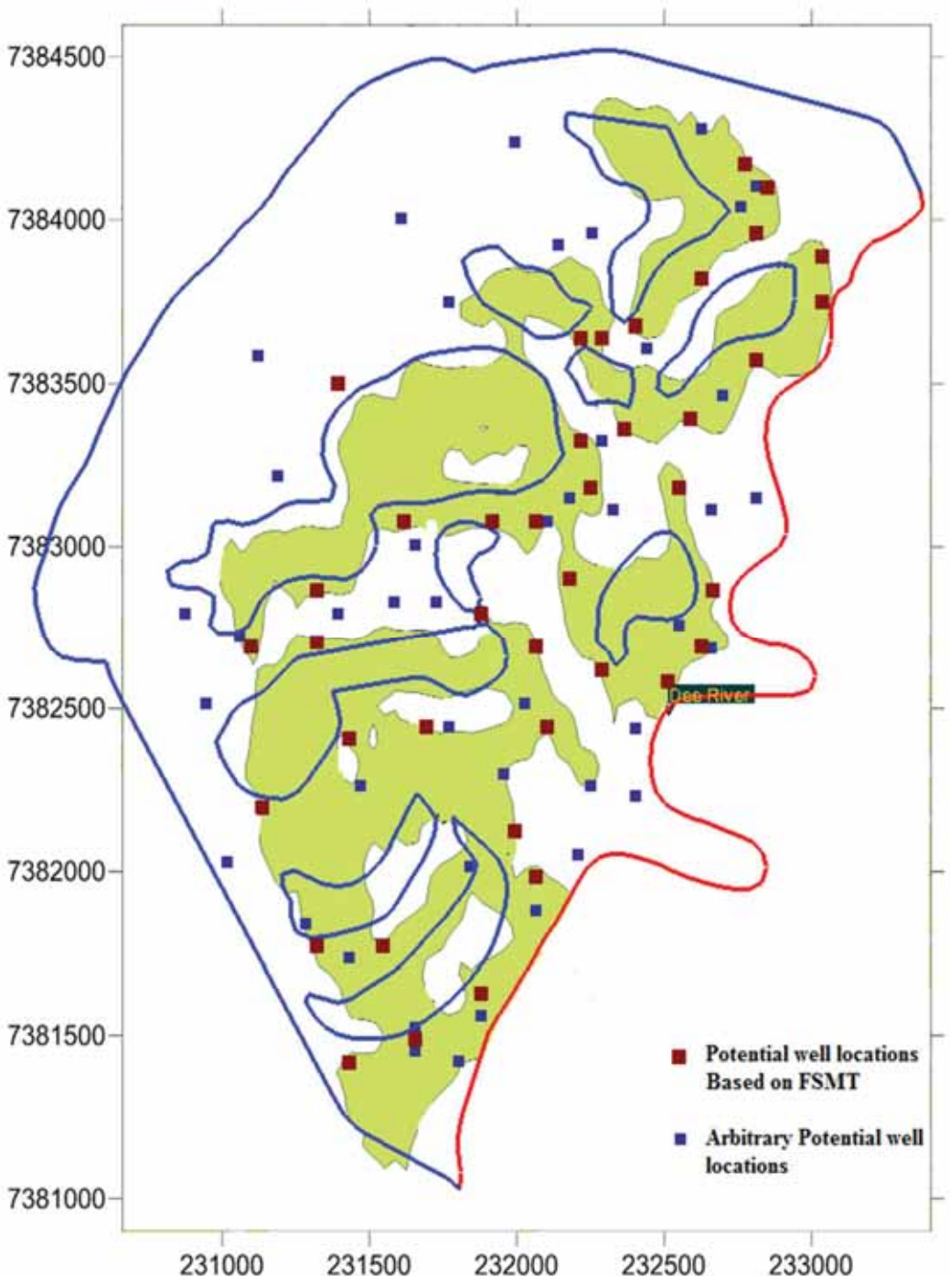


Figure 5.6 Plan view of singularity index based plume boundary and the potential well locations for monitoring network design. The red and blue points are the singularity index-based potential well locations and arbitrary well locations respectively

5.3.2. Optimal Monitoring Networks and Arbitrary Monitoring Networks

The optimal monitoring network is designed based on the Pareto optimal solutions for the two objectives using the criterion of minimizing the maximum normalized error between measured and spatially interpolated concentration values (F1), and maximizing the sum of the estimated concentration gradient values (F2). For this performance evaluation purpose the spatial interpolation is based on kriging. The measured concentration values are based on synthetic concentration measurement data obtained using the simulation model, which is essential for performance evaluation purpose, as explained earlier. Two sets of 10 different Pareto-optimal solutions are obtained as different pairs of F1 and F2 values with and without using singularity index technique. The value of the minimum satisfaction level of the second objective function F2 varies from a minimum 0.47 to a maximum of 5.38 for the first monitoring network group, and also 0.35 and 4.5 are the minimum and maximum value for the second monitoring network group respectively [Eqs. (10) and (11)]. Each of these solutions on the Pareto-optimal front represents different Pareto-optimal monitoring networks represented as MNSI1 to MNSI10 (for the first set) and MN1 to MN10 (for the second set), respectively, for the corresponding values of objective functions F1 and F2 (Table 5.3). A total of fifteen monitoring wells are selected for each Pareto-optimal monitoring network solution.

Table 5.3 Pareto-Optimal Monitoring Networks

	Monitoring network	Objective function value F1	Objective function value F2
Monitoring Network Using Singularity Index	MNSI1	0.56	0.47
	MNSI2	0.76	1.13
	MNSI3	0.87	1.94
	MNSI4	0.92	2.43
	MNSI5	0.99	2.92
	MNSI6	1.05	3.41
	MNSI7	1.04	3.9
	MNSI8	1.05	4.4
	MNSI9	1.1	4.89
	MNSI10	1.14	5.38
Monitoring Network without Singularity Index	MN1	0.65	0.35
	MN2	0.85	0.95
	MN3	0.93	1.43
	MN4	0.95	2.11
	MN5	0.99	2.56
	MN6	1.02	3.14
	MN7	1.09	3.43
	MN8	1.12	3.89
	MN9	1.19	4.25
	MN10	1.2	4.5

5.3.3. Linked Simulation-Optimization Model for Identification of Unknown Source Fluxes Using Pareto-Optimal Monitoring Networks and Arbitrary Monitoring Network

A linked simulation-optimization model [objective function Eq. (5.14) and constraint set Eq. (5.15)] is solved to estimate the pollution sources concentration to evaluate the performance of the proposed methodology. The two sets of 10 Pareto-optimal monitoring networks are designed from the potential well locations using FSMT (MNSI1 to MNSI10), and without utilizing FSMT (MN1 to MN10), respectively. Concentration measurements from each of the monitoring networks corresponding to each of the Pareto-optimal solution are used to estimate the pollution sources concentration. These evaluation results using concentration observations from the two sets of 10 Pareto-optimal monitoring networks are compared to find the efficiency of using FSMT based on the two objectives. For this evaluation purpose, the observed aquifer responses are simulated by solving HYDROGEOCHEM (Sun 2004, Yeh et al. 2004), along with appropriate initial and boundary conditions. In order to evaluate the performance for erroneous concentration measurement data, numerically simulated concentration measurements are perturbed to represent the effect of random measurement errors. The observed pollutant concentration data is perturbed with random measurement error with maximum specified deviation of 10% as shown in Eq. (5.17).

$$^{pert}C^{k}_{est_{iob}} = C^{k}_{est_{iob}}(i + err) \tag{5.18}$$

$$err = \mu_{per} \times rand \tag{5.19}$$

Where ${}^{pert}C^{k}_{est_{iob}}$ is the perturbed numerically simulated concentration values; $C^{k}_{est_{iob}}$ is the numerically simulated concentration value; err is the error term; μ_{per} is the maximum deviation expressed as a percentage; and rand is a random fraction between −1 and +1 generated using Latin hypercube distribution.

To illustrate the efficiency of singularity index technique on the source identification model, the linked optimization algorithm is used with three different monitoring networks. These three monitoring networks design scenarios evaluated are: FSMT-arbitrary networks, FSMT-designed optimal monitoring networks, and optimized monitoring networks without utilizing FSMT for potential well locations. Also, the contaminant source estimates obtained using the totally arbitrary (without any FSMT information) monitoring network is compared with those obtained using other monitoring networks. The linked simulation-optimization model is solved using concentration measurement data from a FSMT-arbitrary monitoring networks and an arbitrary monitoring network. A total of fifteen monitoring wells are selected for each of these two networks. The arbitrary monitoring network is used in the source identification model. In these scenarios, source concentration is estimated using error-free data and then estimated using perturbed concentration measurement data perturbed with random error term.

5.4. Results and Discussion

The fractal singularity index method was used to delineate the contamination plume in a highly nonlinear and complex scenarios such as illustrative mine site aquifer. Choosing the potential well locations based on the singularity index map increase the efficiency of the monitoring network in terms of source identification, especially for multiple

distributed source characterization purposes. To evaluate the performance of the methodology using FSMT, the source characterization results from different monitoring networks are compared for different scenarios with and without utilizing FSMT, and also compared with the scenario of arbitrary monitoring network with and without using FSMT. The two-objective optimal monitoring network design model is solved and the first objective function values (F1) (Equation 5.6) are plotted against the minimum satisfaction level of the second objective function value (F2) (Eq. 5.7) as shown in Figure 5.7. Two sets of ten Pareto-optimal networks (MNSI1-MNSI10 and MN1 to MN10) are chosen for different values of objective function F2 (Table 5.3). To design the MNSI1 to MNSI10 monitoring networks, the singularity index technique was used to select the potential well locations, while arbitrary potential well locations were utilized to design the MN1 to MN10 monitoring locations optimally. The non-inferior solutions show the conflicting nature of the two objective functions. The y axis represents the value of the objective function F1 and the x axis represents value of objective function F2. It is seen that as the satisfaction value of the F2 (objective function two) decreases the satisfaction value of F1 (objective function one) increases and vice versa. This essentially shows the conflicting nature of the two objective functions and their trade-offs. Figure 5.7 shows the actual Pareto optimal solutions. It shows that for a particular obtained value of objective function F1 (minimization), it is possible to achieve a larger value of F2 (maximization) while utilizing the FSMT. Therefore, based on this criterion, the Pareto solutions are more efficient when the singularity index map is utilized.

Lower values of the first objective function F1 decreases the kriging interpolation error based on monitoring network and increases the likelihood of choosing optimal monitoring locations more efficient in estimation of distributed contaminant sources

throughout the aquifer. Therefore, it increases the possibility of more accurate source identification by tracking the plume contours. A large value for the second objective function F2 increases the likelihood of monitoring well locations where the concentration gradient is higher than other potential well locations. Using the results of the multi-objective optimization increases the accuracy of source characterization in complex and highly nonlinear aquifers where overlapping of different pollutant plumes resulting from separate distributed sources can occur. The FSMT provides an approach to take into consideration the plume boundaries for locating more efficient monitoring wells for improved contaminant source identification.

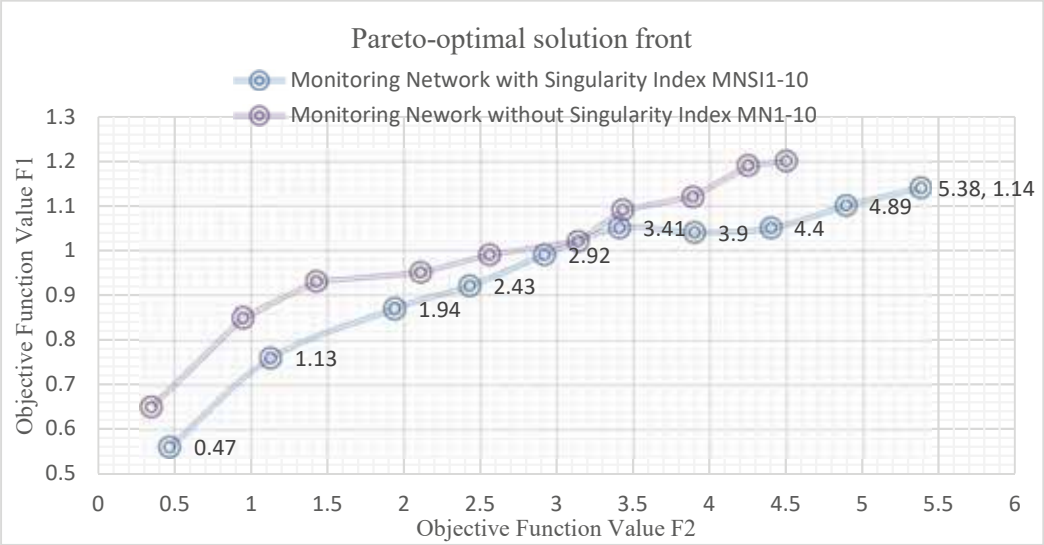


Figure 5.7 Pareto-optimal monitoring networks with two different monitoring well locations with and without utilizing singularity index technique

Linked optimization based algorithm is used to obtain the distributed source characterization in terms of magnitude and locations. The source characterization results for all 20 Pareto-optimal monitoring networks with and without using singularity index (MNSI1- MNSI10 and MN1-MN10) are compared. Figures 5.8(a) and 5.8(b) illustrate

these comparisons using error free, and perturbed-measurement data, respectively, for monitoring networks obtained by using the singularity mapping method. Likewise the comparison results using error free and perturbed–measurement data for arbitrary potential-wells monitoring networks are showed in Figures 5.9(a) and 5.9(b) respectively. The x axis shows the unknown potential contaminant sources (pit, Airfield, Mundic, Shepherd Tailing Dump, Shepherd down, Mundic East, B Dump, k Dump, Horse paddock and No.2 Mill Tailing). The bar graphs corresponding to an unknown source concentration shows the actual concentration value and estimated values using multi objective optimal monitoring networks. In both Pareto-optimal networks, the results of source flux identification for all the 20 Pareto-optimal monitoring network (MNSI1 to MNSI10 and MN1 to MN10) designs are very close to the actual concentration when solved using error-free measurement data. However, when the concentration measurement data is perturbed with random errors to incorporate measurement errors, the results of source flux identification show a greater amount of deviation from the actual flux values in all 20 Pareto-optimal monitoring networks (MNSI1 to MNSI10 and MN1 to MN10) when compared to the results with error-free data.

To select one of the network designs from the Pareto-optimal solution, different criteria should be considered. However, by increasing the satisfaction of one objective function, the other objective function should be relatively scarified. Figure 5.7 shows that in the MNSI optimization line, the F2 increases from MNSI6 to MNSI8 while the low level of F1 is scarified. The marginal F2 improvement is much more than sacrifice level of F1. Therefore, MNSI8 could be a good choice for optimal monitoring network design from this Pareto-optimal solution. In the same way, the F2 increases from MN5 to MN6 effectively while the corresponding decrease in F1 is negligible. Therefore, it can be

argued that without additional preference ordering MN6 can be chosen as an optimal monitoring network in this set.

To investigate the efficiency of the selected monitoring network out of the two sets of 10 Pareto-optimal monitoring networks (MNSI1 to MNSI10 and MN1 to MN10), absolute difference between actual source flux and the estimated source flux for all two sets of 10 Pareto-optimal monitoring networks is calculated. This is possible, because, for this performance evaluation purpose, the measurements are synthetic and the actual errors can be computed, unlike when field measurement of concentrations are used. Figures 5.10(a) and 5.10(b) show the absolute difference between actual source fluxes and the estimated source fluxes using the singularity mapping method. Also Figures 5.11(a) and 5.11(b) show the absolute difference between actual source fluxes and the estimated source fluxes for scenarios without using FSMT to select the potential well locations. Absolute difference of actual source flux and the estimated source flux for all 20 Pareto-optimal monitoring networks (MNSI1 to MNSI10 and MN1 to MN10) using error-free and perturbed-measurement data show similar trends. The average of absolute error in both error free and erroneous measurement data confirm that the choice of the monitoring network 8 (MNSI8) and monitoring network 6 (MN6) can be optimal. Further increase in the value of F_2 increases the average absolute difference between actual source flux and the estimated source flux.

These results show that for effective and efficient characterization of unknown contaminant source concentrations it is important to have monitoring networks ensuring balance between low resulting interpolation errors and large concentration gradient at the same time. These evaluations show that Pareto-optimal monitoring networks MNSI8 and MN6 results in minimum average absolute difference between the actual source

concentration and the estimated source concentration in each Pareto-optimal monitoring networks, both in case of error-free data and erroneous measurement data. However, more accurate source identification results and lower absolute error in both error free and erroneous measured data are achieved using singularity mapping method.

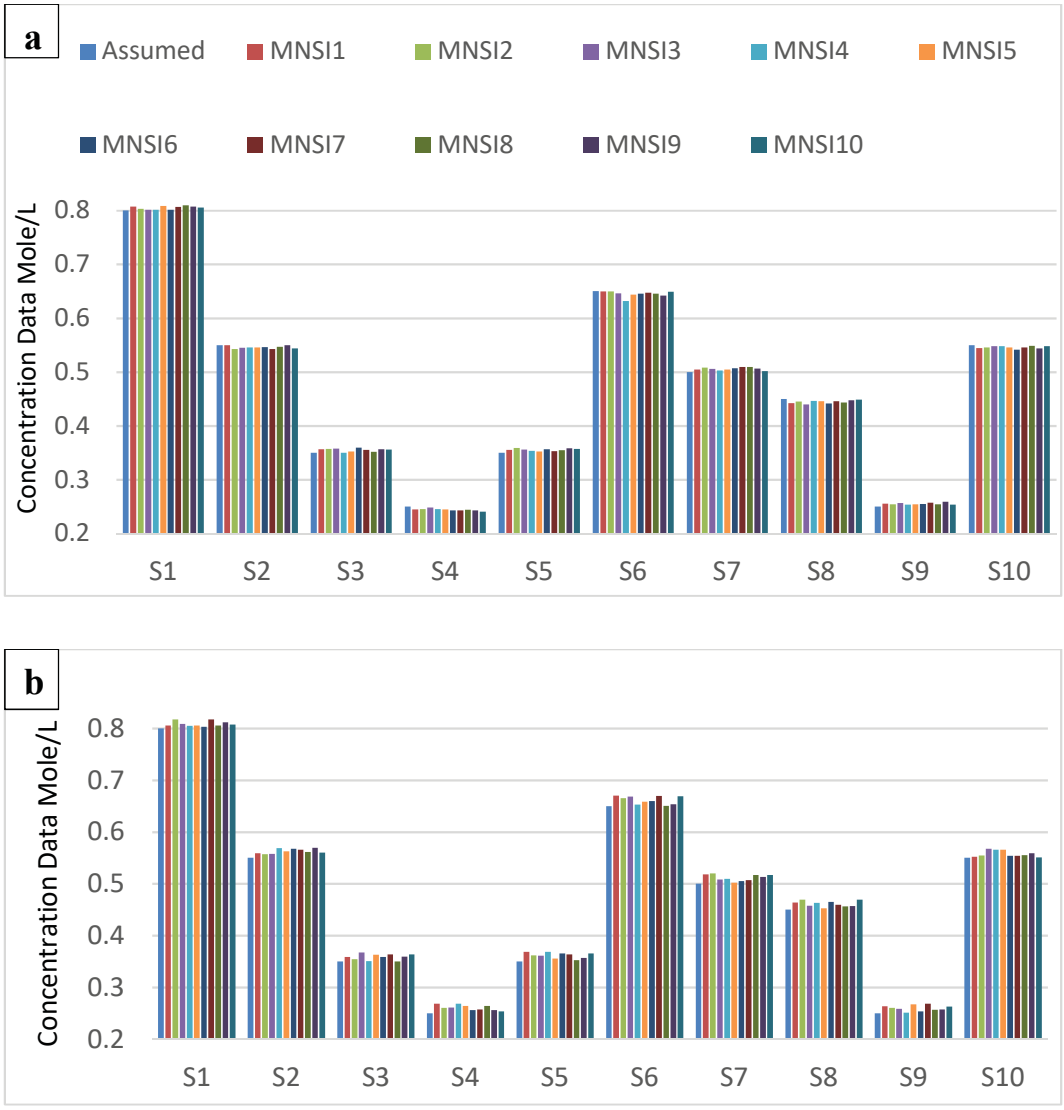


Figure 5.8 Source identification results using error-free measurement data from Pareto-optimal monitoring networks, a) using FSMT; b) without using FSMT

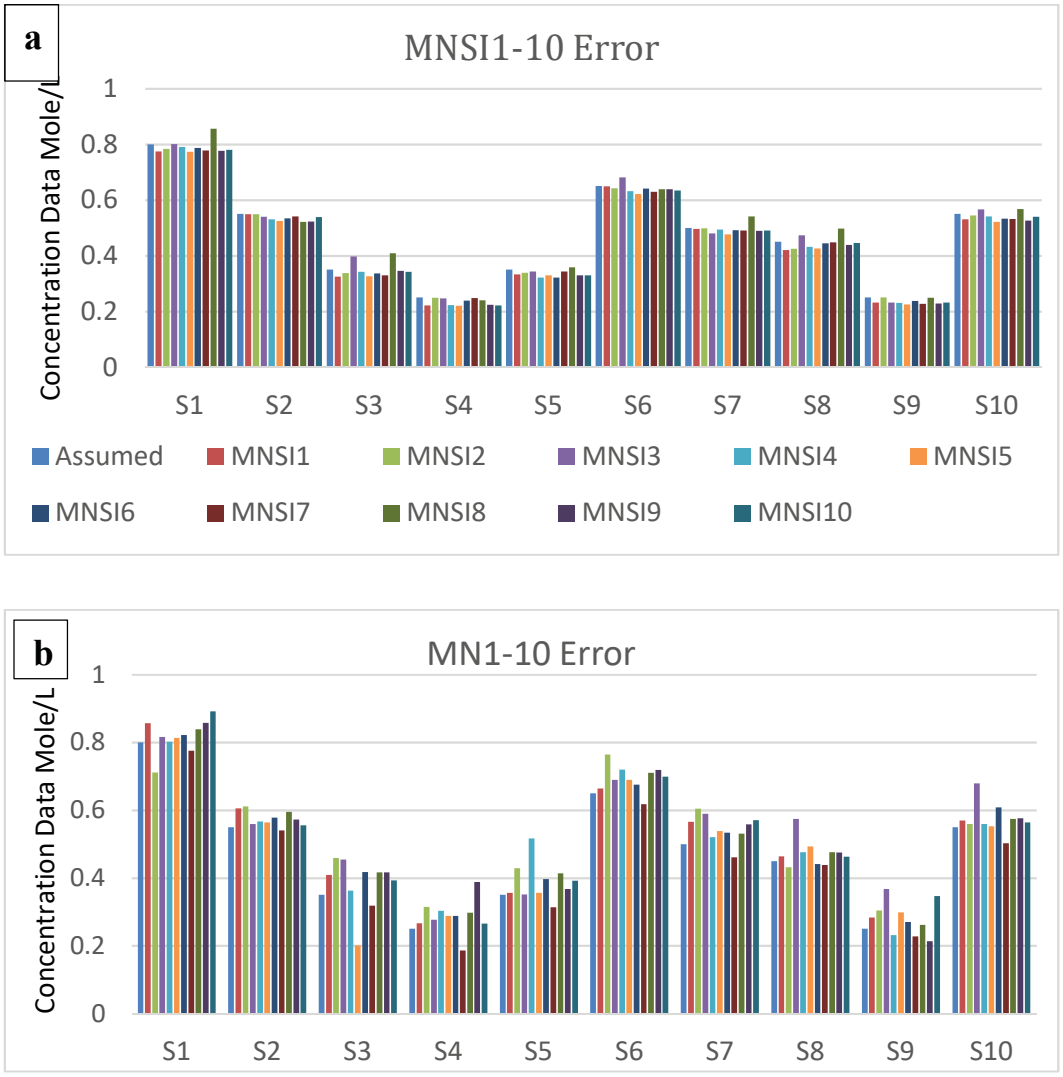


Figure 5.9 Source identification results using erroneous measurement data from Pareto-optimal monitoring networks, a) using FSMT; b) without using FSMT

To evaluate the effects of using singularity mapping index an arbitrary monitoring network is selected using the singularity mapping method. The results of source identification using concentration measurements from arbitrary monitoring networks (AR), arbitrary monitoring networks using FSMT, monitoring network 8 (MNSI8) and

monitoring network 7 (MN5) with error-free data and erroneous data are shown in Figures 5.12(a) and 5.12(b), respectively.

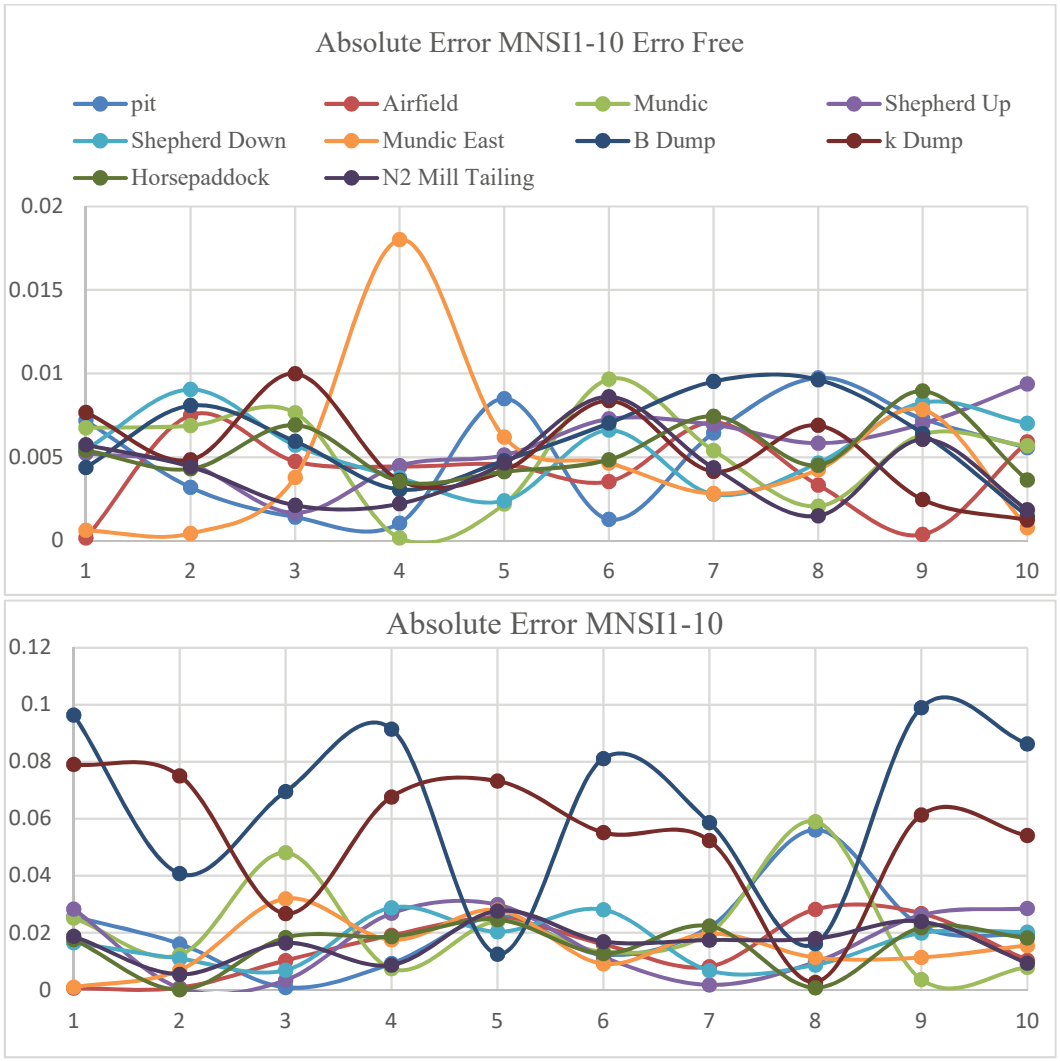


Figure 5.10 Absolute difference between actual fluxes and estimated fluxes using FSMT; a) error-free measurement data; b) erroneous measurement data, from Pareto-optimal monitoring networks

The estimated source concentration values for each scenario, based on Pareto-optimal monitoring networks with FSMT (MNSI1-MNSI10) and without FSMT (MN1-MN10) are averaged, and compared with the actual source concentrations, as well as with

estimated source concentration values obtained using monitoring networks MNSI8 and MN6. It is seen that the results of source concentration estimates when using singularity index is more accurate than using optimization algorithm with arbitrary potential well locations, without using FSMT for all distributed sources except Mundic East, B Dump and No.2 Mill Tailing.

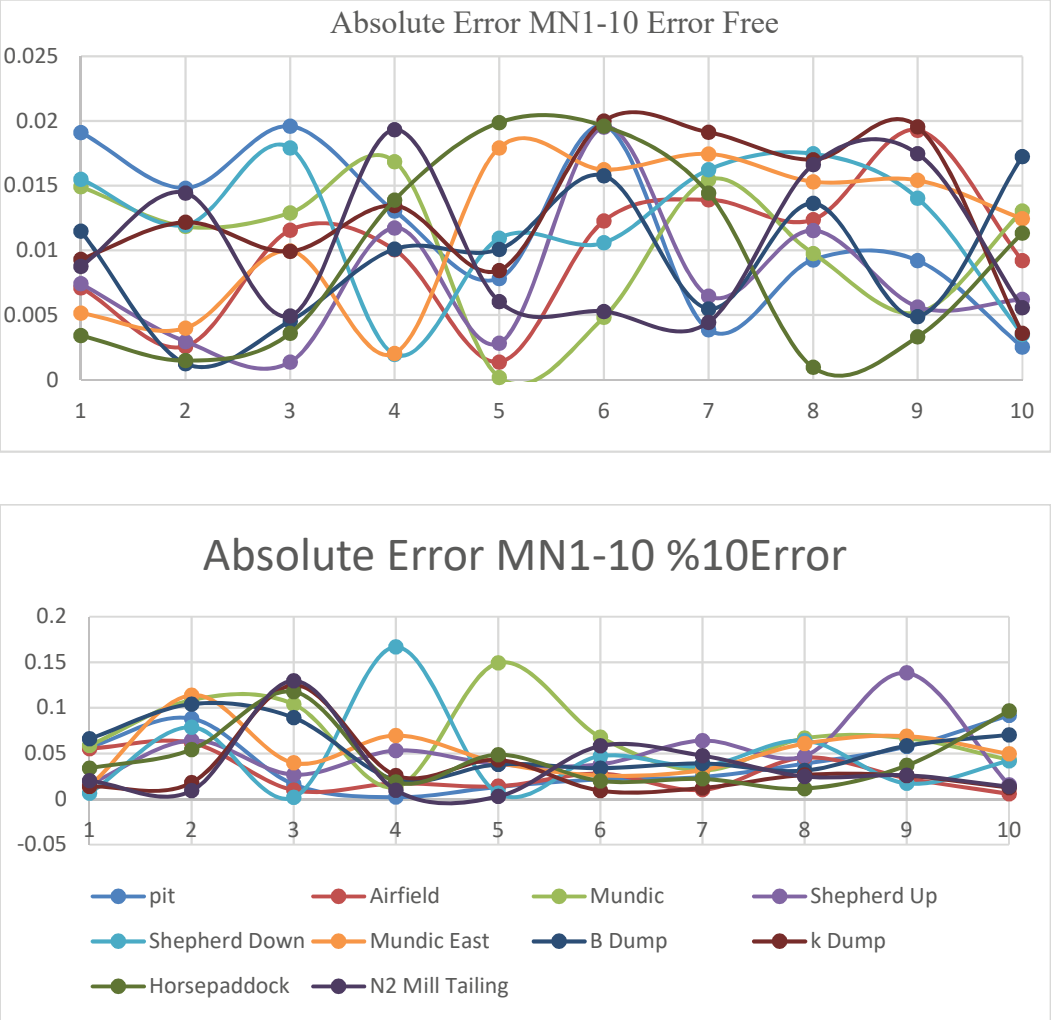


Figure 5.11 Absolute difference between actual fluxes and estimated fluxes without using FSMT; a) error-free measurement data; b) erroneous measurement data, from Pareto-optimal monitoring networks

The absolute difference between the actual flux and the estimated flux for the arbitrary network, FSMT-arbitrary network, MNSI8 network, MNSIs networks averaged, MN6 network and MNs networks averaged are calculated and compared. These comparisons using error-free and erroneous measurement data are shown in Figures 5.13(a) and 5.13(b), respectively. It is seen that the averaged absolute difference between the actual source fluxes and the estimated fluxes obtained using the arbitrary networks is larger compared to the other monitoring networks.

Comparing the results of source concentration estimation using concentration measurements from an optimally designed monitoring network, an FSMT-arbitrary network, and an arbitrary network shows it is more efficient to use a designed network or a network arbitrarily based on FSMT as compared to a completely arbitrary monitoring network When the identification results are compared, then the utility of FSMT designed monitoring network becomes clear. Such designed monitoring network is even more important in the case of aquifers polluted by multiple distributed sources. Placing monitoring wells arbitrarily in such a situation may obtain good results in case of some source fluxes, but the estimates may be very poor in complex aquifers with overlapping multiple distributed sources.

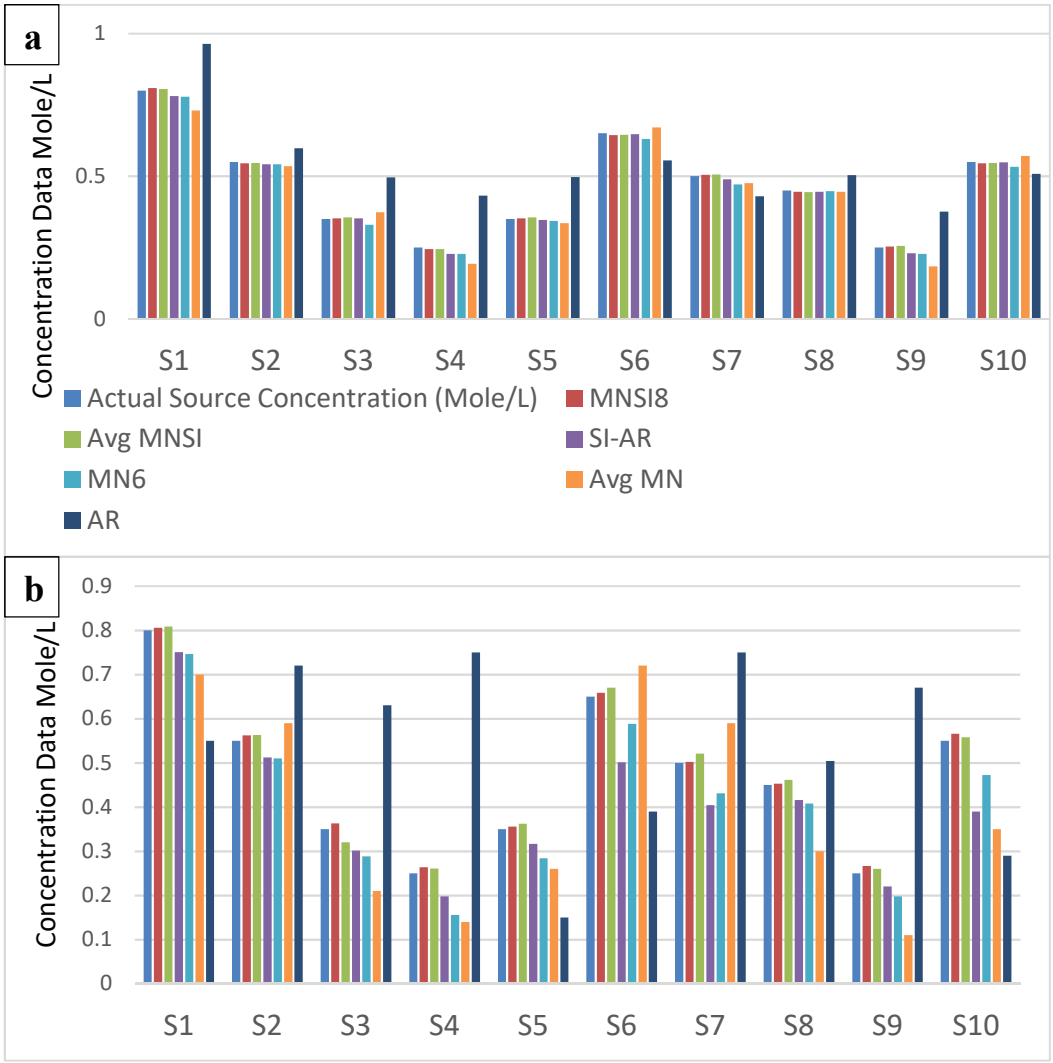


Figure 5.12 Comparison of source identification results using a) error-free measurement data; b) erroneous measurement data

Figure 5.13 Comparison of averaged absolute error for arbitrary networks versus absolute error; a) error-free measurement data; b) erroneous measurement data

5.5. Conclusions

Unknown groundwater contaminant source characterization models utilize spatiotemporal pollutant concentration measurement data for identifying the source characteristics. The spatiotemporal location of pollutant concentration measurements is an important factor which affects the accuracy of solution results. Not all monitoring wells are ideally located for accurate identification of source characteristics. However, in all real life scenarios, the

number of such spatiotemporal locations at which pollutant concentrations are measured is limited due to budgetary constraints. To address this issue of designing an effective and efficient monitoring network to characterize contaminant sources in complex contaminated aquifers, in terms of concentration and location, two techniques have been proposed and applied for designing an optimal monitoring network. It has been shown using limited performance evaluation for an illustrative mine site study area, that the source characterization results can be improved effectively using measured data from these monitoring locations. Two techniques are utilized to obtain the monitoring networks by incorporating fractal Singularity Index Mapping and multi-objective Pareto-optimal monitoring network designs.

FSMT has been widely used in many scientific areas, but very few studies applied this method for delineating the boundary of contamination plumes. In this paper, the window based singularity mapping technique is utilized in an illustrative contaminated site. The singularity indices are used to divide the study area into contaminated and clean characterization by assigning singularity indices smaller and larger than two, respectively. Therefore, contours corresponding to singularity index two are used to delineate the contamination plume boundaries. The contours of singularity index two are suitable for guiding the selection of the potential well locations.

The developed methodology based on a multi-objective optimization algorithm appears to result in efficient identification of unknown groundwater pollution source concentrations. However, the source characterization results are more accurate and precise while incorporating the FSMT to choose the potential well locations for optimal design of the monitoring network. The solution results in the illustrative example problem show that the accuracy of source flux identification varies when using pollutant concentration

measurement data from different monitoring locations. The designed optimal monitoring network results in better source identification compared to those obtained by using an arbitrary network based on FSMT, both with error-free and erroneous measurement data.

In all real-field scenarios of source characterization, the degree of uncertainty in terms of source locations and hydrogeologic properties of the aquifer are high. Large amount of measurement data can decrease some of these uncertainties. The number of wells in monitoring network may be restricted by various constraints. The proposed methodology is shown to improve the source identification efficiency as compared to arbitrary measurements. The designed monitoring locations are important in obtaining accurate unknown source identification. FSMT is demonstrated to provide a useful and effective guideline to select the potential locations for monitoring network design. In addition, the results show that there is a trade-off in selecting the optimal monitoring locations in terms of minimizing the maximum normalized error between assumed and interpolated concentration values, and maximizing the sum of the estimated concentration gradient values. However, the ideal levels of trade-off needs to be studied and may depend on site-specific conditions. This study also demonstrated the feasibility and applicability of using surrogate models based linked simulation optimization models together with fractal singularity index mapping technique, and Pareto optimal monitoring network design for effective characterization of unknown contaminant sources in a complex contaminated aquifer with a reactive geochemical transport process.

The source characterization under complex and uncertain conditions, in a contaminated aquifer has been shown to be further improved if a sequential preliminary source characterization and subsequent design of a monitoring network is utilized, for a few iterations. This approach should be particularly useful for contaminated aquifer sites with

very little initial measurement data. Such an iterative and sequential procedure is discussed in the next chapter.

6. Integrated sequential groundwater contaminant source characterization and Pareto-optimal monitoring network design application for a contaminated aquifer site

6.1. Introduction

This chapter presents an application of developed methodologies to a real life contaminated aquifer. The source characterization and optimal monitoring network design methodologies described in previous chapters are used sequentially for a contaminated aquifer site located in New South Wales, Australia. The location details of this site are not disclosed because of confidentiality requirements.

In the following sections, first, a review of the problem and the site properties is presented. Then, the implemented flow and transport simulation models are explained briefly. Finally, the application results of the integrated sequential source identification and monitoring network design which are discussed.

6.2. Background and the Study Area Description

The contaminated aquifer is located in Macquarie Groundwater Area in New South Wales, Australia. Some reports mentioned the sign of BTEX in the subsurface water in this area, and the polluted region was investigated using the observation data from the

monitoring groundwater wells. However, it did not indicate the first record of the BTEX pollution in this area. The highest BTEX concentration of 320 mg/l was reported in October 2009 in one monitoring well in this area. The affected area was implicitly estimated as over 1 Km^2. Between October 2006 and July 2011, the groundwater level and the contaminant concentration were measured and recorded with seventy-four monitoring wells. These wells were installed and utilized at various times during the monitoring period to realize the pollutant plume and pollutant transport process in the study area.

It was suspected and later determined that the contamination source was an underground leaking petrol tank at a service station. The aim of the methodology evaluation for this illustrative contaminated aquifer site is to test the feasibility of accurately characterizing the contaminant source in terms of location and release history, utilizing a sequential and integrated source characterization and monitoring network design for concentration measurements. For evaluation purposes, the potential source locations included a dummy potential source location, so that it is also determined if the actual location is identified using this methodology. The performance evaluation was based on the assumption that the best estimate of the source characteristics are those obtained using the extensive concentration measurements available for this site. The deviation between these best source characterization estimated and the source characteristics obtained using a much smaller number of monitoring locations based on the sequential optimal design procedure is considered as a measure of the accuracy of the results. This approach was necessary, as the source was detected many years after it become active, and initially the location was also unknown. Therefore, comparison of the source characterization solution results and

actual source characteristics was not possible for a real life site like this, and the indirect method of evaluation was adopted.

In this site, most of the existing wells were installed close to the potential source, and only a few were located to regions further away from the source. However, the time and location of measured observation data were selected arbitrarily. It seems based on the previous investigations, the leaking underground storage tank at a gas station was the potential source of the pollution. However, source characteristics regarding starting time of leakage, the magnitude of flux and the time history of flux releasing out of the source were not specified in the investigation reports.

The aim of this study is to characterize the unknown contaminant source(s) characteristics in terms of source location, starting time and flux release history of any potential source in the polluted aquifer. A source characterization model with unknown source starting time is integrated with a fractal multi-objective optimal monitoring network design model for estimating the unknown source characteristics. The contaminant measured data are available at almost all of the seventy-four monitoring wells every three months during the investigation period (October 2006 to July 2011). However, there are some limitations in collected observation data including: concentration data are not available for all wells and all times, some wells installed later during the investigation period, some wells installed far from the sources and these data are considered less relevant. The polluted aquifer site is a small part of the Upper Macquarie Groundwater Management Area. 292m AHD is the starting ground elevation at East, decreasing until elevation reaches the 251m AHD at West. The aquifer is mainly recharged from the rainfall and the river. The majority recharge comes from the Macquarie River, and minority comes from precipitation (average 583 mm/year) in the wet season, from November to February.

The aquifer water in the study area is mainly extracted for potable water and irrigation usage through pumping wells. There have been substantial variations in the range of pumping rates due to changes in groundwater usage policy from the available aquifer, and also a voluntary limitation of extraction pumping rate from 2010 (Manderson 2011). Another source of losing the water from the aquifer is through evapotranspiration, which peaks to 260mm/month during the dry season (Puech 2010).

The primary goal of this chapter is the integrated source characterisation with an optimal number of spatiotemporal concentration measurement data using Pareto-optimal monitoring network design in sequence. The developed source characterization methodology and monitoring network design are explained in detail in Chapters 4 and 5, respectively.

6.3. Groundwater Flow and Contaminant Transport Simulation Models

The numerical groundwater flow and transport simulation models utilized for flow and transport processes are MODFLOW (Zheng et al. 2001) and MT3DMS (Zheng and Wang 1999), respectively. The aquifer parameters calibrated earlier for the flow process and the flow model developed by Prakash and Datta (2015) are utilized in this study. The utilized aquifer properties and other details in the flow and transport simulation models are described briefly in the next two sections. More details about the simulation models used are described by Prakash and Datta (2015).

6.3.1. Groundwater Flow Modelling of the Contaminated Aquifer

There were no easily detectable natural boundary conditions around the contaminated study area. Therefore, a bigger site area with natural boundary conditions was used for flow simulation and area was designated as the extended study area. This study area has dimensions of 2.187 Km by 2.426 Km, while the dimensions of the modelled smaller contaminated area are 608 m by 864 m, which is called the specified area. The groundwater flow in the aquifer study area is modeled as an unconfined aquifer. The natural boundary condition of the study area at West is the specified constant head condition based on average stage at the Macquarie River (Figure 6.1). Similarly, the ground topography is laid down toward the river in the West. Other boundaries are specified as constant head boundary conditions. The hydrogeologic properties and the boundary conditions used in Prakash and Datta (2015) for the flow model are the same as those used in modeling groundwater flow in the entire Upper Macquarie Groundwater Management Area, developed by Puech (2010).

The groundwater flow process is modeled for 18 years with 18 one-year time steps from 1 January 1995 until 31 December 2012. In the three-dimensional simulation models, the study area is discretized into small grids of size 21.87m by 21.08m in the x and y directions respectively. The soil layers thickness are variable at different places; therefore, the grid size in the z direction is different and matches with the layer thickness. The hydrogeological properties, such as hydraulic conductivity, porosity, specific storage and specific yield, were obtained from previous studies conducted in this study area by Prakash and Datta (2015). These hydrogeological properties are listed in Table 6.1.

The calibration of the flow model was achieved using observed measurement data from the all well locations. The five years calibration period started from October 2006 to July 2011. One metre tolerance from the observed head data with 90 percent confidence was the calibration targets. The satisfaction level of calibration process was obtained with boundary conditions adjustment.

Once the extended study area is modeled and calibrated, the flow model parameters used for the specified area are derived from the calibrated extended model. The GMS7.0 feature, Regional to Local, is used to interpolate the starting head and layer thickness values for the specified area from the extended study area model. Figure 6.2 shows the specified area where the pollutant is estimated to be present. The grid sizes are refined further in the flow model for the specified area, and the area is discretised into 75 rows, 50 columns, and three layers. All of the boundaries are considered as time-varying specified head boundary conditions. The value of the time varying specified heads at the edge of the specified area are extrapolated from the calibrated model for the extended study area. All of the other hydrogeological flow parameters are kept the same as in Table 6.1.

Table 6.1 Hydrogeological Properties used in Flow Modelling of the extended Study Area

Parameter	Unit	Value
Maximum length of study area	m	2187.1
Maximum width of study area	m	2425.6
Saturated thickness, *b*	m	Variable

Number of layers in z-direction		3
Grid spacing in x-direction, Δx	m	21.87
Grid spacing in y-direction, Δy	m	21.08
Grid spacing in z-direction, Δz	m	Variable
Kxx (Layer 1, Layer 2, Layer 3)	m/d	12.37, 16.24, 0.001
Kyy (All Layers)	m/d	0.2
θ (All Layers)	dimensionless	0.27
Longitudinal Dispersivity, αL	m/d	12
Transverse Dispersivity, αT	m/d	6
Horizontal Anisotropy	dimensionless	1.5
Specific Yield *Sy* (All Layers)	dimensionless	0.1
Specific Storage *Ss* (All Layers)	dimensionless	0.000006
Initial pollutant concentration	g/l	0.00

6.3.2. Pollutant Transport Simulation in the Impacted Area

MT3DMS, a three-dimensional numerical transient transport simulation model was developed to study the fate and transport of the petrochemical pollutant BTEX originating from potential point sources. For the purpose of implementation, the pollutant is assumed to be conservative in nature, and the pollutant plume boundary is assumed to be contained within the boundary of the impacted area. The transport simulation model predicts the movement of the contaminants using the results of the flow simulation model in the impacted area of the aquifer, over time. The initial concentration of BTEX is assumed to

be zero at the first time step in the study area. All the other relevant transport parameters used in the transport model are shown in Table 6.1.

Figure 6.1 Plan Views of the Study Area and the Impacted Area

6.4. Sequential Integrated Model

The proposed methodology integrates the source identification model (Chapter 4) and the multi-objectives monitoring network design model (Chapter 5) by sequentially solving for the source characterization followed by monitoring network design. This sequence may be followed for a number of iterations. The monitoring network design is to be implemented for new concentration measurement to be utilized for another sequence of source characterization. This methodology evaluation is performed for the specified study area for contamination transport modelling. Although no actual measurements are

conducted at the designed optimal monitoring locations, the available set of extensive concentration measurement data at already existing monitoring locations are used as the actual measurements. This is possible only because the extensive concentration measurement data is collected over a number of time periods, and also, the potential monitoring locations specified in the monitoring network design model coincides with some of the existing monitoring locations.

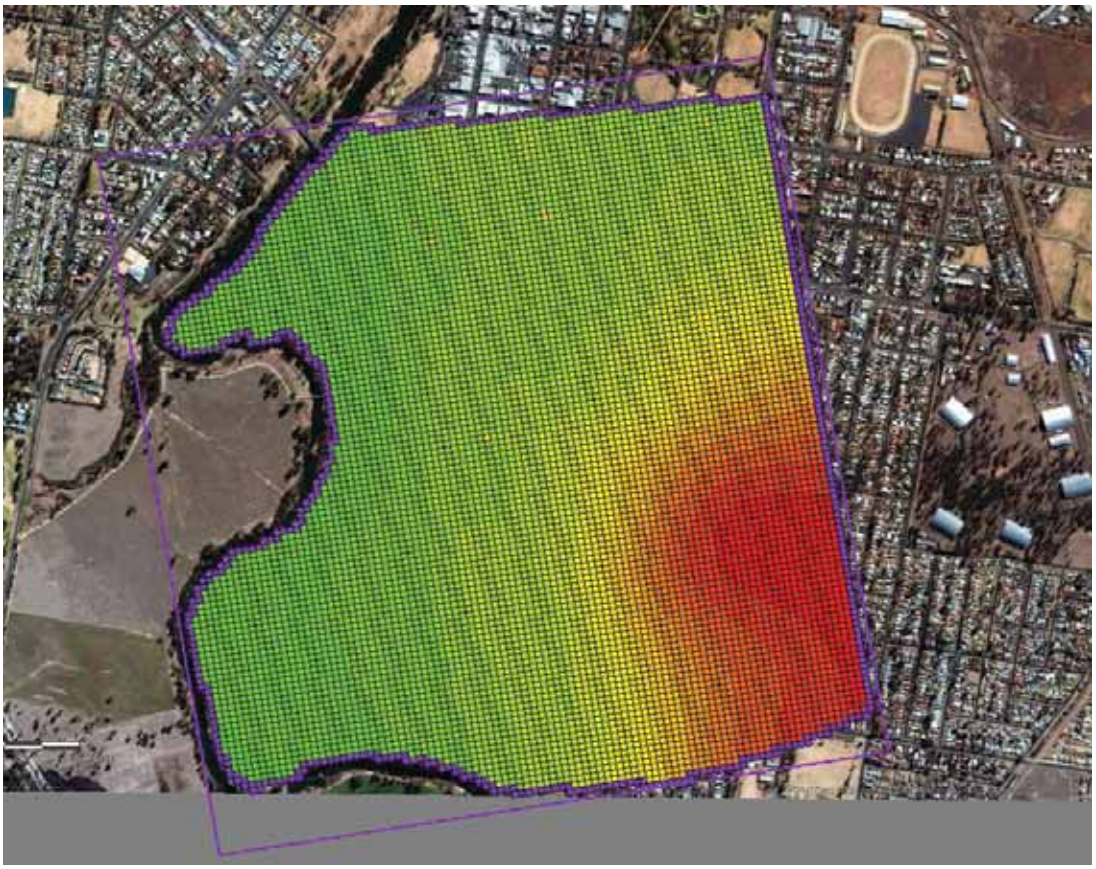


Figure 6.2 Plan view of the extended Study Area (Prakash and Datta 2015)

The unknown groundwater source characterization model uses the observed measurement data at arbitrary well locations to estimate the initial source characteristics. Next, the optimal monitoring network design utilizes the source identification results from the

previous step and determines the new monitoring well locations. The linked optimization source characterization methodology is solved to identify the source characteristics using the observation data from the new monitoring wells in addition to the measured data from the existing monitoring wells. The developed integrated methodology need to be used iteratively until reasonable accuracy of source characterization is obtained. The results of the source characterization model are utilized for obtaining new monitoring well locations using the Pareto-optimal monitoring network design for each sequence.

The following steps present the new optimal monitoring network design process. The first step involves the estimation of spatial concentration values utilizing available concentration data and spatial extrapolation methods. In the illustrative application of the methodology, the initial spatial concentration measurements were simulated using a contaminant transport simulation model. The transport simulation model provides the estimates of contaminant concentration throughout the aquifer using the initial source characterization results. Second, the Fractal Singularity Mapping Technique (FSMT) provides the plume boundary as a guideline for selecting the potential monitoring wells. Section 5.4 shows the efficiency of using FSMT for monitoring network design. Third, the Pareto-optimal two-objectives monitoring network design is implemented for collecting new concentration measurement data (Section 5.3.2). In the next sampling time step, pollutant concentration measurements from these newly implemented monitoring wells and data at already existing monitoring wells are obtained. Subsequently, the pollutant concentration measurements from the current sampling time step and previous sampling time steps are utilized for source identification. Figure 6.3 shows the schematic diagram of the proposed methodology.

6.5. Performance Evaluation of the Applied Methodologies

The groundwater unknown source characterization uses the calibrated contaminated transport simulation model. The unknown source identification methodology developed earlier (chapter 4) is utilized for recreating the source flux release history and the source activity initiation time. To increase the accuracy and efficiency of source identification, the Pareto-optimal monitoring network design methodology to identify the source characteristics is developed using fractal singularity mapping technique (FSMT) (Chapter 5). The integrated monitoring network design and source identification sequentially, increases the efficiency of source characterisation in real life scenarios with sparse spatiotemporal measurement data. The aim of this study is to identify the unknown source characteristics in terms of flux magnitude, contaminant flux release history, and activity starting time. The location of the contaminant source is explicitly known for evaluating the methodology, but unknown to the source characterization model. For evaluating the efficiency of developed source characterization methodology, two possible source locations are assumed as unknown sources. One of these should be identified as a non-active or non-existent source. The two potential sources in the study area are shown in Figure 6.4. The points marked in red circles are the grid locations containing the possible sources, and the yellow points are the observation wells where the concentration of BTEX is observed. A total of seventy-four concentration measurement locations are present in the study area.

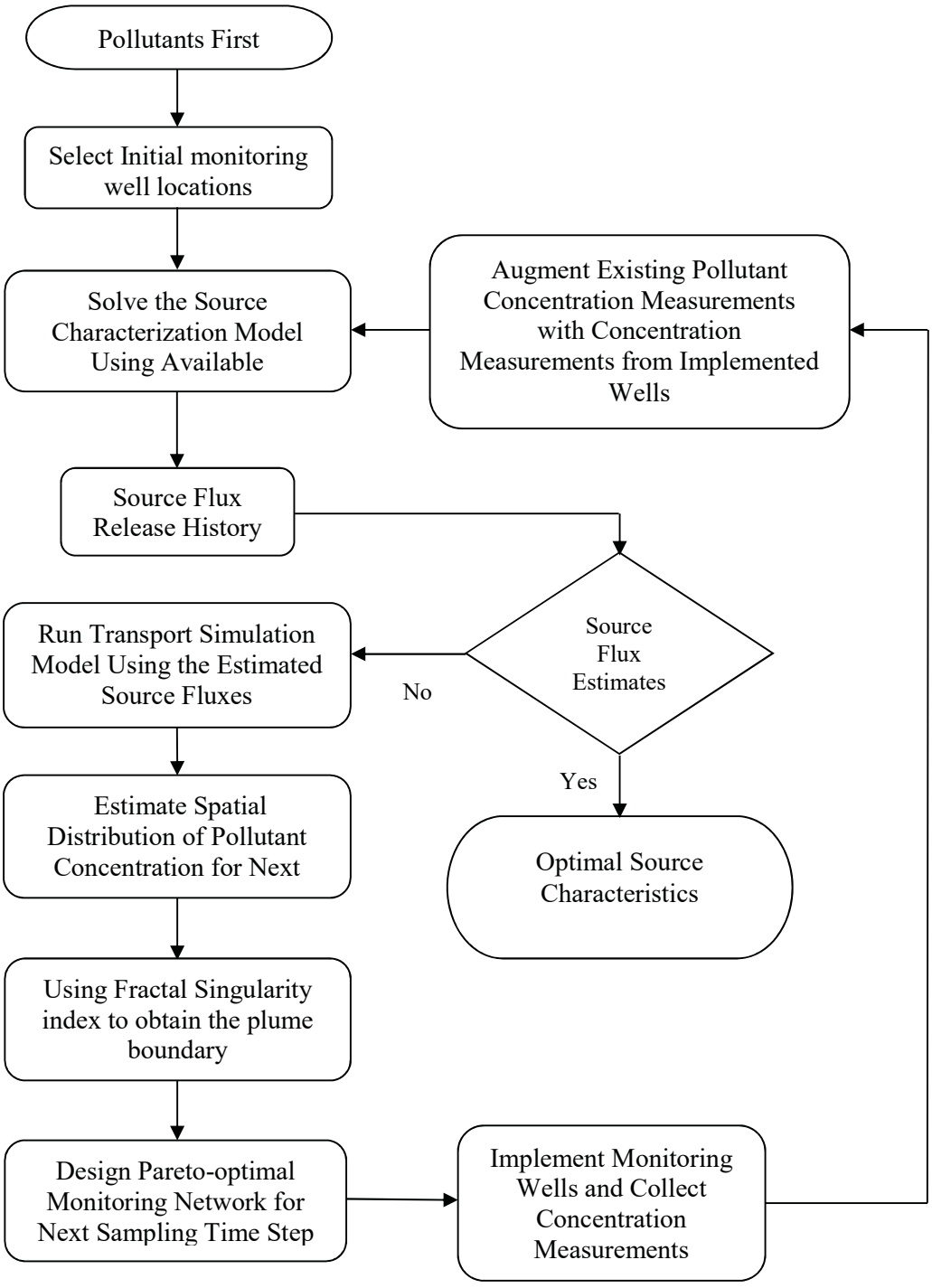


Figure 6.3 Schematic diagram of sequential source characterization and monitoring network design methodology

Figure 6.4 Plan view of the specified area. (Red circle: Potential sources; Yellow Square: Monitoring wells)

6.5.1. Simultaneous Source Flux Release History and Source Activity Initiation Time Identification

ASA optimization linked source characterization methodology is used to identify the source flux magnitude. The potential contaminant sources are named source 1 and source 2 which are located at cell (1, 17, 29), and (1, 16, 24), respectively. The simulation model starts from 1 January 1995. However, the unknown starting time of the source activity can be anywhere between 1 January 1995 and 31 December 2011. Ten equal stress periods (1 year each) cover the ten years activity duration of the sources. The pollutant flux from each of the sources is represented as Si,j (i=1,2, j=1,2,..,10) where i represents the source number, and j accounts for the stress period number. In this case, S1 is the actual source and S2 is the unreal source. Other assumptions of source identification include the contaminant source flux is constant over each stress period, and the contaminant releases of the both sources S1 and S2 starts at the same time. To realize the

starting time of the sources an additional time lag variable ΔT is introduced in the optimization linked program (Prakash and Datta 2015). In source characterization model, observed measurement data are used from 22 January 2009 and continued every three months.

6.5.2. Sequential Optimal Monitoring Network for Efficient Source Characterization

To increase the efficiency of source identification, the Pareto-optimal monitoring network design is integrated with the linked simulation-optimization source identification model. The source identification methodology regarding flux magnitude and source activity starting time is started using the observed measurement data from three arbitrary wells. These wells are randomly selected as a subset of all available monitoring wells within the study area. At each sequence of model running, three new monitoring wells are chosen for the next monitoring time step. The available installed observation wells in the specified area are considered as the potential monitoring wells. The preliminary source characterization model is solved using concentration data measured on 22 January 2009 from the first three arbitrary wells.

The source characterization solution results are utilized in the transport simulation model as inputs to predict the BTEX concentration at the next sampling time step (30 April 2009) in the specified area. The plume concentration in the specified area on 30 April 2009 is utilized to obtain the next monitoring wells locations. The two-steps optimal monitoring network design procedure (Chapter 5) is applied to locate the next three monitoring well locations.

Firstly, FSMT is used to compute the singularity indices based on the initially estimated source fluxes at sources and corresponding aquifer response at the specified area. FSMT is applied to estimate the likely contamination plume boundary which is then used as one of the guidelines for locating potential monitoring well locations. All installed monitoring wells close to the plume boundary are suitable potential monitoring well locations for the monitoring network. These wells are selected out of all available monitoring wells for the next step. In the second phase, a multi-objective optimization methodology is used for Pareto-optimal monitoring network design, with constraints on the total number of new monitoring wells to be installed. In this performance evaluation scenario, the total number of new monitoring wells to be optimally chosen were restricted to three locations out of the potential monitoring well locations specified. The Pareto-optimal solutions obtained from the two-objective model is used to design a set of Pareto-optimal monitoring networks. The monitoring network is chosen based on the following two objective: (1) maximize the summation of the product of estimated concentration gradients, and the simulated intensity at that location, and (2) minimize the maximum normalized error between actual concentration and those estimated with the kriging interpolation models, based on monitoring data from designed monitoring locations.

Once a new monitoring network is designed and implemented, the concentration measurements from all wells in the monitoring network obtained on 30 April 2009, together with measurements obtained from the pre-existing three arbitrarily chosen wells on 22 January 2009, are utilized in the source identification model. This approach of sequential source characterization and monitoring network design and implementation is repeated for the subsequent sampling time steps, until changes in the source flux and starting time estimates are negligible.

6.6. Results and Discussion

Well M03, M14 and M20 are selected as the three arbitrary initial monitoring wells, located at (1, 21, 25), (1, 24, 25), and (1, 20, 29), respectively. After every sequence of monitoring wells implementation, the source characterization results are presented. To reflect the real-life conditions, measurement errors are incorporated with the concentration measurement data. The concentration measurement data are perturbed using random measurement error with a maximum specified deviation of 10 percent of the actually measured value.

6.6.1. First sequence

In the beginning, as the initial step, the linked simulation-optimization model is solved to obtain the source characteristics regarding contaminant release fluxes and the source activity starting times. The objective function of the linked optimization methodology for source characterization is formulated using the observed measurement data for January 2009 at the three arbitrary well locations. Adaptive Simulated Annealing optimization algorithm is used to link the MT3DMS transport simulation model to obtain the best solution results based on the provided information. Figure 6.5 shows the source identification results in this step, using three arbitrary monitoring wells. The x-axis is marked by the source flux variables for two sources at different time steps (Sij i=1,2, j=1,2,..,10), and also time lag variable ΔT. The source flux magnitude (mg/s) is shown on the primary y axis, and the secondary y-axis shows the lag time (day).

The lag time estimated by this methodology indicates that source activity started in the year 2001. The first sequence of source characterization indicates that both sources were active over the time steps. These preliminary solution results based on arbitrary

observation measured data does not appear to be reasonable. Therefore, to determine more accurate source characteristics, the next sequence of the optimal monitoring network is essential. To obtain the next three optimal monitoring well locations, the two-steps FSMT multi-objectives monitoring network design model is solved for the next time step, with concentration measurement data for dated 30 April 2009.

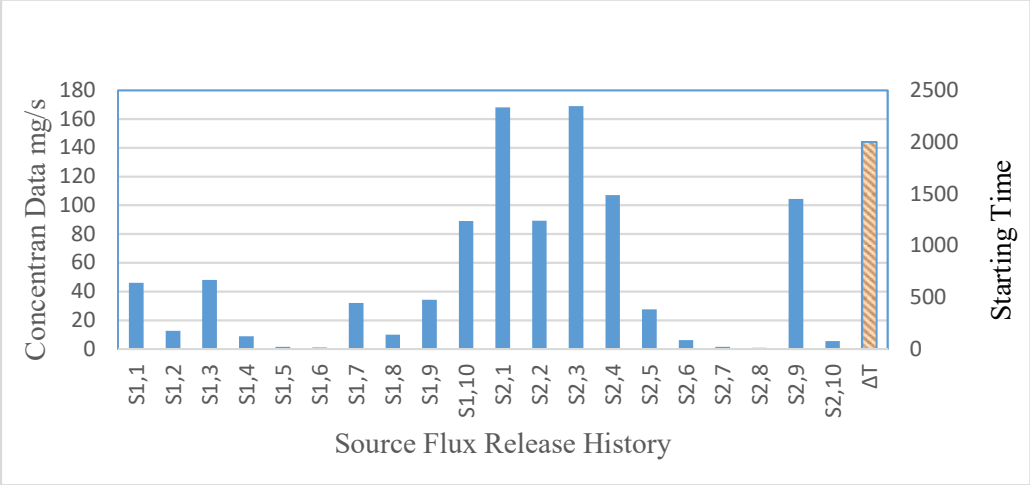


Figure 6.5 Source Identification Result Using Initial Observed Concentration Measurements

6.6.2. Second sequence

The first sequence results of source characterization indicate that the monitoring network for the next monitoring time step (April 2009) should be designed. To increase the feasibility and efficiency of monitoring network design, FSMT is used. Singularity Index contours indicate the plume boundary. The source characterization is more efficient using the wells close to the plume boundary. The wells which are close to the singularity index contours are used as candidates in the two objectives monitoring network design model, and the wells which are far from the plume boundary is eliminated. Figure 6.6 and 6.7

show the singularity analysis results and singularity index contours for a value of 2, respectively.

The next monitoring well locations are chosen using the Pareto-optimal monitoring network design based on the following two objectives: (1) maximize the summation of the product of estimated concentration gradients, and the simulated concentration at that location, and (2) minimize the maximum normalized error between actual concentration and those estimated with the kriging interpolation models, based on monitoring data from designed monitoring locations. In the monitoring network process, three well locations out of 18 potential monitoring wells are selected as observation wells for April 2009. The selected wells from the Pareto-optimal monitoring network design model are M17, M19 and M16 located at (1, 21, 23), (1, 18, 24), and (1, 17, 26), respectively.

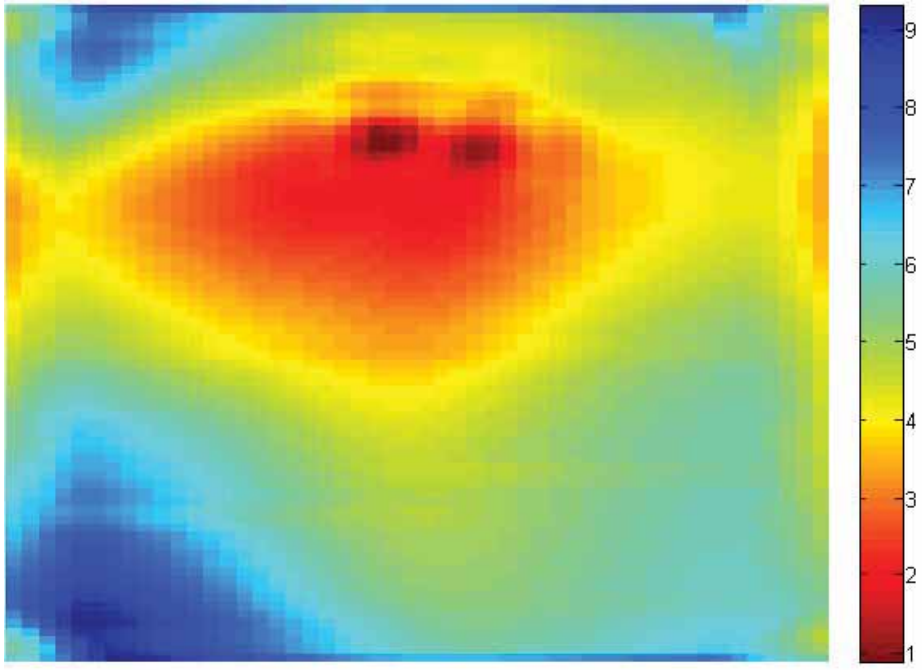


Figure 6.6 Singularity Analysis Result using the First Source Characterization Results

Figure 6.7 Singularity Index using the results of First Source Characterization

The observed measurement data from all six monitoring wells both the first sequence and the second sequence (M03, M14, M20, M17, M19, and M16) are recorded for April 2009. These concentration measurement data, in addition to the concentration data collected in January 2009 (already available from the previous sequence) are utilized in the linked simulation-optimization model to improve the solution for optimal source characterisation in the study area. Figure 6.8 illustrates the source flux release history results and the lag time estimation from the source characterization results using concentration measurements from all wells in the monitoring network, obtained on 30 April 2009, and measurements obtained on 22 January 2009 from the three pre-existing arbitrary wells.

It can be noted that there is an improvement in the estimate of the source flux magnitude for source two (S2). However, the source characterization model could not identify the actual and the dummy (not actually a source) sources completely. The same integrated monitoring network design and source characterization model solution sequence is repeated until the reasonable accurate results are obtained.

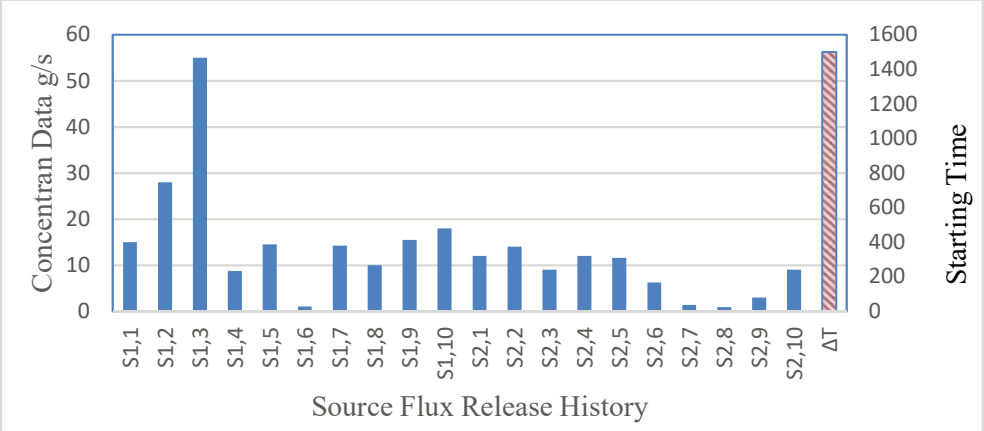


Figure 6.8 Source Identification Results from the Second Sequence

6.6.3. Third sequence

In the third sequence, again new monitoring wells need to be selected for the next monitoring time (July 2009) to further improve the source characterization. Similar to the previous sequence, plume boundary contours are obtained using the FSMT. Figure 6.9 and 6.10 show the results of FSMT methodology solution. The new potential wells are identified, and those wells which are far from the plume boundary are eliminated. Then the two objectives Pareto-optimal monitoring network design is utilized to obtain the location of next three monitoring wells on July 2009. Well locations M02 (1, 20, 25), M05 (1, 22, 25) and M18 (1, 21, 27) are selected as the new monitoring wells for next monitoring time step.

The concentration data at existing monitoring wells (M03, M14, M20, M17, M19 and M16) and the newly selected monitoring wells (M02, M05, and M18) are recorded for July 2009. The source identification model is solved using the concentration observation data in July 2009 in addition to the existing available concentration data (in January and April 2009). Figure 6.11 shows the source flux estimation, and the lag time estimation for two sources for the third sequence of source characterization.

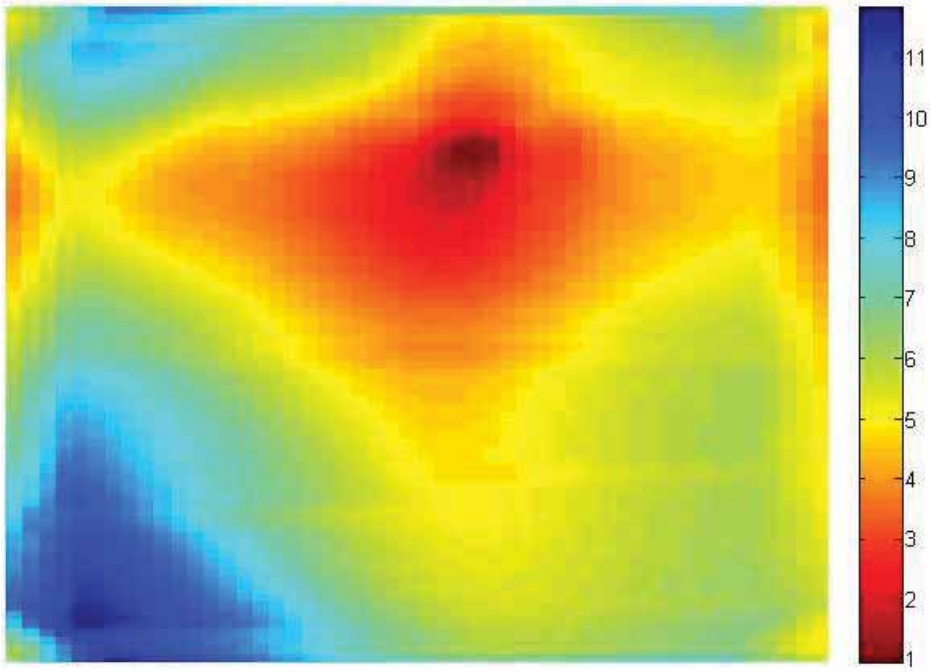


Figure 6.9 Singularity Analysis Result using the Second Source Characterization Results

The starting time estimates do not seem to change from the previous design sequence to this design sequence. It is also evident that the dummy source is identified correctly in this sequence; therefore, the methodology is terminated. Subsequent to this all the available concentration measurement data from all the 74 well locations are utilized for accurate source characterization. These solution results are utilized as the benchmark for evaluating or validating the source characterization solution results obtained with a very

small number of designed monitoring locations. These solution results of source flux characterization and the selected optimal monitoring locations are presented in Table 6.2.

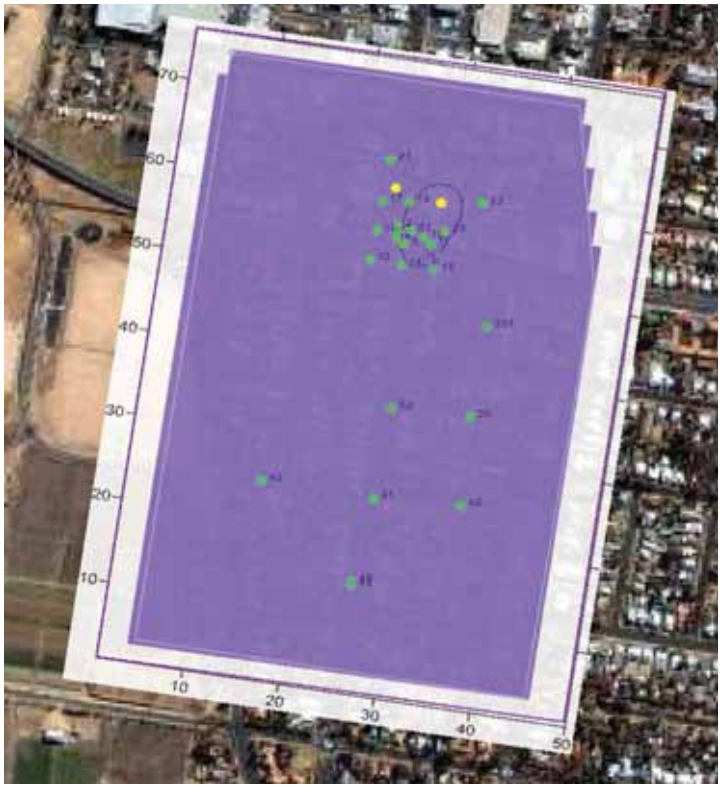


Figure 6.10 Singularity Index using the results of Second Source Characterization

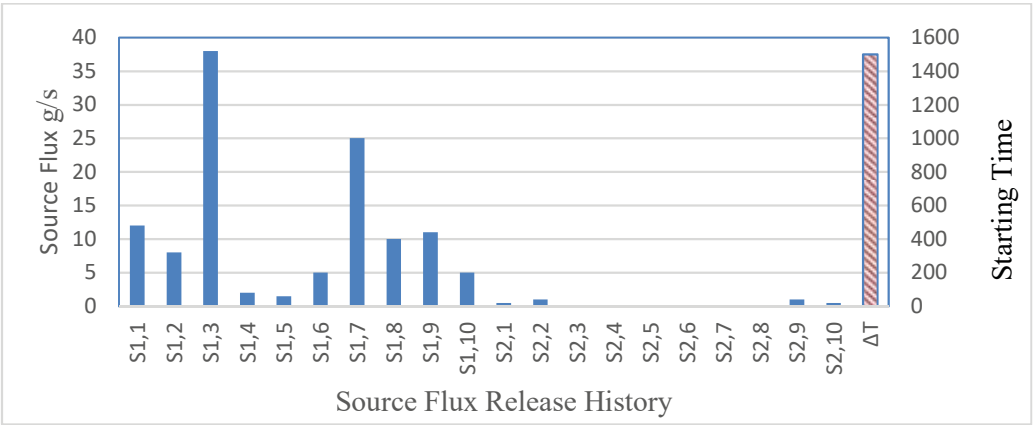


Figure 6.11 Third Source Identification Results

6.6.4. Evaluation

In order to evaluate the efficiency and accuracy of the proposed sequential methodology in estimating the source flux magnitude, release history and source activity starting time, accurate values of the actual fluxes, release history in terms of time including the activity initiation time are necessary. Performance evaluation of the proposed methodology requires these benchmark values for comparison. As in almost all such real life scenarios, actual source flux magnitudes, possibly location, and the release history or the source(s) activity initiation time are not known. Therefore, for evaluation purpose, the performance evaluation is based on the comparison of the estimates obtained using the proposed sequential methodology with the more accurate values obtained by utilizing the extensive concentration measurement network for this polluted urban aquifer site, consisting of 74 monitoring wells covering part of a small urban city in New South Wales, Australia. In fact, the motivation behind choosing this real life site was the availability of extensive concentration and head measurement data, as well as the available information regarding the hydrogeologic parameters. The bench mark values of the source characteristics in terms of location, magnitude, and release history were established using the source characterization based on this extensive monitoring information for this site. Subsequently, these bench mark values were compared with the solution results obtained using the proposed methodology.

Therefore, the linked simulation-optimization model for optimal source characterization, without any monitoring network design component is solved using all available contaminant concentration measurement data for all three observed time steps (January, April, and July 2009). Figure 6.12 presents the flux release history for the evaluation

model. Table 6.2 shows the optimal source flux estimation using all available measured concentration data.

Table 6.2 Results of source characterization and designed monitoring well locations

	Source & Source Location (i, j, k)	Flux Values (g/s) at different time steps									
		S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Evaluation Model	Source 1 (Actual) (17, 29,1)	14.10	11.05	35.08	1.15	2.61	7.90	27.40	7.32	8.14	15.95
	Source 2 (dummy) (16,24, 1)	0	0	0	0	0	0	0	0	0	0
Third Sequence	Source 1 (Actual) (17,29,1)	12	8	38	2	1.5	6	27	6	9	5
	Source 2(dummy) (16,24, 1)	0.2	0.1	0	0	0	0	0	0	0.1	0.1
Second Sequence	Source 1 (Actual) (17,29,1)	15	28	55	8.8	14.5	1	14	10	15.5	18
	Source 2(dummy) (16,24, 1)	12	14	9	12	11.6	6	1.4	0.9	3	9
First Sequence	Source 1 (Actual) (17,29,1)	46	12..6	48	8.7	1.4	1	32	10	34	89
	Source 2(dummy) (16,24, 1)	168	201.2	169	107	27.6	6.3	1.4	0.9	164	5.4

Figure 6.12 shows the source characterization results in the evaluation model using the extensive concentration measurement network for this polluted urban aquifer site, consisting of seventy four monitoring wells in the study area. The estimated source flux magnitude for S10 shows a steep jump in source flux value. As the lag time ΔT estimate indicates that the source activity started in the year 1999, source flux S10 represents the source flux magnitude for 2009. All the concentration measurements used in the identification of source characteristics are from the beginning of the year 2009 (22

January 2009). The observation monitoring wells always obtain the source activity with time delay. Therefore, it seems that the source flux magnitude S10 may not have impacted the concentration measurements taken in January, April, and July 2009. Figure 6.13 shows the concentration breakthrough curves at the selected monitoring locations in the specified study area. These breakthrough curves show the relevance of the selected monitoring locations, as these are impacted in a time varying manner by the source flux. The locations M17 and M19 appear to be minimally affected by the source, and may not have proved very effective in the source characterization process. However, these wells are selected using the Pareto-optimal monitoring network design for identifying the source characteristics and therefore, it is possible that some of these locations do not show large concentrations. However as discussed in Prakash and Datta (2015), a few monitoring locations with very small concentrations may also help in distinguishing multiple overlapping plumes.

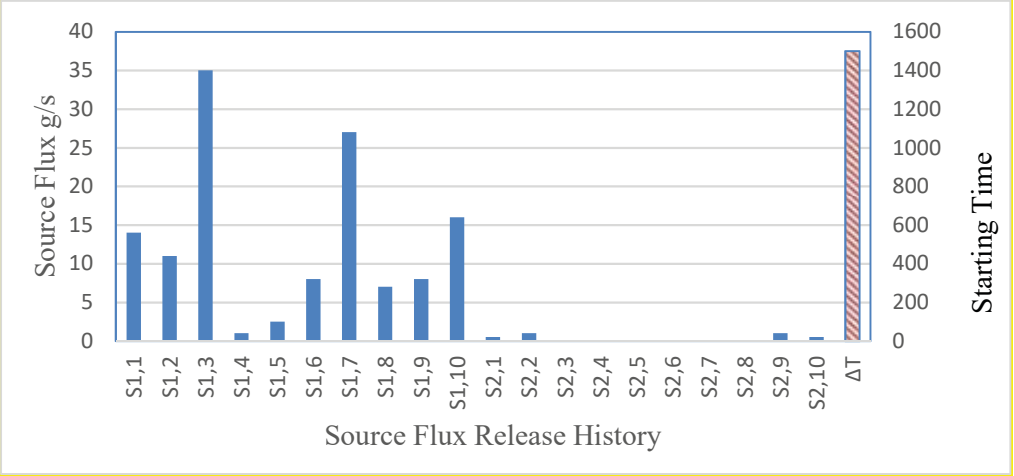


Figure 6.12 Flux release history and lag time in the evaluation source identification

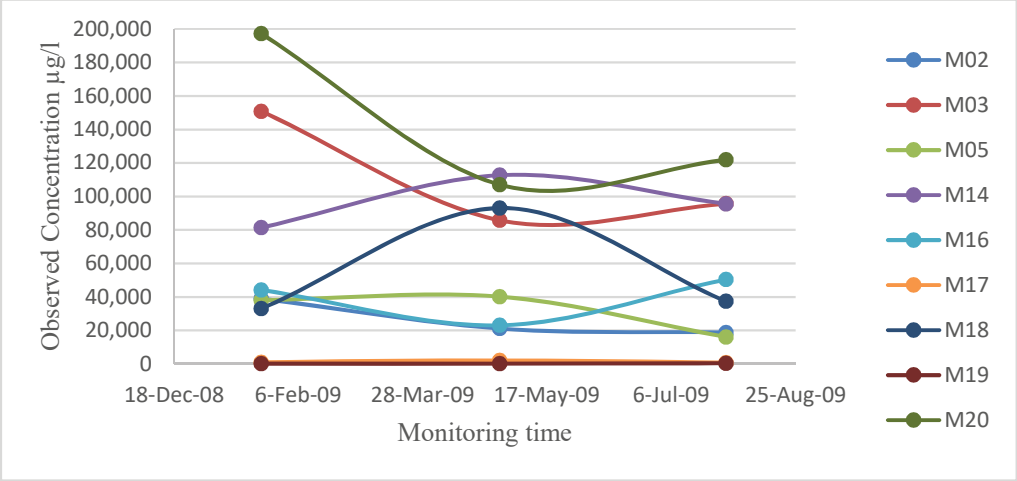


Figure 6.13 Breakthrough curves at monitoring well locations

To evaluate the efficiency of the proposed sequential methodology of monitoring network design and subsequent source characterization, the normalized absolute errors between the estimated temporal release history of source one at different sequences, and corresponding benchmark source fluxes (obtained using extensive information and 74 monitoring location data) are calculated. Figure 6.14 shows the normalized absolute errors for various sequences of the integrated source characterization and monitoring network design methodology.

These limited performance evaluation results for a real-life contaminated aquifer site results show that the developed methodology of sequential source characterization and FSMT based multi-objectives monitoring network design can successfully identify the source characteristics of the unknown contaminant source, and also correctly identify the dummy (not an actual) source in terms of locations, flux release history, and source activity starting time. These reasonably accurate solution results are obtained using a limited number of concentration measurement data, at only a few selected monitoring locations (18 concentration measurement data) in the third sequence of the iterative

procedure. Also, these solution results are comparable to those obtained using a much more expensive and comprehensive concentration monitoring network. These results show the potential application of the methodology to design an economically efficient and effective monitoring network utilizing FSMT for initial determination of the potential monitoring locations.

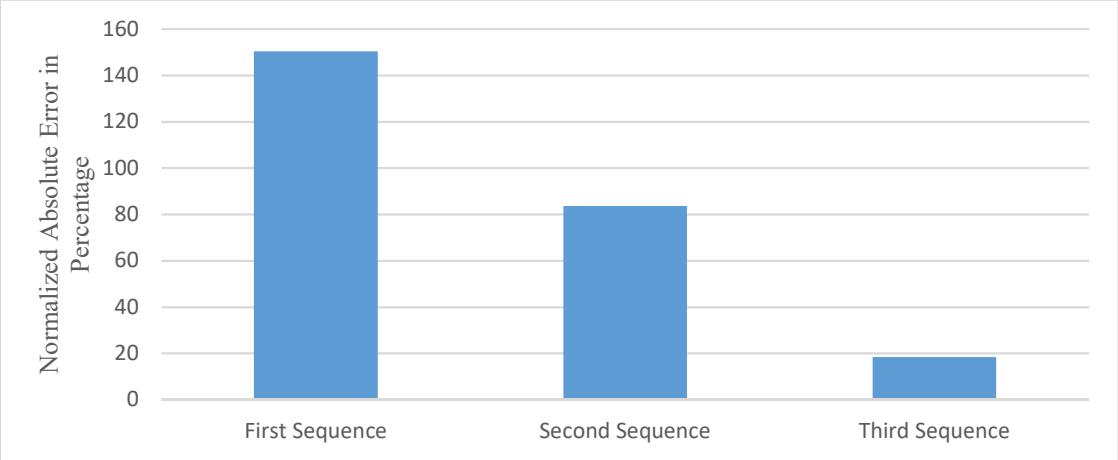


Figure 6.14 Normalized Absolute Errors for Potential Source Location One

6.7. Conclusion

In this chapter, the performance of an integrated sequential source characterization methodology based on the solution of an optimal linked simulation-optimization based source characterization model and a sequential Pareto-optimal monitoring network design methodology is evaluated for a real-life contaminated aquifer in an urban area. The reasonably accurate solution results for source characterization in terms of flux release history, source activity starting time and accurate source locations demonstrate the potential applicability of the proposed methodology to real contaminated aquifer sites. The available hydraulic head data from the observation wells are utilized to obtain the

calibrated flow model, and the transport simulation models for the study area. Three sequences of source identification and monitoring network design are applied to the study area to obtain the final solution results for the unknown source characteristics.

Initially, three arbitrary monitoring wells are utilized at the first round of source characterization. Then the estimated plume concentration data are used to choose the next three monitoring well locations using FSMT based two objectives monitoring network design model. The well locations which are close to the plume boundary, are suitable candidates for monitoring network design for source identification model. A Singularity Index guideline improves the optimal design of the monitoring network by effectively decreasing the number of potential monitoring well locations. The concentration measurement data from the new wells in the designed monitoring network, in addition to the previous monitoring well locations are utilized to identify the source characteristics. In this application, the sequential process is repeated three times to obtain the optimal accuracy in estimating release fluxes and source flux starting time.

The proposed methodology shows efficiency in identifying the unknown source characteristics as only nine monitoring wells are utilized in the final sequence. The designed monitoring network uses less number of well locations as compared to the source identification model using seventy-four available observation wells, with comparable results. Only three temporal readings (January, April, and July 2009) were utilized to estimate the source characteristics satisfactorily. Therefore, the proposed methodology is potentially useful for efficient characterization of unknown contaminant sources in a complex contaminated aquifer site, where very little initial concentration measurement data are available. The proposed sequential procedure helps in designing relevant and efficient monitoring networks which when implemented, provides fresh

concentration measurement data. The illustrative application of the methodology to a real-life contaminated aquifer site demonstrates the capability and efficiency of the proposed methodology.

The next chapter summarizes various aspects of the research undertaken as part of this Thesis. The main conclusions are stated, and limitations as well as future scope is discussed.

7. Summary and Conclusion

This chapter briefly discusses the main components of this thesis. It summarizes the salient features of the methodologies developed, performance evaluation of the methodologies, as well as limited applications of the developed methodologies. The main findings of the developed methodologies, some of the limitations associated with the methodologies and scope for future studies are also highlighted.

7.1. Summary

The main aims of this study were: develop a computationally efficient and computationally feasible unknown aquifer contaminant source characterization methodology applicable on a large regional scale; extend the existing methodologies for this purpose to include multiple species reactive transport process in hydrogeocochemically complex aquifer systems; and develop improved monitoring network design models for more efficient source characterization. The methodologies developed included trained and tested surrogate models as approximate simulator of the flow and transport processes in the aquifer based on Genetic Programing models. These surrogate models were then incorporated in a linked simulation-optimization model for optimal source characterization. The linked optimization-simulation model is solved using the Adaptive Simulated Annealing optimization algorithm, which further improves the search for an optimal solution.

Because the accuracy and reliability of the source characterization process depends on the quality and extent of the spatial and temporal concentration measurement data, a relevant issue is the design and implementation of a suitable and efficient monitoring network

under conditions of various uncertainties. This is especially true, where the initial measurement data available are sparse and obtained from arbitrary monitoring locations. Therefore, a new two objective Pareto optimal monitoring network design methodology was developed. This design methodology utilized Fractal Singularity Mapping Technique to determine plume boundaries, information used to select potential monitoring locations for contaminant concentration monitoring. This approach substantially improved the source characterization efficiency as demonstrated for illustrative study areas. In order to improve the efficiency and accuracy of the source characterization methodology in real life sites where, contamination is evident, but the monitoring data are very sparse and arbitrary, the monitoring network design model was integrated with the source characterization process by sequentially utilizing the source characterization model to estimate the sources, then utilize this information to design and implement a cost effective monitoring network. This sequential and iterative methodology was shown to improve the source characterization efficiency and accuracy, even when dealing with a hydrogeochemically complex aquifer system with multiple reactive species.

The performance of the linked source characterization model was also evaluated by limited application to real life sites, which included a complex, contaminated, previously operating gold and copper mine site in Queensland, Australia. The sequential source characterization and monitoring network design methodology was applied to a contaminated aquifer in an urban area in Australia. The contaminant source was detected to be a leaking underground tank. In this application the contaminant was assumed to be conservative and nonreactive. However, it is possible to utilize this methodology for reactive contaminants as well, by utilizing a different numerical contaminant transport simulation model. Also, it is possible to develop trained and tested surrogate models as

approximated simulators of such reactive contaminants and utilize the surrogate model within the linked simulation optimization model. Several techniques are utilized in the proposed methodology to increase the efficiency of the source characterization including trained and tested Genetic Programming based surrogate models, Adaptive Simulated Annealing optimization algorithm, Fractal Singularity Mapping Technique, and Statistical Kriging interpolation.

7.2. Conclusion

The potential applicability of surrogate models based on ensemble GP models is demonstrated for a chemically reactive multiple species transport process in a complex contaminated aquifer such as an abandoned mine site under input data uncertainties. The developed ensemble GP models result in improving the computation efficiency and computational feasibility while providing acceptable results. The evaluation results demonstrate the viability of replacing the numerical geochemical transport processes simulation model with GP based surrogate models in highly complex contaminated aquifers also incorporating uncertainties in describing the physical system.

The linked simulation-optimization methodology using the GP based surrogate models as an approximate simulator is applied to a real mine site area in central Queensland, Australia. The highly complex area and the sparse available data are the main challenges to calibrate the flow and transport processes with the numerical simulation model. The limitations in the process of implementing adequately calibrated flow and transport simulation models for such a hydrogeologically and geochemically complex contaminated aquifer site with very limited availability of measurement data are also highlighted. The performance evaluation results demonstrate the feasibility of the

proposed methodology for source characterization in contaminated aquifer sites with multiple species reactive transport.

To obtain efficient and accurate source characterization, relevant spatiotemporal concentration measurement data is essential. The arbitrarily located concentration measurements measured data may lead to erroneous source characterization, and installing a vast number of monitoring wells to identify the source characteristic may not be suitable and economically efficient. To address this issue, optimal monitoring network design methodology is applied to ensure economic efficiency, and also increase the source characterization efficiency. In the proposed methodology, fractal singularity mapping technique (FSMT) is utilized and the computed singularity indices are used as guidelines for specifying potential monitoring locations for designing an optimal monitoring network.

FSMT is widely used in different science and engineering fields, but few of them applied this technique to delineate the plume boundary. The window based singularity mapping technique provides singularity indices for the aquifer. The singularity indices smaller and larger than two, indicates contaminated and clean zones of the study area, respectively. Therefore, contours corresponding to singularity index two are used to delineate the contamination plume boundaries. The contours of singularity index equal to two were suitable for guiding the selection of the potential well locations.

Different objective functions are used in the groundwater monitoring network for different purposes. In this study, the monitoring network is chosen based on the following two objectives: (1) maximize the summation of concentration gradients, and the concentration at that location, and (2) minimize the maximum normalized absolute error

at the monitoring locations. Two effective objective functions are applied in the Pareto-optimal design of the monitoring network to obtain the efficient identification of unknown groundwater pollution source characteristics. However, the source characterization results are more accurate while incorporating the FSMT guideline to choose the potential well locations for optimal design of the monitoring network. FSMT is demonstrated to provide a useful and effective guideline to select the potential locations for monitoring network design. The outcomes from the developed methodology demonstrate there is a trade-off in choosing the optimal monitoring locations in terms of minimizing the maximum normalized error between assumed and interpolated concentration values, and maximizing the sum of the estimated concentration gradient and the concentration at location values. However, the ideal levels of trade-off need to be studied and may depend on site-specific conditions and stakeholders policies.

In this study, applicability of the developed methodologies is demonstrated by applying it to polluted aquifer sites, including a real-life contaminated mine site no longer operational with complex geochemical processes present. The integrated sequential source characterization and monitoring network design methodology is applied to a polluted aquifer site in New South Wales, Australia. The integrated methodology is applied using a sequential source characterization with a sequential optimal monitoring network implementation to identify source characteristics in terms of flux release history, source activity starting times and source locations. The source flux magnitudes and source activity starting times are validated based on benchmark best estimates obtained using extensive concentration measurements available for this site. The source magnitude, duration and activity initiation time results obtained appear to be intuitively correct, based on all the site information available.

The comparatively accurate source flux estimations obtained by using the integrated methodology, are obtained using a small number of monitoring wells in sequentially designed monitoring networks. This methodology can potentially result in economically efficient and effective source characterization in sites with very little initial concentration measurements data. Performance evaluation results for all the proposed methodologies illustrate their potential for field application. However, there are some limitations associated with the developed methodologies, which can be used as guidelines for further studies. Some of the major limitations are:

1. The availability of field measurement data and their resolution are critical in determining the capabilities of the calibrated model.
2. The trained and tested GP surrogate models have some limitations such as providing a separate formulation for each output in the input-output pattern. It is possible to improve approximate simulation further by incorporating other surrogate models.
3. The methodologies developed in this study are sensitive to uncertainties in hydrogeological parameters and need to be expanded to incorporate these uncertainties and random heterogeneities explicitly.
4. Some of the performance evaluations are based on the assumption that the calibrated model represents actual field conditions as closely as possible. However, the performance evaluation will depend on the accuracy of the calibration.
5. Further improvement in the fractals based technique is possible, such as using the multifractal technique to improve the monitoring network design.
6. In multi-objective monitoring network design, further studies are required to establish a guideline for obtaining the ideal trade-off considering site-specific conditions, and stakeholder's preferences.

References

Ababou, R., A. C. Bagtzoglou and A. Mallet (2010). "Anti-diffusion and source identification with the RAW scheme: a particle-based censored random walk." J. Environ. Fluid Mech.

Abarca, E., E. Vazquez-Sune, J. Carrera, B. Capino, D. Gamez and F. Battle (2006). "Optimal design of measures to correct seawater intrusion." Water Resource Research. 42.

Afzal, P., A. Y. Fadakar, A. Khakzad, P. Moarefvand and O. N. Rashidnejad (2011). "Delineation of mineralization zones in porphyry Cu deposits by fractal concentration–volume modeling." Journal of Geochemical Exploration 18: 220-232.

Agterberg, F. P. (2012a). "Multifractals and geostatistics." J. Geochem. Explor. 122: 113–122.

Agterberg, F. P. (2012b). " Sampling and analysis of chemical element concentration distribution in rock units and orebodies." Nonlinear Process. Geophys 19: 23–44.

Akcil, A. and S. Koldas (2006). "Acid Mine Drainage (AMD): causes, treatment and case studies." Journal of Cleaner Production 14: 7.

Ala, K. N. and A. P. Domenico (1992). "Inverse analytical techniques applied to coincident contaminant distribution at Otis Air force base, Massachusetts." Groundwater.

Alpati, S. and Z. J. Kabala (2000). "Recovering the release history of a groundwater contaminant using a non-linear least-squares method." Hydrological processes 14(6): 1003-1016.

Aly, A. H. and R. C. Peralta (1999). "Optimal design of aquifer cleanup systems under uncertainty using a neural network and a genetic algorithm." Water Resour. Res. 35(8): 10.

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

Anderson, M. P. and W. W. Woessner (1992). " Applied Groundwater Modeling." Academic Press.

Aral, M. M., J. I. Guan and M. L. Maslia (2001). "Identification of pollutant source location and release history in aquifers." Journal Of Hydrologic Engineering. 6(3): 10.

Atmadja, J. and A. C. Bagtzoglou (2001). "Pollution source identification in heterogeneous porous media." Water Resour. Res. 37(8): 2113-2125.

Atmadja, J. and A. C. Bagtzoglou (2003). "Marching-jury backward beam equation and quasi-reversibility methods for hydrologic inversion: application to contaminant plume spatial distribution recovery." Water Resour. Res. 39(2): 10-14.

Ayvaz, M. T. (2010). "A linked simulation–optimization model for solving the unknown groundwater pollution source identification problems." Journal of Contaminant Hydrology 117(1-4): 46-59.

Babovic, V. and M. B. Abbott (1997). "The evolution of equations from hydraulic data Part I: Theory." Journal of Hydraulic Research 35(3): 397-410.

Babovic, V. and M. Keijzer (2002). "Rainfall runoff modelling based on genetic programming." Hydrology Research 33(5): 331-346.

Bagtzoglou, A. C. (2003). "On the non-locality of reversed time particle tracking methods." Environ. Forensics 4(3): 215-225.

Bagtzoglou, A. C. and J. Atmadja (2005). Mathematical methods for hydrologic inversion: the case of pollution source identification. Chapter in Environmental Impact Assessment of Recycled Wastes on Surface and Ground Waters. Engineering Modeling and Sustainability. The Handbook of Environmental Chemistry, Water Pollution Series. T. A. Kassim, Springer-Verlag, Heidelberg-New York, ISBN 3-540-00268-5, pp. 65-96.Volume 5, Part F.

Bagtzoglou, A. C., A. F. B. Tompson and D. E. Dougherty (1991). Probabilistic simulation for reliable solute source identification in heterogeneous porous media. Water resources engineering risk assesment. J. Ganoulis. Heidelberg, Springer-verlag. 38: 189-201.

Bagtzoglou, A. C., A. F. B. Tompson and D. E. Dougherty (1992). "Application of particle methods to reliable identification of groundwater pollution sources." water Resource management 6: 15-23.

Bain, J. G., D. W. Blowes, W. D. Robertson and E. O. Frind (2000). "Modelling of sulfide oxidation with reactive transport at a mine drainage site." J. Contam. Hydrol 41: 20.

Banzhaf, W., P. Nordin, R. E. Keller and F. D. Francone (1998). "Genetic Programming: An Introduction." Morgan Kaufmann, Inc., San Francisco, USA.

Bear, J. (1979). Hydraulics of groundwater. New York, N.Y., 241: McGraw-Hill.

Bhattacharjya, R. and B. Datta (2005). "Optimal management of coastal aquifers using linked simulation optimization approach." Water Resour.Manage. 19(3): 295-320.

Bhattacharjya, R. K. and B. Datta (2009). "ANN-GA-based model for multiple objective management of coastal aquifers." Water Resour. Planning Manage. 135(5): 314-322.

Bhattacharjya, R. K., B. Datta and M. Satish (2007). "Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers." Journal of Hydrologic Engineering 12(3): 273-283.

Cannon, J. R. (1966). Some numerical results for th solution of the heat equation backwards in time. Numerical Solutions of Non Linear Differential Equation. New York, John Wiley and Sons, Inc. MR 34, 7037: 21-54.

Carey, G. R., P. J. Van Geel and J. R. Murphy (1999). "BioRedox-MT3DMS V2.0: A Coupled Biodegradation-Redox Model for Simulating Natural and Enhanced Bioremediation of Organic Pollutants." User's Guide ConestogaRovers and Associates, Waterloo, Ontario, Canada.

Chandalavada, S. and B. Datta (2007). "Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers." Water Resource Management 22: 651-670.

Chandalavada, S., B. Datta and R. Naidu (2011). "Optimisation approach for pollution source identification in groundwater: an overview. International of Environment and Waste Management." International of Environment and Waste Management 8(1): 40-61.

Chandalavada, S., B. Datta and R. Naidu (2011b). "Uncertainty based optimal monitoring network design for chlorinated hydrocarbon contaminated site." Environment Monitoring Assess. 173(12): 929-946.
Chandalavada, S., B. Datta and R. Naidu (2012). "Optimal Identification of Groundwater Pollution Sources Using Feedback Monitoring Information: A Case Study." Environmental Forensics 13(2): 140-154.
Charlton, S. R. and D. L. Parkhurst (2011). "Modules based on the geochemical model PHREEQC for use in scripting and programming languages." Computers and Geosciences 37: 11.
Chen, G., Q. Cheng, R. Zuo, T. Liu and a. Y. Xi (2015). "Identifying gravity anomalies caused by granitic intrusions in Nanling mineral district, China: a multifractal perspective." Geophysical Prospecting 63: 256–270.
Chen, Z., Q. Cheng, J. Chen and S. Xie (2007). "A novel iterative approach for mapping local singularities from geochemical data." Nonlinear Processes in Geophysics 14: 317–324.
Cheng, Q. (2007). "Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China." Ore Geology Reviews 32: 314–324.
Cheng, Q. (2008). "A combined power-law and exponential model for streamflow recessions." Journal of Hydrology 352(1): 157-167.
Cheng, Q. (2012). "Singularity theory and methods for mapping geochemical anomalies caused by buried sources and for predicting undiscovered mineral deposits in covered areas." Journal of Geochemical Exploration 122: 55–70.
Cheng, Q., F. P. Agterberg and S. B. Ballantyne (1994). "The separation of geochemical anomalies from background by fractal methods." Journal of Geochemical Exploration 51: 109-130.
Cheng, Q., Y. Xu and E. Grunsky (1999). Integrated spatial and spectral analysis for geochemical anomaly separation. Proceedings of the Fifth Annual Conference of the International Association for Mathematical Geology. N. A. Lippard S.J., Sinding–Larsen R (Eds.). Trondheim, Norway. 1: 87-92.
Cheng, Q., Y. Xu and E. Grunsky (2000). "Integrated Spatial and Spectrum Method for Geochemical Anomaly Separation." Natural Resources Research 9: 43-52.
Clement, T. P. (1997). "RT3D—A modular computer code for simulating reactive multi-species transport in 3dimensional groundwater systems." Battelle Pacific Northwest National Laboratory Research Report, PNNL-SA-28967.
COMSOL (2010). "Multiphysics " http://www.comsol.com/ (accessed 22.06.10).
Coulibaly, P. (2004). "Downscaling daily extreme temperatures with genetic programming." Geophy Research Letters 31(L16203): 1-4.
Cramer, N. L. (1985). "A representation for the Adaptive Generation of Simple Sequential Programs." In: Proceedings of an International Conference on Genetic Algorithms and the Applications, Grefenstette, John J. (ed.), Carnegie Mellon University.
Cressie, N. (1990). "The origin of kriging." Mathematical Geology, 22p 239-252.
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.

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.

Datta, B., M. Amirabdollahian, R. Zuo and O. Prakash (2016). "GROUNDWATER CONTAMINATION PLUME DELINEATION USING LOCAL SINGULARITY MAPPING TECHNIQUE." International Journal of GEOMATE 11(25): 2435-2441.

Datta, B., M. Amirabdollahian, R. Zuo and O. Prakash (2016). "GROUNDWATER CONTAMINATION PLUME DELINEATION USINGLOCAL SINGULARITY MAPPING TECHNIQUE." International Journal of GEOMATE 11(25): 2435-2441.

Datta, B., J. E. Beegle, M. L. Kavvas and G. T. Orlob (1989). "Development of an expert system embedding pattern recognition techniques for pollution source identification, Completion Rep." for U.S.G.S. Grant No. 14-08-0001-G1500, Dept. of Civil Engineering, Univ. of California, Davis, Calif.

Datta, B., J. E. Beegle, M. L. Kavvas and G. T. Orlob (1989). "Development of an expert system embedding pattern recognition techniques for pollution source identification, Completion Rep. ." for U.S.G.S. Grant No. 14-08-0001-G1500, Dept. of Civil Engineering, Univ. of California, Davis, Calif.

Datta, B., D. Chakrabarty and A. Dhar (2009). " Optimal dynamic monitoring network design and identification of unknown groundwater pollution sources." Water Resour Manage Springer 23(10): 2031-2049.

Datta, B., D. Chakrabarty and A. Dhar (2009). "Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters." Journal of Hydrology 376(1-2): 48-57.

Datta, B. and S. D. Dhiman (1996). "Chance-constrained optimal monitoring network design for pollutants in groundwater " Journal of Water Resource Planning & Management.

Datta, B., F. Durand, S. Laforge, O. Prakash, H. K. Esfahani, S. Chandalavada and R. Naidu (2016). "Preliminary Hydrogeologic Modeling and Optimal Monitoring Network Design for a Contaminated Abandoned Mine Site Area: Application of a Developed Monitoring Net-work Design Software." Journal of Water Resource and Protection 8(1): 46-64.

Datta, B. and G. Kourakos (2015). "Preface: Optimization for groundwater characterization and management." Hydrogeology Journal 23(6): 1043-1049.

Datta, B. and R. C. Peralta (1986). "INTERACTIVE COMPUTER GRAPHICS-BASED MULTIOBJECTIVE DECISION-MAKING FOR REGIONAL GROUNDWATER MANAGEMENT." Agricultural Water Management 11: 91--116.

Datta, B., O. Prakash, S. Campbell and G. Escalada (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.

Datta, B., O. Prakash and J. Sreekanth (2014). "Application of genetic programming models incorporated in optimization models for contaminated groundwater systems management." In: Tantar, Alexandru-Adrian, Tantar, Emilia, Sun, Jian-Qiao, Zhang, Wei, Ding, Qian, Schütze, Oliver, Emmerich, Michael, Legrand, Pierrick, Del Moral, Pierre, and Coello Coello, Carlos A., (eds.) EVOLVE:

a bridge between probability, set oriented numerics, and evolutionary computation V. Advances in Intelligent Systems and Computing, 288 . Springer, New York, NY, USA. 288: 183-199.
Datta, B., O. Prakash and J. Sreekanth (2014). Application of Genetic Programming Models Incorporated in Optimization Models for Contaminated Groundwater Systems Management. EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V. A.-A. Tantar, E. Tantar, J.-Q. Sun et al., Springer International Publishing. 288: 183-199.
Deb, K. (2001). "Multi-objective optimization using evolutionary algorithms, Wiley.".
Demers, I., J. Molson, B. Bussière and D. Laflamme (2013). "Numerical modeling of contaminated neutral drainage from a waste-rock field test cell." Applied Geochemistry 33: 11.
Deutsch, C. V. and A. G. Journel (1998). " GSLIB: Geostatistical Software Library and user's guide." New York: Oxford University Press.
Deutsch, C. V. and A. G. Journel (1998). "GSLIB: Geostatistical Software Library and user's guide." New York: Oxford University Press.
Dhar, A. and B. Datta (2007). "Multi-objective design of dynamic monitoring networks for detection of groundwater pollution." Journal of Water Resource Planning and Management. 133(4): 329-338.
Dhar, A. and B. Datta (2009). "Saltwater Intrusion Management of Coastal Aquifers. I: Linked Simulation-Optimization." Journal Hydrology Engineering 14: 1263-1272.
Dhar, A. and B. Datta (2010). "Logic-Based Design of Groundwater Monitoring Network for redundancy Reduction." Journal of Water Resource Planning and Management. 136.
Domenico, P. A. and F. W. Schwartz (1998). "Physical and Chemical Hydrogeology." 2nd Ed., NewYork: John Wiley & Sons, Inc.
Dorado, J., J. R. Rabunal, J. Puertas, A. Santos and D. Rivero (2002). " Prediction and modelling of the flow of a typical urban basin through genetic programming." Appl. Artifici. Intel. 17(4): 329-343.
Edraki, M., S. D. Golding, K. A. Baublys and M. G. Lawrence (2005). "Hydrochemistry, mineralogy and sulfur isotope geochemistry of acid mine drainage at the Mt. Morgan mine environment, Queensland, Australia." Applied Geochemistry 20(4): 789–805.
Edraki, M., Golding, S.D., Baublys, K.A., Lawrence, M.G. (2005). "Hydrochemistry, mineralogy and sulfur isotope geochemistry of acid mine drainage at the Mt. Morgan mine environment, Queensland, Australia." Applied Geochemistry 20: 789–805.
Elberling, B., R. V. Nicholson and J. M. Scharer (1994). "A combined kinetic and diffusion model for pyrite oxidation in tailings: a change in controls with time." J. Hydrol 157: 14.
Elçi, A. and M. T. Ayvaz (2014). "Differential-Evolution algorithm based optimization for the site selection of groundwater production wells with the consideration of the vulnerability concept." Journal of Hydrology 511: 736–749.
Esfahani, K. H. and B. Datta (2015a). "Simulation of Reactive Geochemical Transport Processes in Contaminated Aquifers Using Surrogate Models." Int. J. of GEOMATE 8(15): 1190-1196.

Esfahani, K. H. and B. Datta (2015b). Use of Genetic Programming Based Surrogate Models to Simulate Complex Geochemical Transport Processes in Contaminated Mine Sites. Handbook of Genetic Programming Application. A. H. Gandomi, Alavi, A. H., Ryan, C.,, Springer International Publishing, Switzerland: 359-379.

Esfahani, K. H. and B. Datta (2016). "Linked Optimal Reactive Contaminant Source Characterization in Contaminated Mine Sites: A Case Study." Journal of Water Resource Planning and Management ASCE Accepted.

Feng-guang, Y., C. Shu-you, L. Xing-nian and Y. Ke-jun (2008). " Design of groundwater level monitoring network with ordinary kriging." J. Hydrodyn. 20(2008): 339-346.

Fethi, B. J., A. H. Loaiciga and A. M. Marino (1994). " Multivariate geostatistical design of groundwater monitoring networks." Journal of Water Resource Planning and Management ASCE 120(4): 505-522.

Forsyth, R. and R. Roy (1986). "Machine Learning Applications in Expert Systems and Information Retrieval, Ellis Horwood series." In: Artificial intelligence, Chichester, UK.

Francone, F. D. (1998). "Discipulus Software Owner's Manual, version 3.0 draft." Machine Learning Technologies Inc., Littleton, CO, USA.

Freedman, V. and M. Ibaraki (2002). "Effects of chemical reactions on density-dependent fluid flow: on the numerical formulation and the development of instabilities." Advances in Water Resources 25((4)): 15.

Gaur, S. and M. C. Deo (2008). "Real-time wave forecasting using genetic programming." Ocean Eng. 35(11-12): 1166-1172.

Goldberg, D. (1989). "Genetic Algorithms in Search and Optimization, Addison-wesley.".

Gorelick, S. M., B. Evans and I. Remson (1983). "Identifying sources of groundwater pollution: An optimization approach." WATER RESOURCES RESEARCH 19(3).

Gurarslan, G. and H. Karahan (2015). "Solving inverse problems of groundwater-pollution-source identification using a differential evolution algorithm." Hydrogeology Journal 23(6): 1109–1119.

Hansen, T. M. (2004). "mGstat V 0.99:MATLAB code, http://sourceforge.net/projects/mgstat/files/."

Herzer, j. and W. Kinzelbach (1989). "Coupling of Transport and Chemical Processes in Numerical Transport Models." Geoderrna 44: 13.

Hong, Y. S. and M. R. Rosen (2002). "Identification of an urban fracturedrock aquifer dynamics using an evolutionary self-organizing modelling." J. Hydrol. 259: 15.

Hong, Y. S., P. A. White and D. M. Scott (2005). " Automatic rainfall recharge model induction by evolutionary computational intelligence." Water Resour. Res. 41(W08422).

Ingber, L. (1993). ""Adaptive Simulated Annealing (ASA),'' Global optimization C-code." Caltech Alumni Association, Pasadena, CA, .

Javandel, I., C. Doughty and T. Chin-Fu (1984). " Groundwater Transport: Handbook of Mathematical Models." Water Resources Monograph No. 10. Washington, D.C.: American Geophysical Union.

Jaynes, D. B., A. S. Rodowski and H. B. Pionke (1984). " Acid mine drainage from reclaimed coal strip mines 1. Model description." Water Resour. Res. 20: 233-243.

Jha, M. and B. Datta (2012). "Application of simulated annealing in water resources management: optional solution of groundwater contamination source characterization problem and monitoring network design problems." In: Sales Guera Tsuzuki, Marcos de, (ed.) Simulated Annealing- Single and Multiple Objective Problems. InTech, Rijeka, Croatia: 157-174.

Jha, M. and B. Datta (2013). "Three-Dimensional Groundwater Contamination Source Identification Using Adaptive Simulated Annealing." JOURNAL OF HYDROLOGIC ENGINEERING 18: 307-318.

Jha, M. and B. Datta (2015a). "Application of Dedicated Monitoring–Network Design for Unknown Pollutant-Source Identification Based on Dynamic Time Warping." J. Water Resour. Plann. Manage. 10.1061/(ASCE)WR.1943-5452.0000513 , 04015022.

Jha, M. and B. Datta (2015b). "Application of Unknown Groundwater Pollution Source Release History Estimation Methodology to Distributed Sources Incorporating Surface-Groundwater Interactions." Environmental Forensics 16(2): 143-162.

Jha, M. K. and B. Datta (2011). "Simulated Annealing Based Simulation-Optimization Approach for Identification of Unknown Contaminant Sources in Groundwater Aquifers." Desalination and Water Treatment 32: 79-85.

Johnson, D. B. and K. B. Hallberg (2005). "Acid mine drainage remediation options: a review." Science of the Total Environment 338: 12.

Jones, D. (2001). "Contaminant Source Study, Mount Morgan mine." EWL Sciences Pty Ltd., Darwin.

Journel, A. G. and C. J. Huijbregts (1978). "Mining Geostatistics." London: Academic Press, 600p.

Kalin, M., A. Fyson and W. N. Wheeler (2006). "Review The chemistry of conventional and alternative treatment systems for the neutralization of acid mine drainage." Science of the Total Environment 366: 14.

Ketabchi, H. and B. Ataie-Ashtiani (2015). "Review: Coastal groundwater optimization—advances, challenges, and practical solutions." Hydrogeology Journal 23(6): 1129–1154.

Khu, S. T., S. Y. Liong, V. Babovic, H. Madsen and N. Muttil (2001). "Genetic programming and its application in real-time runoff forecasting." Journal Of The American Water Resources Association 37(2): 439-451.

Kirkpatrick, S., C. D. Gelatt Jr. and M. P. Vecchi (1983). "Optimization by Simulated Annealing." Science 22(4598): 671-680.

Kollat, J. B., P. M. Reed and R. M. Maxwell (2011). "Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics." WATER RESOURCES RESEARCH 47(W02529).

Koza, J. R. (1994). "Genetic Programming as a Means for Programming Computers by Natural-Selection." Statistics and Computing 4(2): 87-112.

Koza, J. R. (1994). "Genetic programming as a means for programming computers by natural-selection." Statistics and computing 4(2): 26.

LAU, J. E., D. P. COMMANDER and G. JACOBSON (1987). Hydrogeology of Australia. DEPARTMENT OF RESOURCES AND ENERGY AUSTRALIAN GOVERNMENT PUBLISHING SERVICE CANBERRA

Li, Y. (2003). "A Coupled Model of Fluid Flow, Thermal Transport, and Reactive Chemical Transport through Variably saturated Media. M. S. Thesis." Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816.

Lichtner, P. C. (1996). "Continuum formulation of multicomponent-multiphase reactive transport." Mineralogical Soc. Am., Washington, DC, USA Chapter 1: 81.

Liu, C. W. and W. P. Ball (1999). "Application of inverse methods to contaminant source identification from aquitard diffusion profile at Dover AFB, Delaware." Water Resource Research. 35(7): 1975-1985.

Loaiciga, H. A. (1989). "An optimization approach for groundwater quality monitoring network design." Water Resource Research 25(8): 1771-1782.

Loaiciga, H. A., R. J. Charbeneau, L. G. Everett, G. E. Fogg, B. F. Hobbs and S. Rouhani (1992). "Review of Ground-Water Quality Monitoring Network Design." Journal of Hydraulic Engineering, 118(1): 11-38.

Loaiciga, H. A. and P. F. Hudak (1993). "An optimization method for network design in multilayered groundwater flow systems." Water Resource Research. 29(8): 2835-2846.

Luoa, Q., J. Wub, Y. Yangb, J. Qian and J. Wu (2016). "Multi-objective optimization of long-term groundwater monitoring network design using a probabilistic Pareto genetic algorithm under uncertainty." Journal of Hydrology 534: 352–363.

Magee, J. W. (2009). "Palaeovalley Groundwater Resources in Arid and Semi-Arid Australia, A Literature Review." Geoscience Australia, Australian government national water comission.

Mahar, P. S. and B. Datta (1997). "Optimal monitoring network and ground-water-pollution source identification." Journal of Water Resource Planning and Management. 123(4): 199-207.

Mahar, P. S. and B. Datta (2000). "Identification of pollution sources in transient groundwater system." Water Resource Management. 14(6): 209-227.

Mahar, P. S. and B. Datta (2001). "Optimal identification of ground-water pollution sources and parameter estimation." Journal of Water Resource Planning and Management. 127(1): 20-30.

Makkeasorn, A., N.-B. Chang, M. Beaman, C. Wyatt and C. Slater (2006). "Soil moisture estimation in a semiarid watershed using RADARSAT-1 satellite imagery and genetic programming." Water Resources Research 42(9).

Makkeasorn, A., N. B. Chang and X. Zhou (2008). "Short-term streamflow forecasting with global climate change implications—A comparative study between genetic programming and neural network models." J. Hydrol. 352(3-4): 336-354.

Mandelbrot, B. (1989). "Multifractal measures, especially for the geophysicist." Pure and Applied Geophysics 131: 5-42.

Manderson, J. (2011). "Report for strengthening basin communitie- planning study bussiness case- enhancing Dubbo's irrigation system." Dubbo City Council.

Mao, X., H. Prommer, D. A. Barry, C. D. Langevin, B. Panteleit and L. Li (2006). "Threedimensional model for multi-component reactive transport with variable density groundwater flow." Environmental Modelling & Software 21 ((5)): 14.

Matheron, G. (1963). " Principles of Geostatistics." Economic Geology 58: 1246-1266.

Mayer, K. U., E. O. Frind and D. W. Blowes (2002). "Multicomponent reactive transport modeling in variably saturated porous media using a generalized formulation for kinetically controlled reactions." Water Resour. Res. 38(13): 21.
Mayer, K. U., E. O. Frind and D. W. Blowes (2003). "Multicomponent reactive transport modelling in variably saturated porous media using a generalized formulation for kinetically controlled reactions." Water. Resour. Res. 38.
McPhee, J. and W. G. Yeh (2006). "Experimental design for groundwater modeling and management." WATER RESOURCES RESEARCH 42(2): 1-13.
Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller (1953). "Equation of state calculations by fast computing machines." J. Chem. Phys. 21: 1087–1092.
Meyer, P. D. and E. D. Brill Jr (1988). "A method for locating wells in a groundwater pollution monitoring network under conditions of uncertainty." Water Resource Research. 24(8): 1277-1284.
Meyer, P. D., A. J. Valocchi and J. W. Eheart (1994). "Monitoring network design to provide initial detection of groundwater pollution Water Resource Research." Water Resource Research. 30(9): 2647-2660.
MLA (2008). "Meat & livestock Australia." Mount Morgan climate history 113.
Molson, J., M. Aubertin, B. Bussie`re and M. Benzaazoua (2008). "Geochemical transport modelling of drainage from experimental mine tailings cells covered by capillary barriers." Applied Geochemistry 23: 24.
NCGRT, N. C. f. G. R. a. T. (2013). Economic Value of Groundwater in Australia, Deloitte Access Economics.
Neupauer, R. M., B. Borchers and J. L. Wilson (2000). "Comparison of inverse methods for reconstructing the release history of a groundwater contamination source." Water Resour. Res. 36(9): 2469-2475.
Neupaur, R. M. and J. L. Wilson (1999). "Adjoint method for obtaining backward in-time location and travel probabilities of a conservative groundwater contaminant." Water Resource Research. 35(11): 3389-3398.
Neupaur, R. M. and J. L. Wilson (2005). "Backward probability model using multiple observations of contamination to identify groundwater contamination sources at the Massachusetts military reservation." Water Resource Research. 41.
Nordstrom, D. K. and C. N. Alpers (1999). "Geochemistry of acid mine waters, in Plumlee, G.S., and Logsdon, M.J., eds., The environmental geochemistry of mineral deposits, Part A: Processes, techniques, and health issues, Reviews." Economic Geology, 6A: 28.
Parasuraman, K., A. Elshorbagy and S. K. Carey (2007a). " Modelling the dynamics of evapotranspiration process using genetic programming." Hydrol. Sci. J. 52(3): 15.
Parasuraman, K., A. Elshorbagy and B. C. Si (2007b). "Estimating saturated hydraulic conductivity using genetic programming." Soil Sci. Soc. Am. J. 71(6): 9.
Parkhurst, D. L. and C. A. J. Appelo (1999). " User's guide to PHREEQC—A computer program for speciation, reac-tion-path, 1D-transport, and inverse geochemical calculations. ." Technical Report 99-4259, US Geol. Survey Water-Resources Investigations Report.
Parkhurst, D. L., K. L. Kipp, P. Engesgaard and S. R. Charlton (2004). "PHAST e A Program for Simulating Ground-water Flow, Solute transport, and Multicomponent Geochemical Reactions." U.S. Geological Survey, Denver, Colorado.

Parkhurst, D. L., D. C. Thorstenson and L. N. Plummet (1982). " PHREEQE - A computer program for geochemical calculations." Water Resour. Invest. 210: 80-96.
Prakash, O. and B. Datta (2013). "Sequential optimal monitoring network design and iterative spatial estimation of pollutant concentration for identification of unknown groundwater pollution source locations." Environmental Monitoring and Assessment 185(7): 5611-5626.
Prakash, O. and B. Datta (2014). "Multi-Objective Monitoring Network Design for Efficient Identification of Unknown Groundwater Pollution Sources Incorporating Genetic Programming Based Monitoring." J. Hydrol. Eng. 19(11).
Prakash, O. and B. Datta (2014a). "Characterization of Groundwater Pollution Sources with Unknown Release Time History." Journal of Water Resource and Protection 6(4): 337-350.
Prakash, O. and B. Datta (2014b). "Optimal monitoring network design for efficient identification of unknown groundwater pollution sources." Int. J. of GEOMATE 6(11): 785-790.
Prakash, O. and B. Datta (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(6): 1089-1107.
Prommer, H. (2002b). "PHT3D—A reactive multi-component transport model for saturated porous media." Version 1.0 User's Manual, Technical report, Contaminated Land Assessment and Remediation Research Centre, The University of Edinburgh.
Prommer, H., D. A. Barry and G. B. Davis (2002a). "Modelling of physical and reactive processes during biodegradation of a hydrocarbon plume under transient groundwater flow conditions." Journal of Contaminant Hydrology 59: 19.
Puech, V. (2010). Upper macquarie groundwater model. Technical Report VW04680, Office of Water, NSW Government and National Water Commission, Australia.
Ranjithan, S., J. W. Eheart and J. H. Garrett (1993). "Neural network based screening for groundwater reclamation under uncertainty." Water Resour. Res. 29(3): 563-574.
Reed, P. and B. S. Minsker (2004). "Striking the balance: long-term groundwater monitoring design for conflicting objective." Journal of Water Resource Planning and Management. 130(2): 140-150.
Reed, P. M. and J. B. Kollat (2013). "Visual analytics clarify the scalability and effectiveness of massively parallel many-objective optimization: A groundwater monitoring design example." Advances in Water Resources 56: 1–13.
Reed, P. M., B. S. Minsker and D. E. Goldberg (2000). "Designing a competent simple genetic algorithm for search and optimization." Water Resources Research 36(23): 3757-3761.
RML Technologies Inc. "Discipulus 5.2." RML Technologies, Inc.
Rogers, L. L., F. U. Dowla and V. M. Johnson (1995). "Optimal field-scale groundwater remediation using neural networks and the genetic algorithm." Environmnetal Science & Technology 29: 11.
Rubin, j. (1983). "Transport of reacting solutes in porous media: relation between mathematical nature of problem formulation and chemical nature of reactions." Water Resour. Res 19(5): 1231-1253.

Sadeghi, B., N. Madani and E. J. M. Carranza (2015). "Combination of geostatistical simulation and fractalmodeling for mineral resource classificatio." Journal of Geochemical Exploration 149: 59–73.
Salmon, S. U. and M. E. Malmstrom (2004). "Geochemical processes in mill tailings deposits: modelling of groundwater composition." Applied Geochemistry 19: 17.
Savic, D. A., G. A. Walters and J. W. Davidson (1999). "Genetic programming approach to rainfall-runoff modelling." Water Resour. Manage. 13(3): 219-231.
Scharer, J. M., R. V. Nicholson, B. Halbert and W. J. Snodgrass (1994). "A computer program to assess acid generation in pyritic tailings." In: Alpers, C.N., Blowes, D.W. (Eds.), Environmental Geochemistry of Sulfide Oxidation. ACS symposium series, Washington DC, USA, (Chapter 11). 21.
Sheta, A. F. and A. Mahmoud (2001). " Forecasting using genetic programming." In: Proceedings of the 33rd Southeastern Symposium on System Theory, Athens, OH, USA,: 347-352.
Sidauruk, R. M., A. H. D. Cheng and D. Ouazar (1998). "Groundwater contaminant source and transport parameter identification by correction coefficient optimization." Groundwater 36: 208-214.
Simunek, J., M. ejna, H. Saito, M. Sakai and M. T. Van Genuchten (2009). "The HYDRUS-1D Software Package for Simulating the Movement of Water, Heat, and Multiple Solutes in Variably Saturated Media, Version 4.0." Department of Environmental Sciences, University of California Riverside, Riverside, California.
Singh, R. M. and B. Datta (2004). "Groundwater pollution source identification and simultaneous parameter estimation using pattern matching by artificial neural network." Environment Forensics 5(3): 143-159.
Singh, R. M. and B. Datta (2006). "Identification of Groundwater Pollution Sources Using GA-based Linked Simulation Optimization Model." JOURNAL OF HYDROLOGIC ENGINEERING 11: 101-110.
Skaggs, T. H. and K. Z.J. (1994). "Recovering the release history of a groundwater contaminant." Water Resource Research. 30(1): 71-79.
Smith, K. S. and H. L. O. Huyck (1999). "An overview of the abundance, relative mobility, bioavailability, and human toxicity of metals, in Plumlee, G.S., and Logsdon, M.J., eds., The environmental geochemistry of mineral deposits, Part A: Processes, techniques, and health issues." Reviews in Economic Geology 6A(Chapter 2: Littleton, Colorado, Society of Economic Geologists, Inc): 42.
Sreekanth, J. and B. Datta (2010). "Multi-objective management models for optimal and sustainable use of coastal aquifers." Journal of Hydrology 393(3-4): 245-256.
Sreekanth, J. and B. Datta (2011). "Comparative Evaluation of Genetic Programming and Neural Network as Potential Surrogate Models for Coastal AquiferManagement." Water Resour Manage(25): 3201-3219.
Sreekanth, J. and B. Datta (2011b). "Coupled simiulation-optimization model for coastal aquifer management using genetic programming-based ensembel surrogate models and multiple-realization optimization." Water Resources Research 47(W04516): 1-17.
Sreekanth, J. and B. Datta (2012). Genetic programming: efficient modelling tool in hydrology and groundwater management. Genetic Programming - New Approaches and Successful Applications D. S. V. Soto, InTech.: 225-238.

Sreekanth, J. and B. Datta (2015). "Review: Simulation-optimization models for the management and monitoring of coastal aquifers." Hydrogeology Journal 23(6): 1155-1166.

Sun, J. (2004). "A Three-Dimensional Model of Fluid Flow, Thermal Transport, and Hydrogeochemical Transport through Variably Saturated Conditions. ." M. S. Thesis. Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816.

Taube, A. (1986). "The Mount MorganG old-CopperM ine andE nvironmentQ, ueensland: A VolcanogeniMc assiveS ulfideD epositA ssociatewd ith PenecontemporaneouFsa ulting." Economic Geology Vol. 81(6): 1322-1340.

Taube, A. (2000). "Dumps and tailings on the Mt Morgan mine lease." In: Paddon B, Unger C (eds) Proceedings Mt Morgan Rehabilitation Planning Workshop,, Dept of Mines and Energy Central Region, Rockhampton.

Tebes-Stevensa, C., A. J. Valocchia, V. J. M. and B. E. Rittmannb (1998). "Multicomponent transport with coupled geochemical and microbiological reactions: model description and example simulations." Journal of Hydrology 209: 16.

Turcotte, D. L. (2002). "Fractals in petrology." Lithos 65(3): 261-271.

Ulrich, T., S. D. Golding, B. S. Kamber, Z. Khin and A. Taube (2003). "Different mineralization styles in a volcanichosted ore deposit; the fluid and isotopic signatures of the Mt. Morgan Au–Cu deposit Australia." Ore Geol. Rev. 22(1-2): 61–90.

Waddill, D. W. and M. A. Widdowson (1998). "SEAM3D: A numerical model for three-dimensional solute transport and sequential electron acceptor-based bioremediation in groundwater,." Technical report, Virginia Tech., Blacksburg, Virginia.

Wagner, B. J. (1992). "Simultaneous parameter estimation and pollutant source characterization for coupled groundwater flow and pollutant transport modeling." Journal of Hydrology (Amst) 135(1-4): 275-303.

Wang, G., E. J. M. Carranza, R. Zuo, Y. Hao, Y. Du, Z. Pang, Y. Sun and J. Qu (2012). "Mapping of district-scale potential targets using fractal models." J. Geochem. Explor. 122: 34–46.

Wang, H. F. and M. P. Anderson (1982). "Introduction to Groundwater Modeling: Finite Difference and Finite Element Methods." San Francisco: W. H. Freeman and Company, 237 pp.

Wang, W., J. Zhao and Q. Cheng (2013). "Fault trace-oriented singularity mapping technique to characterize anisotropic geochemical signatures in Gejiu mineral district, China." J. Geochem. Explor. 134: 27–37.

Wang, W. C., K. W. Chau, C. T. Cheng and L. Qiu (2009). "A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series." J. Hydrol. 374(3-4): 294-306.

Wels, C., L. Findlater and C. McCombe (2006). ASSESSMENT OF GROUNDWATER IMPACTS AT THE HISTORIC MOUNT MORGAN MINE SITE, QUEENSLAND, AUSTRALIA. The 7th International Conference on Acid Rock Drainage (ICARD). Lexington: 2311-2331.

Wels, C., L. Findlater and C. McCombe (2009). "Contaminant Load Balance Study for Mount Morgan Mine, QLD, Australia." presented at Securing the Future and 8th ICARD, June 23-26, 2009, Skellefteå, Sweden.

Werner, K. and S. Berglund (1999). "Effects of spatial variability on oxygen transport in sulphidic mining waste deposits. ." In: Proc. Sudbury '99, Conf. Mining and the Environment II, Sudbury, Canada, : 10.
Whigham, P. A. and P. Craper (2001). "Modelling rainfall-runoff using genetic programming." Math. Comput. Modell. 33: 14.
Wissmeier, L. and D. A. Barry (2010). "Implementation of variably saturated flow into PHREEQC for the simulation of biogeochemical reactions in the vadose zone." Environmental Modelling & Software 25(4): 13.
Wissmeier, L. and D. A. Barry (2011). "Simulation tool for variably saturated flow with comprehensive geochemical
reactions in two- and three-dimensional domains." Environmental Modelling & Software 26: 9.
Wu, J., C. Zheng and C. C. Chien (2005). "Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions." Journal of Contaminant Hydrology 77: 41–65.
Wu, J., C. Zheng and C. C. Chien (2005). "Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions." Journal of Contaminant Hydrology 77: 41-65.
Wunderly, M. D., D. W. Blowes, E. O. Frind and C. J. Ptacek (1996). "Sulfide mineral oxidation and subsequent reactive transport of oxidation products in mine tailing impoundments: a numerical model." Water Resour. Res. 32: 14.
Xiao, F., J. Chen, F. P. Agterberg and C. Wang (2014). "Element behavior analysis and its implications for geochemical anomaly identification: A case study for porphyry Cu–Mo deposits in Eastern Tianshan, China." J. Geochem. Explor. 145: 1–11.
Yeh, G. T., S. L. Carpenter, P. L. Hopkins and M. D. Siegel (1995). "User's Manual for LEHGC: A Lagrangian-Eulerian Finite Element Model of Hydrogeochemical Transport Through SaturatedUnsaturated Meida - Version 1.1. SAND95-1121." Sandia National Laboratory, Albuquerque, NM.
Yeh, G. T., J. R. Cheng and H. C. Lin (1994). "3DFEMFAT: User's Manual of a 3-Dimensional Finite Element Model of Density Dependent Flow and Transport through Variably saturated Media." Technical Report submitted to WES, U.S. Corps of Engineers, Vicksburg, Mississippi. Department of Civil and Environmental Engineering, Penn State University, University Park, PA 16802.
Yeh, G. T., J. R. Cheng and H. C. Lin (2000). " Computational Subsurface Hydrology Reactions, Transport, and Fate of Chemicals and icrobes." Kluwer Academic Publishers, New York, USA.
Yeh, G. T. and K. M. Salvage (1997). " HYDROBIOGEOCHEM: a Coupled Model of HYDROlogic transport and Mixed BIOGEOCHEMical Kinetic/Equilibrium Reactions in Saturated-Unsaturated Media. ." Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, PA 16802.
Yeh, G. T., J. Sun, P. M. Jardine, W. D. Burgos, Y. Fang, M. H. Li and M. D. Siegel (2004). "HYDROGEOCHEM 5.0: A Three-Dimensional Model of Coupled Fluid Flow, Thermal Transport, and HYDROGEOCHEMical Transport through Variably Saturated Conditions - Version 5.0." OAK RIDGE NATIONAL LABORATORY Oak Ridge, Tennessee 37831-6302.

Yeh, G. T. and V. S. Tripathi (1991). "HYDROGEOCHEM: A coupled model of hydrologic transport and geochemical multicomponent equilibria in reactive systems." Environmental Science Division Publication No. 3170, Oak Ridge National Laboratory, Oak Ridge, TN,.

Yeh, M. S., Y. P. Lin and L. C. Chang (2006). "Designing an optimal multivariate Geostatistical groundwater quality monitoring network using factorial Kriging and genetic algorithm." Journal of Environmental Geology 50: 101-121.

Yeh, W. W.-G. (2015). "Review: Optimization methods for groundwater modeling and management." Hydrogeology Journal 23: 1051–1065.

Yeh, W. W. G. (1986). "Review of parameter identification procedure in groundwater hydrology: the inverse problem." Water Resource Research. 22(2): 95-108.

Yoon, K. Y., J. Hoon-Byeon, J. H. Park and J. Hwang (2007). " Susceptibility constants of Escherichia coli and Bacillus subtilis to silver and copper nanoparticles " J. Sci. Total Environ. 373: 572–575.

Zechman, E., M. Baha, G. Mahinthakumar and S. R. Ranjithan (2005). "A genetic programming based surrogate model development and its application to a groundwater source identification problem." ASCE Conf. 173(341).

Zhang, F., G. T. Yeh, J. C. Parker, S. C. Brooks, M. N. Pace, Y. Kim, P. M. Jardine and D. Watson (2007). "A reaction-based paradigm to model reactive chemical transport in groundwater with general kinetic and equilibrium reactions." Journal of Contaminant Hydrology 92: 23.

Zhao, H., B. Xia, J. Qin and J. Zhang (2012). "Hydrogeochemical and mineralogical characteristics related to heavy metal attenuation in a stream polluted by acid mine drainage: A case study in Dabaoshan Mine, China." Journal of Environmental Sciences 24(6): 11.

Zheng, C. and G. D. Bennett (1995). "Applied Contaminant Transport Modeling: Theory and Practice." 115 Fifth Avenue, New York, NY 10003,USA Van Nostrand Reinhold, 440 + xxiii pp.

Zheng, C., M. C. Hill and P. A. Hsieh (2001). MODFLOW-2000, the U.S. Geological Survey modular ground-water model : user guide to the LMT6 package, the linkage with MT3DMS for multi-species mass transport modeling. Open-File Report. USA, USGS Numbered Series.

Zheng, C. and P. P. Wang (1999). " MT3DMS: A modular three-dimensional multispecies model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems; Documentation and User's Guide. ." Contract Report SERDP-99-1, US Army Engineer Research and Development Center, Vicksburg,MS.

Zhou, X., M. Chen and C. Liang (2003). "Optimal schemes of groundwater exploitation for prevention of seawater intrusion in the Leizhou Peninsula in southern China." Environmental Geology 43(8): 978-985.

Zuo, R. and J. Wang (2016). "Fractal/multifractal modeling of geochemical data: A review." Journal of Geochemical Exploration 164: 33–41.

Zuo, R., J. Wang, G. Chen and M. Yang (2015). " Identification of weak anomalies: Amultifractal perspective." J. Geochem. Explor. 148: 12–24.