Linked Simulation-Optimization Based Methodologies for Unknown Groundwater Pollutant Source Identification in Managed and Unmanaged Contaminated Sites

Thesis submitted by

Manish Kumar Jha, B.Tech, M.Tech

for the degree of Doctor of Philosophy (PhD)

in the School of Engineering and Physical Sciences

James Cook University

November 2012

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I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and I wish this work to be embargoed until March 2013.

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. Scholarship from Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (CRC CARE), Salisbury South, SA 5106, Australia.

2. School of Engineering and Physical Sciences, James Cook University, Graduate Research Scheme funding for travel and equipment in 2010.

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 conceptualize the problem, suggested several ways of solving it and provided insights to the tools and technologies used in this thesis.
- **Department of Environment and Resource Management, Queensland** : Provided field data for validation of some of the methodologies developed in this 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 CRC CARE 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 Engineering and Physical Sciences 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. I am grateful to all my colleagues who made my stay at JCU a lot more enjoyable and fun-filled.

Abstract

Groundwater is the primary source for irrigation and drinking in many parts of the world. Anthropogenic activities such as mining; large scale production, storage and transport of various chemicals; improper waste management practices; and unsustainable intensive agricultural practices have resulted in the contamination of many groundwater aquifers. The identification of the exact location and release history of contributing sources, which are often unknown, is very important in planning effective remediation measures as well as in determining the liability on the polluter. Contamination of ground water aquifers may be caused by a combination of pollutant sources varying in time of release, flux and location. In situations where they are unknown, location and release histories have to be estimated by inversion. Inversion of the equations governing flow and transport over time and space is an ill-posed problem.

Estimation of unknown groundwater pollutant source characteristics from measured pollutant concentrations at several monitoring locations is generally an ill-posed and sometimes non-unique inverse problem. Linked simulation-optimization based methodologies have evolved as effective tools capable of solving this problem.

One of the important issues in estimation of unknown contamination sources is the release history reconstruction. It is generally assumed that reliable information on potential source locations and their time of activity is available from background studies of anthropogenic activities on a contaminated site. In such cases, only the release history of the pollutant sources is unknown.

Some of the main limitations in accurate source characterization are:

1. Sparsity of concentration measurement data.
2. Inefficient monitoring network for concentration measurements.
3. Difficulty in establishing the time of pollutant source activity initiation.
4. Applicability of optimal source characterization to distributed sources.
5. Problems associated with achieving a global optimal solution efficiently.

In order to address some of these limitations, initially a linked simulation-optimization model for optimal source characterization is developed using adaptive simulated annealing (ASA) as the optimization algorithm. Performance of the ASA based methodology was compared with a source characterization method using genetic algorithm (GA) for optimization, in terms of their ability to handle uncertainties and efficiency of convergence. Using illustrative aquifer examples, it was shown that ASA converges faster and produces better results even with erroneous measurement data and with uncertainties in hydraulic conductivity and porosity.

A more complex scenario exists when no reliable information is available on the potential location or initial time of activity of sources. Apart from this, the frequency of measurement at monitoring wells may not be uniform and some measurements might be missing in practical situations.

A methodology is developed to generate initial estimates of source characteristics such as source location and to estimate the initial time of activity from pollutant concentration measurements obtained from a single location where the contamination was first detected. dynamic time warping (DTW) distance is used to minimize errors in estimation of source characteristics arising from improper alignment of estimated and observed concentration data on the temporal axis.

Performance of this methodology is evaluated using data obtained from both an illustrative site and an actual contaminated site. Based on these estimates, a methodology is developed to design a monitoring network to generate concentration measurement information aimed at obtaining more reliable estimates of source characteristics. This methodology is implemented for a real contaminated site and it was found that the use of developed methodology results in reliable estimates of source characteristics with a far lesser number of monitoring wells.

The source characterization methodology is then extended for estimation of release history of distributed pollutant sources in a realistic scenario. Distributed sources in an abandoned mine site were considered for this purpose. A conceptual flow model is developed and calibrated for an abandoned mine site in South-East Queensland. Various illustrative scenarios of contamination are considered for evaluating the performance of this developed methodology. It was shown that the developed methodology is potentially applicable for estimation of distributed source characteristics.

When management measures are implemented to control contamination in a groundwater aquifer, measured concentration values are the resultant effect of natural transport and control measures. This can produce incorrect estimates for source characteristics. The methodology for release history reconstruction can be applied to managed contaminated sites by incorporating the proposed management strategy into the groundwater flow or transport model. This is illustrated by incorporating contamination management strategies already in place into the groundwater flow and transport model for an abandoned mine site with some degree of existing control measures.

Contents

List of Figures

List of Tables

Partial List of Symbols and Abbreviations

$\Delta\mathbf{x}$	Finite Difference grid spacing in x-direction
$\Delta\mathbf{y}$	Finite Difference grid spacing in y-direction
$\Delta\mathbf{z}$	Finite Difference grid spacing in z-direction
θ	Porosity of the porous medium, dimensionless
ϑ_i	Seepage or linear pore water velocity, (LT^{-1})
C_{ns}	Simulated concentration
C_{pert}	Perturbed contaminant concentration value
D_{ij}	Hydrodynamic dispersion coefficient, (L^2T^{-1})
K_{ij}	The hydraulic conductivity tensor (LT^{-1})
S_s	The specific storage of the porous media (L^{-1})
S_{ud}	A uniform random number between -1 and +1
ASA	Adaptive Simulated Annealing
C	Concentration of pollutants dissolved in groundwater, (ML^{-3})
DERM	Department of Environmental Resources Management, Govt. of Queensland, Australia.
DTW	Dynamic Time Warping
GA	Genetic Algorithm
GMU	Groundwater Management Unit
h	Potentiomentric Head (L)
NAEE	Normalized absolute error of estimation
NSW	New South Wales, Australia
SA	Simulated Annealing
SRTM	Shuttle Radar Topography Mission
t	The time (T)
UNEP	United Nations Environment Programme

Chapter 1

Introduction

Groundwater resources are susceptible to contamination from pollutants generated by anthropogenic agricultural and industrial activities. Often, when the contamination is first detected, little is known about the various characteristics of the pollutant source. This study presents methodologies for unknown groundwater pollutant source characterisation using linked simulation-optimization approach. The performance of these methodologies are evaluated for practical scenarios of point and distributed sources of groundwater pollution.

Water is essential to support life and for the preservation of environment. Due to growing populations of human beings and associated increase in anthropogenic activities such as agriculture and industrialisation, water demand has risen sharply. According to the recent estimates published by UNEP, global water consumption has tripled over the past 50 years (Gaddis et al., 2012). In many parts of the world, groundwater is the primary source for irrigation and drinking water supply. With increasing dependency of some regions on groundwater, global groundwater abstraction rate has more than doubled between 1960 and 2000 (Wada et al., 2010). Total global groundwater abstraction was estimated to be about 600-700 km^3yr^{-1} in 2003. In the same year, groundwater was estimated to meet at least 50% of potable water supplies; 40% of the demand from those industries that do not use mains water, and 20% of water use in irrigated agriculture. Dependence on groundwater is particularly becoming prominent in expanding urban communities. It has been estimated that about 1500 million urban dwellers worldwide depend on well, borehole and spring sources (Foster and Chilton, 2003).

Given the importance of groundwater in meeting current global water demand, sustainability of groundwater resources is vital for ensuring long-term water security. A major constraint to the sustainability of groundwater resources is quality deterioration. Among various reasons for groundwater quality degradation, anthropogenic contamination and salination of groundwater are the most important (Morris et al., 2003). Groundwater resources are polluted mainly because of pollutant streams generated by rapidly increasing industrial activities and use of chemicals in agriculture. Some of the more severe incidents of groundwater pollution with large plumes of high concentration pollutants are associated with industrial point sources from major accidental spillage or casual discharge in highly vulnerable areas (Foster and Chilton, 2003).

The realization of increasing vulnerability of groundwater resources to pollution has necessitated development of efficient techniques for prevention, detection and remediation of contaminated groundwater aquifers. Contamination of groundwater resources can often remain undetected for significant periods of time. In order to develop methodologies for effective and economical remediation of groundwater contamination, it is necessary to locate the source and predict the future course of groundwater contamination.

The first step in groundwater aquifer remediation should therefore be the characterisation of unknown pollutant sources. Reliable information about sources of pollution in terms of their location, release history and time of initial activity is highly important in planning effective remediation strategies. It is also important for estimating the extent, and assigning the liability for pollution.

1.1 Unknown Groundwater Pollutant Source Characterisation

The characteristics that define a groundwater source include:

1. Type of source (point, areal, etc.)
2. Spatial location and extent of the source
3. Point of time when the source first became active
4. Pollutant flux released as a function of time elapsed since start time.

Non-reactive groundwater flow and transport processes in a three-dimensional aquifer can be represented mathematically by the advection dispersion equation (ADE). This equation can be solved using numerical techniques. When groundwater source characteristics are known and the flow and transport parameters (i.e. hydraulic conductivity, porosity, etc.) of the porous media can be measured accurately, numerical simulation models can be used for predicting pollutant concentration at any given point in the study area with respect to time. When the sources of pollution are unknown, the measured pollutant concentration at various locations in the study area over a period of time is used to estimate the unknown source characteristics, by solving the advection dispersion equation backwards in time and space. Therefore, ascertaining various characteristics of the pollutant sources from available pollutant concentration measurements is an inverse problem.

Unknown groundwater pollutant source characterisation is an ill-posed inverse problem. An inverse problem is well posed if the following conditions are satisfied (Tikhonov and Arsenin, 1977):

- A solution exists
- The solution is unique, and
- The solution is stable.

The solution to this problem exists theoretically as one or more sources of contamination must be physically present in the study area to have caused the contamination. However, aquifer parameters used in governing equations of groundwater flow and transport are often not known precisely. This makes it difficult to solve for the characteristics of the pollutant source mathematically. Hence, the existence of a solution for inverse mathematical model of groundwater transport is not guaranteed. Moreover, similar pollutant plumes can be produced by several different combinations of source characteristics. This means a groundwater pollutant source identification problem has non-unique solutions. Groundwater transport equations are solved using numerical methods. These methods are not stable in reverse time. Hence, the stability of solution cannot be guaranteed as well.

Identification of groundwater pollutant sources and their characterization is complicated because of inadequacy of measured concentration information and due to uncertainty in available information. Most often, the only information available is the pollutant concentration measured in one or more affected wells, average porous media properties, and some possible guesses about the location of the pollutant source. Out of all the characteristics listed above, type of source is often obvious. In some cases the groundwater pollutant source location could be obvious from preliminary investigations. If an exhaustive record of pollutant inventory and industrial activities of the area is available, it may be possible to infer other characteristics of the source, mainly the start time and release pattern. However, in a number of instances, one or more of the listed characteristics of the source remains unknown at the time of detection of contamination in an aquifer, either due to inaccessibility of the pollutant source or due to lack of any previous information. In such cases the source characterization has to be undertaken by using measured information from a set of monitoring wells. Monitoring wells can provide point information about the pollutant concentration, potentiometric head and hydraulic conductivity. Solution of this problem is highly sensitive to measurement errors either in the observation data or model parameters (Sun, 1994).

1.2 Linked Simulation-Optimization Approach

One of the earlier methods used to identify unknown pollution source was to run forward simulations and try to match the results with observed data. This is an inefficient and exhaustive approach which may not perform satisfactorily. A more direct and efficient approach is to use an optimization approach. However, any optimum decision based on inadequate simulation of the physical processes in the groundwater system is almost meaningless. Therefore, a proper optimization based methodology for groundwater pollution source identification should incorporate a simulation of the physical process. This method is known as linked simulation-optimization approach.

Earlier implementations of this approach used linear programming and response matrix along with forward simulations (Gorelick et al., 1983). More recent developments incorporated the nonlinear nature of the aquifer processes, and use more efficient optimization algorithms, and linked simulation-optimization approaches (Mahar and Datta, 1997, 2000, 2001; Singh and Datta, 2006; Atmadja and Bagtzoglou, 2001b; Chadalavada et al., 2011a). Recently, however, evolutionary algorithms such as genetic algorithm and simulated annealing have been used for optimization (Mahinthakumar and Sayeed, 2005; Singh and Datta, 2006; Yeh et al., 2007; Chadalavada et al., 2011b). Use of evolutionary algorithms makes it easier and computationally more efficient to link the optimization algorithm with a simulation model.

1.3 Research Objectives

Existing methodologies for unknown groundwater pollution source characterization have several limitations. Some of the main limitations are:

1. Sparsity of concentration measurement data.
2. Inefficient monitoring network for concentration measurements.
3. Difficulty in establishing the time of pollutant source activity initiation.
4. Applicability of optimal source characterization to distributed sources.
5. Problems associated with achieving a global optimal solution efficiently.

This study aims to develop a methodology based on linked simulation-optimization approach for faster and efficient characterization of unknown groundwater pollutant sources. Estimation of characteristics of the unknown groundwater pollutant sources is even more difficult when the time at which the source first became active is unknown. Another problem is the sparsity of measured concentration data. Often, in real monitoring scenarios, some measurements are missing or the time interval between measurements is not uniform. This study presents a methodology utilizing dynamic time warping (DTW) distance to address both these limitations. The developed methodology has been extended for application to distributed sources on managed sites. Specific objectives of this study are:

1. Develop a simulation-optimization based methodology for unknown groundwater contamination source identification.

2. Incorporate uncertainties and measurement errors in the developed simulation-optimization approach to handle uncertainty in measurements or in estimated hydrogeological parameters.

3. Develop a methodology to obtain initial estimates of source characteristics such as source location and initial time of source activity using concentration measurements from a single location.

4. Design of a dedicated monitoring network for more efficient source identification using the initial estimates of source characteristics.

5. Extend the developed methodology for application to sites where some form of remedial management strategies have already been implemented.

6. Performance evaluation of the developed methodologies using aquifer data.

7. Evaluate the applicability of developed methodology for specific application to mining sites with distributed pollutant sources.

1.4 Organization of the Thesis

This thesis contains seven chapters including the introduction. Chapter 2 of the thesis presents the state-of-the-art on various techniques used in this study.

Chapter 3 presents a linked simulation-optimization model for optimal source characterization using adaptive simulated annealing (ASA) as the optimization algorithm. Performance of the ASA based methodology is compared with a source characterization method using genetic algorithm (GA) for optimization, in terms of their ability to handle uncertainties and efficiency of convergence. Using illustrative aquifer examples, it is shown that ASA converges faster and produces better results, even with erroneous concentration measurement data and with uncertainties in hydraulic conductivity and porosity.

Chapter 4 presents a more complex scenario when no reliable information is available on the potential location or initial time of activity of sources. Apart from this, the frequency of measurement at monitoring wells may not be uniform and some measurements might be missing in practical situations. A methodology is developed to obtain initial estimates of source characteristics such as source location and to estimate the initial time of activity from pollutant concentration measurements obtained from a single location where the contamination was first detected. Dynamic time warping (DTW) distance is used to minimize errors in estimation of source characteristics arising from improper alignment of estimated and observed concentration data on the temporal axis.

Chapter 5 presents the performance evaluation of the methodology developed in Chapter 4 using simulated data for an illustrative site, and using data obtained from an actual contaminated site. Based on these estimates, a methodology is developed to design a monitoring network to generate concentration measurement information aimed at obtaining more efficient and reliable estimates of source characteristics. This methodology is evaluated for a real contaminated site and it was found that the use of the developed methodology results in reliable estimates of source characteristics with a far lesser number of monitoring wells.

Chapter 6 presents an extension of the source characterization methodology developed in Chapter 3 for estimation of release history of distributed pollutant sources to a managed contaminated site in a realistic scenario. Distributed sources in an abandoned mine site are considered for this purpose.

Chapter 7 presents a summary of the salient points that have been addressed, and the major conclusions of this study.

Chapter 2

Review of Literature

This chapter briefly discusses the body of literature that is relevant to solving unknown groundwater pollutant source characterisation problem. The first part of this chapter describes various approaches to solve the problem of unknown groundwater pollutant source characterisation. It also discusses the variations of this problem based on pollutant source characteristics that are unknown and need to be estimated.

The second section of this chapter describes various methodologies developed for monitoring network design, particularly in the context of unknown pollutant source identification.

The final section presents an overview of literature on various tools and techniques used for groundwater flow and transport simulation, representation of uncertainty, optimization and pattern recognition in this study.

2.1 Unknown Groundwater Pollutant Source Characterisation

When a contamination event is initially detected in a groundwater aquifer, several characteristics of the pollutant sources are unknown. Groundwater pollutant source characterisation involves the identification of the magnitude, location and duration of unknown pollutant sources. Estimation of unknown characteristics of an unknown groundwater pollutant source from measured contaminant concentrations at various locations in the study area over a period of time involves solving the advection dispersion equation backwards in time and space. Therefore, it is an inverse problem.

Several methods have been suggested to solve this inverse problem. These methods can be broadly classified as heat transport inversion, analytical solutions and regression, deterministic direct methods, probabilistic and geo-statistical simulation approaches and optimization approaches. A detailed review of these methodologies can be found in Atmadja and Bagtzoglou (2001b); Michalak and Kitanidis (2004); Bagtzoglou and Atmadja (2005) and Sun et al. (2006a,b).

Since the mathematical model of heat and mass transfer is similar to that of groundwater flow and transport, solutions for ill-posed inverse problems are applicable to such problems in groundwater flow and transport as well. Most of the solutions in this category involve using approximations for inverse modelling and sometimes a method to eliminate ill-posedness. One of the methods for eliminating ill-posedness is Tikhonov-Regularization (Tikhonov and Arsenin, 1977). Parameters used in mathematical models of heat and mass transfer are homogeneous and they can be accurately measured. This is not always possible in a groundwater aquifer which is mostly non-homogeneous and the parameters such as hydraulic conductivity and porosity are not easily measurable at every point in the aquifer. Hence, the use of heat transport inversion methods has been limited.

Analytical solutions and regression as well as deterministic direct methods rely on inversing the mathematical solutions of the governing equations of groundwater flow and transport. Some of the significant studies using these methods are Skaggs and Kabala (1995); Sidauruk et al. (1998); Alapati and Kabala (2000). However, these methods have limited application as they assume a homogeneous aquifer with simple geometry and flow conditions.

Probabilistic and geo-statistical simulation approaches represent the groundwater transport process as a stochastic model. These methods aim at solving stochastic differential equations backward in time.

Geo-statistical techniques are used to better represent the heterogeneity in porous media properties. Atmadja and Bagtzoglou (2001a); Bagtzoglou and Atmadja (2005) used this method to present a probabilistic framework to identify solute sources in heterogeneous media. Snodgrass and Kitanidis (1997) used probabilistic approach based on Bayesian theory and geo-statistical techniques. This study assumed the source locations to be known. Probabilistic and geo-statistical approaches can address the problem of non-homogeneity in the porous media parameters. However, this involves solving the governing stochastic equations backward in time and requires extensive computational resources.

2.1.1 Linked Simulation-Optimization Approach of Unknown Groundwater Contaminant Source Characterisation

A number of optimization based methodologies for unknown groundwater contaminant source identification have been proposed by a number of researchers. First attempt in this regard was made by Gorelick et al. (1983). In this attempt, they formulated the source identification problem as forward-time simulations coupled with an optimization model using linear programming and response matrix approach. The solute transport model was implemented as a series of constraints in the form of a concentration response matrix. Aquifer parameters were assumed to have no uncertainty. The main limitation of this method is that it is applicable generally to linear systems.

Wagner and Gorelick (1986) used nonlinear multiple-regression to estimate aquifer parameters and coefficient of zero order production for a one-dimensional hypothetical system. Estimation of the linear source term was found to be highly sensitive to the introduction of measurement errors.

The first attempt to estimate model parameters along with source characterization was implemented by Wagner (1992). Wagner used an inverse model as a non-linear maximum likelihood estimation problem. Estimates of hydro-geological and source parameters were based on measurements of hydraulic head and contaminant concentration. Steady confined groundwater flow and transient, non-reactive, single species transport was assumed for the example problem. When the contaminant flux was assumed to be unknown along with model parameters, this method estimated the model parameters within 30% of their actual values and the source fluxes were overestimated by about 20%.

Mahar and Datta (1997, 2001) developed an embedded simulation-optimization approach combining optimal identification of a pollutant source with the design of a groundwater quality monitoring network in order to improve on the efficiency of the identification process. The method was applied to a 2-D homogeneous, isotropic, and saturated aquifer. They proposed a two-step methodology in which an optimization model was used to identify an unknown pollution source based on observation data. In the next step, different realizations of pollutant plumes were simulated using perturbed sources. On obtaining these realizations, integer programming was used to determine the optimal locations of the monitoring wells. The concentrations measured in these wells were used in the nonlinear optimization model to obtain a more accurate estimation of sources. Mahar and Datta (2000) were also able to estimate the magnitude, location and duration of pollutant sources using nonlinear optimization technique.

Aral et al. (2001) formulated a contaminant source characterization problem as a nonlinear optimization model, in which contaminant source locations and release histories were defined as explicit unknown variables. The optimization model selected was the standard model, in which the residuals between the simulated and measured contaminant concentrations at observation sites were minimized. Simulated concentrations at the observation locations were implicitly embedded into the optimization model through the solution of groundwater flow and contaminant fate and transport simulation models. To simplify this computationally intensive process, they used progressive genetic algorithm (PGA) for the solution of the nonlinear optimization model.

Singh et al. (2004); Singh and Datta (2004) used a trained artificial neural network (ANN) to simultaneously solve the problems of estimating unknown pollution sources and estimating hydrogeological parameters. The universal function approximation property of a multilayer, feed-forward ANN was utilized to estimate temporally and spatially varying unknown pollution sources, as well as to provide a reliable estimation of unknown flow and transport parameters. ANN was trained on patterns of simulated data using a back-propagation algorithm. A set of source fluxes and temporally varying simulated concentration measurements constituted the pattern for training. Performance of this methodology was evaluated under varying concentration measurement errors.

Mahinthakumar and Sayeed (2005) investigated and compared several hybrid optimization approaches that combine genetic algorithms with a number of local search approaches for solving these problems. The example problems used contained both single and multiple-source releases in three-dimensional heterogeneous flow fields. A parallel computing environment was used to handle the heavy computational needs of these problems. The results indicate that hybrid optimization methods, especially those that combine an initial global heuristic approach (e.g. genetic algorithms) with a subsequent gradient-based local search approach (e.g. conjugate gradients) are very effective in solving these problems.

A genetic algorithm (GA) based simulation optimization approach was used for optimal identification of unknown groundwater pollution sources by Singh and Datta (2006). A flow and transport numerical simulation model was externally linked to the GA-based optimization model to simulate the physical processes involved. The simulation model used potential pollution source characteristics that are evolved by the GA and simulates the resulting concentration measurement values at observation locations. These simulated spatial and temporal pollutant concentration measurement values were used to evaluate the fitness function value of the GA. This approach makes it feasible to solve the source-identification problems for complex aquifer study areas with multiple unknown pollution sources.

Yeh et al. (2007) proposed an approach that combines simulated annealing

(SA), tabu search (TS), and the three-dimensional groundwater flow and solute transport model (MODFLOW-GWT). It was used to estimate source location, release concentration, and release period of the source. The sampling concentrations at monitoring points were simulated by the MODFLOW-GWT with an assumed release concentration and release period at a known source location. In the source estimation process, the source location was selected by TS within the suspected source area, and the trials for release concentrations and release periods were generated by SA. MODFLOW-GWT was utilized to compute the simulated concentrations at the monitoring points with the trial solution. The above mentioned procedures were repeated until the stopping criterion regarding the differences of objective function value (OFV) was met. Six studies on a homogeneous site, two studies on the heterogeneous site, and one study on the transient flow problem were conducted in this study.

He et al. (2009) studied a coupled simulation-optimization approach for optimal design of petroleum contaminated groundwater remediation under uncertainty. Compared to the previous approaches, it had the advantages of: (1) addressing the stochasticity of the modelling parameters in simulating the flow and transport of NAPLs in groundwater, (2) providing a direct and response-rapid bridge between remediation strategies (pumping rates) and remediation performance (contaminant concentrations) through the created proxy models, (3) alleviating the computational cost in searching for optimal solutions, and (4) giving confidence levels for the obtained optimal remediation strategies. The approach was applied to a site in Canada for demonstrating its performance.

Datta et al. (2009c) developed a methodology for simultaneous pollution source identification and parameter estimation in groundwater systems. The groundwater flow and transport simulator that serves as a binding constraint was linked to the nonlinear optimization model as an external module. This methodology was aimed at addressing some of the computational limitations of using the embedded optimization techniques. Performance of the proposed methodology using spatio-temporal hydraulic head values and pollutant concentration measurements was evaluated by solving illustrative problems. They found that the solution results obtained using the proposed methodology were better than those obtained using the embedded optimization technique.

Datta et al. (2011) proposed a linked simulation-optimization based source identification methodology using a classical nonlinear optimization model linked to a groundwater flow and transport simulation model. The essential link between the simulator and the optimization method were the derivatives or gradient information required for the optimization algorithm. They concluded that the proposed methodology was potentially applicable to large scale study areas with multiple unknown pollution sources and eliminates some of the computational limitations of embedded optimization techniques.

Based on the characteristics that are unknown, groundwater contaminant source characterization problems can be classified into various categories (Pinder, 2009):

1. Reconstruction of source release history problems
2. Identification of source location or release time of contaminant.
3. Identification of source location and magnitude.
4. Identification of source location and release time of contaminant.

5. Identification of location, magnitude of source and release time of contaminant.

The last category is the most challenging (Pinder, 2009). In this study, this is the problem that has been addressed.

2.2 Monitoring Network Design for Unknown Groundwater Contaminant Source Characterisation

Monitoring network design problems for groundwater management have been widely studied with different objectives. Design objectives of monitoring networks vary widely depending on the need for design. Fundamental approaches of monitoring network design for groundwater quality management are a natural extension of observation network design for meteorology. Detailed reviews of methods implemented for monitoring network design are reported in Loaiciga et al. (1992), Minsker (2003), and Kollat et al. (2011). Most of the research on this subject focuses on monitoring networks for leak detection from known contaminant sources such as landfills, large spills and historically contaminated lands.

Optimization based methodologies for monitoring network design have considered a wide range of objectives. Massmann and Freeze (1987) proposed a methodology for designing a monitoring network for contamination detection to be located between the source and the regulatory compliance surface of a landfill site. They used stochastic contaminant transport simulations to calculate the probability of detection of the monitoring network. It was assumed that the contamination is brought about by the release of a single, inorganic nonradioactive species into a saturated, high-permeability, advective, steady state horizontal flow system which can be analyzed with a two-dimensional analysis.

Loaiciga (1989) proposed a mixed integer programming formulation based approach. The proposed model was tested on a real aquifer and it was concluded that it results in optimal monitoring policy. Meyer and Brill (1988) presented a method for the optimal monitoring design network using the maximum covering location problem (MCLP) formulation. The MCLP maximizes the demand served within the maximal service distance given a fixed number of facilities. Knopman and Voss (1989) proposed a multi-objective formulation of sampling network design for site characterization studies. They considered optimal design of a sampling network as a sequential process in which the next phase of sampling is designed on the basis of all available physical knowledge of the system. They considered three objectives: model discrimination, parameter estimation and cost minimization.

McKinney and Loucks (1992) proposed a new network design algorithm for improving the reliability of groundwater simulation model predictions. Their objective was to minimize the simulation model prediction variance choosing optimal monitoring locations. Variance of predicted state variables, hydraulic head and contaminant concentration was used as a measure of model prediction reliability in this study. This method was implemented to design a monitoring network and the authors showed that a significant increase in simulation model prediction reliability is achieved by measuring aquifer properties at locations selected by the algorithm. Meyer et al. (1994) proposed a method that incorporates system uncertainty in monitoring network design and provides network alternatives that are optimal with respect to several objectives of designing monitoring networks. They considered three design objectives. (1) minimize the number of monitoring wells, (2) maximize the probability of detecting a contaminant leak, and (3) minimize the expected area of contamination at the time of detection. Yenigül et al. (2005) presented a reliability assessment to estimate the performance of groundwater monitoring systems at landfill sites. They presented a hypothetical problem where the detection probabilities of several monitoring systems are compared. Using a Monte Carlo approach to incorporate uncertainties due to subsurface heterogeneity and the leak location they showed that lateral dispersivity of the medium has a significant influence on the reliability of the monitoring system. They also demonstrate that the number and the location of the monitoring wells is dependent on the heterogeneity of the medium and size of the contaminant leak.

Cieniawski et al. (1995) extended the work of Meyer and Brill (1988) on the optimal location of a network of groundwater monitoring wells under conditions of uncertainty using genetic algorithms (GAs).

Datta and Dhiman (1996) developed a mathematical model for designing a groundwater quality monitoring network using a linked simulation-optimization model. They formulated the model using chance constraints and solved it by using a mixed-integer programming algorithm. Their model incorporates uncertainties in the prediction of pollutant movement in the saturated zone. Nonlinearities due to the inclusion of cumulative distribution functions (CDFs) of actual spatial concentrations were accommodated in the optimization model through a piecewise linearization scheme.

The first attempt to link optimal groundwater pollution source identification with the measurement data collected from a designed monitoring network was reported in Mahar and Datta (1997). They developed a methodology combining an optimal groundwater quality monitoring network design with an optimal source identification model. They used a three-step methodology. In the first step an embedded nonlinear optimization model was utilized for preliminary identification of pollutant sources. The second step utilized these preliminary identification results and a linked simulation-optimization approach to design an optimal monitoring network that could be implemented in the subsequent time periods. In the third step, the observed concentration data at the designed monitoring well locations were utilized for more accurate identification of the pollutant sources.

Hudak (1998) developed a method for designing configurations of monitoring wells, consisting of vertically nested intakes in boreholes. This methodology was tested on a 32 ha solid waste landfill in Tarrant County, Texas, USA. The objective of investigation was to design a monitoring network that is able to minimize the un-detection of contaminant plumes in the study area. Results of this study illustrated a practical need for structured approaches to designing detection-based groundwater monitoring configurations.

Reed and Minsker (2004) used high-order Pareto optimization (i.e. optimizing a system for more than two objectives) on a long-term monitoring (LTM) application. Their application combined quantile kriging and the non-dominated sorted genetic algorithm-II (NSGA-II) to successfully balance four objectives: (1) minimizing sampling costs, (2) maximizing the accuracy of interpolated plume maps, (3) maximizing the relative accuracy of contaminant mass estimates, and (4) minimizing estimation uncertainty.

Herrera and Pinder (2005) proposed a method for the space-time optimization of monitoring networks for groundwater quality. The objective of their study was to minimize the total cost of sampling groundwater contaminants by estimating optimal monitoring locations and optimal sampling frequency. This method used Kalman filter coupled with a stochastic transport model in which velocity and dispersion are spatially correlated random fields to consider the combined spatial and temporal redundancy of the sampling network. The objective of optimization was to minimize the estimated variance of monitored parameters. Kalman filter was used again to obtain real-time update of the estimates. Synthetic examples were presented to show that for a contaminant plume in motion this method can obtain cost-effective sampling networks.

Kollat and Reed (2007) presented a detailed assessment of how increasing problem sizes (measured in terms of the number of decision variables being considered) impacts the computational complexity of using multiple objective evolutionary algorithms (MOEAs) to solve long-term groundwater monitoring (LTM) applications.

Dhar and Datta (2007) proposed a methodology for optimal design of a time varying monitoring network that has wells installed in stages. Their optimization model incorporates uncertainties in prediction or estimation of some of the aquifer parameters such as hydraulic conductivity and dispersivity.

Chadalavada and Datta (2008) developed optimal groundwater pollution monitoring network design models to prescribe optimal and efficient sampling locations for detecting pollution in groundwater aquifers. Multiple realizations of pollutant plumes in a two-dimensional aquifer were generated incorporating the uncertainty in both source and aquifer parameters. These concentration realizations were incorporated in the optimal monitoring network design models. Two different objectives were considered separately. The first objective function minimizes the summation of unmonitored concentrations at different potential monitoring locations and the second minimizes estimation variances of pollutant concentrations at various unmonitored locations. The first objective function minimizes the probability of choosing monitoring locations with low concentrations and the second results in a design that chooses optimal monitoring locations where the uncertainties in simulated concentrations are large. The developed optimization models were solved using genetic algorithm. The variances of estimated concentrations at potential monitoring locations were computed using kriging. The solution results were evaluated for an illustrative study area and performance evaluation results established the potential applicability of this methodology.

Dhar and Datta (2010) developed a methodology based on an optimization model solution for optimal design of a groundwater quality monitoring network. The developed methodology addressed the issue of redundancy in monitoring network results in the optimal design of a monitoring network. The methodology interpolates concentration data spatially using inverse distance weighting method. A logic-based mixed-integer linear optimization model was formulated and solved using the branch-and-bound algorithm. The proposed methodology was tested for a real world problem and its performance was evaluated for different scenarios using available historical concentration data. These performance evaluation results showed that the proposed methodology performs satisfactorily when compared with other existing methodologies.

Chadalavada et al. (2011a) proposed models for the design of monitoring networks to improve efficiency of source identification. In this study, a new approach of monitoring network design is presented assuming no prior information on location, release history or magnitudes of the contaminant source. The developed methodology also assumed that the contaminant had only been detected at one location and uses the observation recorded over several periods at this location to design an optimal monitoring network.

2.3 Relevant Tools and Techniques

This section is intended to present relevant literature on several tools for groundwater flow and transport simulation, optimization, pattern recognition and uncertainty representation that have been used through various stages in this study.

2.3.1 Flow and Transport Modelling

Fundamental mathematical models representing groundwater flow and transport have been discussed in detail by Javandel et al. (1984) and Fetter (1994). A number of numerical simulation models using finite difference and finite element methods have been developed to solve these governing equations. A detailed discussion of the developed methods has been presented by Anderson and Woessner (1992) and Zheng and Bennett (1995). McDonald and Harbaugh (1988) developed a finite difference based modular three-dimensional groundwater flow model. This model was named MODFLOW and has been widely used in groundwater flow simulations. MODFLOW has continuously been updated and the most recent version was released in 2005. Zheng (1990) developed a modular 3-D transport model for simulation of various transport processes such as advection, dispersion and chemical reactions of contaminants in groundwater systems. This model was called MT3D and it has a comprehensive set of solution options including method of characteristics (MOC), the modified method of characteristics (MMOC), a hybrid of these two methods (HMOC), and the standard finite difference method (FDM). Zheng and Wang (1999) extended the capabilities of MT3D to include a multi-component program structure which can accommodate add-on reaction packages for modelling various biological and geochemical reactions. The solving methods were augmented and an option to include non-equilibrium sorption and dual-domain advection-diffusion mass transport. Clement (1998) presented another modular computer code for simulating reactive multi-species transport in three-dimensional groundwater systems. The model is called RT3D and it provides a flexible framework to simulate natural attenuation, accelerated bio-remediation or other reactive transport modelling scenarios. The program also has an option to add any reaction kinetics for multiple aqueous and sorbed phase species.

2.3.2 Techniques for Parameter Uncertainty Representation

Generally, average values of aquifer parameters such as hydraulic conductivity and porosity are used in the groundwater flow and transport models. In reality, however, these parameters can have a different value at each location in space. Hence, they can be represented most realistically by a stochastic set of values defined by a probability distribution. One of the earliest investigations into stochastic-conceptual analysis of one-dimensional groundwater flow was carried out by Freeze (1975). This study concluded that values of hydraulic conductivity follow a log-normal distribution whereas those of porosity follow normal distribution. This study analyzed groundwater flow in non-uniform media using stochastic-conceptual approach in which the effects of stochastic parameter distributions on predicted hydraulic heads were analyzed with the aid of a set of Monte Carlo solutions to the pertinent boundary value problems.

Dagan (1982) presented a methodology to solve the inverse problem of determining transmissivity at various points, given the shape and boundary of the aquifer and recharge intensity and given a set of measured log-transmissivity Y and head H values at a few points.

Kitanidis (1986) examined the effect of parameter uncertainty in a Bayesian framework with emphasis on the derivation of the Bayesian distribution (and its first two moments) of unknown quantities given some measurements. This distribution accounts not only for natural variability but also for parameter uncertainty. It was shown that when both drift and covariance function parameters are uncertain, the Bayesian distribution is generally not Gaussian, and the Bayesian conditional mean is a nonlinear estimator. The case of diffuse priors was examined in some detail; it was shown that the posterior distribution of the covariance function parameters is given by the restricted likelihood function, i.e. the likelihood function of generalized increments. The results provided insight into the applicability of maximum likelihood versus restricted maximum likelihood parameter estimation, and conventional linear versus kriging estimation.

Andricevic and Kitanidis (1990) presented an optimization methodology for aquifer remediation using differential dynamic programming. This method accounts for and reduces parameter uncertainty. The methodology uses a dual-control method in which system parameters are improved and the aquifer parameter is managed to achieve the specified objectives at minimal cost. The methodology was applied to a hypothetical one-dimensional system. This methodology was extended to the case of two-dimensional groundwater systems by Lee and Kitanidis (1991).

Detailed characterization of the spatial distribution of hydrogeological parameter values in an aquifer is also described in Yeh (1992) and Gelhar (1993).

2.3.3 Optimization Algorithms

Choice of optimization algorithm largely depends on the type of problems to be solved. Groundwater pollutant source characterization is a complex multi-variate optimization problem. Hence, heuristics based methods were chosen so as to obtain maximum computational efficiency without sacrificing accuracy of obtained solutions. Two of the most popular optimization algorithms in this category are simulated annealing and genetic algorithm.

In simulated annealing, a current solution may be replaced by a random "neighbourhood" solution chosen with a probability that depends on the difference between corresponding function values and on a global parameter T (called temperature) that is gradually decreased in the process (Kirkpatrick, 1984). Of the various simulated annealing implementations, it is evident in literature that the adaptive simulated annealing algorithm converges faster (Ingber and Rosen, 1992) while maintaining the reliability of results and hence it was preferred over traditional Boltzmann annealing implementation (Kirkpatrick, 1984). Yeh et al. (2007) used simulated annealing (SA), Tabu Search (TS), and the three-dimensional groundwater flow and solute transport model (MODFLOW-GWT) to estimate source location, release concentration, and release period of the source. In the source estimation process, the source location was selected by TS within the suspected source area, and the trials for release concentrations and release periods were generated by SA. This study attempts to use adaptive simulated annealing for enhanced computational efficiency for source characterisation. However, use of combination of TS with SA, in which TS is used for screening purposes, was computationally not efficient.

Genetic algorithms (GAs) are population based search strategies which are popular for many difficult to solve optimization problems including inverse problems. GAs emulate the natural evolutionary process in a population where the fittest survive and reproduce (Holland, 1975). GA-based search performs well because of its ability to combine aspects of solutions from different parts of the search space.

Simulated annealing, as an algorithm, is very efficient in solving non-convex optimization problems by ensuring that it does not always move downhill on a complex non-convex search space and hence avoids getting trapped in local minimum. Simulated annealing also differs significantly from conventional iterative optimization algorithms in that gross features of the final state of the system are seen at higher temperatures, whereas the finer details of the state appear at lower temperatures (Haykin, 1999). The fact that simulated annealing ensures a global optimal solution enhances its suitability for solving ill-posed inverse problems in general, and the problem of unknown groundwater pollutant source characterization in particular.

Its ease of use and remarkable efficiency in handling complex objective functions and constraints has made simulated annealing an attractive choice for solving a wide range of complex optimization problems. However, the slow convergence and hence long time of execution of standard Boltzmann-type simulated annealing has been a constraint (Ingber, 1996). Adaptive simulated annealing removes that constraint by making the annealing schedules decrease exponentially in annealing time, thereby making the convergence much faster. A major difference between ASA and traditional Boltzmann annealing algorithms is that the ergodic sampling takes place in terms of n parameters and the cost function. In ASA the exponential annealing schedules permit resources to be spent adaptively on re-annealing and on pacing the convergence in all dimensions, ensuring ample global searching in the first phases of search and ample quick convergence in the final phases (Ingber, 1996).

Another major advantages of using adaptive simulated annealing is also the fact that the parameters of algorithm are adjusted adaptively and hence the solutions do not vary widely if parameter values are changed within reasonable limits. This is in contrast with genetic algorithm, where even minor changes to parameters such as mutation probability, crossover probability or population size cause a significant difference in the solutions.

2.3.4 Techniques for Pattern Recognition and Classification

Contaminant concentration observations obtained from monitoring locations in a study area can be represented as a time series. In linked simulation-optimization methods, candidate values of various parameters are used to obtain estimated characteristic curve at the monitored locations. Pattern recognition techniques can be used to match the observed and estimated concentration time series. Datta et al. (1989) presented one of the earliest attempts to use pattern recognition for groundwater source characterization. They used statistical pattern recognition techniques to identify groundwater pollution source magnitudes. They also incorporated the simulation equations as response matrix in the model. They investigated the effects of parameter uncertainty and measurement errors on the source identification. The optimal groundwater pollution source identification model was used as a screening model to limit the number of pattern classes to be incorporated. The final pollution source characteristics was estimated using the expert system to accommodate imprecise knowledge regarding data reliability. The performance was found encouraging in general and specifically good under conditions of missing observed concentration data.

One of the main aims of the present study is to address one of the difficult scenarios of unknown groundwater pollution source characterization, i.e. to estimate the unknown sources when the exact time a source becomes active is completely unknown. To address this, a new methodology is developed using dynamic time warping as a distance measure to compute the difference between estimated and observed concentrations at various monitoring locations.

Dynamic time warping (DTW) distance is used as a measure of dissimilarity between two time series. It is a widely used pattern recognition tool, first proposed in the 1960s (Bellman and Kalaba, 1959) as a measure of speech signal dissimilarity. Since then, it has been used in a variety of applications and has been particularly popular in time series clustering and data mining applications (Rabiner and Juang, 1993). An in-depth review of dynamic time warping is presented in Senin (2008).

2.4 Motivation for this Study

Existing methodologies for unknown groundwater pollution source characterization have several limitations. Some of the main limitations are their inability to handle sparsity of concentration measurement data, lack of consideration for specific monitoring network design for concentration measurements, difficulty in establishing the time of contaminant source activity initiation, and problems associated with achieving a global optimal solution efficiently.

This study aims to develop a methodology based on linked simulation-optimization approach for faster and efficient characterization of unknown groundwater pollutant sources. Estimation of characteristics of the unknown groundwater pollutant sources is even more difficult when the time at which the source first became active is unknown. Another problem is the sparsity of measured concentration data. Often, in real-life monitoring scenarios, some measurements are missing or the time interval between measurements is not uniform. This study presents a methodology utilizing dynamic time warping (DTW) distance to address both these limitations. The developed methodology has been extended for potential application to distributed pollution sources in managed sites.

The issue of systematic and planned monitoring of pollutant concentration over time and space is also an important factor in efficient estimation of unknown sources of groundwater pollutants. This area needs further attention. This study therefore addresses the issue of designing monitoring networks to enhance the efficiency of source identification. Another issue that needs attention is the applicability of some of the developed methodologies to distributed pollution sources and managed sites. These issues have also been addressed in this study. Performance evaluation of the methodologies developed is carried out to estimate the potential applicability of these proposed methodologies. The next chapters present the methodologies utilized in this study.

Chapter 3

Three-Dimensional Groundwater Contamination Source Identification Using Adaptive Simulated Annealing

A similar version of this chapter has been published and copyrighted in the Journal of Hydrologic Engineering.

Jha, M., & Datta, B. (2012). Three dimensional groundwater pollution source identification using adaptive simulated annealing, to appear in: Journal of Hydrologic Engineering (ASCE), doi: 10.1061/(ASCE)HE.1943-5584.0000624.

In the event of detection of pollution in a groundwater aquifer, the first step generally is a detailed reconnaissance or audit of all available information on the history of pollution. It involves creating an inventory of past anthropogenic activities, particularly involving the chemical substances that were found to have contaminated the aquifer. In several cases, it is possible to generate reliable estimates on potential source locations and the time at which the source activity began. In such instances, only the magnitude of pollutant flux needs to be estimated as a function of time over the entire study period, utilizing available measurement information obtained after pollution detection. This chapter addresses the problem of estimating unknown groundwater source flux magnitude as a function of time given the potential source locations are known and that a reliable estimate of the time when the sources first became active is available. This chapter proposes a linked simulation-optimization based methodology, utilizing adaptive simulated annealing (ASA) for optimization, to solve this problem. Performance of the ASA based methodology is compared with that based on genetic algorithm (GA). The ASA based solution algorithm is shown to be computationally efficient for optimal identification of the source characteristics in terms of execution time and accuracy. This computational efficiency appears to prevail even with moderate levels of errors in estimated parameters and concentration measurement errors. Also, the pollutant concentration monitoring locations are shown to be critical in the efficient characterization of the unknown pollutant sources. Optimal identification results for different monitoring networks are presented to demonstrate the relevance of a suitable network to efficient source identification.

In linked simulation-optimization approaches a numerical groundwater flow and transport simulation model is linked to the optimization model. Most of these approaches, such as linear programming with response matrix approach (Gorelick et al., 1983), nonlinear optimization with embedding technique (Mahar and Datta, 1997, 2000, 2001), artificial neural network approach (Singh et al., 2004; Singh and Datta, 2004, 2007), constrained robust least square approach (Sun et al., 2006a,b), classical optimization based approach (Datta et al., 2009a,b, 2011) and genetic algorithm based approach (Aral et al., 2001; Mahinthakumar and Sayeed, 2005; Singh and

Datta, 2006) minimize an objective function representing the difference in measured concentration and simulated concentration at various monitoring locations.

Most of the recent implementations of simulation-optimization method adopt evolutionary global search approaches as the optimization model. In recent years, genetic algorithm (GA) and its variants have been widely used in solving groundwater problems, particularly unknown pollutant source identification (Hilton and Culver, 2005; Mahinthakumar and Sayeed, 2005; Singh and Datta, 2006). However, most of the implementations of simulation-optimization approaches assume a two-dimensional groundwater transport, or availability of observation data throughout the study time period or both.

Simulation-optimization based approaches are also computationally intensive owing to the fact that the simulation model has to be run thousands of times before an acceptable solution is produced. This is a major roadblock to any desktop implementation of the simulation-optimization approach. This chapter compares the performance of adaptive simulated annealing (ASA) (Ingber, 1993, 1996) based simulation-optimization approach to the one based on genetic algorithm in order to evaluate the performance using ASA especially in a time constrained scenario.

The ASA code was first developed in 1987 as very fast simulated re-annealing (VFSR) (Ingber, 1989). Ingber and Rosen (1992) showed that VFSR is at least an order of magnitude superior to genetic algorithms in convergence speed and is more likely to find the global optima during a time limited search. The performance evaluation of competing simulation-optimization approaches is based on a realistic scenario of missing measurement data, where pollutant concentration measurements are available a few years after the sources have ceased to exist. An illustrative three-dimensional aquifer is used for performance evaluation. Apart from the convergence speed, the two algorithms are compared for their ability to produce accurate source release histories with moderately erroneous data and with uncertainty in estimation of hydrogeological parameters.

Another important factor that affects the execution time and accuracy of solutions generated by linked simulation-optimization approaches is the choice of observation locations. Poorly chosen pollutant observation locations often produce misleading results and hence it becomes important that, after the initial estimation of the pollutant sources, a monitoring network is designed and implemented. Some researchers have attempted to link the pollutant monitoring locations with the efficiency of source identification (Bagtzoglou et al., 1991; Mahar and Datta, 1997; Dhar and Datta, 2010; Chadalavada et al., 2011a). Bagtzoglou et al. (1992) presented a method for probabilistic estimation of source locations and spill time histories. Mahar and Datta (1997) and Chadalavada et al. (2011a) proposed models for design of monitoring networks to improve efficiency of source identification. Limited evaluation results presented in this work attempt to demonstrate the relevance of a designed monitoring network and the shortcomings of arbitrary observation locations.

3.1 Methodology

The linked simulation-optimization approach consists of two parts. An optimization algorithm generates the candidate solutions corresponding to various unknown groundwater source characteristics. The candidate solutions are used as input in the numerical groundwater transport simulation model to generate the concentration of pollutant in the study area. The generated concentration at designated monitoring locations is matched with the observed values of pollutant concentrations at various time intervals at the same locations. The difference between simulated and observed concentration is then used to calculate the objective function value which is utilized by the optimization algorithm to improve the candidate solution. The process continues until an optimal solution is obtained.

A schematic representation of this process of using ASA as the optimization algorithm in a linked simulation-optimization model is presented in Figure 3.1. The classical simulated annealing (SA) algorithm has many associated guiding parameters such as the initial parameter temperature, annealing schedule, acceptance probability function, goal function, etc. Effective application of classical simulated annealing to a particular optimization problem normally involves a lot of trials and adjustments to achieve ideal values for all or most of these parameters. ASA, which is a variant of classical SA, helps overcome this difficulty to a certain extent by automating the adjustments of parameters controlling temperature schedule and random step selection, thereby making the algorithm less sensitive to user-defined parameters compared with classical SA. This additional ability of ASA, combined with inherent efficiency of classical

SA to find the global optimal solution even when multiple local optimums exist, makes it a natural choice for solving the groundwater pollutant source identification problem.

3.2 Simulation of Groundwater Flow and Transport

Physical processes of groundwater flow and transport have been modelled as mathematical equations and a number of methods exist for solving these governing equations. A discretized numerical method is used to solve the governing equations of groundwater flow and transport in this study.

3.2.1 Mathematical Representation of Groundwater Flow and Transport

Groundwater flow follows the same physical principles as any other fluid flowing through a porous media. It can be described by differential equations. As several variables affect the groundwater flow, it is generally described as a partial differential equation in space and time. Hence, spatial co-ordinates such as x, y, z, in Cartesian system and time (t) are independent variables. The governing equations of a fluid flow in porous media are derived by employing laws of mass and energy conservation. The partial differential equation governing groundwater flow is given by Equation 3.1 (Fetter, 1994).

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x}\right)+\frac{\partial}{\partial x}\left(K_{yy}\frac{\partial h}{\partial y}\right)+\frac{\partial}{\partial x}\left(K_{zz}\frac{\partial h}{\partial z}\right)+W=S_s\frac{\partial h}{\partial t} \tag{3.1}$$

Where

K_{xx}, K_{yy}, and K_{zz} are the values of hydraulic conductivity (LT^{-1}) along the x, y and z co-ordinate axes respectively;

h is the potentiometric head (L);

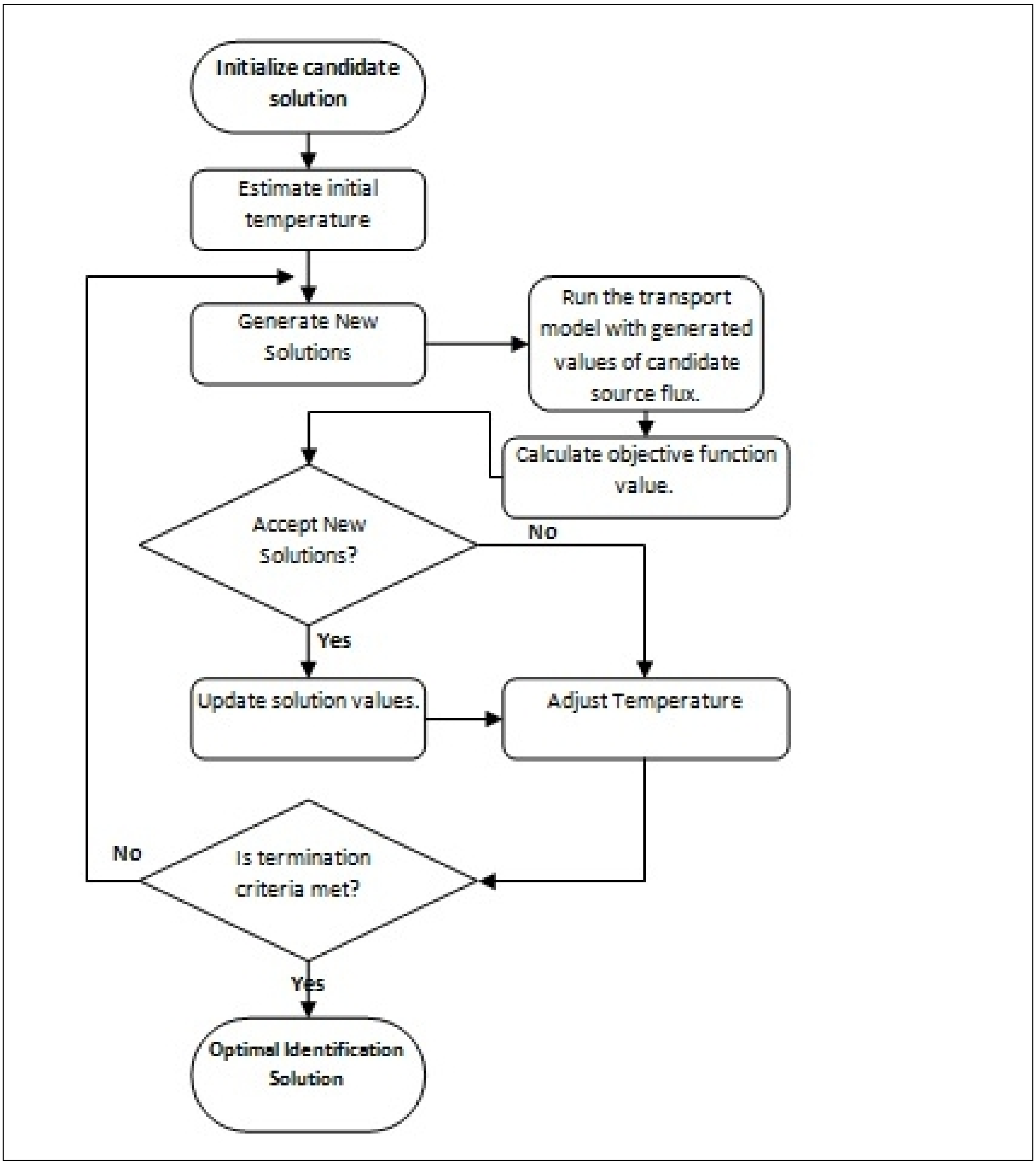


Figure 3.1: Schematic Representation of Linked Simulation-Optimization Model using Adaptive Simulated Annealing

W is the volumetric flux per unit volume representing sources and/or sinks of water (T^{-1});

S_s is the specific storage of the porous media (L^{-1}); and

t is time (T).

Detailed derivation of this equation is presented in Freeze and Cherry (1977), Bear (1988) and Fetter (1994).

Processes involved in movement of solutes contained in groundwater are complex. However, two processes dominate such solute movement. One of these is diffusion and the other is advection. Diffusion is the process by which solutes move from areas of high concentration to areas of lower concentration. Advection, on the other hand, is the process by which groundwater carries the dissolved solutes with it. Apart from these two processes, dispersion acts on the solutes to dilute them and lower their concentration. There may also be processes such as adsorption and chemical reaction that retard the solutes and cause them to move in the porous media at a rate slower than that suggested by advection process. All these processes can be modelled by partial differential equations. Detailed derivation of the governing equation of groundwater solute transport has been presented in Freeze and Cherry (1977), Javandel et al. (1984) and Fetter (1994). Transport of solutes or contaminants in a three-dimensional saturated aquifer can be represented by the partial differential Equation 3.2 (Javandel et al., 1984).

$$\frac{\partial C}{\partial t} = D_{ij}\frac{\partial}{\partial x_i}\left(\frac{\partial C}{\partial x_j}\right) - \frac{\partial}{\partial x_i}\left(\vartheta_i C\right) + \frac{q_s}{\theta}C_s + \sum_{k=1}^{N} R_k \tag{3.2}$$

Where

C is the concentration of pollutants dissolved in groundwater,

(ML^{-3});

t is time, (T);

x_i, x_j is the distance along the respective Cartesian coordinate axis, (L);

D_{ij} is the hydrodynamic dispersion coefficient, (L^2T^{-1});

ϑ_i is the seepage or linear pore water velocity, (LT^{-1});

q_s is the volumetric flux of water per unit volume of aquifer representing sources (positive) and sinks (negative), (T^{-1});

C_s is the concentration of the sources or sinks, (ML^{-3});

θ is the porosity of the porous medium, dimensionless;

N is the number of chemical species considered;

$\sum_{k=1}^{N} R_k$ is the chemical reaction term for each of the N species considered, $(ML^{-3}T^{-1})$.

In order to solve this 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.

The distribution of pollutant at all locations in a study area for any specific point in time after the release of pollutants is obtained as the solution of groundwater transport equation with known aquifer parameters and source characteristics. For a simple case of a non-decaying continuous pollutant source, leaking into a one-dimensional aquifer at a rate of C_0, the concentration of pollutant at some length L away from the source at time t is given by Equation 3.3 (Ogata and Banks, 1961).

$$C = \frac{C_0}{2}\left[erfc\left(\frac{L - v_x t}{2\sqrt{D_L t}}\right) + exp\left(\frac{v_x L}{D_L}\right) erfc\left(\frac{L + v_x t}{2\sqrt{D_L t}}\right)\right] \tag{3.3}$$

Where

C is pollutant concentration (ML^{-3}, mg/L);

C_0 is initial pollutant concentration (ML^{-3}, mg/L);

L is the length of flow path (L, m)

v_x is the average linear groundwater velocity (LT^{-1}, m/day);

t is time elapsed since the release of pollutant (T, days);

D_L is the longitudinal dispersion coefficient (L^2T^{-1}, m^2/day).

3.2.2 Numerical Solution of Groundwater Flow and Transport Equations

In the mathematical model of groundwater flow or transport, the variables are assumed to be continuous. Numerical solutions aim at approximately solving the governing partial differential equations by discretizing the independent variables. In order to solve the equations of groundwater flow or transport, boundary conditions need to be known. Two basic types of boundary conditions are used (Anderson and Woessner, 1992): *Dirichlet condition* and *Neumann condition*. When the head or solute concentration is known around a boundary it is referred to as Dirichlets condition, and when the flux or concentration gradient is known around a boundary it is called Neumann condition. It is also possible to use a combination of these two boundary conditions. In order to solve for transient conditions, it is also essential to know the initial conditions. Since the groundwater transport model is inherently transient, it cannot be solved without knowing initial condition. Initial condition is essentially the starting or background concentration of a pollutant in a groundwater aquifer. In this study, governing equations of groundwater flow and transport are solved using computer models that use block-centred finite difference spatial discretization. In a non-homogeneous aquifer, parameters such as hydraulic conductivity, porosity and dispersivity vary continuously throughout the matrix of study area. In order to discretize the distribution of such parameters, the entire study area is divided into a number of cuboidal, block-centred, non-overlapping three dimensional finite-difference cells. Aquifer parameters are associated to the centre of each cell in the finite difference grid. The governing equations are solved by what are known as *iterative methods*.

3.2.2.1 MODFLOW

MODFLOW (McDonald and Harbaugh, 1988; Harbaugh, 2005), is one of the most versatile finite difference groundwater flow models. It was developed by United States Geological Survey to simulate groundwater flow in confined, leaky confined and unconfined aquifers. It is made up of a series of separate, independent modules that simulate recharge, evapo-transpiration, areal recharge, flow to wells, flow to drains and flow through riverbeds. Transient flow conditions can also be simulated. In such cases, the time domain is discretized into a number of stress periods of finite time length. It also gives the user an option to use several methods of solution. MODFLOW has been used in this study to simulate groundwater flow in the study area.

3.2.2.2 MT3DMS

The MT3DMS groundwater transport model (Zheng and Wang, 1999) uses a mixed Eulerian-Lagrangian approach for the solution of the governing equation. The Lagrangian part of the method, used for solving the advection term, employs the forward-tracking method of characteristics (MOC), the backward-tracking modified method of characteristics (MMOC), or a hybrid of these two methods. The Eulerian part of the method, used for solving the dispersion and chemical reaction terms, utilizes a conventional block-centred finite difference method. MT3DMS is used along with the flow model (MODFLOW) since it utilizes the flow field generated by the flow model (MODFLOW) to compute the velocity field. MT3DMS is used in this study to simulate pollutant transport in groundwater.

3.2.3 Formulation of the Optimization Problem

It is assumed in this study that information on a set of potential source locations is available. The objective of simulation-optimization method then reduces to regenerating the source release histories at these potential source locations. Spatial and temporal pollutant concentration (C) is known at specific monitoring locations at various points of time. Candidate source fluxes are generated by the optimization algorithm. These values are used for forward transport simulations in MT3DMS. The difference between simulated and observed pollutant concentrations are then used to calculate the objective function. The objective function for this optimization problem is defined as:

$$MinimizeF1 = \sum_{k=1}^{nk} \sum_{iob=1}^{nob} \left(cest_{iob}^{k} - cobs_{iob}^{k} \right)^2 .w_{iob}^{k} \tag{3.4}$$

Subject to the constraints:

$$cest_{iob}^{k} = f\left(x, y, z, v_x, D_L, D_T, t, \theta\right) and$$
$$C_{min}^{i} \leq \theta^{i} \leq C_{max}^{i} \tag{3.5}$$

Where

$cest_{iob}^{k}$ = concentration estimated by the identification model at observation well location 'iob' and at the end of time period 'k'.

nk = total number of concentration observation time periods;

nob = total number of observation wells;

$cobs_{iob}^{k}$ = observed concentration at well 'iob' and at the end of time period 'k';

w_{iob}^{k} = weight corresponding to observation location 'iob', and the time period 'k';

x,y,z = cartesian co-ordinates of the monitoring location with pollutant source as the origin;

v_x = groundwater velocity (Darcy's velocity) in horizontal direction;

t = time elapsed since the release of pollutant in groundwater;

D_L, D_T = dispersivity in longitudinal and transverse direction;

θ = pollutant source flux (in terms of mass per unit time) released into groundwater;

θ^{i} = candidate solution for source flux in stress period 'i';

C_{min}^{i} = minimum pollutant source flux in stress period 'i';

C_{max}^{i} = maximum pollutant source flux in stress period 'i'.

The first constraint to this optimization function in Equation 3.5 means that $cest_{iob}^{k}$ is obtained by solving the linked simulation models, and is a function of various hydrogeologic parameters and source characteristics. The implicit constraints represent the linkage between the optimization and the groundwater flow and transport simulation model. The second constraint points to the fact that values of source flux in every stress period are bounded by a maximum and a minimum value.

The weight w_{iob}^{k} can be defined as follows:

$$w_{iob}^{k} = \frac{1}{(cobs_{iob}^{k} + n)^2} \tag{3.6}$$

Where n is a constant, sufficiently large, so that errors at low concentrations do not dominate the solution (Keidser and Rosbjerg, 1991). It is also possible to include other forms of this weight.

3.2.4 Optimization Algorithms

Solution results obtained by using adaptive simulated annealing based simulation-optimization method are compared to the solution results obtained by using genetic algorithm in this study. In simulated annealing, a current solution may be replaced by a random neighbourhood solution chosen with a probability that depends on the difference between corresponding function values and on a global parameter T (called temperature). The temperature parameter is gradually decreased in the search process (Kirkpatrick, 1984). Of the various simulated annealing implementations, it is evident in literature that the adaptive simulated annealing algorithm converges faster (Ingber and Rosen, 1992) while maintaining the reliability of results and hence it was preferred over traditional Boltzmann annealing implementation (Kirkpatrick, 1984). Its application to unknown pollutant source identification has been limited but it is potentially a good alternative because its convergence curve is steep, thereby producing better results when execution time is limited. Genetic algorithms (GAs) are population based search strategies which are popular for many difficult to solve optimization problems, including inverse problems. GAs emulate the natural evolutionary process in a population where the fittest survive and reproduce (Holland, 1975). GA-based search performs well because of its ability to combine aspects of solutions from different parts of the search space. Real coded genetic algorithm was used with a population size of 100, crossover probability of 0.85 and a mutation probability of 0.05. The values were chosen based on a series of numerical experiments.

3.2.5 Suitability and Sensitivity of Adaptive Simulated Annealing

In the application discussed here, simulated annealing is utilized for finding the global minimum of an objective function that characterizes large and complex systems such as transport of pollutants in groundwater. Simulated annealing, as an algorithm, is very efficient in solving non-convex optimization problems by ensuring that it does not always move downhill on a complex non-convex search space and hence avoids getting trapped in local minimum. Simulated annealing also differs significantly from conventional iterative optimization algorithms in that gross features of the final state of the system are seen at higher temperatures, whereas the finer details of the state appear at lower temperatures (Haykin, 1999). The fact that simulated annealing ensures a global optimal solution enhances its suitability for solving ill-posed inverse problems in general, and the problem of unknown groundwater pollutant source characterization in particular.

Its ease of use and efficiency in handling complex objective functions and constraints has made simulated annealing an attractive choice for solving a wide range of complex optimization problems (Ingber, 1996). However, the slow convergence and hence long time of execution of standard Boltzmann-type simulated annealing has been a constraint. Adaptive simulated annealing removes that constraint by making the annealing schedules decrease exponentially in annealing time, thereby making the convergence much faster. A major difference between ASA and traditional Boltzmann annealing algorithms is that the ergodic sampling takes place in terms of n parameters and the cost function. In ASA the exponential annealing schedules permit resources to be spent adaptively on re-annealing and on pacing the convergence in all dimensions, ensuring ample global searching in the first phases of search and ample quick convergence in the final phases (Ingber, 1996).

Another major advantage of using adaptive simulated annealing is also the fact that the parameters of algorithm are adjusted adaptively and hence the solutions do not vary widely if parameter values are changed within reasonable limits. This is in contrast with genetic algorithm where even minor changes to parameters such as mutation probability, crossover probability, or population size cause a significant difference in the solutions.

3.3 Performance Evaluation

In order to evaluate the performance of two different optimization algorithms based on the solutions obtained, it is vital to first ensure that only one solution exists. In other words, a unique solution has to be guaranteed. This is possible only under the following idealized assumptions (Sun, 1994):

1. The numerical models used for simulation of groundwater flow and transport are able to provide exact solution of the governing equations in forward runs.

2. All the model parameters and concentration measurements are known without any associated errors.

3. The unknown parameter is piecewise constant.

The first assumption is valid for cases where grid size and time step used in the numerical solution tend to zero. As the groundwater simulation models used in this study have been proven to be stable and convergent, this assumption approximately holds. The second assumption, however, cannot hold in real-life scenarios. Hence, it becomes necessary to use synthetically generated observation values initially which can be considered free of measurement errors. The third condition is implemented by assuming that the unknown fluxes are constant in every stress period. In such conditions it approximately resembles a well-posed problem. Therefore, these evaluations are initially carried out for synthetic data (simulated data) with known parameter values. There is another related issue of unique solutions. Whenever numerical simulation and optimization models are used, the convergence of the solutions may be another issue related to unique solutions. These issues are discussed in Datta (2002). In this study the use of synthetic observation data, with known hydrogeologic parameter values reduces the ill-posed nature of the problem. The uniqueness of the solution cannot be guaranteed. However, sufficient iterations were allowed to ensure convergence to the optimal solution. Performance of the source identification methodology is evaluated using synthetic data from a three-dimensional aquifer study area. The synthetic pollutant concentration data are obtained by solving the numerical flow and transport simulation models.

3.3.1 Simulating Errors in Concentration Measurement Data

Once the global optimal solution has been obtained for the idealistic assumption, the performance evaluation of developed methodology can take into account the effects of pollutant concentration measurement errors as well as uncertainty associated with the determination of hydrogeological parameters. To test the performance for realistic scenarios, concentration measurement errors are incorporated by introducing varied amounts of synthetically generated statistical noise in the simulated concentration values. The perturbed simulated concentration represents erroneous measurement and is defined as follows:

$$C_{pert} = C_{ns} + S_{ud} \times a \times C_{ns} \tag{3.7}$$

Where

C_{pert} = perturbed concentration value;

C_{ns} = simulated concentration;

S_{ud} = a uniform random number between -1 and +1;

a = a fraction between 0 and 1.0.

3.3.2 Incorporating Uncertainty in Hydrogeologic Parameters

Many times the groundwater flow models use layers with several zones that are assigned representative average values of hydraulic conductivity to generate the head distribution in the study area. This information is used by the transport model to calculate pollutant concentrations at different points in time and space. Hydraulic conductivity, however, is represented most realistically by a stochastic set of values defined by a log-normal probability distribution. If the study area has several non-homogeneous layers, then the distribution is different in each layer. Even within a homogeneous layer the hydraulic conductivity values are not unique. They show random variations in space. The measure of this random variation is represented by the standard deviation of the distribution.

In order to test the performance of ASA based simulation-optimization methodology, the observed pollutant concentrations are generated using non-homogeneous non-uniform hydraulic conductivity values whereas the transport simulation model linked to optimization algorithm still uses the head distribution generated by a flow model that uses average values of hydraulic conductivity. The values of hydraulic conductivity (K) are assumed to follow a log-normal distribution (Freeze, 1975). If we define another variable Y such that Y = $log_{10}K$, then Y is distributed normally with a mean μ_y and a standard deviation σ_y. The values of σ_y generally vary between 0.2 and 20 (Freeze, 1975). And the value of μ_y is close to the log of average hydraulic conductivity.

Spatial variation of hydrologic parameters is also an important consideration for simulating realistic groundwater flow and solute transport processes. Detailed characterization of the spatial distribution of hydrologic parameter values in an aquifer is described in Yeh (1992) and Gelhar (1993).

In order to create a quasi three-dimensional hydraulic conductivity field in this study, different values for mean and standard deviation are chosen for each layer. Furthermore, in order to incorporate geo-spatial correlations for the hydraulic conductivity in each layer, a small number of sample hydraulic conductivity values are first sampled from the statistical distributions using Latin hypercube sampling (Pebesma and Heuvelink, 1999). Values from this sample are then interpolated to the entire layer using ordinary kriging (OK) (Cressie, 1988). Interpolation of a defined variogram simulates the geo-spatial correlation in hydraulic conductivity values. An example of hydraulic conductivity distribution generated using this method is shown in Figure 3.2. The point value of porosity is assumed to follow a normal distribution (Freeze, 1975). In this study, porosity was assumed to have a mean of 0.3-0.4 and a standard deviation of 0.005-0.008. Similar to the generation of hydraulic conductivity values, different values of mean and standard deviation are used for each layer of the aquifer to generate quasi three-dimensional values of porosity in a non-uniform, heterogeneous media. Geo-spatial correlations are simulated as described above for hydraulic conductivity values.

3.3.3 Performance Evaluation Criteria

The execution times of the algorithms are compared based on convergence curves which represent the value of objective function achieved versus

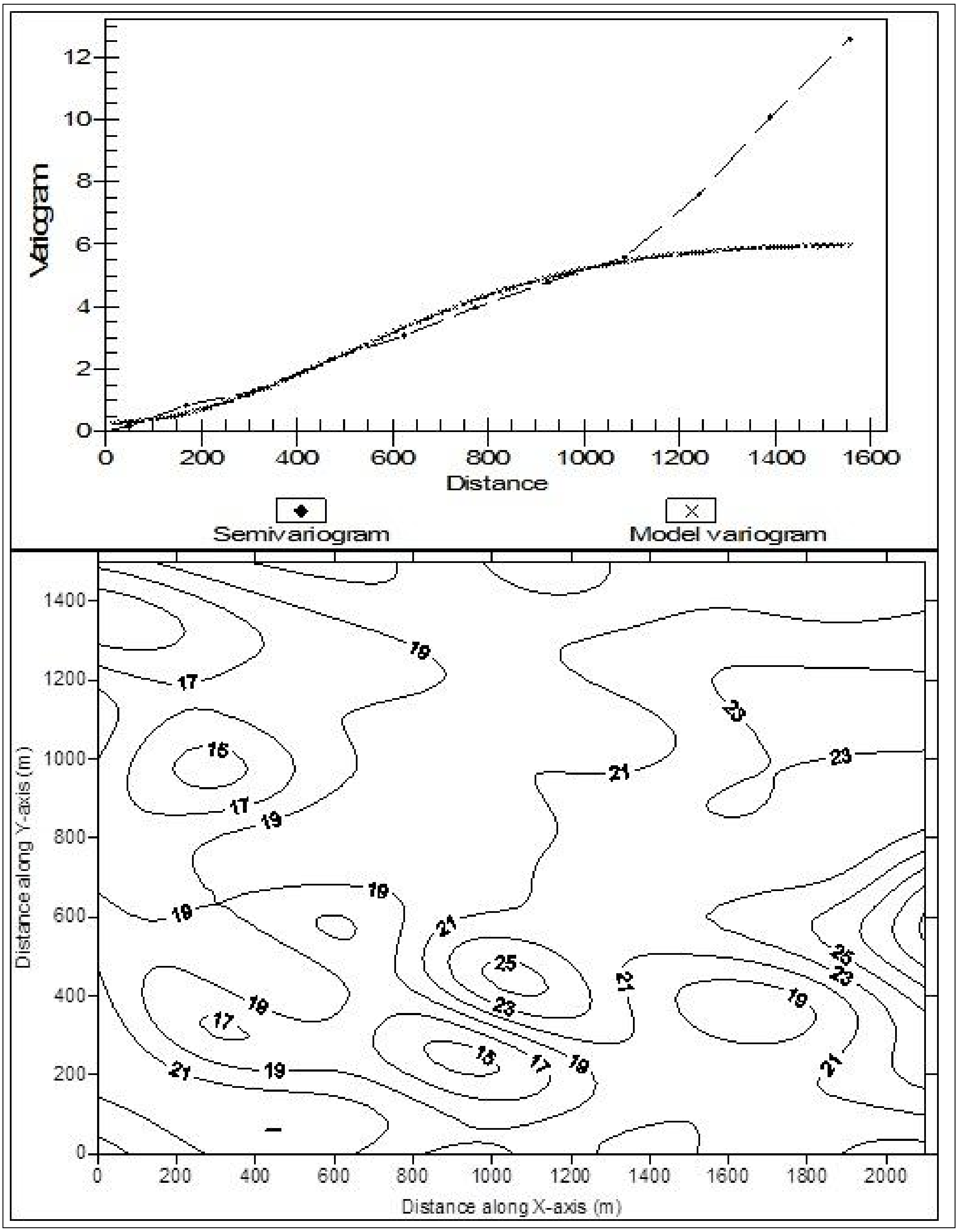


Figure 3.2: Model Variogram and Spatially Correlated Hydraulic Conductivity Values Generated for the First Layer

time. To compare the ability of competing linked simulation-optimization approaches to produce accurate source histories, the error in estimating source fluxes accurately is also used as a performance criterion. Normalized absolute error of estimation (NAEE) is used as the measure of errors in estimation of the sources. It can be represented as:

$$NAEE(\%) = \frac{\sum_{i=1}^{S}\sum_{j=1}^{N}\left|\left(q_i^j\right)_{est} - \left(q_i^j\right)_{act}\right|}{\sum_{i=1}^{S}\sum_{j=1}^{N}\left(q_i^j\right)_{act}} \times 100 \tag{3.8}$$

Where

NAEE = normalized absolute error of estimation;

S = number of sources = 2 in this case;

N = number of transport stress periods = 5 in this case;

$\left(q_i^j\right)_{act}$ = actual source flux for source number i in stress period j;

$\left(q_i^j\right)_{est}$ = estimated source flux for source number i in stress period j.

3.3.4 Incorporation of Different Concentration Monitoring Scenarios

In order to evaluate the relevance of any existing monitoring network in efficient estimation of the unknown pollutant sources, a number of plausible monitoring scenarios are incorporated. The performance of developed methodology is evaluated and compared for each of these monitoring scenarios.

A set of five different monitoring networks, each consisting of five individual monitoring locations are separately used to reconstruct source release histories. The effectiveness of each monitoring location is also

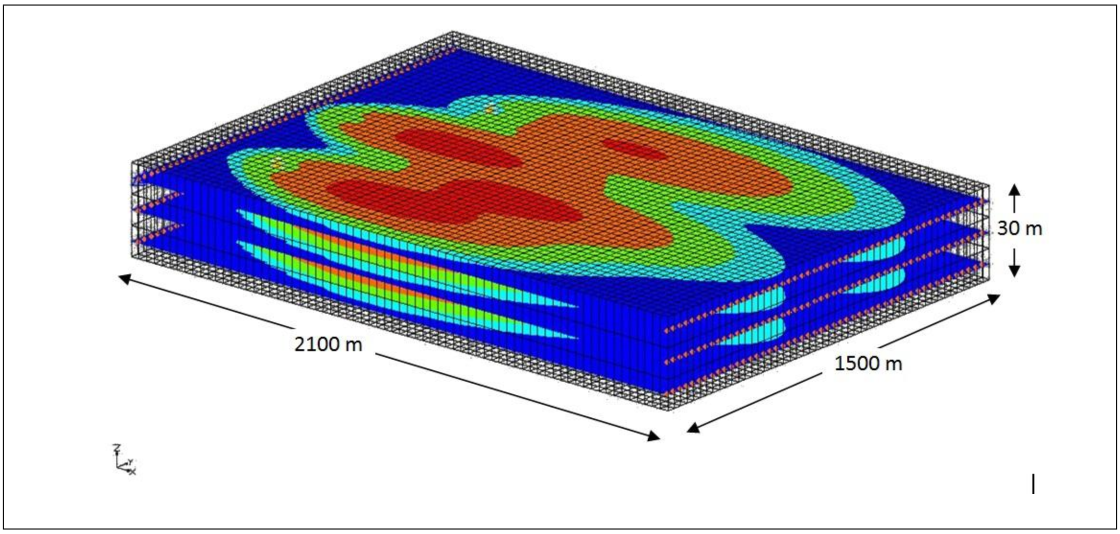


Figure 3.3: Illustrative Study Area

compared based on the normalized absolute error of estimation.

3.4 Discussion of Solution Results

The developed methodology was applied to a hypothetical illustrative study area with synthetically generated concentration measurements over space and time. The advantage of using a hypothetical study area lies in the fact that unknown data errors do not distort the performance evaluation of the methodology. This helps in understanding the drawbacks of developed methodology independent of input data error.

3.4.1 Study Area

The illustrative study area is a heterogeneous aquifer measuring 2100 m x 1500 m x 30 m and consisting of three unconfined layers as shown in Figure 3.3.

The east and west boundaries are constant head boundaries, whereas the

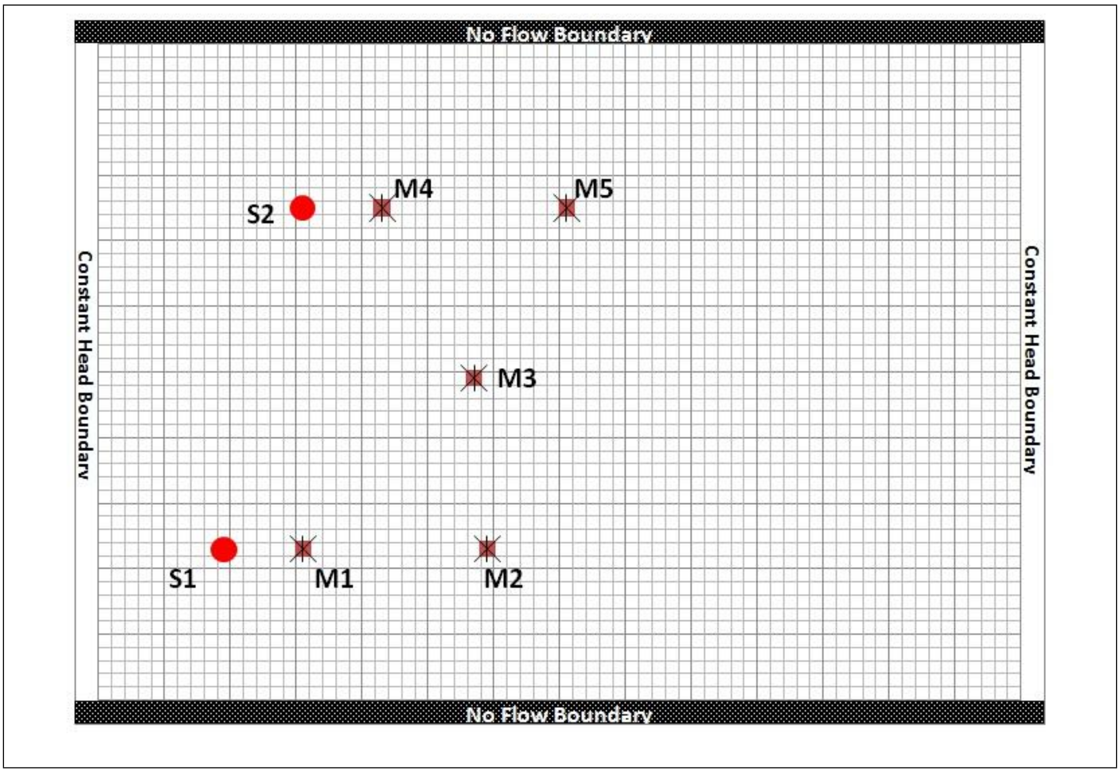


Figure 3.4: Top View of Study Area Showing Sources and Monitoring Locations

north and south boundaries are no flow boundaries. There are two sources (S1 and S2) of pollution. S1 is located in the top layer and S2 in the middle layer. Five monitoring locations (M1 through M5) are located in the first layer as shown in Figure 3.4. A grid size of 30 m x 30 m x 10 m is used for finite difference based numerical solution of groundwater flow and transport equations. Model parameters are listed in Table 3.1.

Only a conservative pollutant is considered. There are two point sources of pollutants. One in the top layer and another one in the middle layer. The study area is discretized into a number of rows, columns and layers along the x, y and z axes respectively. Hence, the location of sources can be expressed

Table 3.1: Model Parameters

Parameter	Value
Length of study area (m)	2100
Width of study area (m)	1500
Saturated thickness, b(m)	30
Grid spacing in x-direction, Δx (m)	30
Grid spacing in y-direction, Δy (m)	30
Grid spacing in z-direction, Δz (m)	10
Hydraulic conductivity in x-direction, Kxx (m/day)	20
Hydraulic conductivity in y-direction, Kyy (m/day)	20
Vertical anisotropy	5
Hydraulic gradient (m/m)	0.00238
Effective porosity, θ	0.3
Longitudinal dispersivity, α_L (m)	15
Transverse dispersivity, α_T (m)	3
Initial pollutant concentration (mg/l)	0.00

as row, column and layer number of the discretized cell that contains the physical source location. A time horizon of 20 years is considered. The entire time horizon is divided into five different stress periods. The first four stress periods are each two years long and the final stress period is of 12 years duration. Sources are assumed to be active only in the first four stress periods or in the initial eight years. Actual source fluxes are presented in Table 3.2. Source location, in terms of the row, column and layer number in the discretized space is also mentioned. It is assumed that groundwater pollution is detected at five different locations in the study area at the end of 10th year, that is two years after the sources had ceased to exist. The observation wells are monitored for a period of 10 years starting from year 11 at an interval of 73 days. Observed pollutant concentration measurements at the designated monitoring locations are generated using MT3DMS as transport simulation model followed by perturbation as per Equation 3.7.

Table 3.2: Actual Source Fluxes

Sources	Layer	Row	Column	Contaminant Flux (g/s)		
				Stress Period 1	2 years	6.250
				Stress Period 2	2 years	4.630
Source 1	1	12	15	Stress Period 3	2 years	9.028
				Stress Period 4	2 years	5.556
				Stress Period 5	12 years	0.000
				Stress Period 1	2 years	6.690
				Stress Period 2	2 years	9.346
Source 2	2	38	9	Stress Period 3	2 years	6.100
				Stress Period 4	2 years	7.280
				Stress Period 5	12 years	0.000

3.4.2 Source Flux Magnitude Estimation with Error Free Data

A set of error free observation data is generated. These observations are then used to evaluate the developed linked simulation-optimization methodology based on both GA and ASA. Input parameters used for GA and ASA are presented in Table 3.3. Every iteration of ASA based method uses one run of the groundwater transport simulation model (MT3DMS) whereas every generation of GA based method uses 100 (population size) runs of the same simulation model. Irrespective of the method, one run of the groundwater transport simulation model takes 3.784 seconds to run on a Dell Optiplex® running an Intel®$Core^{TM}$2 Duo Processor at 2.93 GHz. The execution time for one transport simulation run is, however, dependent on the computing platform.

In order to keep the comparison independent of computing platform, both the methods were compared based on number of transport simulation runs, which is directly proportional to the execution time. Both the methods were used to estimate source release histories using the error free data. In order to verify the convergence of each optimization method, time of run was made

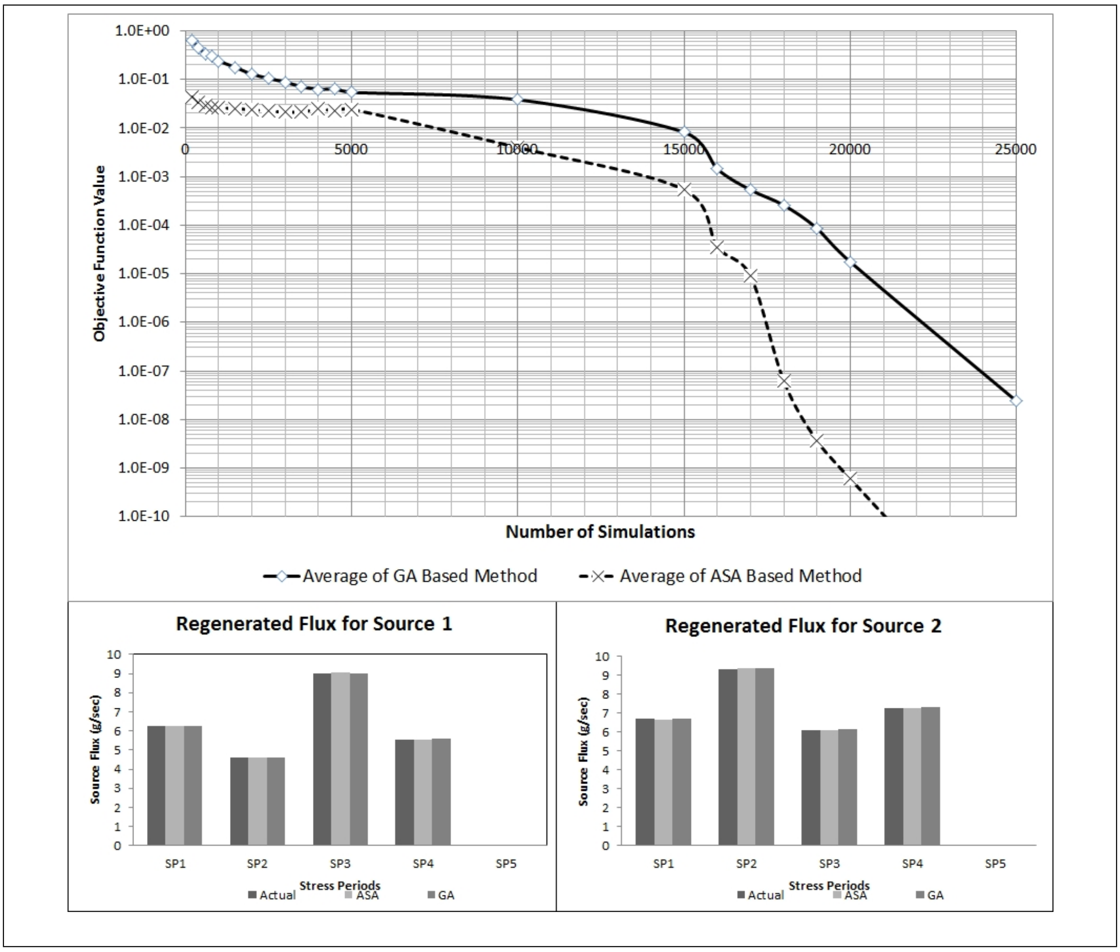


Figure 3.5: Estimated Release History with Error Free Data

practically unconstrained. It was found that eventually both the optimization algorithms were able to achieve an objective function value very close to zero and identified the release history accurately. The objective function convergence profile as well as estimated fluxes are plotted at the end of 25,000 simulation runs of the groundwater transport model. Minimum value of objective function achieved is plotted against number of runs of the transport simulation model. The estimated flux values for both the sources in each stress period are also plotted against actual source fluxes. Convergence profile and source flux estimates are shown in Figure 3.5. Convergence profile shows

Table 3.3: Parameters used in Optimization Algorithms

Parameters of ASA		Parameters of GA	
Accepted to generated ratio	1.00E-06	Mutation strategy: Polynomial mutation	
		Variable boundaries : Rigid	
Cost precision	1.00E-10	Population size	100
Maximum cost repeat	5	Total no. of generations	400
Temperature ratio scale	1.00E-05	Crossover probability	0.778
Temp. anneal scale	100	Mutation probability	0.0512
		Exponent (n for SBX)	2
		Exponent (n for Mutation)	20

that the objective function value for the source identification model converges to a value very close to zero with about 5,000 simulation runs. However, further convergence is accelerated when using ASA algorithm. The entire range of convergence beyond 25,000 simulation runs is not shown here. From these results, it can be concluded that the developed methodology is able to achieve optimal solution for an ideal error free scenario which resembles a well-posed problem.

3.4.3 Source Flux Magnitude Estimation with Erroneous Data

Five sets of erroneous observation data are generated with the formulation described in Equation 3.7. The value of fraction 'a' is specified as 0.1. These erroneous observations are used to reconstruct the release histories of pollutant sources. Linked simulation optimization method using ASA is compared with the method using GA as the optimization algorithm. Parameters used for both the optimization algorithms is presented in Table 3.3. Unlike the case with error free measurement data, in this case both the methods were used to reconstruct source release histories using the erroneous data with a limit on execution time. In order to make the comparison consistent by ensuring same number of simulation runs in

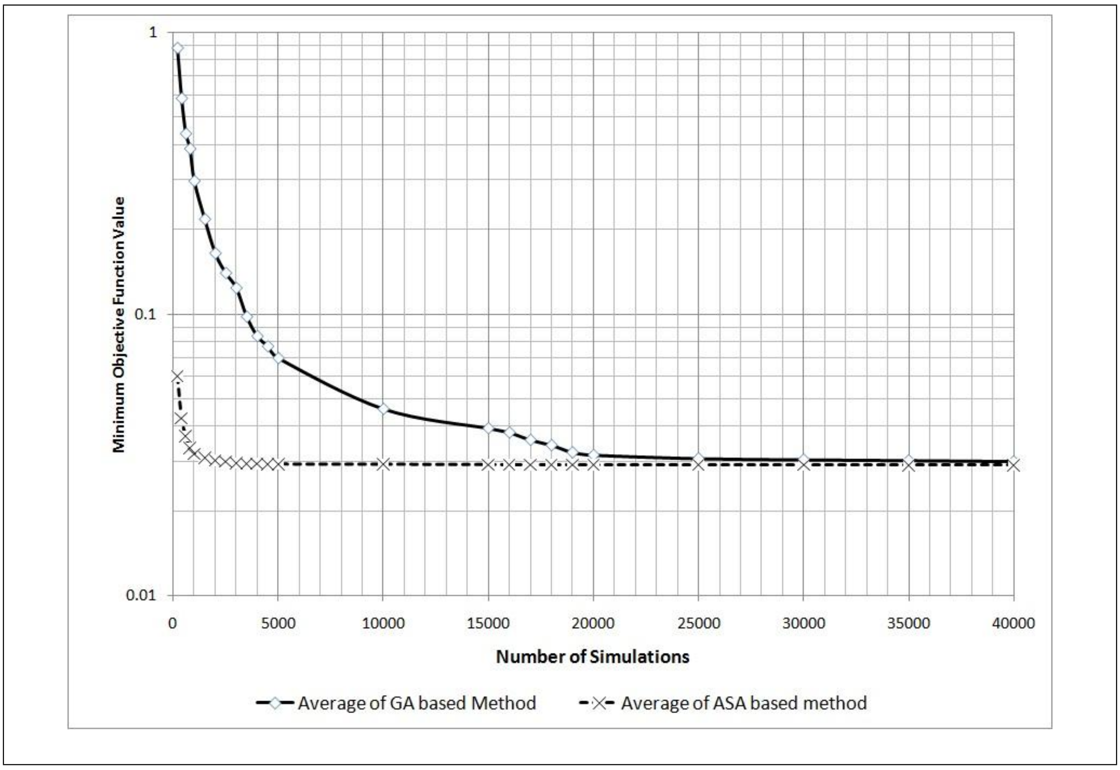


Figure 3.6: Convergence Plot

the ASA and GA based methodologies, the number of simulation runs is restricted to 40,000. This restriction was based on the fact that increasing the number of simulation runs even to 80,000 resulted in very little improvement in the objective function value. Minimum value of objective function achieved is averaged over five solutions and is plotted against number of runs of the transport simulation model. The plot is presented in Figure 3.6. This plot clearly shows that the ASA based method converges much faster in the beginning. The GA based method is able to achieve comparable objective function values only after a much larger number of simulation runs. Because of the erroneous measurement data this problem may be ill-posed and the solution may not be unique. Therefore, lower objective function values do not always mean accurate reconstruction of the release histories.

Table 3.4: Normalized Absolute Error of Estimation

No. Of Simulation Runs	NAEE (%)	
	GA	ASA
10000	5.35	2.94
20000	5.79	3.23
30000	4.49	3.39
40000	3.66	3.50

In order to test the effectiveness of the competing methods based on accuracy of solutions produced, reconstructed release histories were compared to the actual release history after every set of 10,000 transport simulation runs. The results are shown in Figure 3.7. It can be seen that the ASA based method is more efficient compared to the GA based method after 10,000 and 20,000 simulation runs. However, as the execution time increases further with increase in number of simulation runs, the release histories produced by both methods become similar. This is also confirmed from the calculated values of NAEE presented in Table 3.4. As the execution time increases, the NAEE of ASA based method appears to increase only slightly. This could be due to statistical variation in the five different solutions and may be attributed to the input data error. Averaging over a larger number of solutions may modify this inference. NAEE of GA based method consistently improves. However, the NAEE values obtained using ASA are still better in comparison.

3.4.4 Source Flux Magnitude Estimation with Uncertainty in Hydrogeologic Parameters

Any attempt to estimate release history of pollutant sources is susceptible to the uncertainty in estimation of hydrogeological parameters of the aquifer. Most often, an average value of the hydrogeological parameter such as hydraulic conductivity is used in the groundwater flow and transport models.

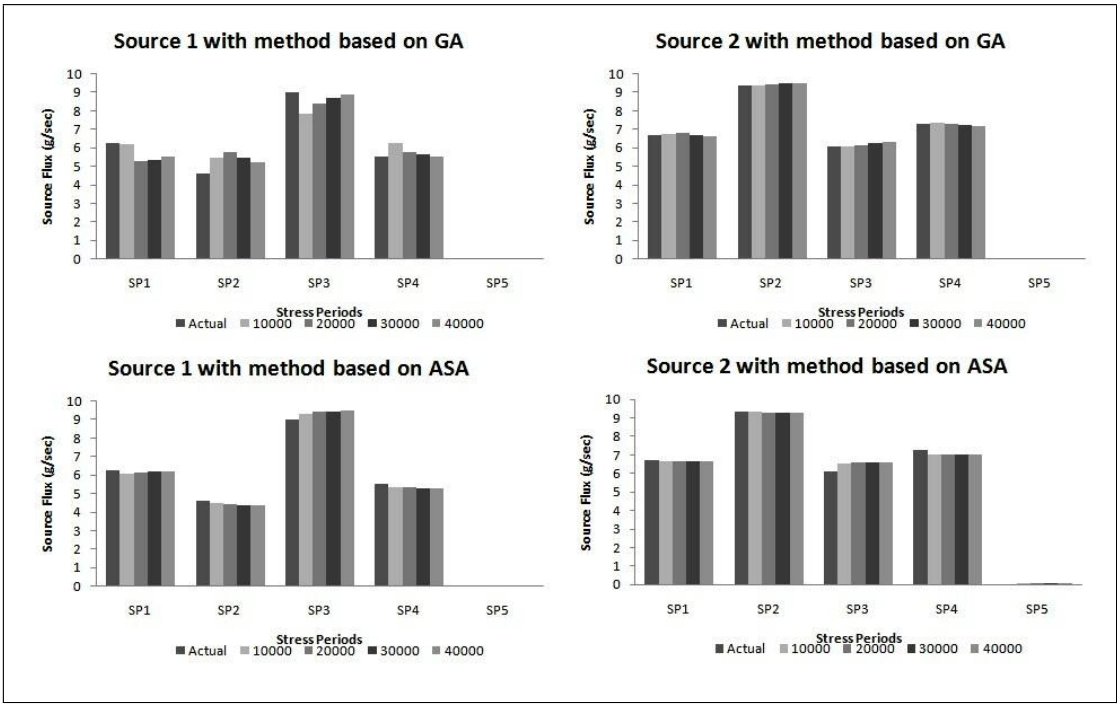


Figure 3.7: Reconstructed Release Histories using the Competing Methods

However, in the real world this is not true as the hydrogeological parameter values are not uniform even in a homogeneous layer. The real world values of hydro-geological parameter can be closely approximated by certain statistical distribution (Freeze, 1975). It has been reported that source release history reconstruction problems are particularly sensitive to hydraulic conductivity and porosity (Datta et al., 2009a,b).

In this study, three sets of numerical experiments were carried out to study the effects of uncertainty in estimation of hydrogeological parameters. Contaminant concentration observation data were generated using a distribution of i) hydraulic conductivity, ii) porosity, and iii) both i) and ii). The distributions for hydraulic conductivity in each of the three layers were generated using log20, log17 and log21 as mean and 0.1, 0.08 and 0.12 as standard deviations respectively. Similarly, the distributions for porosity were generated using 0.30, 0.32 and 0.30 as mean and 0.006, 0.008 and 0.006 as standard deviation. The simulation model used in the linked simulation-optimization methods based on ASA and GA that were used to reconstruct the release histories of the pollutant sources, however, used spatially averaged values of hydraulic conductivity and porosity. The reason is that field measurements reflect actual hydrogeologic conditions. However, for modelling purposes, often average values of the hydrogeologic parameters are used. While the non-uniformity in the hydrogeologic parameters is incorporated in generating actual field measurement, these uncertainties are not included in the linked simulation-optimization model. Therefore, the modelling uncertainties are also incorporated in these evaluations.

Three sets of numerical experiments are carried out, each containing five different runs with different realization of point values of hydrogeological parameters. In the first set, pollutant concentration observation values were generated after incorporating non-uniformity in hydraulic conductivity alone. In the second set, randomly generated values of porosity, as discussed earlier, were used to generate the pollutant concentration measurements. In the third set, randomly generated values of both hydraulic conductivity and porosity values were used. The execution time is limited to 20,000 runs of the groundwater transport model. The solution results are presented in Table 3.5.

It can be inferred from these solution results that the reconstruction of release histories becomes difficult when unmodelled non-uniformity in both hydraulic conductivity and porosity are present. However, the optimal source identification model does not incorporate these non-uniformities.

Table 3.5: Performance Evaluation for Uncertainty in Hydrogeologic Parameters

			Uncertainty in Hydraulic conductivity				Uncertainty in Porosity				Uncertainty in Both			
Fluxes (g/s)		Actual	ASA Based Method		GA Based Method		ASA Based Method		GA Based Method		ASA Based Method		GA Based Method	
Source	Stress Period		Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
Source 1	1	6.25	5.32	0.31	6.47	0.42	5.65	0.54	5.99	0.56	4.08	1.81	5.20	2.11
	2	4.63	5.20	0.43	5.12	0.78	5.24	0.65	5.83	0.39	7.26	2.03	8.78	2.30
	3	9.03	8.67	0.51	9.87	0.41	8.47	0.32	9.22	0.47	7.54	1.96	7.31	1.72
	4	5.56	4.81	0.29	4.86	0.42	5.34	0.27	5.17	0.25	5.03	1.43	5.64	1.18
	5	0.00	0.06	0.02	0.08	0.03	0.09	0.03	0.08	0.03	0.61	0.2	0.67	0.34
Source 2	1	6.69	8.30	0.27	8.24	0.44	8.60	0.28	9.01	0.53	8.90	1.13	9.82	1.62
	2	9.35	10.80	0.49	11.43	0.63	10.81	0.43	12.12	0.47	9.22	2.32	9.56	2.63
	3	6.10	5.67	0.47	7.04	0.76	6.27	0.63	6.01	0.29	6.95	1.21	8.54	1.32
	4	7.28	7.04	0.37	7.64	0.48	8.39	0.31	8.94	0.22	6.83	0.92	8.26	1.36
	5	0.00	0.07	0.03	0.09	0.04	0.09	0.03	0.10	0.05	0.71	.12	0.83	0.18
NAEE (%)			11.79		13.39		12.42		16.51		21.88		28.33	

It is also apparent that, within the restricted execution time, adaptive simulated annealing based methodology produces better results compared to those produced by GA based methodologies. This is the inference even when the modelled values of hydrogeological parameters used in the simulation-optimization model are average of the actual spatially varied values used to generate synthetic pollutant concentration observation values for the performance evaluation.

3.4.5 Effects of Monitoring Network

The selection of a pollutant concentration monitoring network directly affects the results obtained using linked simulation-optimization models. In order to study the effects of monitoring locations, a set of four arbitrary monitoring networks was used apart from the one used in the rest of this study. The observation data obtained from each of these was used to reconstruct the release histories of pollutant sources using linked simulation-optimization method based on ASA. All the four monitoring networks (labeled MN2 through MN5) are shown in Figure 3.8. Two of the monitoring networks have all the observation wells placed perpendicular to the general direction of groundwater flow while the other two have wells chosen on the vertices and centre of a virtual rectangle. The characteristic curves for each of the monitoring networks are shown in Figure 3.9. It is worthwhile to note that no more than three arbitrary observation wells chosen show any appreciable pollutant concentration. This implies that, in effect, at most only three chosen monitoring locations are able to capture the pollutant plume. ASA based simulation-optimization method was used to reconstruct the release histories of sources. The simulation runs were limited to 30,000.

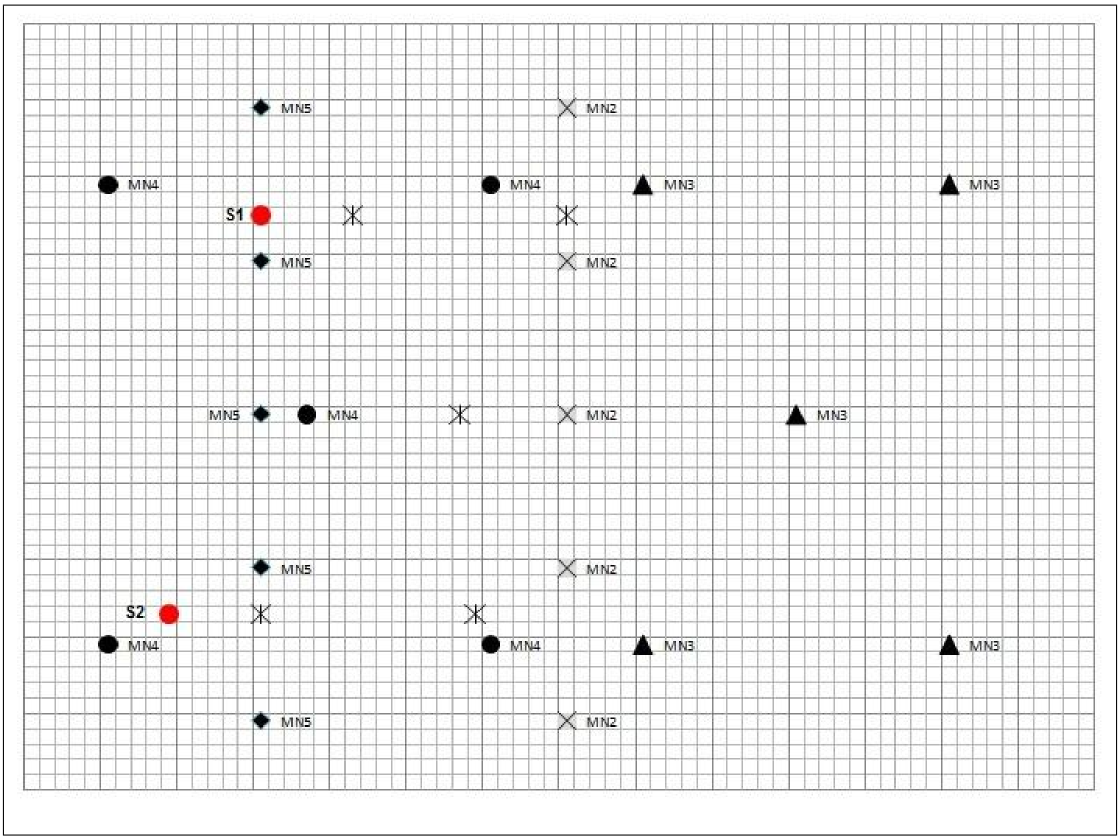


Figure 3.8: Various Monitoring Networks

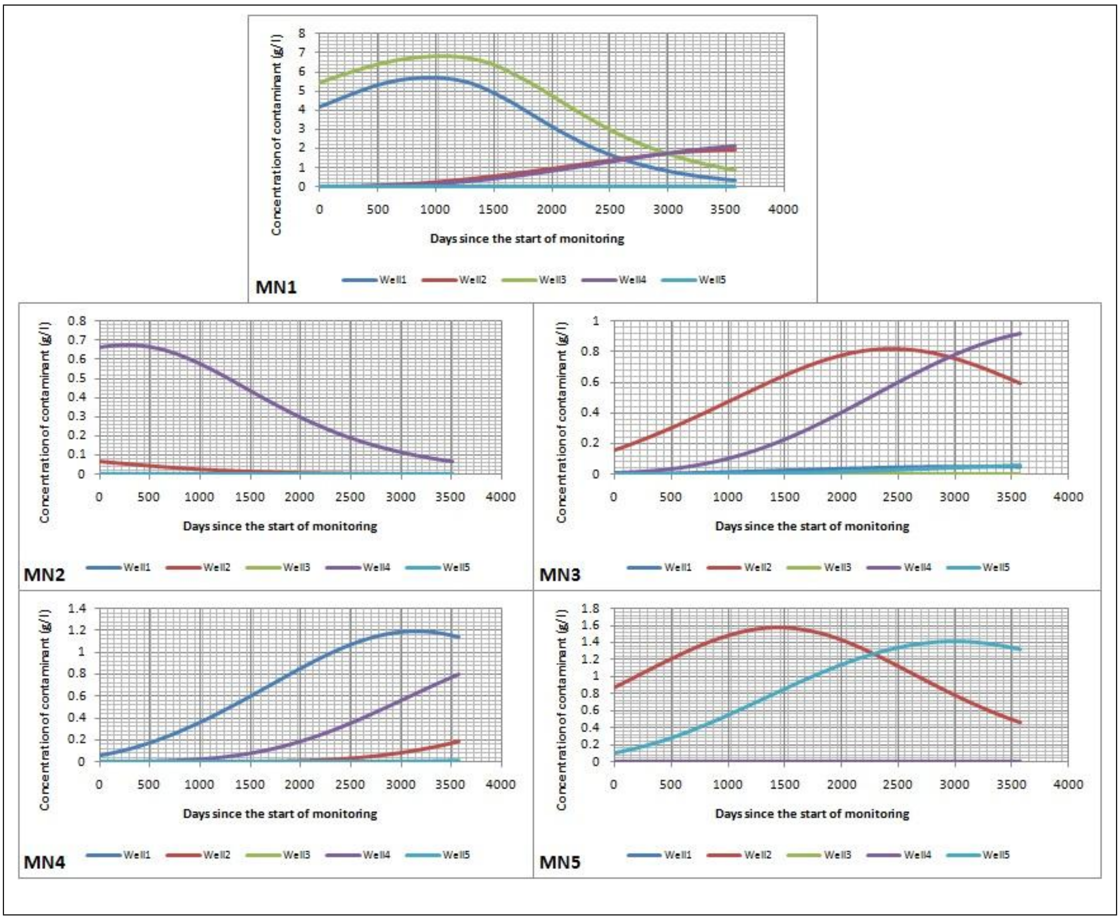


Figure 3.9: Characteristic Curves of Wells on Chosen Monitoring Networks

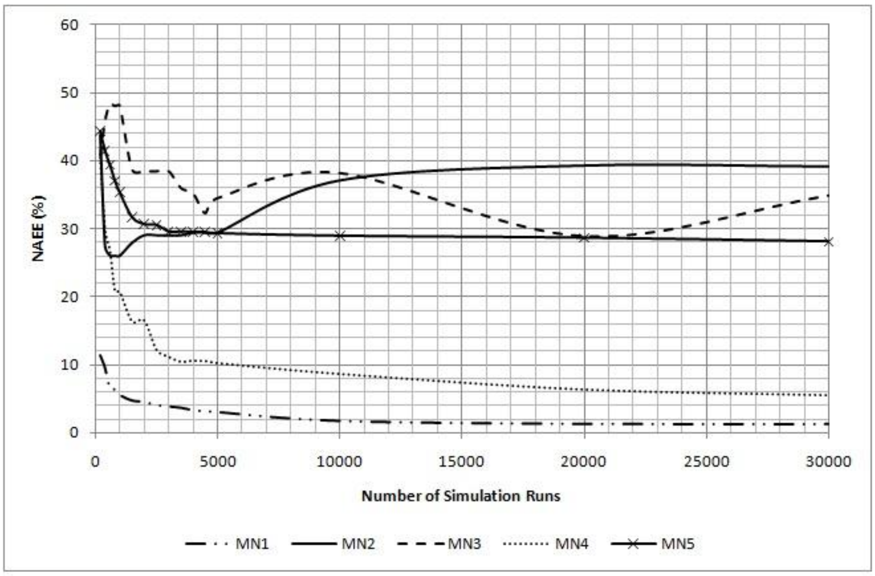


Figure 3.10: Source Release History Reconstruction using Different Monitoring Networks

NAEE was calculated and plotted as shown in Figure 3.10 as a function of number of transport simulation runs.

It is evident that MN4 produces the most accurate results. This could be due to the fact that more wells in MN4 capture the pollutant plume and are located in the direction of flow downstream of the pollutant sources. These results, however, are not rigorous enough to pinpoint a single criterion that could be used for designing a monitoring network dedicated to source identification. Further study on the criteria for designing dedicated monitoring networks for source identification is now being carried out.

3.5 Conclusion

A linked simulation-optimization method for source identification was developed based on adaptive simulated annealing. It was applied to an illustrative study area. The results obtained were compared with those obtained using genetic algorithm, a more widely used optimization approach. It is evident from the limited numerical experiments that adaptive simulated annealing algorithm based solutions converge to the actual source fluxes faster than genetic algorithm based solutions. This results in substantial saving in computational time. The source fluxes identified by using adaptive simulated annealing are closer to actual fluxes when compared to the results obtained using genetic algorithm, even when the observation data are erroneous and the hydrogeological parameters are uncertain. It can be concluded that adaptive simulated annealing is computationally more efficient for use in simulation-optimization based methods for identification of unknown groundwater pollutant sources, especially in a time constrained environment.

Use of ASA has the potential to reduce CPU time required for solution by an order of magnitude. In some cases, with a very large number of iterations in the linked simulation-optimization approach, it is possible that the solutions obtained using GA could converge to a marginally better solution compared to that of an ASA based algorithm. However, it appears that ASA based solutions converge very close to the optimal solution using only a small fraction of the iterations required while using GA. For a much larger scale and complex study area, this computational efficiency may be vital. Although computational time may not be the most important factor in choosing the optimization algorithm, use of ASA may result in an optimal solution within a reasonable number of search iterations.

The next chapter presents a methodology for iterative design of a monitoring network and efficient source estimation. The relevance of designing a pollutant monitoring network to enhance the efficiency of source identification is also discussed.

Chapter 4

Methodology for Initial Estimation of Unknown Pollutant Source Characteristics and Design of Monitoring Network

Any unknown groundwater pollutant source identification problem is very difficult to solve if the time of first activity of the source is unspecified. Most of the existing methodologies and the methodologies presented in the previous chapter assume that reliable estimates of potential source location and the time when these sources start activity exist. However, this may not always be the case. In this chapter, an attempt has been made to solve this problem by matching the time-indexed sequence of pollutant concentration observed at every location in the monitoring network to the entire estimated pollutant breakthrough curve at that location. Apart from this, missing pollutant concentration observations and misaligned estimated and observed pollutant concentration sequences may also affect the solution results of methodology developed in the previous chapter. Since the primary objective of this study is to develop an efficient linked simulation-optimization based methodology for characterizing or identifying unknown pollutant sources in any groundwater aquifer, it is essential to address these issues. Dynamic time warping (DTW) distance has been used as the measure of dissimilarity between the observed and estimated sequences of pollutant concentration. Use of DTW distance in place of Euclidean distance helps limit the influence of missing observation data and those of misaligned estimated and observed pollutant concentration sequences on the estimated unknown pollutant source characteristics.

This chapter discusses a methodology which can be used for reliable estimation of source characteristics in situations where background information on pollutant source characteristics is either non-existent or unreliable. In order to generate more observation information for accurate estimation of source characteristics, a designed monitoring network is essential. Therefore, a methodology is developed utilizing the initial estimates of various source characteristics to design an efficient monitoring network, exclusively for enhancing the efficiency and accuracy of pollutant source flux magnitude reconstruction. The performance of the developed methodology is evaluated for a hypothetical contamination scenario and also for a real-life contaminated aquifer site.

Very often, contamination of groundwater is initially detected in one or more arbitrarily located wells. These wells are referred to as "*detection wells*" in this study. Figure 4.1 illustrates a typical study area at the time of initial detection of a contamination event in a multilayered groundwater aquifer. It is assumed that the pollutant was first observed at a single detection well. At this point, from preliminary investigation, it might be possible to locate a finite number of potential source locations as shown in Figure 4.1. Information related to the likelihood of any of these potential source locations being the actual source location is not available. The

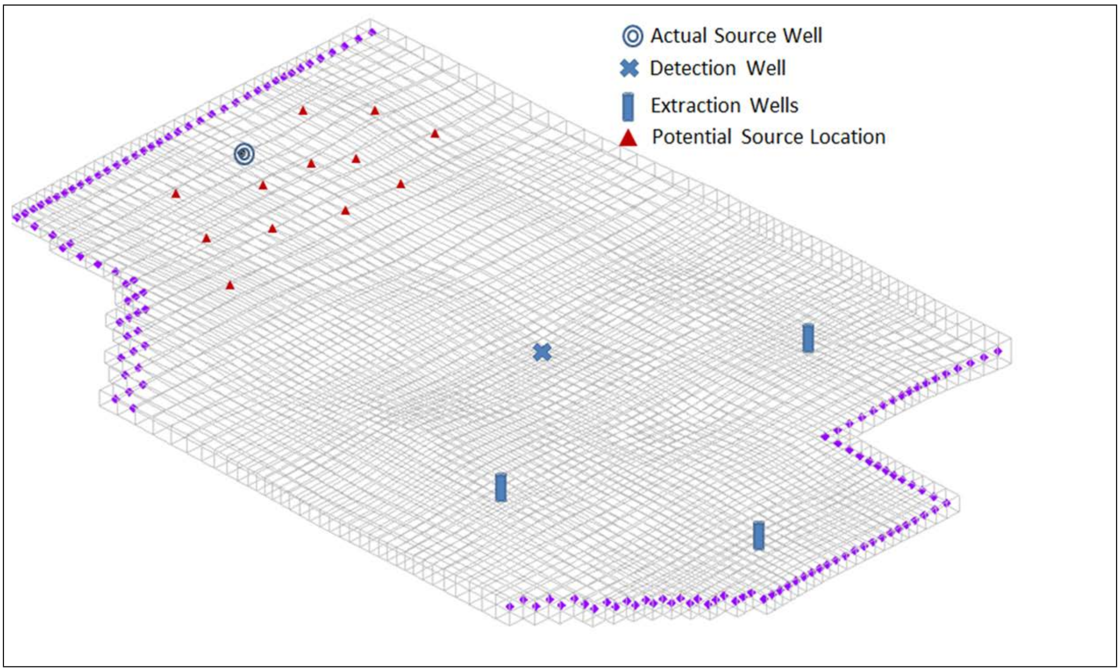


Figure 4.1: Illustrative Example of Initial Pollutant Detection

pollutant plume boundary is not known and the time at which source activity began cannot be ascertained at this stage. This problem of preliminary source characteristics estimation is formulated as an optimization problem with integer valued decision variables. Candidate solutions for unknown source characteristics are generated by the optimization module. In order to incorporate the physical processes governing groundwater flow and transport, this optimization module is linked with a numerical groundwater flow and transport simulation module. The candidate solutions generated by the search mechanism of the optimization module are used as an input for the groundwater flow and transport simulation module to produce estimated pollutant breakthrough curves at the detection wells. The objective of the optimization module is to minimize dissimilarity, or maximize the matching, between observed pollutant concentration sequence at the detection well, and the estimated sequence generated by simulation. DTW distance is used as a measure of dissimilarity between two time sequences of concentrations at a location. The preliminary source characteristics estimated in this step are then used to ascertain best location of a finite number of monitoring wells in a network such that it ensures maximum detection of pollutant and hence most effective source characterization. In order to test its performance, the developed methodology is applied to an illustrative example problem.

4.1 Preliminary Estimation of Unknown Groundwater Pollutant Source Characteristics

The first step in the proposed methodology is to generate preliminary estimates of the following groundwater pollutant characteristics for each of the potential source locations:

1. The likelihood of a potential source location being an actual source location.
2. Duration of activity of source.
3. Lag time between first activity at the source and first detection of contamination.

The developed methodology is based on several assumptions:

1. Contaminant has been detected in at least one arbitrarily located monitoring well and concentration of pollutant in this well has been measured at specified time intervals.
2. Sufficient information exists to set up and calibrate a groundwater flow model of the site.

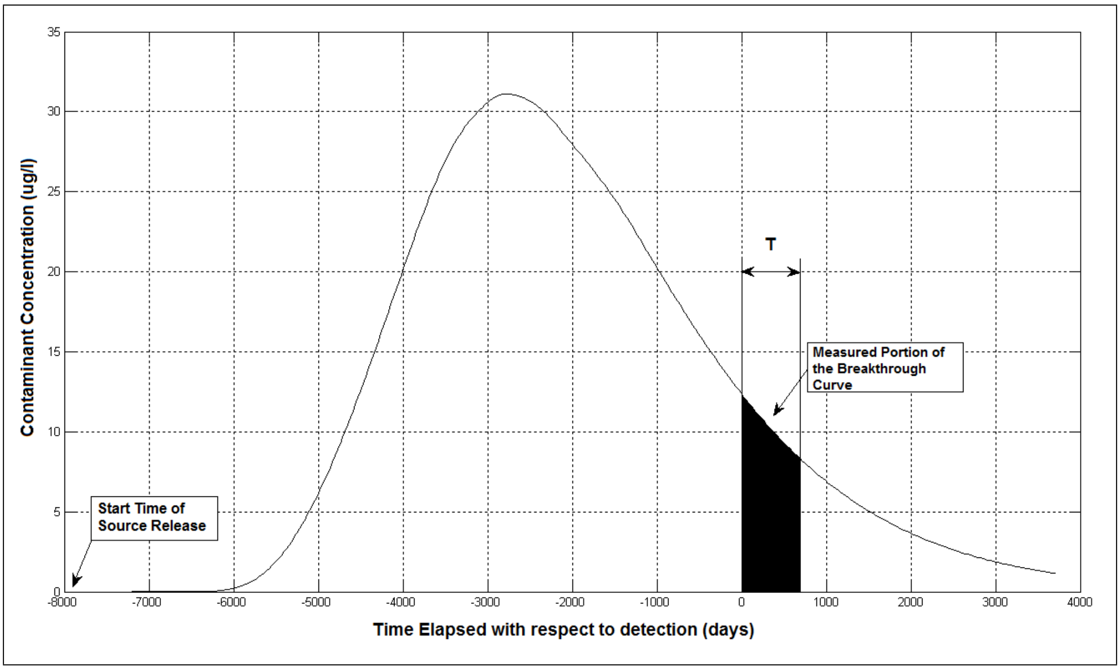


Figure 4.2: Breakthrough Curve at a Monitoring Location

3. There is only one pollutant plume and the plume is not affected by any other pollutant source outside the boundaries of the study area.

4. A finite set of suitable candidate locations for monitoring wells is available.

5. Search domain in space and time is discretized.

6. The source flux has a very low volumetric flow rate (although concentration can be high) and does not affect the hydraulic head distribution in the study area.

If a well is monitored over a specified period of time and the measured pollutant concentration data are plotted against time, it will form a small part of the entire actual pollutant breakthrough curve at the monitored location. In Figure 4.2, the breakthrough curve is shaded in the portion representing measured concentration values. A contaminant breakthrough curve at any given location in the study area can be estimated by solving governing mathematical equations of groundwater flow and transport. These equations have already been discussed in Chapter 3. It can be deduced from Equation 3.3 that, at a given monitoring location M(x,y,z) in a three-dimensional space, the breakthrough curve for pollutant concentration C(t) can be expressed as a function of various unknown parameters as shown in Equation 4.1.

$$C\left(t\right) = f\left(x, y, z, v_x, D_L, D_T, t, \theta\right) \tag{4.1}$$

Where

x,y,z = cartesian co-ordinates of the monitoring location with pollutant source as the origin;
v_x= groundwater velocity (Darcy's velocity) in horizontal direction;
D_L, D_T= dispersivity in longitudinal and transverse direction; and
θ = contaminant source flux (in terms of mass per unit time) released to groundwater.

If all flow and transport parameters are known precisely at every point in the study area, and the pollutant concentrations at a large number of monitoring wells can be measured without any errors, then these error free observations can be utilized by solving the inverse problem to determine the relative spatial location of the monitoring well with respect to the source, the temporal pattern of release at the source, magnitude of pollutant released over time, and the time lag between beginning of source release and pollutant detection in the initial monitoring well. However, such ideal conditions do not exist in real-life contamination scenarios. Flow and transport parameters are known only at a few points and measurements of pollutant concentration cannot be error free.

Reliable estimates of spatial location, release history and magnitude of pollutant release can generate a near-ideal breakthrough curve. In such a case, only the time lag between beginning of source release and pollutant detection will be unknown. Since the observed values from the detection well form a portion of this near-ideal breakthrough curve, the point on the temporal axis where observed pattern fits the ideal breakthrough curve is indicative of this lag.

Initial estimation of pollutant source characteristics can be based on the similarity between a measured concentration sequence of fixed time length to same-sized (in terms of time) portion of a candidate breakthrough curve. Traditionally, Euclidean distance has been used to measure this similarity. However, this distance measure has several disadvantages when applied to this case:

1. Euclidean distance works well only when the frequency of observed and estimated concentrations match.

2. Any missing data in observed concentration sequence can compromise the effectiveness of this measure.

In this methodology, no prior information on any of the source characteristics is assumed to exist at this stage. It is more appropriate to use a pattern comparison technique which can estimate the similarity of two time series very efficiently even if their sampling frequencies don't match, or if some measurements are missing. This can be achieved by using dynamic time warping (DTW) distance.

4.1.1 Pattern Comparison using Dynamic Time Warping Distance

Pattern comparison techniques have been widely used for speech processing and recognition. In this study, dissimilarity or distance between a test pattern and a set of reference patterns is used as a measure of comparison. The test pattern can be represented as:

$$\mathbf{T} = \{\mathbf{t_1}, \mathbf{t_2}, \mathbf{t_3}, ..., \mathbf{t_N}\}$$

where each $\mathbf{t_i}$ is a vector consisting of measured concentrations at one or more detection wells at time period i, and N is the total number of measurements at any given well over the initial monitoring period 'T'. The set of reference patterns can be represented as $\{\mathbf{R^1}, \mathbf{R^2}, ..., \mathbf{R^V}\}$ where each $\mathbf{R^j}$ is a sequence on the estimated breakthrough curve of equal time length (T) as the test pattern.

$$\mathbf{R^j} = \left\{\mathbf{r^j_1}, \mathbf{r^j_2}, \mathbf{r^j_3}, ..., \mathbf{r^j_M}\right\}$$

where M is the total number of concentration values on the estimated breakthrough curve in a time duration 'T'. The goal of pattern comparison in this study is to identify the reference pattern of time length 'T' on the estimated breakthrough curve that has minimum dissimilarity (or distance) with the test pattern. In order to determine the global similarity between the test pattern $\mathbf{T}$ and any reference pattern $\mathbf{R^j}$, the following aspects need to be taken into account:

1. Although $\mathbf{T}$ and $\mathbf{R^j}$ are of equal time length, the number of samples in each of these sequences may be different.

2. $\mathbf{T}$ and $\mathbf{R^j}$ need not line up in time in a well-prescribed manner. This is because the estimated breakthrough curve is being generated with

approximations of actual source characteristics and the test pattern is being compared to all same-sized portions of the breakthrough curve.

3. Pairs of vectors need to be compared for ascertaining local dissimilarity and facilitating temporal lineup between **T** and $\mathbf{R^j}$.

Therefore, a method to solve this pattern comparison problem must be able to use a local dissimilarity measure and a global method of time alignment. This can be achieved in a number of ways. A detailed discussion of these methods is presented in Rabiner and Juang (1993). In this study, dynamic time warping (DTW) distance was chosen as the appropriate method of pattern comparison. One of the most important reasons for using this technique is that it has embedded time alignment.

The aim of dynamic time warping is to find a warping path such that the local dissimilarity or distance between the test sequence **T** and reference sequence $\mathbf{R^j}$ is minimum. In other words, DTW picks the deformation of time axes of test and reference sequences such that it brings the two time series as close to each other as possible. A warping path 'p' can be represented as a sequence $p = (p_1, p_2, ...p_L)$ where each individual member p_l of the sequence consists of a pair of integer indices $(n_l, m_l)\epsilon[1 : N]x[1 : M]$ where N and M represents the size of test and reference sequence respectively. The local dissimilarity for any warping path can be defined as shown in Equation 4.2. The dynamic time warping (DTW) distance between the test and reference sequence is given by Equation 4.3.

$$d_p\left(\mathbf{T}, \mathbf{R^j}\right) = \sum_{l=1}^{L} d\left(\mathbf{t_{n_l}}, \mathbf{r^j_{m_l}}\right) \tag{4.2}$$

$$DTW\left(\mathbf{T},\mathbf{R^j}\right) = min\left\{d_p\left(\mathbf{T},\mathbf{R^j}\right)\right\} \tag{4.3}$$

where 'd' is a local dissimilarity function. In this study, Euclidean distance or the l_2 norm is used as the local dissimilarity measure.

The advantage of using DTW distance as a similarity measure instead of Euclidean distance lies in the fact that exact time alignment between observed and estimated sequences is not necessary and hence frequency at which concentration is observed at the monitoring wells need not match the frequency at which estimated concentrations are available on the candidate characteristic curve. Also, this distance measure is not very much affected by a few missing data instances in the observation sequence.

4.1.2 Pattern Comparison using DTW Distance to Estimate the Time of First Activity of Unknown Pollutant Source

An illustrative study area is used to illustrate the methodology and explain how DTW distance can be used as a similarity measure. The observation sequence consists of 16 measurements covering a 900-day period with readings taken every 60 days from the beginning of detection. The location of the pollutant source is assumed to be known. Therefore, for this specified location, it is possible to simulate a template breakthrough curve, covering a span of time since the start of activity of the source. The observed sequence is compared to all similar sized portions of the template breakthrough curve and corresponding dynamic time warping distances are calculated. A few of these comparisons are shown in Figure 4.3. Figure

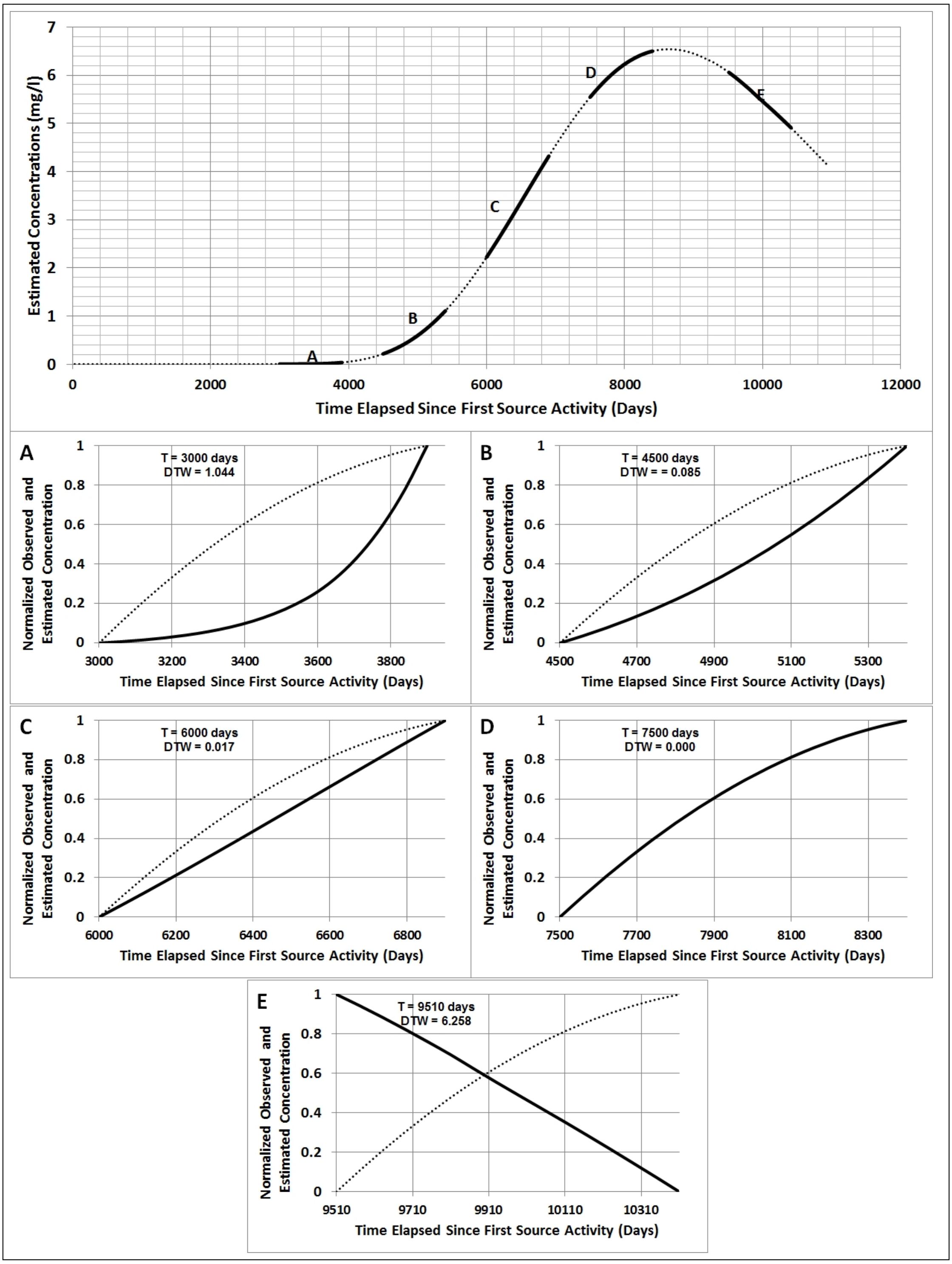


Figure 4.3: Illustrative Example of Pattern Comparison using Dynamic Time Warping

4.3 shows the ideal breakthrough curve and pattern comparison between observed and estimated sequences at various positions in the time domain. The pattern comparison process begins at time $t = 0$ on the time scale. Since in this case the observations are available over 900 days, the observed sequence is compared with the template sequence on the breakthrough curve for intervals of 90 days. The estimated sequence is normalized and dynamic time warping distance is calculated. The process then moves forward by shifting to the next observation (or by one time step) on the breakthrough curve. In this case, each successive observation on the breakthrough curve is available at a 30-day interval. Hence, in each iteration, the comparison window moves 30 days. The time horizon of this study is assumed to be 30 years.

The observed concentration values are compared with 365 individual estimated concentration sequences. In Figure 4.3, five of the possible 365 instances are shown. It should be noted that if the time lag between each successive observed and estimated concentration is the same, and if there are no missing values in the observed sequence, dynamic time warping distance is the same as Euclidean distance. However, the advantage of using dynamic time warping distance is the fact that it aligns observed and estimated sequences on the time scale and hence it is much more robust in dealing with time mismatches and missing data. In order to illustrate this, time sequence comparison was done with all the 16 values of observed concentration and with four values for $t = 120, 360, 540$ and 720 missing. The resulting DTW distances are plotted in Figure 4.4. It can be observed that the missing observed data do not have any significant impact on the point in time domain where DTW achieves minimum distance. It can be seen in Figure 4.4

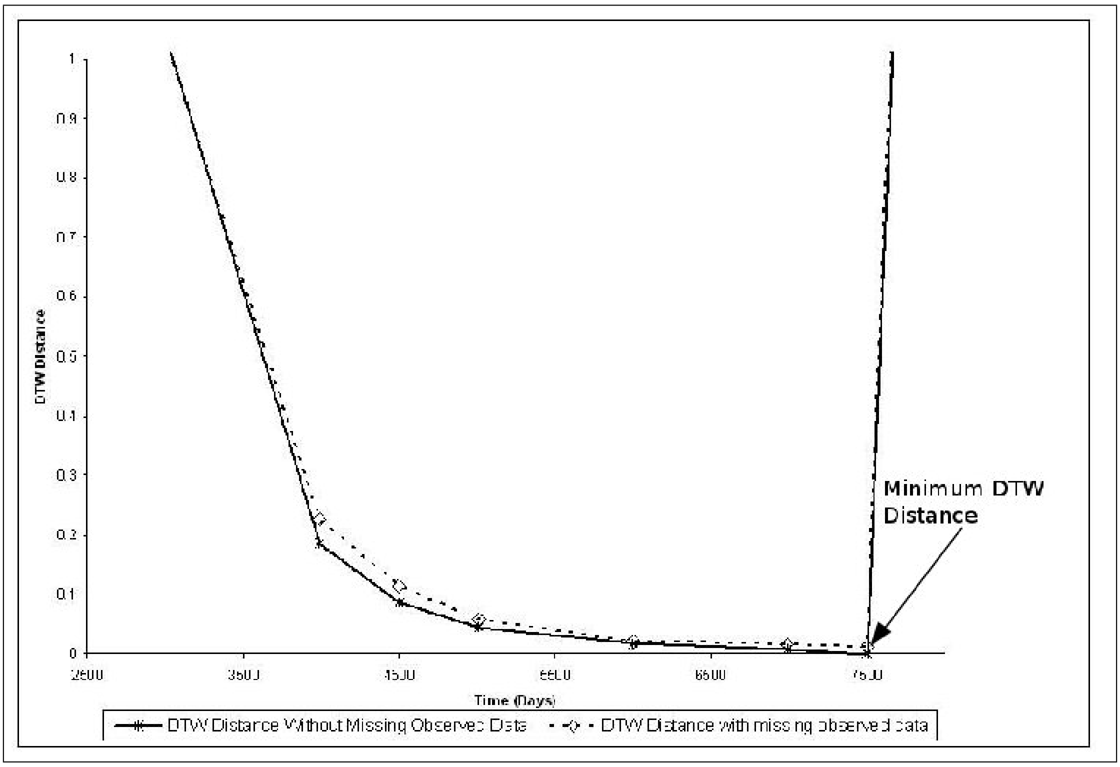


Figure 4.4: Computed DTW Distance over Time

that the DTW distance reaches its minimum at t = 7500. This indicates that the source activity began approximately 7500 days before the contamination was detected.

4.1.3 Initial Source Characteristics Estimation

Unknown pollutant source characteristics are real valued. However, the numerical modelling techniques used to solve the groundwater flow and transport equations in this study are based on finite difference method, where spatial and temporal domain is discretized. Hence, the unknown source characteristics can be mapped to a set of integers that represent this discretization. As an example, the source location can be represented by an integer corresponding to the identity of the discretized cell in the three-dimensional finite difference grid. Similarly, release history can be discretized into a number of stress periods of finite length, and a single integer can represent the number of stress periods since the beginning of the study period in which the source is active. Magnitude of source does not have a significant impact at initial estimation stage, primarily due to the fact that both the test sequence and the reference sequence are normalized to the range (0,1) before being compared. Hence, at this stage, essentially the pattern of increasing or decreasing sequences are being matched and not their precise magnitudes. Integer approximations shift the peak of the breakthrough curve slightly and they also have some impact on the pattern of breakthrough curve. However, it is not very significant. This is illustrated in Figure 4.5. Several breakthrough curves were generated at a given monitoring location using:

1. Actual source flux magnitudes

2. Source flux magnitudes rounded off to nearest 100s.
3. Source flux magnitudes averaged over four stress periods.
4. Source flux magnitudes averaged over the entire study period.

It can be seen from Figure 4.5 that the peak of the breakthrough curve is shifted most by averaging the source flux values over the entire study period. For all other approximations, the breakthrough curve maintains more or less the same pattern as the original one.

Since integer approximation of the source characteristics does not seem to change the pattern of the breakthrough curve drastically, the problem of initial estimation can be cast into an optimization problem with integer valued decision variables. For every potential source location, the decision variables are:

1. Release history: This variable can take an integer value between 1 and the maximum number of stress periods in the transport model. It indicates the number of stress periods since the beginning in which the given source has been active.
2. Source magnitude: Since the pattern of breakthrough curves and not the exact values of pollutant concentrations are being matched, it is sufficient at this stage to estimate the relative source strength in each of the active source periods. Hence, the values can be discretized and mapped to an integer domain. As an example, if the maximum possible limit of source magnitude is 570 mg/l and the level of accuracy is to the nearest 100 then source strength can take any value between 1 and 6.

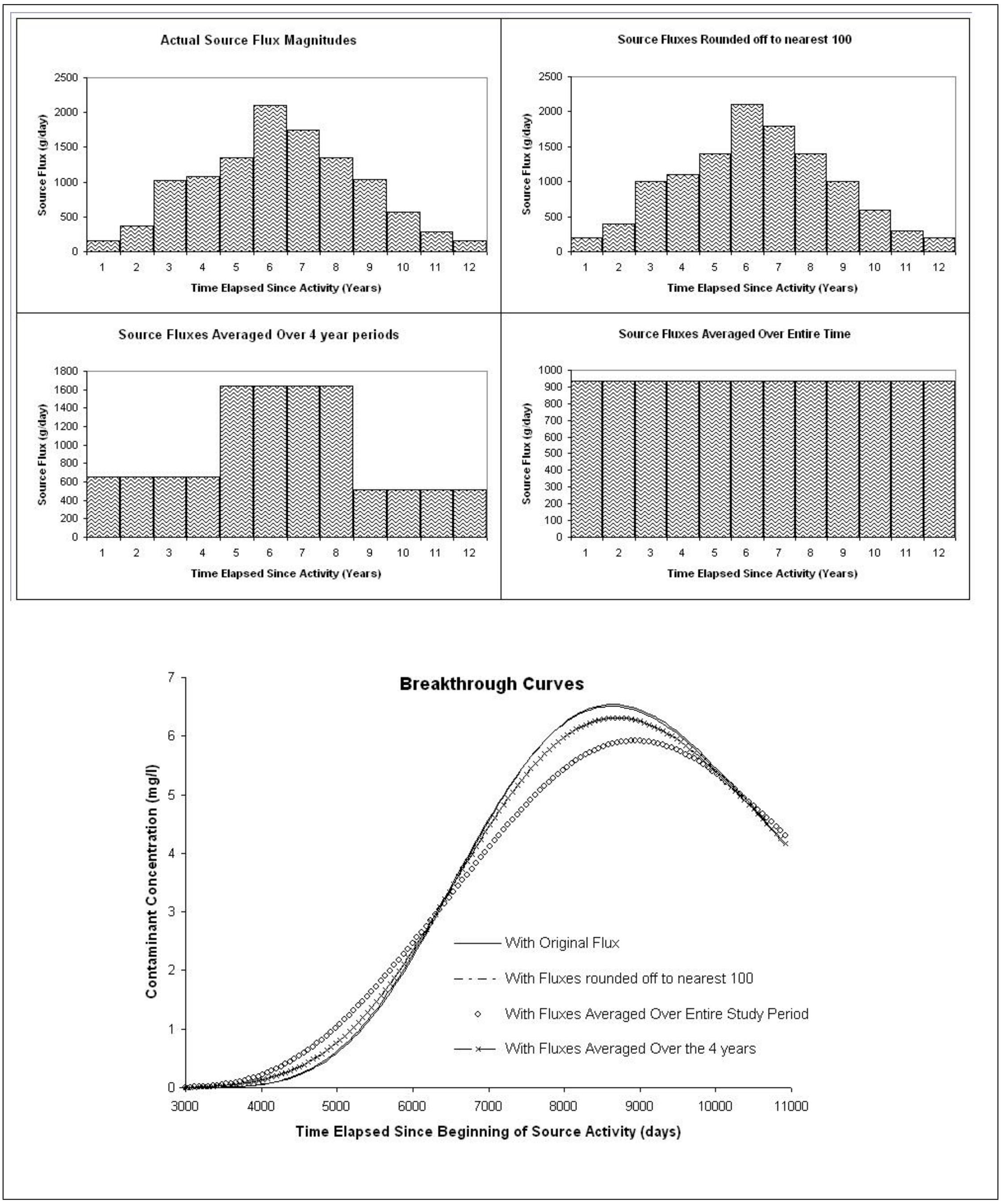


Figure 4.5: Effects of Approximation of Source Flux Magnitudes on Breakthrough Curve

In this methodology, a set of discrete candidate source characteristic values is generated by the optimization algorithm and these values are used as input to the forward flow and transport model in order to generate a candidate characteristic curve. The cost function of this candidate solution set is the minimum DTW distance achieved while comparing the observed values to same time-sized portions of candidate breakthrough curve.

Initial estimation of candidate source characteristics can now be represented as an optimization problem as shown in Equation 4.4.

$$Minimize : F_1 = d_{DTW}\left(\mathbf{C_{obs}}, \mathbf{C^k_{est}}\right), k\epsilon[1, 2, ...n] \tag{4.4}$$

Subject to the constraint:

$$C^k_{est} = f\left(x, y, z, v_x, D_L, D_T, t, \theta\right)$$

Where

$d_{DTW}\left(\mathbf{a}, \mathbf{b}\right)$ = DTW distance between time sequences **'a'** and **'b'**;

C_{obs} = concentration time series observed at detection well;

C^k_{est} = k^{th} concentration time series of equal duration as C_{obs} on the estimated candidate characteristic curve;

n = total number of reference sequences equal in duration to the observed time series;

x,y,z = Cartesian co-ordinates of the monitoring location with pollutant source as the origin;

v_x = groundwater velocity (Darcy's velocity) in horizontal direction;

t = time elapsed since the beginning of pollutant release in

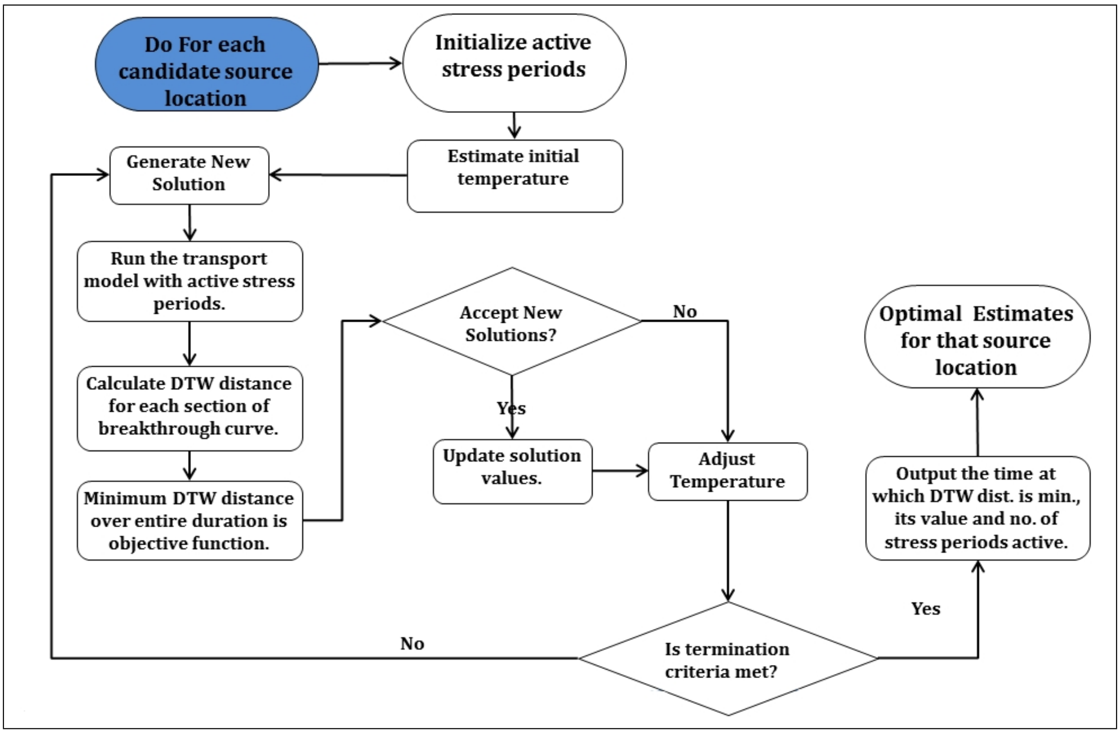


Figure 4.6: Flowchart Showing the Steps in Initial Estimation

groundwater;

D_L, D_T = dispersivity in longitudinal and transverse direction; and

θ = pollutant source flux (in terms of mass per unit time) released into groundwater.

All the steps involved in initial estimation of source characteristics are presented as a flow chart in Figure 4.6.

4.2 Monitoring Network Design for Efficient Unknown Pollutant Source Characterisation

Once the initial estimates of various unknown source characteristics become available, the next step is to choose optimal locations on a monitoring network. The objective of designing a monitoring network for pollutant source identification is to maximize the concentration observed at monitoring locations (Datta and Purwar, 1992; Mahar and Datta, 1997). Since monitoring locations can be represented by integer values, the monitoring network design problem can be expressed as an integer programming problem. The objective function for the monitoring network design linked simulation-optimization model in this case is given in Equation 4.5.

$$Minimize : F_2 = \sum_{s=1}^{S} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{a}{(C_{est} + b)} \tag{4.5}$$

Where

C_{est} = estimated contaminant concentration at candidate monitoring location 'i', in time period 'j', assuming the 's^{th}' candidate source location. Values of C_{est} are computed by groundwater transport simulation model and this simulation model acts as a binding constraint.

a = a specified constant (taken as 1 in this study);

b = a small constant to avoid undefined values for zero concentration (10^{-10} in this case;

S = maximum number of candidate source locations selected in previous step;

I = maximum number of monitoring locations in the network;

J = total number of observations to be recorded at a monitoring well in the observation period.

With this objective function, the optimization model chooses those monitoring locations where the value of pollutant concentration observed is expected to be maximum for all candidate source locations and associated release history and magnitude.

4.3 Conclusion

This chapter outlined the developed methodology for generating initial estimates of unknown groundwater pollutant source characteristics in situations where the time of first activity of the source is unspecified. These initial estimates are utilized for designing an optimal monitoring network specifically for the purpose of unknown groundwater pollutant source identification. This methodology for designing a dedicated monitoring network for more efficient identification of pollution sources was described. It is assumed that measurement data is initially available only at one well where the contamination event has been detected. From the data obtained from this single *"detection well"*, initial estimation of various unknown source characteristics is generated using the principles of time series matching. In order to quantify the dissimilarity between two time series, dynamic time warping distance is utilized. This approach is useful especially if there are missing data in the observed time series, or if the observed and estimated data series are not aligned on the temporal axis. Solution results obtained show that the developed methodology is capable of producing reliable estimates and that the monitoring network designed based on such preliminary estimation has the potential to improve the efficiency of unknown groundwater pollutant source identification results. Limitation of the methodology is that, in its present state, it is applicable only to contaminated sites affected with a single pollutant plume, i.e. one continuous plume. The performance evaluation of the methodologies proposed in this chapter is presented in Chapter 5.

Chapter 5

Performance Evaluation of Methodology for Initial Estimation of Unknown Pollutant Source Characteristics and Design of Monitoring Network

This chapter discusses the performance evaluation of the developed methodology for initial estimation of unknown pollutant source characteristics, and subsequent utilization of these estimates to design an optimal monitoring network. In order to evaluate the applicability of this methodology, it is applied to an illustrative study area with synthetically generated concentration measurements over space and time, as well as to a contaminated groundwater aquifer.

5.1 Performance Evaluation Criteria for Initial Estimation of Unknown Pollutant Source Characteristics

Exact characterisation of source characteristics becomes very difficult in groundwater aquifer contamination scenarios where no prior information exists on any of the unknown pollutant source characteristics, and the pollutant concentration data is available only from one or a very few detection wells. In such situations, reliable initial estimates of unknown pollutant source characteristics can be generated by using the methodology proposed in Section 4.1. The measure of effectiveness of the methodology for initial estimation of source characteristics should be its ability to generate an estimated range of source characteristics that contains the actual source characteristics, even if the observed concentration series contains moderate levels of errors. In order to test this, several sets of erroneous observation data were generated using the method mentioned in Section 3.3.1. The methodology for initial source characteristics estimation was then applied to an illustrative study area using error free as well as erroneous concentration measurement data. The effectiveness of this methodology was assessed based on its ability to estimate the actual source characteristics.

5.1.1 Performance Evaluation Criteria for Monitoring Network Design

Efficiency of the methodology for monitoring network design is tested on the premise that the optimal monitoring network should produce better solutions of source characteristics when compared with arbitrary monitoring networks consisting of the same number of wells. The methodology for source release history reconstruction, as described in Section 3.1, is applied using the concentration measurement data obtained from a set of optimal monitoring locations, and then to three sets of arbitrarily located monitoring networks consisting of the same number of wells. Effectiveness of the monitoring network design methodology is evaluated based on the comparison of results obtained using each set of monitoring network locations.

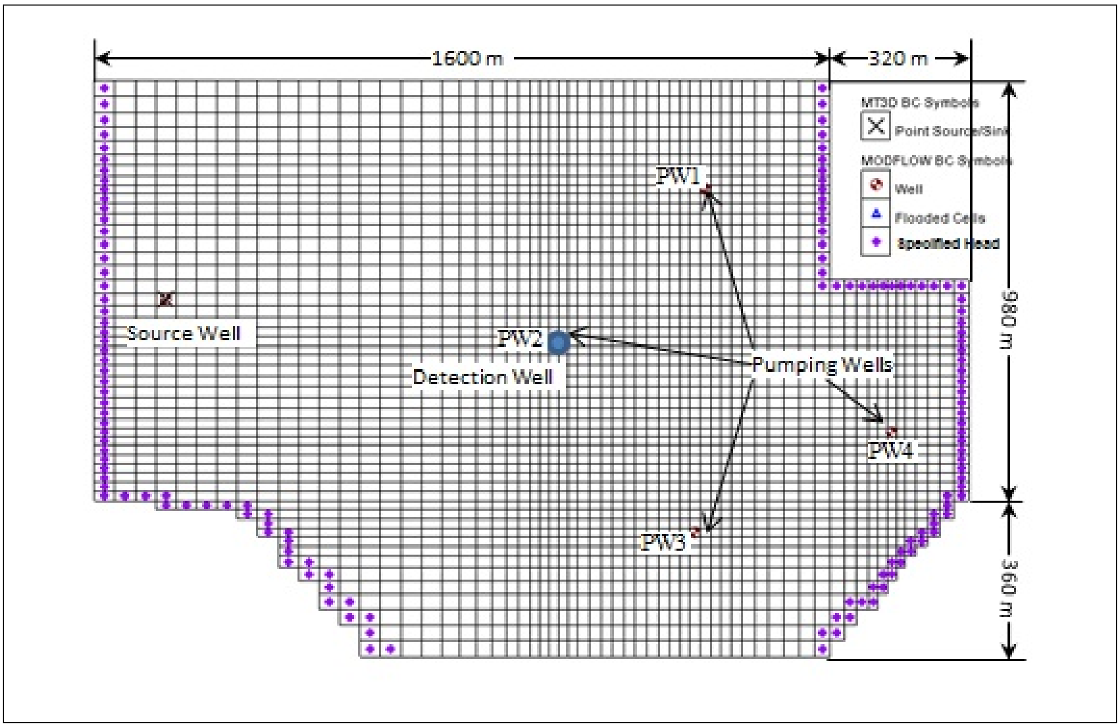


Figure 5.1: Illustrative Study Area

5.2 Results and Discussion

This section presents the results of application of the developed methodology for initial estimation of source characteristics and subsequent monitoring network design to an illustrative study area. Results obtained under various scenarios are discussed in detail.

5.2.1 Study Area

The hypothetical study area is a heterogeneous two-dimensional aquifer measuring 1920 m x 1340 m but irregular in shape as shown in Figure 5.1. The east and west boundaries are constant head boundaries, whereas the north and south boundaries are no flow boundaries. There is a pollutant source well and four pumping wells in the area as shown in Figure 5.1. The

Table 5.1: Model Parameters

Parameter	Value
Length of study area (m)	1920
Width of study area (m)	1320
Saturated thickness, b(m)	10
Grid spacing in x-direction, x (m)	25 to 50
Grid spacing in y-direction, y (m)	25 to 50
Grid Spacing in z-direction, z (m)	10
Hydraulic conductivity in x-direction, Kxx (m/day)	11.2
Hydraulic conductivity in y-direction, Kyy (m/day)	8
Vertical anisotropy	5
Hydraulic gradient (m/m)	0.00463
Effective porosity, θ	0.3
Longitudinal dispersivity, α_L (m)	15
Transverse dispersivity, α_T (m)	3
Initial pollutant concentration (mg/l)	0.00

pollutant is first detected at a pumping well named PW2. A variable grid size ranging from 50 m $\times$ 50 m to 25 m $\times$ 25 m is used for finite difference based numerical solution of groundwater flow and transport equations. Grid size is smaller close to the pumping wells and it expands to regular size further away from pumping wells. This is to enhance the accuracy of numerical flow and transport modelling results. Other important model parameters are listed in Table 5.1. A non-reactive pollutant is assumed to originate from a point source in this study.

5.2.2 Initial Estimation of Source Characteristics

In order to test the proposed methodology, an observed concentration sequence was generated at the detection well (PW2) by simulating groundwater pollutant transport using MT3DMS with known source characteristics. Actual source characteristics are listed in Table 5.2 and the

Table 5.2: Actual Source Characteristics

Source location (Row, Column)	(19,6)
Time of first activity (days before detection)	6720
Duration of source activity (years)	12
Detection well location (Row, Column)	(28,30)

actual release history is shown in Figure 5.2. Total time of activity of the source of pollution is 12 years and Figure 5.2 shows the average source flux for each year. The observation sequence contains 16 observations at an interval of 30 days each after 6720 days of initial source release. Since real world applications invariably involve erroneous observations, it becomes essential to evaluate the proposed methodology using erroneous data as well. Observed pollutant concentration measurements at the designated detection well are generated using MT3DMS as transport simulation model followed by perturbation as per Equation 3.7. Original and perturbed concentration measurements are plotted in Figure 5.3. While solving this illustrative example, the entire study period is taken to be 30 years. In actual contamination detection events, the duration of study period should be either based on information related to the beginning of anthropogenic/contaminating activities in the study area gathered during reconnaissance or taken to be sufficiently large. The entire time horizon is divided into 30 stress periods of one year length each. Since there is no prior information on the actual location of source, regularly spaced locations are chosen to be the potential source locations.

In this study, 101 potential source locations each covering a space of 150 m x 150 m were chosen in the entire study area. These potential source locations are shown in Figure 5.4. The total study period of 30 years is divided into 10

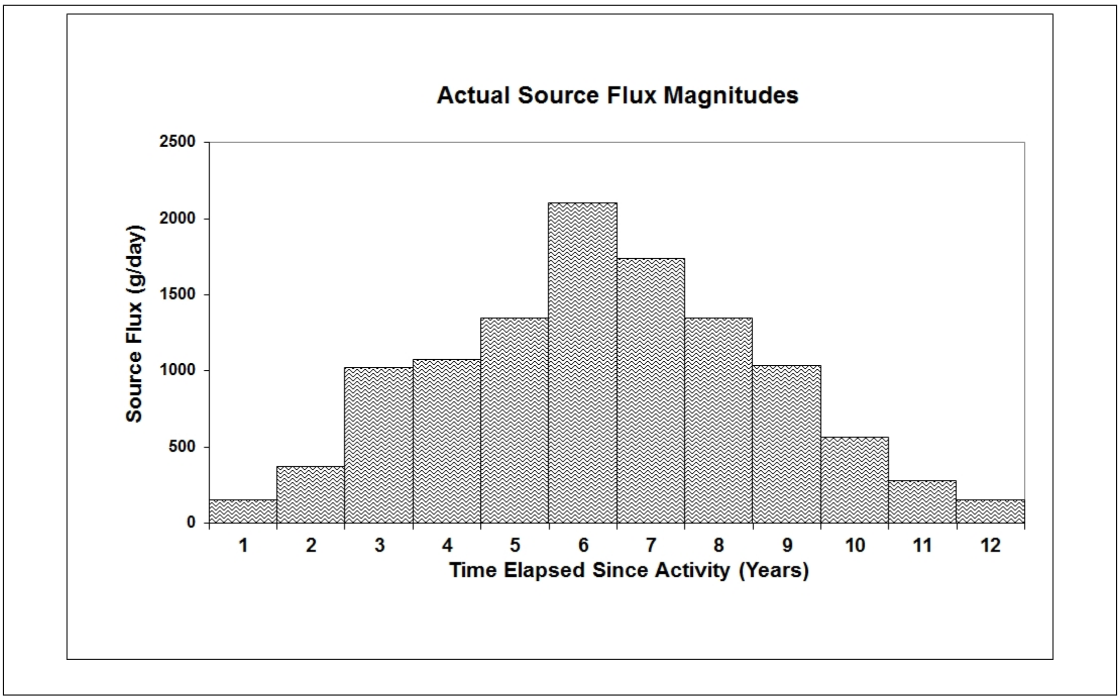


Figure 5.2: Actual Release History of the Source

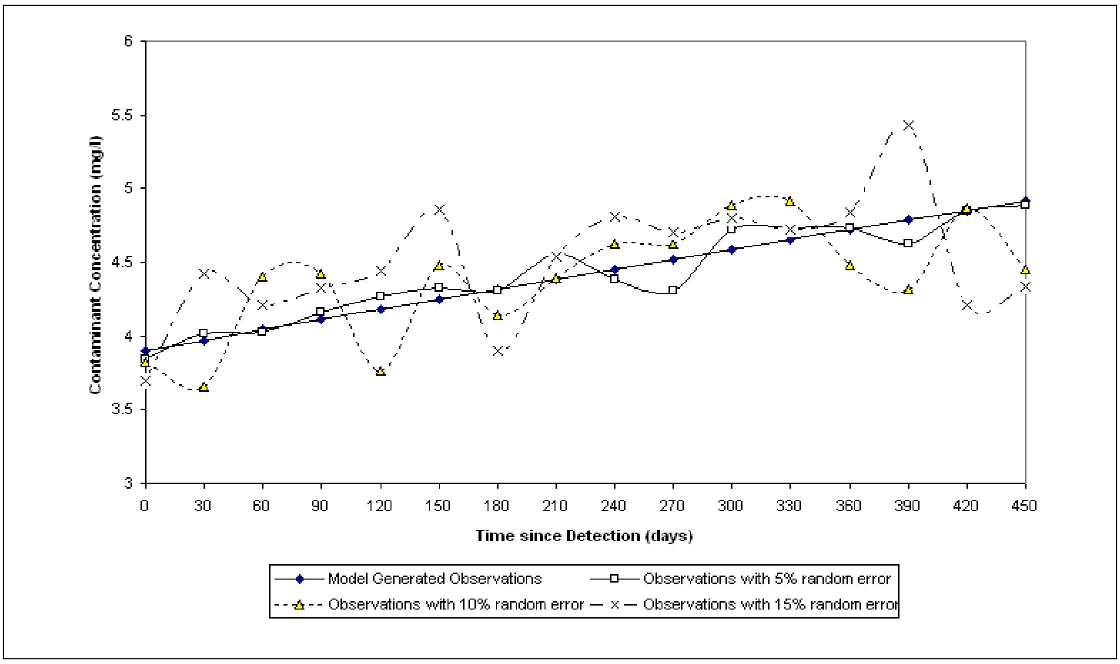


Figure 5.3: Model Generated Observation Sequences with Synthetic Errors

stress periods of three years each for the purpose of initial estimation of source release history. Since the observed concentration sequence is matched with estimated sequence after normalization, it is not very important to estimate exact source release magnitudes. It is sufficient to estimate the relative order of source release magnitudes in each stress period for use in the later stage of accurate release history reconstruction.

Initial source characteristics estimation methodology was applied to all potential sources in order to determine their optimal release history and optimal time lag between detection and source activity. Optimal source characteristics, the lowest DTW distance and time lag between first source activity and its detection have been presented for the 10 best potential source locations, estimated using both error free and erroneous observation sequences in Table 5.3.

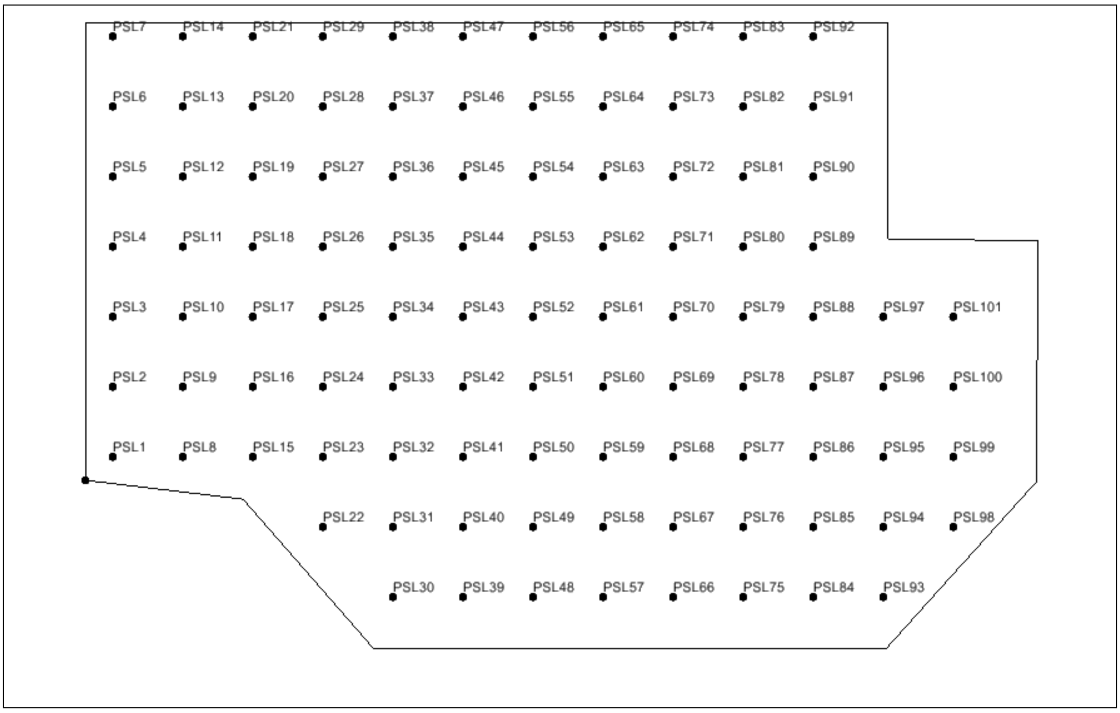


Figure 5.4: Discretized Potential Source Locations

It can be inferred from the results in Table 5.3 that a multitude of combinations of the source characteristics can produce similar effects at the detection location. For example, sources located at PSL11, PSL26 and PSL10 can each produce a concentration measurement sequence that closely matches the observed sequence with different duration of activity and different values of time lag. This is evident by the objective function values achieved in the first three rows of Table 5.3. This is an indication of the inherent non-uniqueness of this problem. It was noted during the study that the quality of estimation deteriorates as the random error added to the model-generated synthetic observations increases. For this reason, ranges of various source characteristics have been estimated using erroneous observation sequence containing 5% random error. These estimations are used in the next step to design an efficient monitoring network specifically

Table 5.3: Estimation of Preliminary Source Characteristics using Error Free Observed Data

Error Free Observation				5% Standard Error in Observation				10% Standard Error in Observation				15% Standard Error in Observation			
PSL ID	ASP	OF Value	Time Lag	PSL ID	ASP	OF Value	Lag Time	PSL ID	ASP	OF Value	Time Lag	PSL ID	ASP	OF Value	Time Lag
PSL11	7	0.0406	5460	PSL12	12	0.1213	6210	PSL4	15	0.1816	6330	PSL33	15	0.2414	5430
PSL26	7	0.0415	6060	PSL34	14	0.1220	5970	PSL5	9	0.1860	5670	PSL17	12	0.2425	5400
PSL10	10	0.0417	6420	PSL16	15	0.1222	5850	PSL25	14	0.1908	6030	PSL24	14	0.2456	6570
PSL24	8	0.0425	6630	PSL4	12	0.1240	6270	PSL11	7	0.1909	5670	PSL35	14	0.2461	5460
PSL17	11	0.0457	5640	PSL24	7	0.1272	6030	PSL33	11	0.1917	5430	PSL44	14	0.2461	6240
PSL12	11	0.0556	6000	PSL17	7	0.1298	6240	PSL34	10	0.1919	5730	PSL12	11	0.2490	5340
PSL18	10	0.0587	6240	PSL25	7	0.1303	6390	PSL26	12	0.1927	5430	PSL18	10	0.2560	5970
PSL19	7	0.0603	5940	PSL11	10	0.1303	6120	PSL27	13	0.1947	5340	PSL10	13	0.2568	5970
PSL25	9	0.0645	6300	PSL5	12	0.1314	5730	PSL36	10	0.1964	6660	PSL27	13	0.2596	5760
PSL4	13	0.0697	6660	PSL33	8	0.1408	6300	PSL17	9	0.2522	5580	PSL19	14	0.2624	6420

PSL ID = potential source location ID; ASP = no. of active stress periods; SP = stress period;
OF = minimum objective function; Lag Time = time of lag between source activity and detection

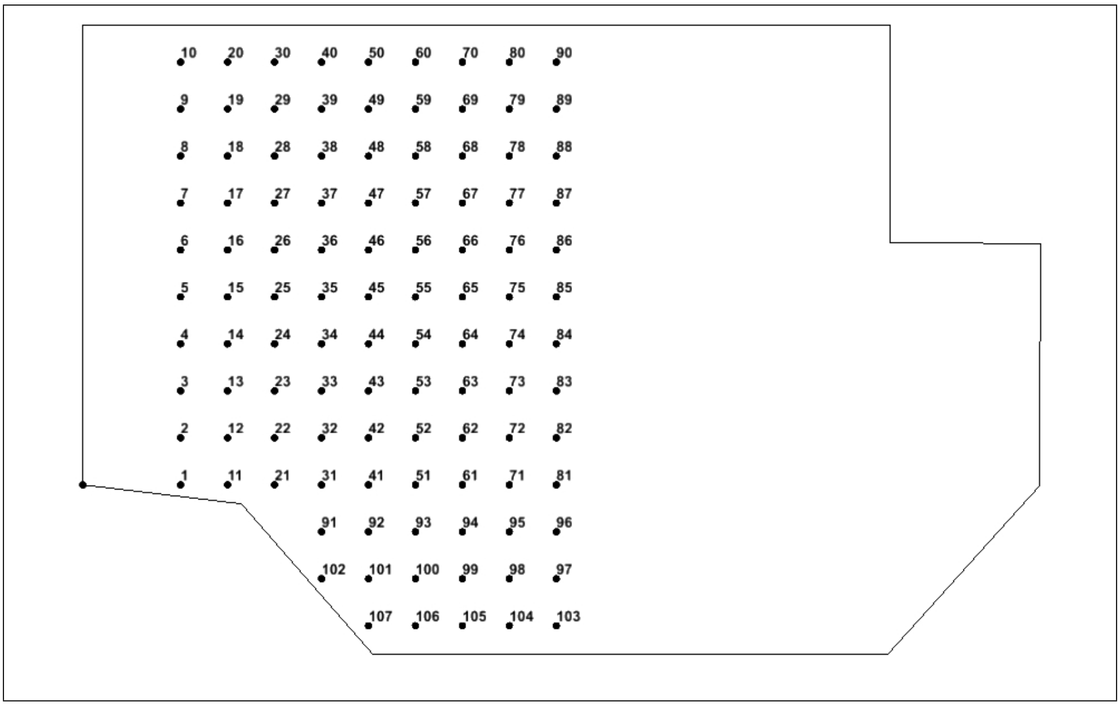


Figure 5.5: Potential Monitoring Locations

for the purpose of precise determination of unknown source characteristics.

5.2.3 Monitoring Network Design

The estimated release history for each source location from the previous step is used to design an optimal monitoring network. Since at this stage, only initial estimates of source characteristics are available, the objective of designing a monitoring network is to capture maximum possible pollutant concentrations in all estimated scenarios of source release. Candidate monitoring locations were chosen over discretized areas of 100 m x 100 m, between the farthest estimated source location and the detection location. These locations are shown in Figure 5.5. It is desired to choose 10 best monitoring locations. In order to show the efficiency of this method, the source flux regeneration algorithm presented in Chapter 3 is implemented using the observation data collected over the next two-year period from:

1. The optimal monitoring network, and
2. Three different sets of arbitrary monitoring networks each containing 10 wells.

The well locations chosen for optimal and arbitrary monitoring networks are shown in Table 5.4. These results are compared after the first 5000

Table 5.4: Monitoring Locations Chosen in the Optimal Monitoring Network and Arbitrary Monitoring Networks for Comparison

	Optimal	Arbitrary 1	Arbitrary 2	Arbitrary 3
Well 1	23	68	67	26
Well 2	24	11	30	50
Well 3	25	78	94	36
Well 4	34	69	61	32
Well 5	35	3	52	84
Well 6	45	51	23	5
Well 7	46	76	103	58
Well 8	55	54	56	78
Well 9	56	23	21	49
Well 10	57	19	75	59

iterations of the optimization algorithm. Convergence profiles for each set of observation data, and the respective solutions achieved after 5000 iterations of adaptive simulated annealing algorithm, are shown in Figure 5.6. As seen from Figure 5.6, the arbitrary monitoring networks do not achieve the same level of accuracy in regenerating source release magnitudes over time as the optimal monitoring network. From the convergence profile it may be also inferred that the measurement data obtained from the optimal monitoring network helps the optimization algorithm converge marginally faster than that using the data obtained from an arbitrary network.

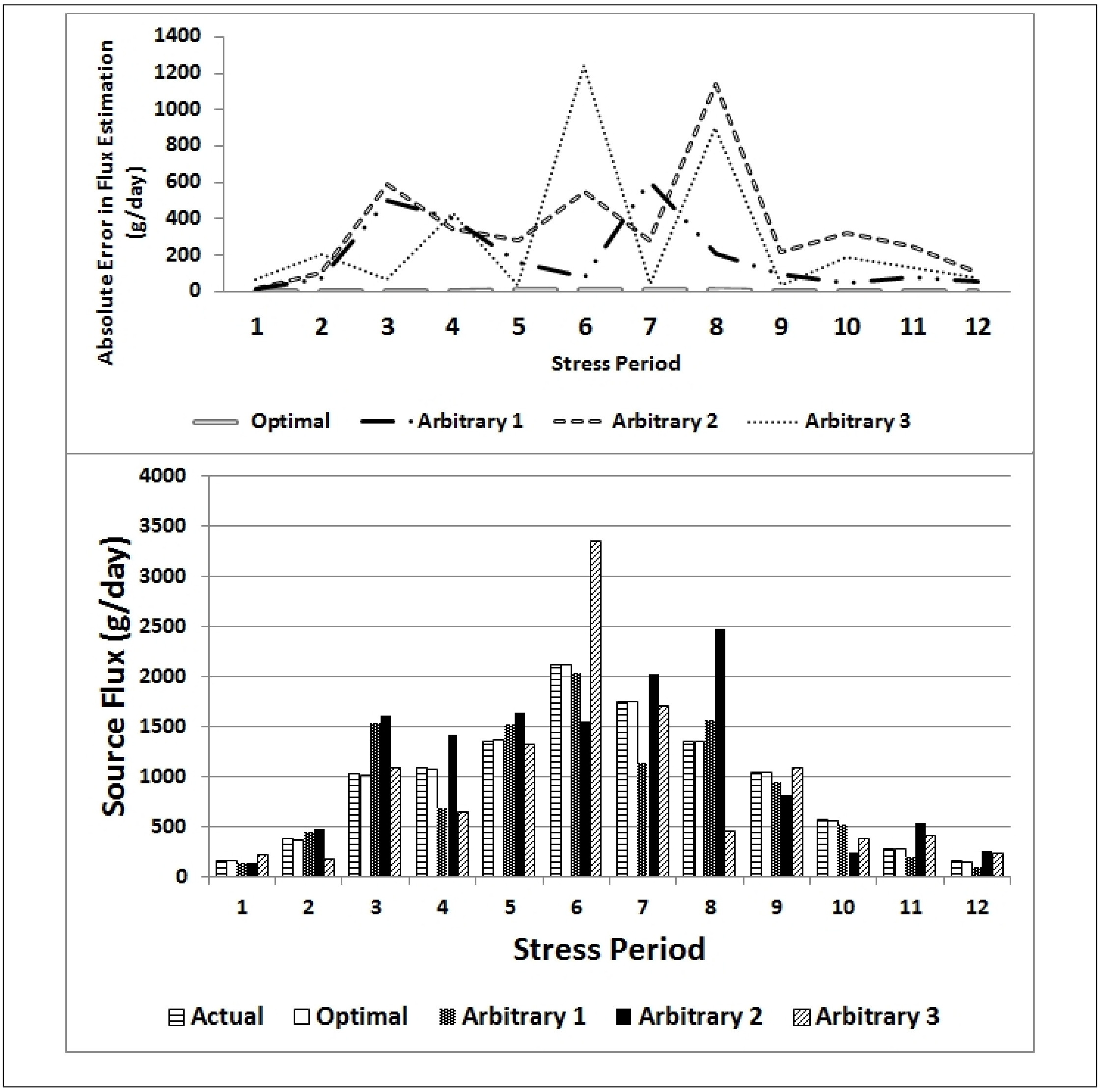


Figure 5.6: Estimated Release History

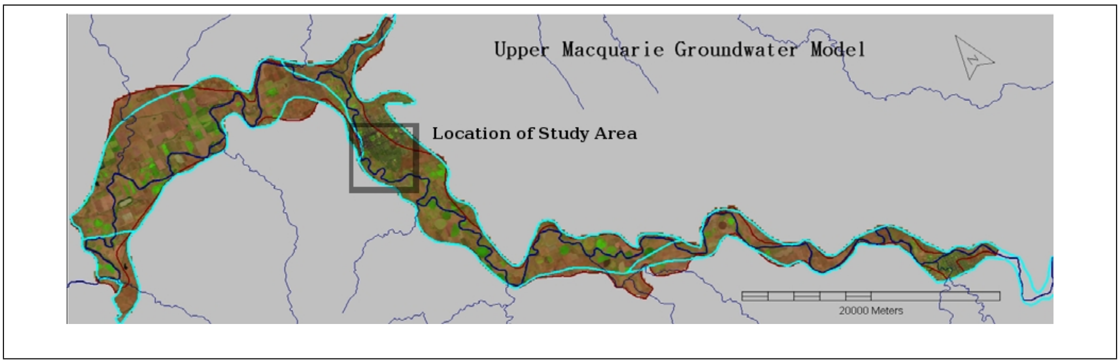


Figure 5.7: Location of the Study Area within Upper Macquarie Groundwater Model

5.3 Application to a Contaminated Aquifer

In order to evaluate the potential applicability of this methodology to actual contaminated sites, it was utilized to find the unknown source characteristics of a real-life petrochemical fuel pollutant source in a shallow unconfined aquifer.

5.3.1 Site Description

The contaminated site is located in Upper Macquarie Groundwater Management Area in New South Wales, Australia. Its exact location cannot be disclosed due to confidentiality agreements with the data provider. However, the position of this site in Upper Macquarie Groundwater model is shown in Figure 5.7. Major sources of groundwater recharge in this study area are the rainfall and contribution from the Macquarie River, which generally flows from south-east to north-west. Major sinks include extraction wells for agricultural and municipal water supply. The aquifer is unconfined and is formed by quaternary and tertiary alluvial deposits over the bedrock.

The contaminated study area falls in an urban neighbourhood and the contamination event was detected at a number of locations. After a series of complaints and the associated event of BTEX vapour emanating from building basements, an investigation into the contamination was ordered by the regulatory authority. Pollutant concentration was monitored in 74 wells over a period spanning four years with an aim to delineate the plume boundary and pinpoint the source of contamination for planning speedy and effective remediation measures. The source of contamination was attributed to a leaking fuel tank. However, the source release history was unknown. Data collected during this investigation was obtained, and is used for the performance evaluation of developed methodology for initial estimation of unknown pollutant source characteristics and subsequent utilization of these estimates to design an optimal monitoring network.

The extent of contamination was limited to about 1 km^2. Available information, however, was not sufficient to describe the boundary conditions reliably in the immediate vicinity of the contaminated area. Hence, to model the groundwater flow and transport in the contaminated study area, a much larger study area was considered. The study area is roughly 3 km $\times$ 3 km. It is bordered by Macquarie River on the western side and by impermeable bedrocks on the eastern side. The elevation of the study area ranges from nearly 245 m with respect to the Australian Height Datum (mAHD) towards the river to 283 mAHD on the north eastern side. Figure 5.8 shows the boundaries of the modelled area as well as the extent of the contaminated study area. It also shows the elevation profile and location of monitoring wells in the contaminated area.

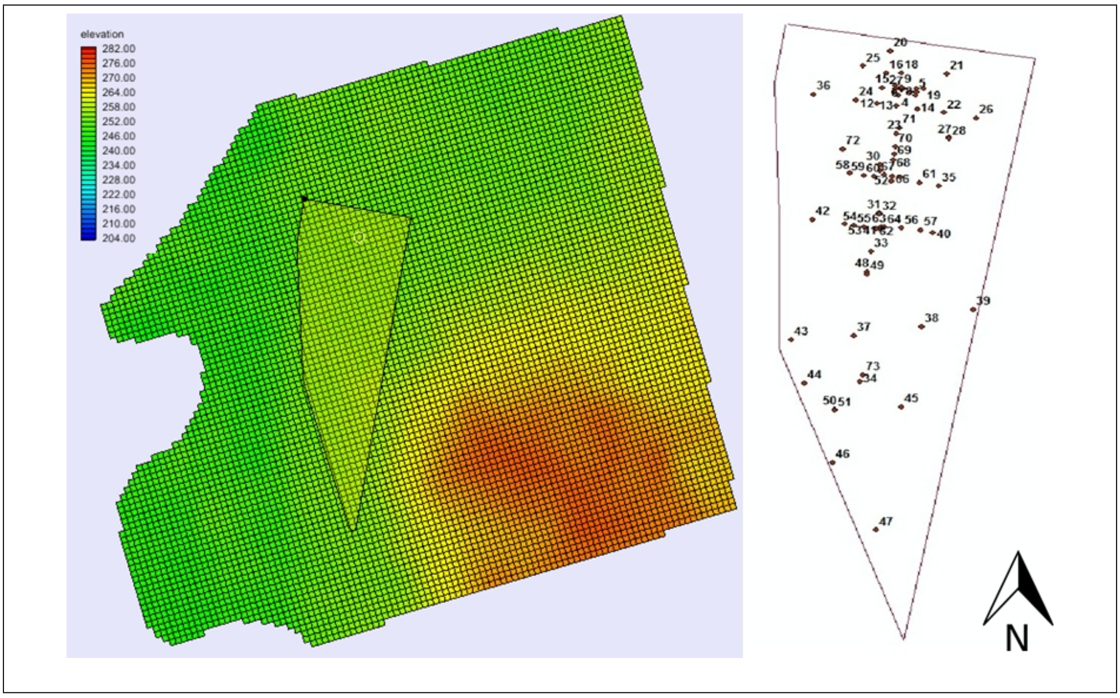


Figure 5.8: Extent of Study Area and Contaminated Area, Elevation Profile and Location of Monitoring Wells

5.3.2 Groundwater Flow Model and its Calibration

A groundwater flow model of the entire Upper Macquarie Groundwater Management Area was developed by Puech (2010). Based on the information available from this report, as well as from the information collected during contaminated site investigation, a conceptual groundwater flow model was developed for the study area. Based on the geologic information and the borehole logs available at the site, the conceptual model of the site was divided into three layers, representing the tertiary alluvium, the quaternary alluvium and the bedrock. Cross-sections of the conceptual model depicting these layers are shown in Figure 5.9. The conceptual model is shown in Figure 5.10.

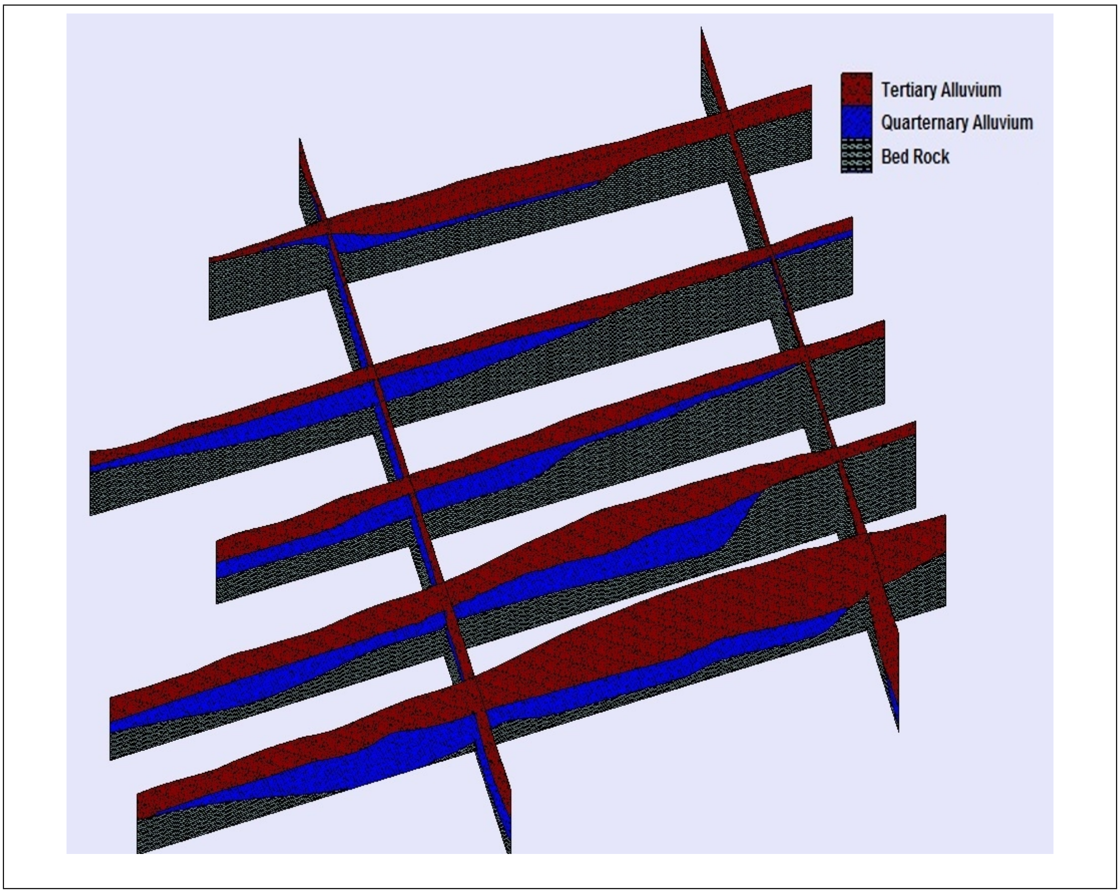


Figure 5.9: Layers of the Developed Conceptual Model

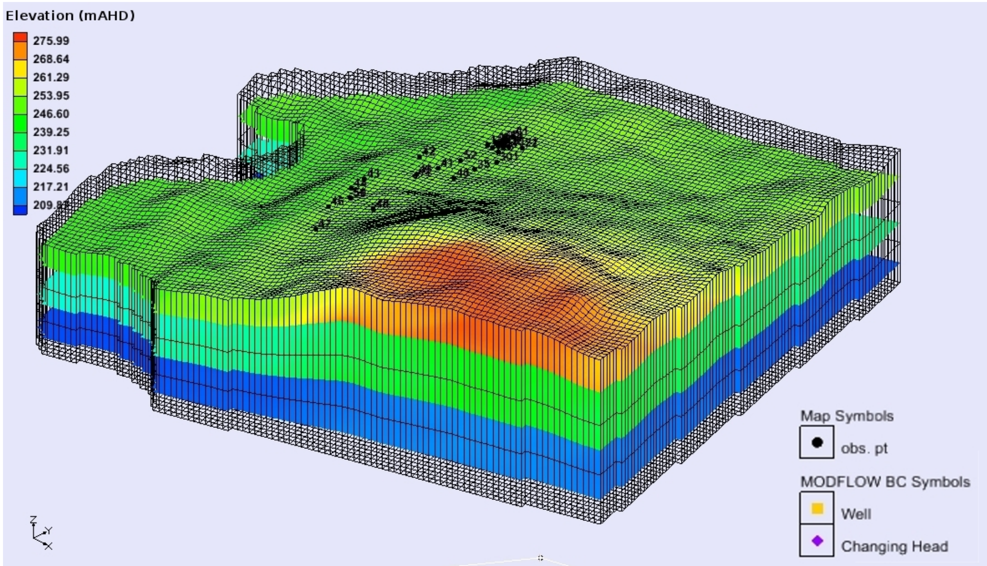


Figure 5.10: Model of the Study Area

5.3.2.1 Boundary Conditions

Macquarie River has been represented as a time varying specified head boundary. The portion of Macquarie river considered in this model falls between two weirs. Since the data pertaining to standing water level at both these weirs are available from NSW Office of Water, specified heads along this boundary have been represented by the river stage at a given point in time. The northern and southern boundaries have also been considered as time varying specified head boundaries. Specified heads along these boundaries were calculated by interpolating observed heads at a few wells lying along these boundaries. The eastern boundary is a no flow boundary as mentioned in the Upper Macquarie Groundwater Model (Puech, 2010).

5.3.2.2 Sources and Sinks

Macquarie River, that has been modelled as a time varying specified head boundary condition, is a major source of groundwater in this study area. Apart from this, rainfall also contributes to the groundwater recharge. The study area receives moderate rainfall with a long-term average of 583 mm/yr. Evaporation rate in the study area peaks at about 260 mm/month during the months of December and January. The study area falls in the irrigated recharge zone of Upper Macquarie Groundwater Model and the initial value of recharge is calculated from the formula presented in Puech (2010). Extraction of groundwater in the study area is mainly through wells for the purpose of drinking water supply and agriculture. Extraction wells included in the model have been presented in Table 5.5.

Table 5.5: Extraction Wells in the Study Area

Name	Row	Column	Layer	Extraction (m^3/day)
Agricultural 1	28	77	1	0.8767
Agricultural 2	66	38	1	4.6
Agricultural 3	56	44	1	8.11
Municipal Supply 1	57	46	1	1118.88
Municipal Supply 2	68	58	1	756.77
Agricultural 4	76	54	1	503.1
Agricultural 5	76	54	2	503.1
Municipal Supply 3	85	46	1	987.62

5.3.2.3 Model Calibration

Hydraulic conductivity in the study area varies spatially, even within the same layer. In order to represent this heterogeneity, each layer in the conceptual model was divided into a number of hydraulic conductivity zones. The extent, locations and values of hydraulic conductivity for each

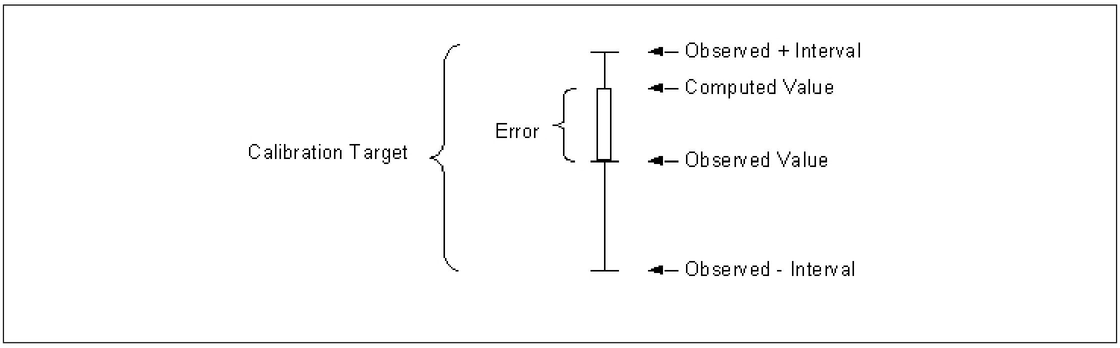


Figure 5.11: Components of a Calibration Target Box Plot

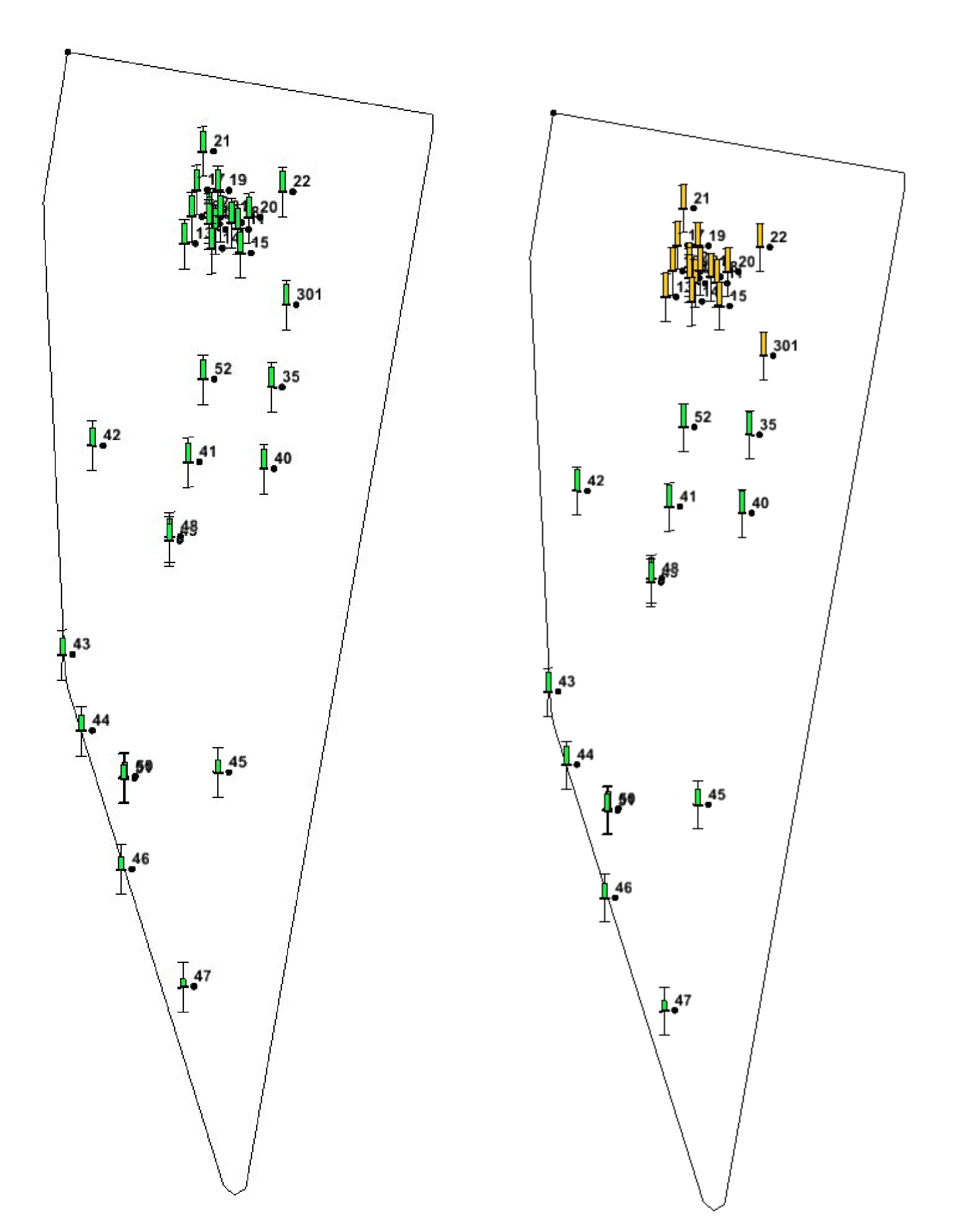
of these zones were obtained from the detailed model proposed in Puech (2010). Recharge into the aquifer was considered unknown and its values were obtained through calibration.

The developed model was calibrated using groundwater levels of monitoring wells obtained from Puech (2010) and from all 74 monitoring wells in the contaminated zone of study area. Automatic calibration of the model was carried out using PEST following the guidelines mentioned in Doherty and Hunt (2010). The calibration process aimed at ensuring that the deviation between the actual observed heads and the heads simulated by the calibrated model was within 2 m with a confidence level of 90%, at all head monitoring well locations. Deviation between the observed heads and those simulated by the calibrated model can be plotted as calibration target box plots. The components of a calibration target box plot are illustrated in Figure 5.11. A set of calibration targets provides useful feedback on the magnitude, direction (high, low), and spatial distribution of the calibration error. These calibration target box plots can be plotted at each monitoring location for every stress period of the flow model. In this flow model, 50 stress periods of six months each were considered and hence it is not possible to show all the results. Calibration target box plots for stress periods ending in June 1997 and December 2001 are shown in Figure 5.12.

The box plots are colour-coded to indicate relative deviation of the observed and estimated heads. Green box plots indicate that the estimated values are within the calibration target of head deviations lying within 2 m. Yellow boxes indicate slight over-reach from the calibration target and the red boxes indicate that estimated values do not meet the calibration target. It may be noted that most of the deviations are positive which means that the heads simulated by the calibrated model are higher than those observed. This may have been caused by the fact that actual draw-down from agricultural and other wells in the study area is slightly higher than those modelled. This is also evident from the fact that, during the stress period ending in December 2001, some of the observations are slightly off the calibration target. July to December is a relatively dry period and groundwater withdrawal is generally higher. However, it is very difficult to get precise pumping data for a very accurate calibration. Also, the instances of calibration target breach are only a few and hence the calibration can be considered satisfactory.

Figure 5.13 shows the plot of observed and estimated groundwater heads at the monitoring locations for the last four years of the modelling period. It shows a linear correlation between observed and estimated groundwater head values with a correlation co-efficient of 0.97. Hence, it can be inferred from Figure 5.13 that the calibration was satisfactory and the calibration targets were achieved.

Groundwater heads simulated using the calibrated model towards the end of the study period are shown in Figures 5.14 through 5.16. It can be noted



(a) June 1997 (b) December 2001

Figure 5.12: Calibration Results of Groundwater Flow Model

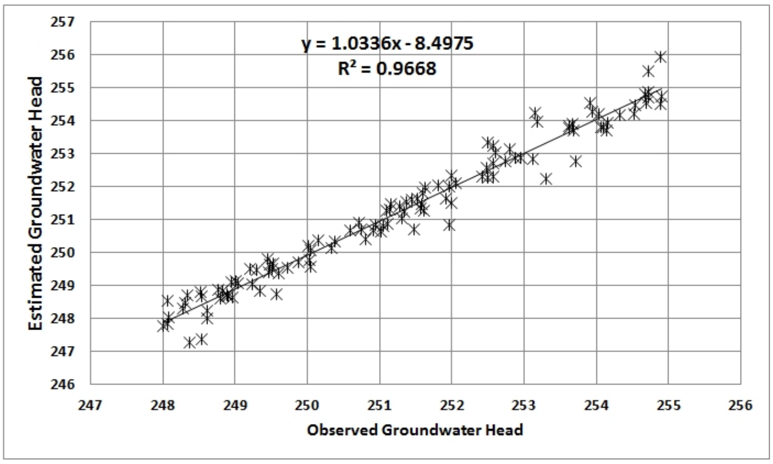


Figure 5.13: Estimated vs Observed Heads after Calibration

that the extraction or pumping wells, shown as yellow squares in Figure 5.14, have a pronounced effect on the groundwater flow regime in this study area. The large amount of pumping from these wells as presented in Table 5.5 has a significant impact on the local direction of groundwater flow which is not in agreement with the regional groundwater flow direction mentioned in Puech (2010).

5.3.3 Groundwater Transport Model

In order to develop a groundwater transport model of the affected area, it was assumed that sorption is negligible and can be ignored. Dispersivity values depend on scale of discretization and values for this were estimated based on Fetter (1994). Rate constants for aerobic natural attenuation of BTEX were obtained from experimentation results as reported in Lu et al. (1999). A detailed list of salient transport model parameters is presented in Table 5.6.

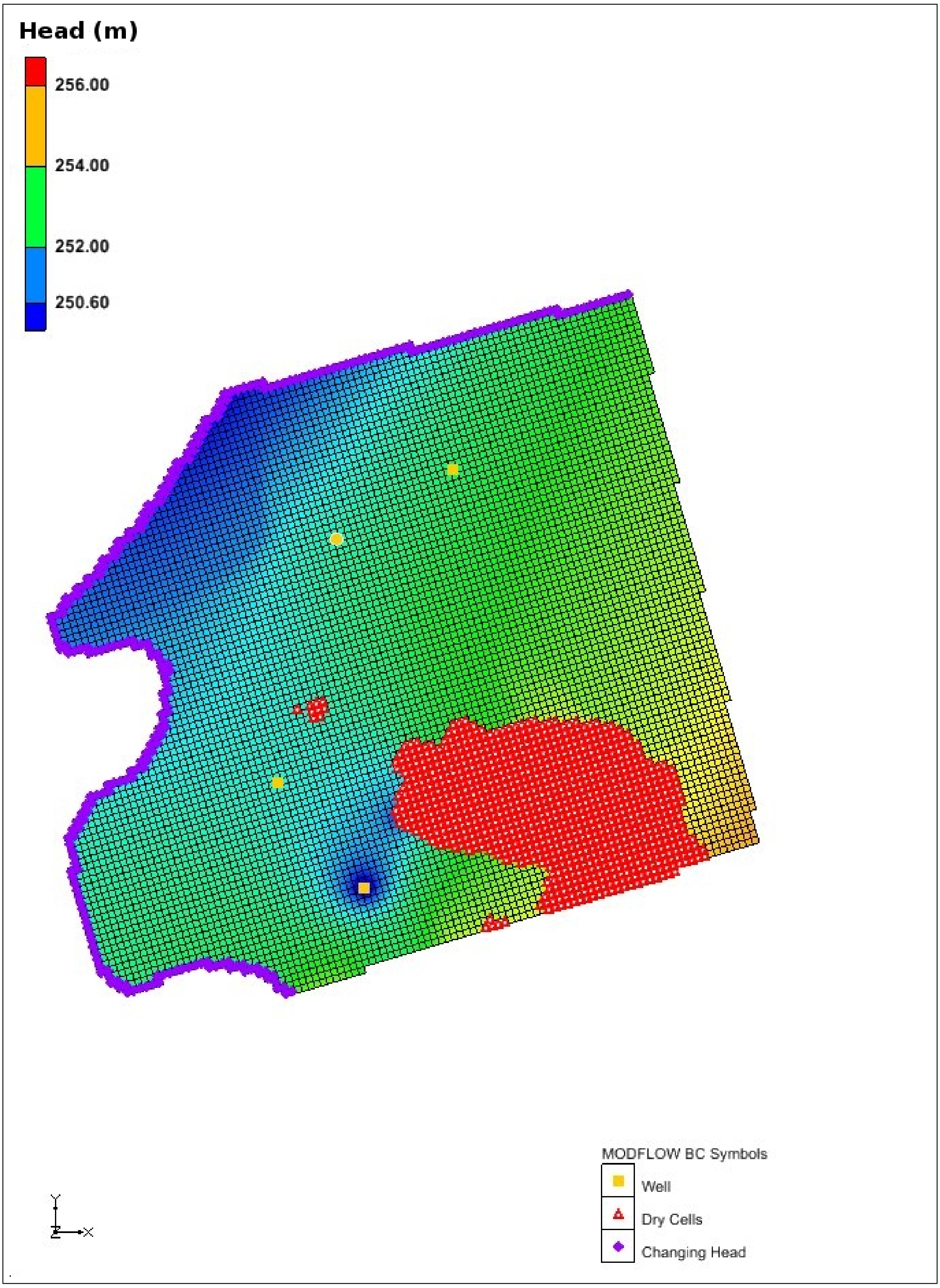


Figure 5.14: Simulated Heads in Layer 1

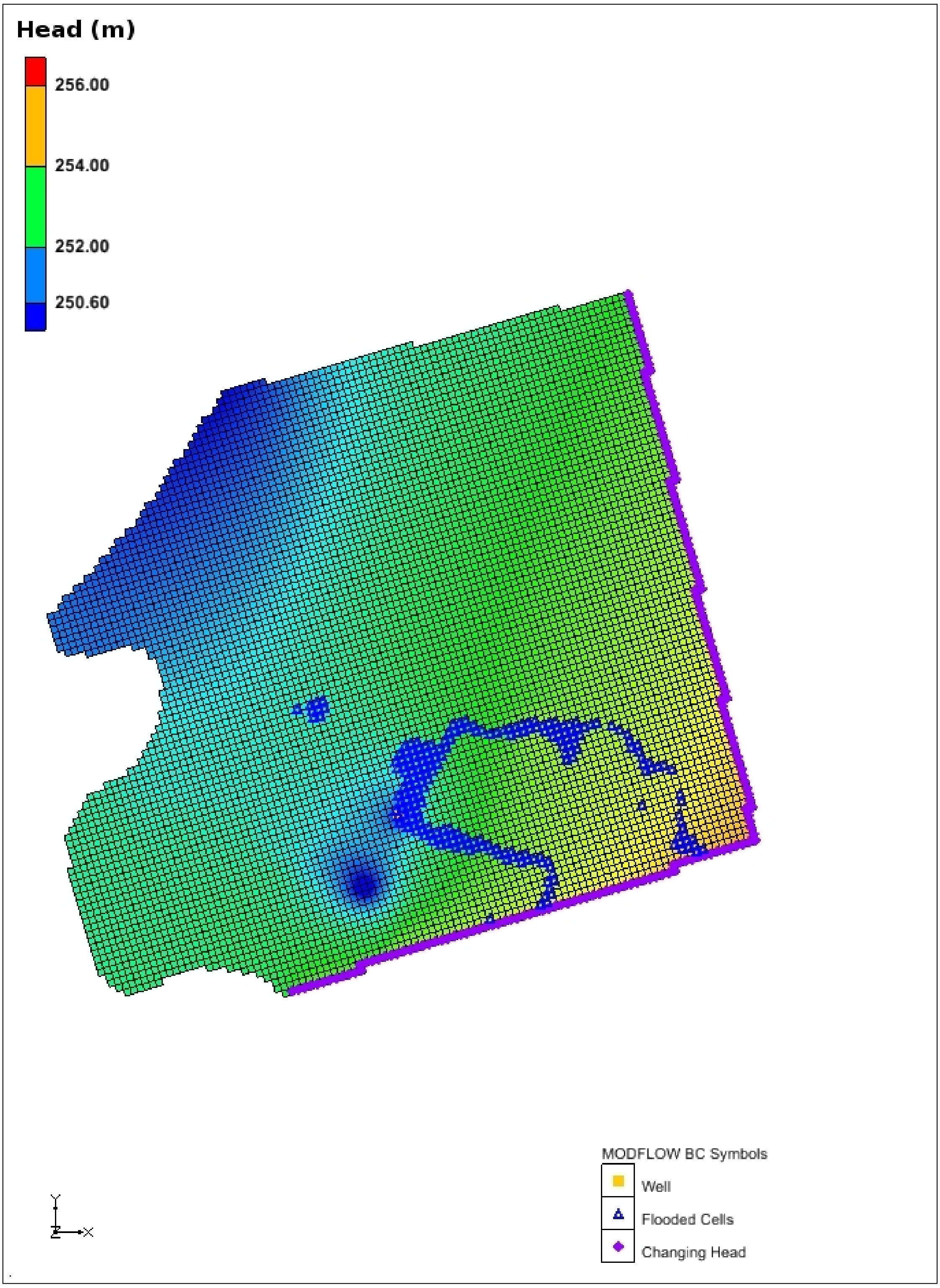


Figure 5.15: Simulated Heads in Layer 2

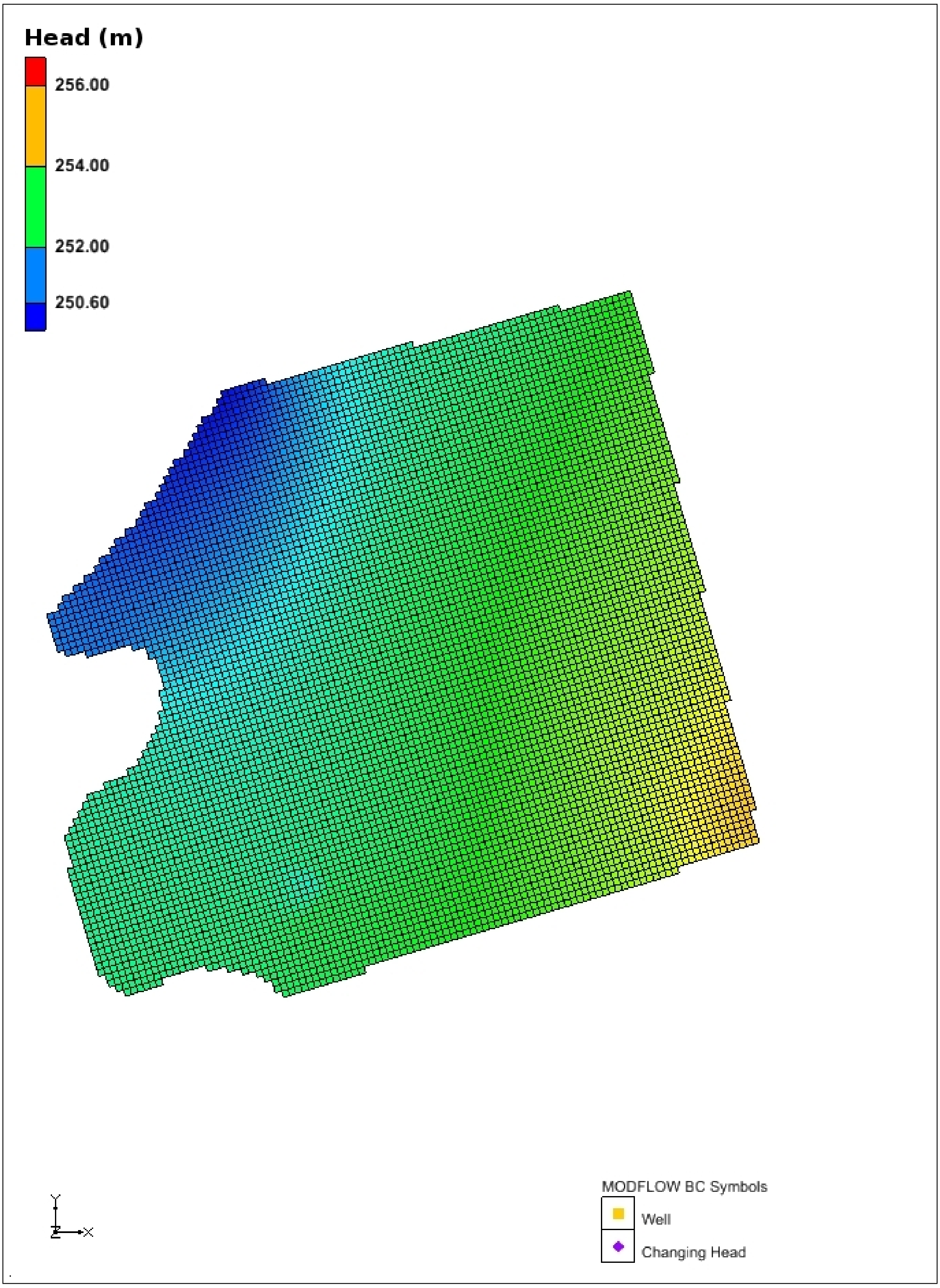


Figure 5.16: Simulated Heads in Layer 3

Table 5.6: Parameters Used for Flow and Transport Model of BTEX Affected Study Area

Parameter	Value
Length of study area (m)	3300
Width of study area (m)	3000
Saturated thickness, b(m)	Variable
Grid spacing in x-direction, Δx (m)	30
Grid spacing in y-direction, Δy (m)	30
Number of layers in z-direction	3
Average horizontal hydraulic conductivity (m/day)	
Layer 1 (tertiary alluvium)	12.37
Layer 2 (quaternary alluvium)	16.24
Layer 3 (bedrock)	0.001
Vertical hydraulic conductivity (m/day)	
all layers	0.2
Effective porosity for all layers, θ	0.28
Longitudinal dispersivity, α_L (m)	11.34
Transverse dispersivity, α_T (m)	1.2
First order decay rate constant (day^{-1})	0.051
Initial pollutant concentration (mg/l)	0.00

The methodology was implemented only to the contaminated area and not to the entire study area modelled. The source of contamination is known to be within the boundaries of the affected area and the monitoring wells are to be confined within the same boundaries as well. The affected area contains a total of 1262 discretized cells distributed in three layers.

5.3.4 Performance Evaluation of Initial Estimation of Unknown Pollutant Source Characteristics

For the purpose of evaluation of this methodology, well no. 37 as shown in Figure 5.8 is considered to be the first detection well, or the well where contamination was first observed. From preliminary studies at the site, it is estimated that the initial source release could not have taken place more than 35 years before detection. Since the monitoring at the detection well lasted

1 year, a time horizon of 36 years is considered. The entire time horizon is divided into 18 stress periods of two years length each. In this case, the actual source release history and the time when this source first became active is unknown. Hence, the performance of the methodology for initial estimation of unknown pollutant source characteristics cannot be verified. However, the estimates can be used to design a monitoring network and effectiveness of the monitoring network design methodology can be verified by comparing source characteristics obtained from using all available data from all the wells and those obtained from using just the data obtained from the designed network.

All the potential source locations are shown in Figure 5.17. Initial source

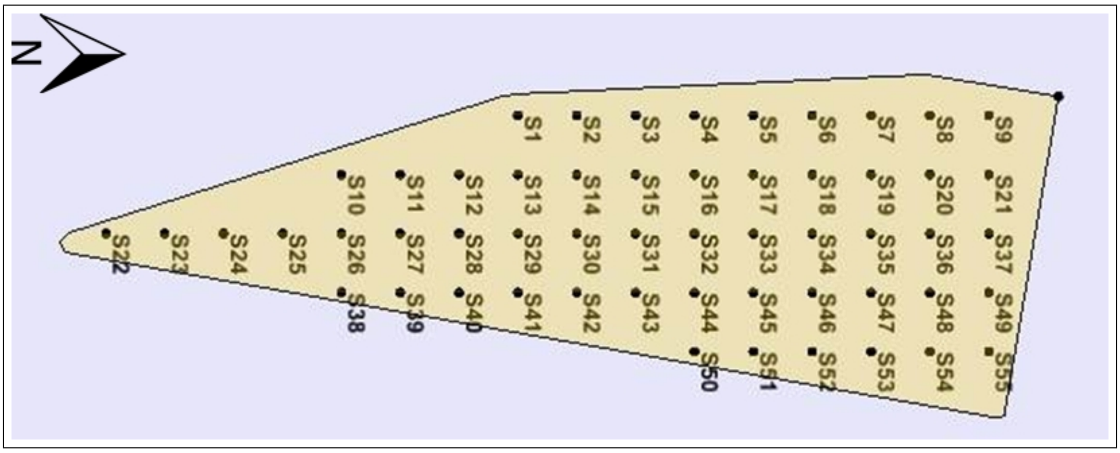


Figure 5.17: Potential Source Locations

characteristics estimation methodology was applied to all potential sources in order to determine their optimal release history and optimal time lag between detection and source activity. Optimal source characteristics, the lowest DTW distance and time lag between first source activity and its detection are presented for the 10 best potential source locations in Table 5.7.

Table 5.7: Initial Estimates of Source Characteristics

PSL	ASP	OF Value	Time Lag
S36	7	1.11E-06	7440
S21	6	1.45E-06	6450
S48	7	3.59E-06	6810
S49	7	4.19E-06	6450
S47	6	2.82E-05	7080
S51	5	0.000219	7440
S53	8	0.002165	6810
S17	5	0.006932	6420
S34	6	0.002481	7130
S35	6	0.012548	6900

5.3.5 Performance Evaluation of Monitoring Network Design

The initial estimates of source location, active stress period and lag time from the previous step are used to design an optimal monitoring network. All existing monitoring locations as shown in Figure 5.8 are considered potential locations. It is assumed that a maximum of 15 monitoring wells can be used in the monitoring network. Optimal monitoring locations are shown in Figure 5.18.

Pollutant concentration measurements recorded for the next two years obtained from the optimal monitoring network are used to estimate the pollutant sources. The same flux estimation algorithm presented in Chapter 3 is implemented. Estimated source characteristics are shown in Table 5.8. The actual location of source for this study area is known and a fairly accurate guess on the time of initiation is available. However, source release history is unknown. For this reason, the optimal source release history estimated using the measurement available from the designed monitoring network (Scenario 1) is compared with the ones obtained using all available monitoring data from all the 74 wells (Scenario 2).

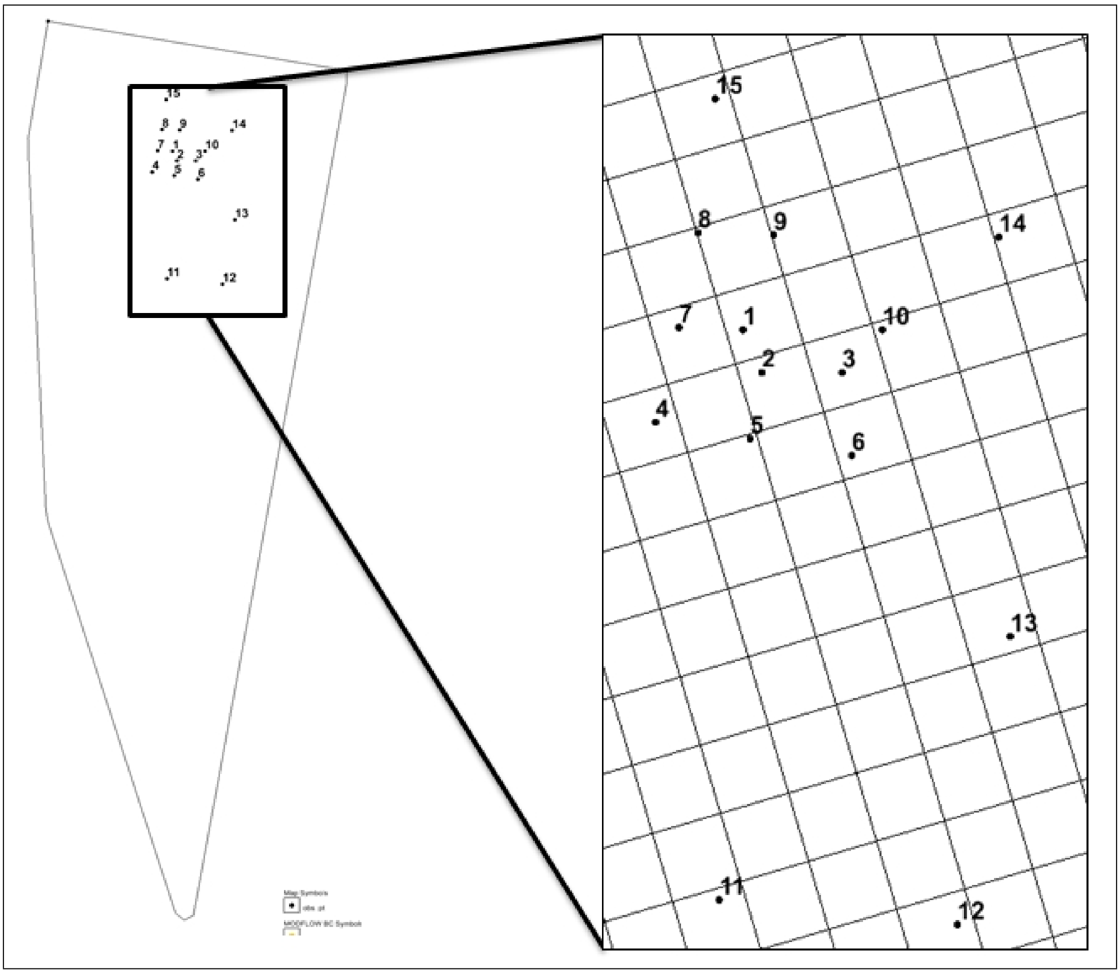


Figure 5.18: Optimal Monitoring Well Locations

It can be noticed in Table 5.8 that source characteristics reconstructed using monitoring data obtained from the monitoring network designed using the developed methodology is very close to those obtained using all the available monitoring data from all wells in the original monitoring network. Although the actual release history for this case study is not exactly known, these solution results demonstrate how monitoring networks can be specifically designed for improved source characterization. This results in the use of only a fraction of resources compared to those that could be used if a designed

Table 5.8: Source Characteristics Obtained using Linked Simulation-Optimization Method

Source Characteristics	Actual	Estimated	
		Scenario 1	Scenario 2
Location			
Row	34	34	34
Column	55	55	55
Layer	1	1	1
Start time (days before detection)	Approx. 7300	6810	6600
Duration of activity (years)		14	12
Release history (g/day)			
SP 1	–	63.31	67.27
SP 2	–	79.20	71.61
SP 3	–	67.91	48.82
SP 4	–	72.86	55.28
SP 5	–	54.32	47.82
SP 6	–	46.28	42.17
SP 7	–	48.86	2.37

dedicated monitoring network is not utilized.

5.4 Conclusion

Performance evaluation of a new approach for initial source characteristics estimation and monitoring network design was presented in this chapter. It was applied to an illustrative aquifer as well as to a real-life unconfined contaminated aquifer located in Upper Macquarie Groundwater Zone in New South Wales, Australia. The real contaminated aquifer was already under management while undertaking this study, and pollutant concentration had been recorded over nearly four years time span at 74 arbitrarily chosen monitoring locations. Using the methodology developed, a subset of most significant monitoring locations was chosen. Source characterization results obtained using information from the designed monitoring network were compared with those obtained using all available data from all 74 wells.

It was evident that the source characterization results compare well when using extensive concentration measurement data from a very large number of monitoring wells, and when using only a fraction of the large number of wells based on a designed monitoring network. This validates the purpose of designing a dedicated monitoring network for efficient pollutant source identification.

Chapter 6

Application of Release History Estimation Methodology to Distributed Sources incorporating Surface-Groundwater Interactions

Unknown sources of groundwater contamination are not always point sources. Distributed or non-point sources can also cause widespread and long-lasting contamination. Some examples of groundwater contamination from distributed sources are:

- Contamination from agricultural chemical use.
- Contamination due to deposits from recharge such as rain, snow, and dry atmospheric fallout.
- Contamination from large scale overland waste dumps such as those in hazardous waste or mining waste disposal sites.

This chapter demonstrates the potential for using the developed methodology for the estimation of release history in aquifers contaminated by distributed sources. The developed methodology is applied to an abandoned mine site that has several sources of pollutant in the form of mining waste dumps, tailings ponds, and a lake formed by flooding of the open pit.

Unsafe storage or disposal of chemical pollutants used in various ore dressing processes and inappropriate management of wastes are among some major reasons for mine site contamination. Such contamination is particularly predominant at abandoned mining sites because of loopholes in historical environmental regulations, and due to slow process of natural remediation. Contamination at abandoned mining sites adversely impacts natural water resources in its vicinity. It has the potential to affect any groundwater resources and render them unusable for consumptive use for prolonged periods. Apart from introducing polluting chemicals into the groundwater ecosystem, mining operations also tend to modify the surface hydrologic features that permanently modify groundwater flow regime at the mine site. Construction of tailings dams and prolonged ponding in the mine pits of open cast mines create new sources and sinks, thereby changing the entire surface and subsurface hydrology as well as water quality at the mine site.

Effective remediation of any contaminated groundwater aquifer is highly dependent on accurate characterization of the pollutant sources. In order to begin the remediation of a contaminated mine site, it is very important to first estimate fluxes of pollutants from various possible sources. Often, locations of potential sources are known and estimates of the starting time of pollutant release are also available in such situations. In this scenario, the source characterization problem can be treated as an unknown pollutant source flux magnitude reconstruction problem.

Often several management measures are already in place at sites affected by contamination. All such measures need to be incorporated into the numerical flow and transport model. If the effects of implemented management measures are not incorporated in the simulation of flow and transport processes, the pollutant release history reconstruction methodology may produce inaccurate results, as the pollutant concentrations measured at observation wells will be lower after the implementation of management measures. Therefore, the source characterization methodology applied to this site also incorporates some management measures, i.e. presence of seepage drains existing in the site.

6.1 Site Description

The abandoned mine site discussed in this chapter is located in the coastal ranges of central Queensland. Over a lifespan of about 100 years, it produced significant amounts of gold, silver and copper. By the time this mine ceased its operations, it was very severely affected by acid rock drainage (ARD). Mining operations exposed sulphate-containing minerals to erosion for a considerable duration resulting in widespread leaching of acidic chemicals from the minerals to the water resources around the study area. Rehabilitation efforts are being undertaken at this study area to minimize the impact of contamination on the water resources in and around it. One management measure implemented as part of the rehabilitation plan is to treat the water impounded in the open cut and arrest subsurface infiltration from the upper reaches to the lower reaches (Unger et al., 2003). However, this limited remediation measure has not been successful due to uncertainties regarding the actual spatially varied sources and the pathways of pollution. Therefore, a preliminary attempt is made to apply the source characterization methodology to this site, to gain insight into its potential applicability.

6.1.1 Topography and Climate

The abandoned mine is located in hilly terrain and the topography at the study area is highly varying with the altitude ranging between 250 and 300 m. Natural topography has been modified to a large extent due to open pit mining and deposition of waste rock dumps, tailings, etc. The detailed topography is shown in Figure 6.1. Shuttle Radar Topography Mission (SRTM) obtained elevation data was used in this study to digitally represent the topography of the entire study area (Jarvis et al., 2008). This elevation database was sampled at 3 arc-seconds, which is 1/1200th of a degree of latitude and longitude, or about 90 metres. The study area is located in the sub-tropical, sub-humid climatic region with an average annual rainfall of 680 mm. The wet period is generally from October to March. Mean minimum temperature is 9.5 °C whereas the mean maximum temperature is 32.1 °C (MLA, 2008).

6.1.2 Hydrology

The mine is located in the Don and Dee River Groundwater Management Unit (GMU) which is drained by the Don River and its major tributaries, the Dee River and Alma Creek. The extent of the entire Don and Dee River groundwater management unit is shown in Figure 6.2. In Figure 6.1, it may be noted that this mine site is very close to Dee River. Contamination from the mine site has affected the water quality in Dee River. Since the alluvial

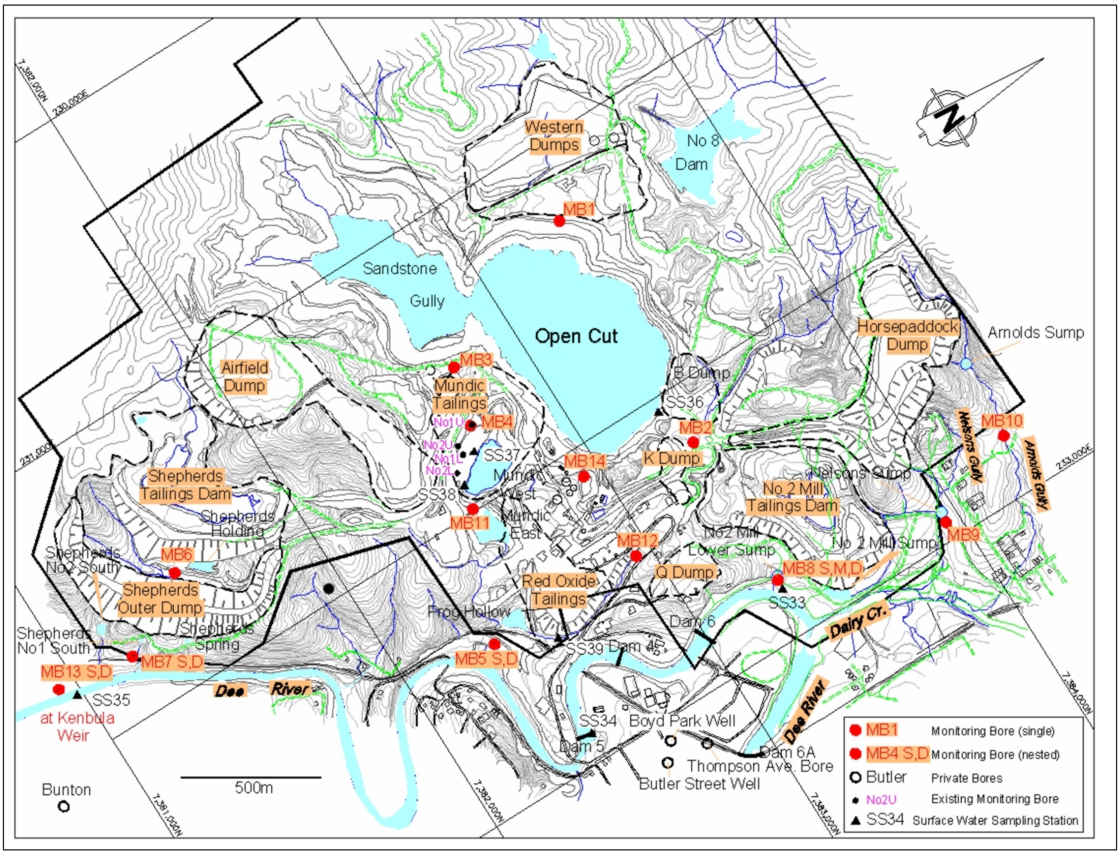


Figure 6.1: Topographical Features of the Study Area. Adapted from: Wels et al. (2006)

bed of Dee River extends to form the Dee River aquifer, this contamination has the potential to impact groundwater quality in the entire groundwater management unit. Natural drainage before the mining activities began has been documented in Jesson and Bamber (1959). These historical drainage paths are important because even after mineral processing wastes have been dumped on natural surface, these historic drainage channels provide a preferential path for subsurface water flow because the material forming these channels has a lower hydraulic conductivity when compared with the waste dumps and tailings. The entire site was divided into four major catchments based on the pre-mining topography. Each of these catchments drains into the Dee River. These are shown in Figure 6.3. Although the topography has been modified by anthropogenic features created during the mineral extraction and processing, the catchment boundaries are still valid.

6.2 Numerical Groundwater Flow Modelling

A conceptual flow model for the mine site was developed using MODFLOW-2005 (Harbaugh, 2005). Field investigation carried out for a previous model by Wels et al. (2006) was taken as the basis for this model in terms of hydrogeological features, boundary conditions and model orientation. The conceptual model discussed here utilizes a different elevation dataset, and the subsurface layers are obtained using available well log data for the site. The flow model discussed here simulates the groundwater flow for dry season assuming a steady-state condition. It was assumed that there is negligible change in groundwater storage and that the groundwater flow is maintained by a constant recharge from the unsaturated zone. It was also

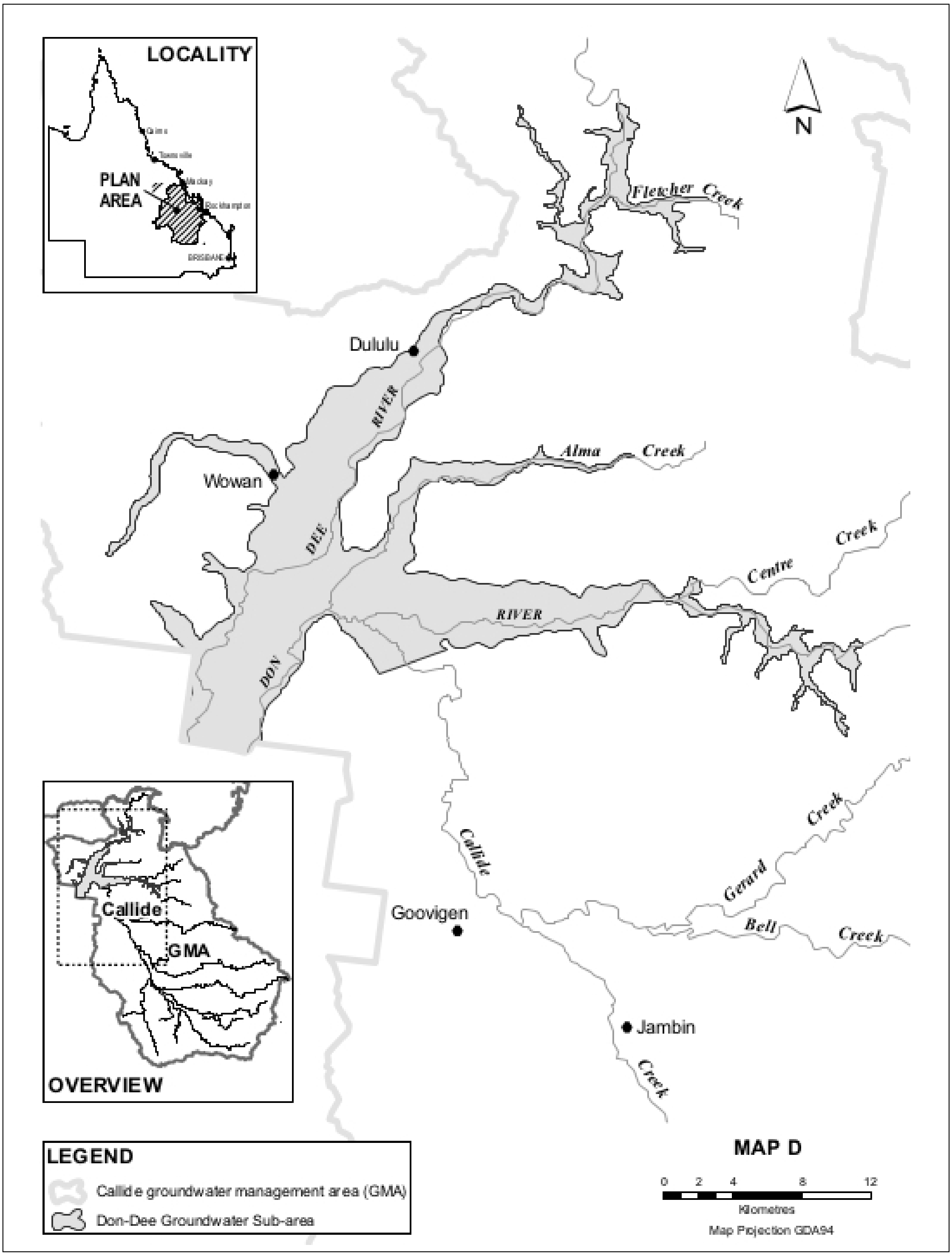


Figure 6.2: The Don and Dee River Groundwater Management Unit Boundaries. Adapted from: Government of Queensland (2011)

Figure 6.3: Historical Catchment Boundaries. Adapted from: Unger et al. (2003)

assumed that no groundwater flow takes place from beyond the geographical boundaries of this model.

6.2.1 Geology and Hydrogeology

Geology of the site has been described in detail in Taube (1986) and is shown in Figure 6.4. A good deal of information on subsurface cross-sections for this study area is available. This information helps in understanding the hydrogeology of this site. In order to get a better understanding of the subsurface, well logs were used to generate hydrogeologic cross-sections. Well log data was obtained from the groundwater database provided by the Department of Environment and Resource Management (DERM), Government of Queensland, Australia. These cross-sections combined with the elevation data were used to generate a three-dimensional representation of the entire site.

The deposit was formed in the late Devonian geologic period from intrusive igneous rocks, mainly tonalite which has almost no permeability. The site investigation carried out by Wels et al. (2006) showed that the groundwater flow at this site occurs mainly in the permeable mine waste dumps and in shallow bedrocks that have been fractured in due course of mining, and have thereby become more permeable. It was also confirmed that historical surface drainage channels create preferable conduits for groundwater flow. This is due to the presence of more permeable deposits from historical erosion leading to washing away of fine permeable material before the commencement of mining as well as in early stages of mining. As the open cut has been flooded, it acts as a major source/sink of water thereby influencing groundwater flow regime in the study area. As a contamination

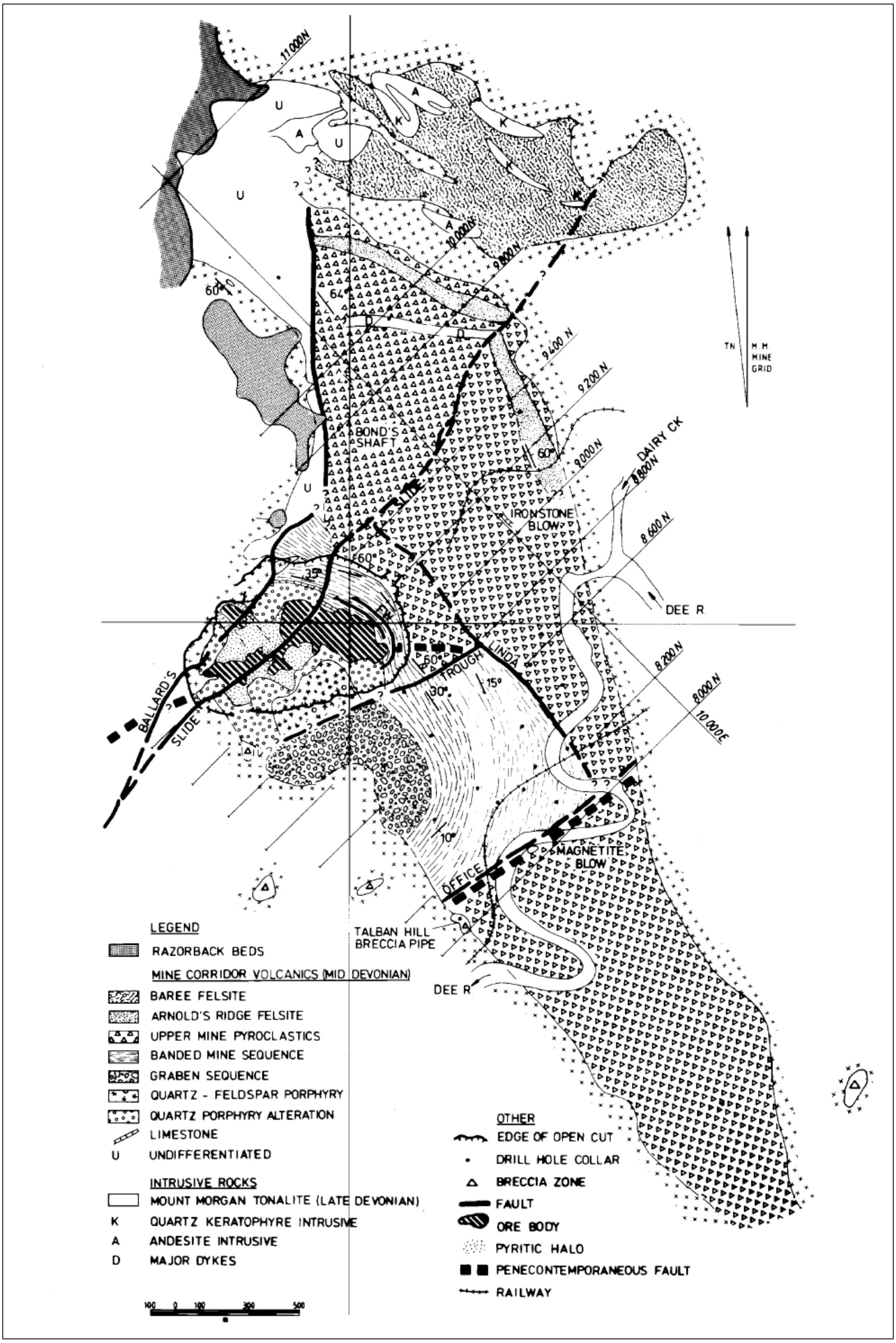


Figure 6.4: Geology of the Mine Site Adapted from: Taube (1986)

management measure, a seepage interception system has been implemented in the study area to prevent the shallow underflow from reaching into the Dee River. Groundwater flow from deeper layers reaches the Dee aquifer that consists of alluvial sediments and fractured igneous rocks.

6.2.2 Model Layers

Mining activities at the site have modified its hydrogeology. Nearly impermeable igneous rocks were removed from the open cut and waste rocks and tailings produced as a result of mineral processing were piled over the natural ground surface. This changed the recharge pattern in the area and also provided a source/sink in the form of a flooded open cut that resulted in the establishment of present groundwater paths.

Based on the hydrogeological information (Taube, 1986), aquifer system at the mine site can be sub-divided into four layers for the purpose of flow modelling. These are listed hereunder:

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

Each of these layers have variable thicknesses that are based on cross-sections generated using the well log data. The LPF package in MODFLOW 2000 was used to specify properties controlling flow between cells. All layers were defined as convertible layers. Cells were allowed to become dry depending upon the calculated hydraulic head.

6.2.3 Hydrogeological Properties

Hydrogeological properties such as the hydraulic conductivity and porosity of the layers was estimated during the field investigation carried out by Wels et al. (2006). They conducted pumping tests at several locations in the study area to obtain representative hydraulic conductivity. In modelling groundwater flow for this site, it was assumed that each model layer has a uniform representative value for hydrogeologic parameters (i.e. hydraulic conductivity and porosity). It was further assumed that the groundwater movement follows Darcy's Law and that there are no fractures or fissures in the porous media that can lead to violation of this assumption. Representative values of hydraulic conductivity were obtained through the calibration process of the flow model, using dry-season heads at several monitoring locations.

6.2.4 Sources, Sinks and Boundary Conditions

Major sources of groundwater at this mine site include the constant recharge from the unsaturated zone. A constant recharge has been assumed for the entire dry-season period that is simulated in this model. Recharge occurs only through the disturbed and loosened portion of the top layer. This is represented as the top layer in this model. The recharge is estimated during calibration process of this model.

The flooded open pit also acts as a significant source/sink of groundwater. This has been represented as a constant head boundary condition in layers 2 and 3 of this model with the constant head equal to standing water level in the lake at the time of model calibration. Thus, the hydraulic interaction between the aquifer and the pit is incorporated into the simulation model.

Several drainage channels were present in the site before mining activities began, although these are now buried by waste dumps. These drains provide preferential pathways of infiltration and have been represented as drains in the MODFLOW flow model. Seepage faces at the site were also represented as drains in the MODFLOW flow model. The Dee River aquifer is also represented as a constant head boundary condition in layer 2 of the flow model with the head equal to standing water level in the Dee River at the time of calibration. Sources, sinks and various other boundary conditions are illustrated in Figures 6.5 to 6.8. The first layer, as shown in Figure 6.5, mainly represents mining waste rock dumps. Preferential flow paths in this model have been represented as drains. In the second layer, the open pit, tailings pond and the Dee River are all represented as constant head boundary conditions. Preferential flow paths in this layer are also represented as drains. This is shown in Figure 6.6. The open pit is also represented as a constant head boundary condition in the third layer and this is shown in Figure 6.7. Figure 6.8 shows the bottommost layer, which is the base rock and has no source or sinks of groundwater.

A summary of model parameters used in the flow and transport model is presented in Table 6.1.

6.2.5 Model Calibration

The developed model was calibrated for the dry-season flow using observed head data obtained at observation wells. Calibration was done

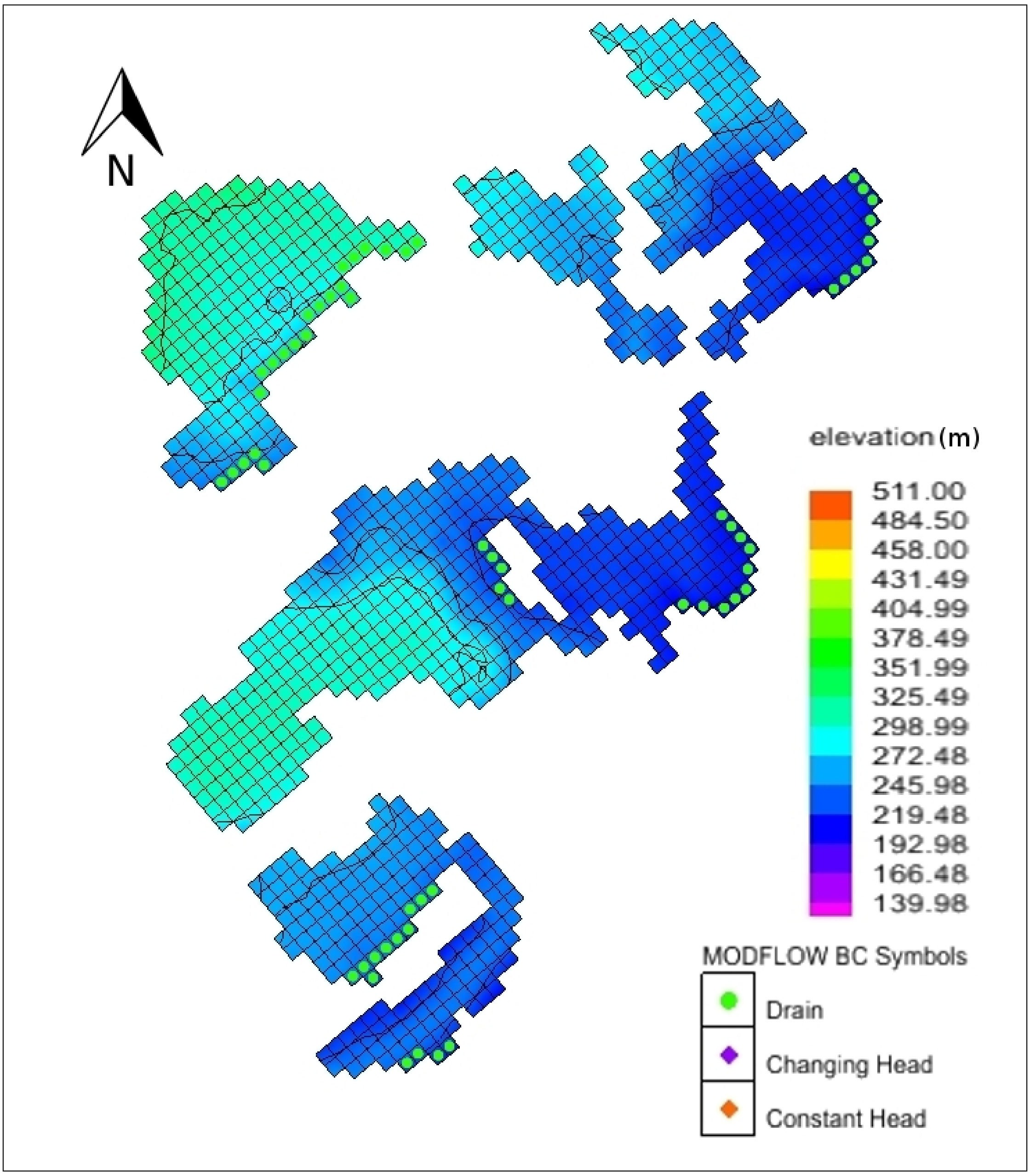


Figure 6.5: Top Elevation Contour Map and MODFLOW Boundary Conditions in Layer 1

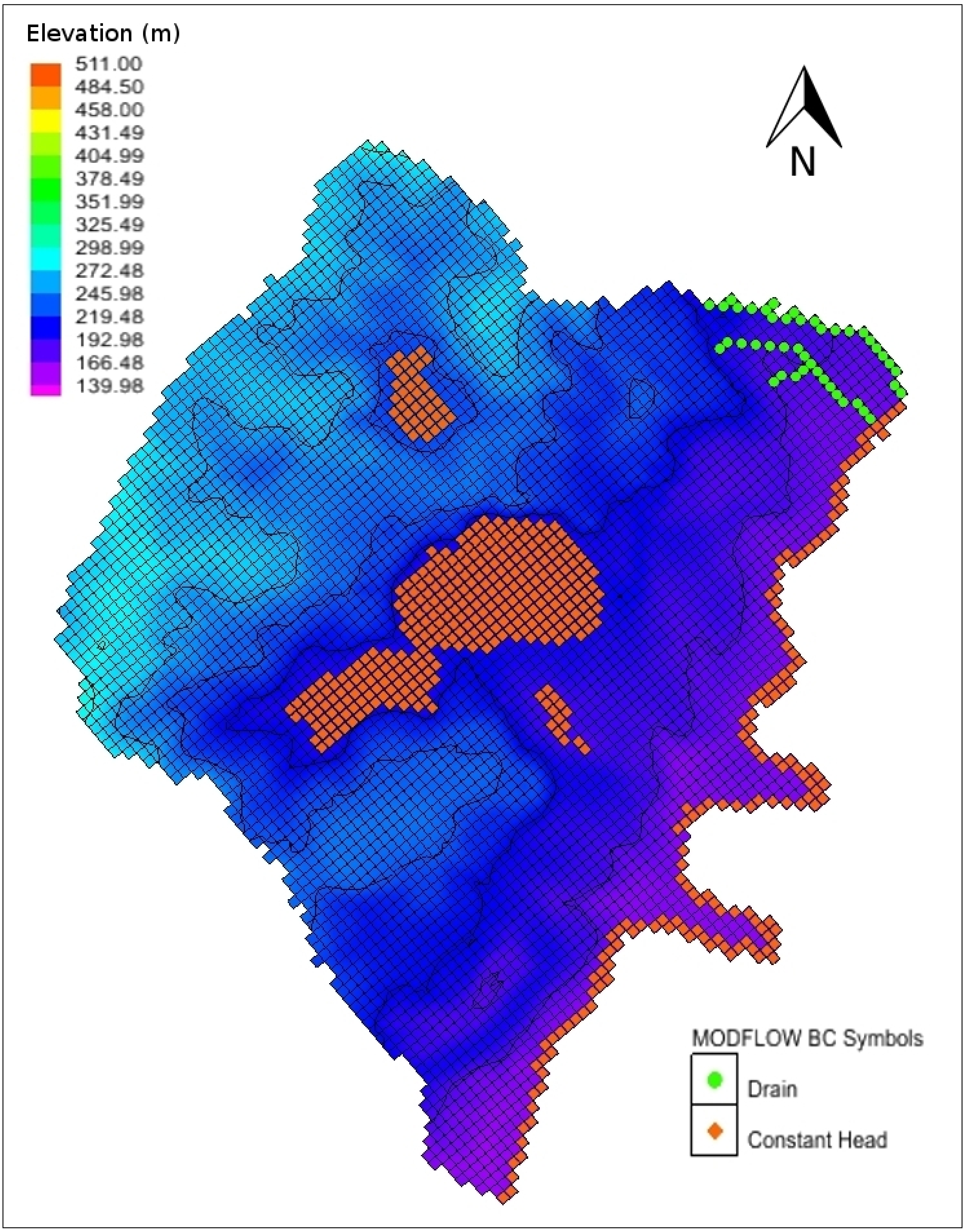


Figure 6.6: Top Elevation Contour Map and MODFLOW Boundary Conditions in Layer 2

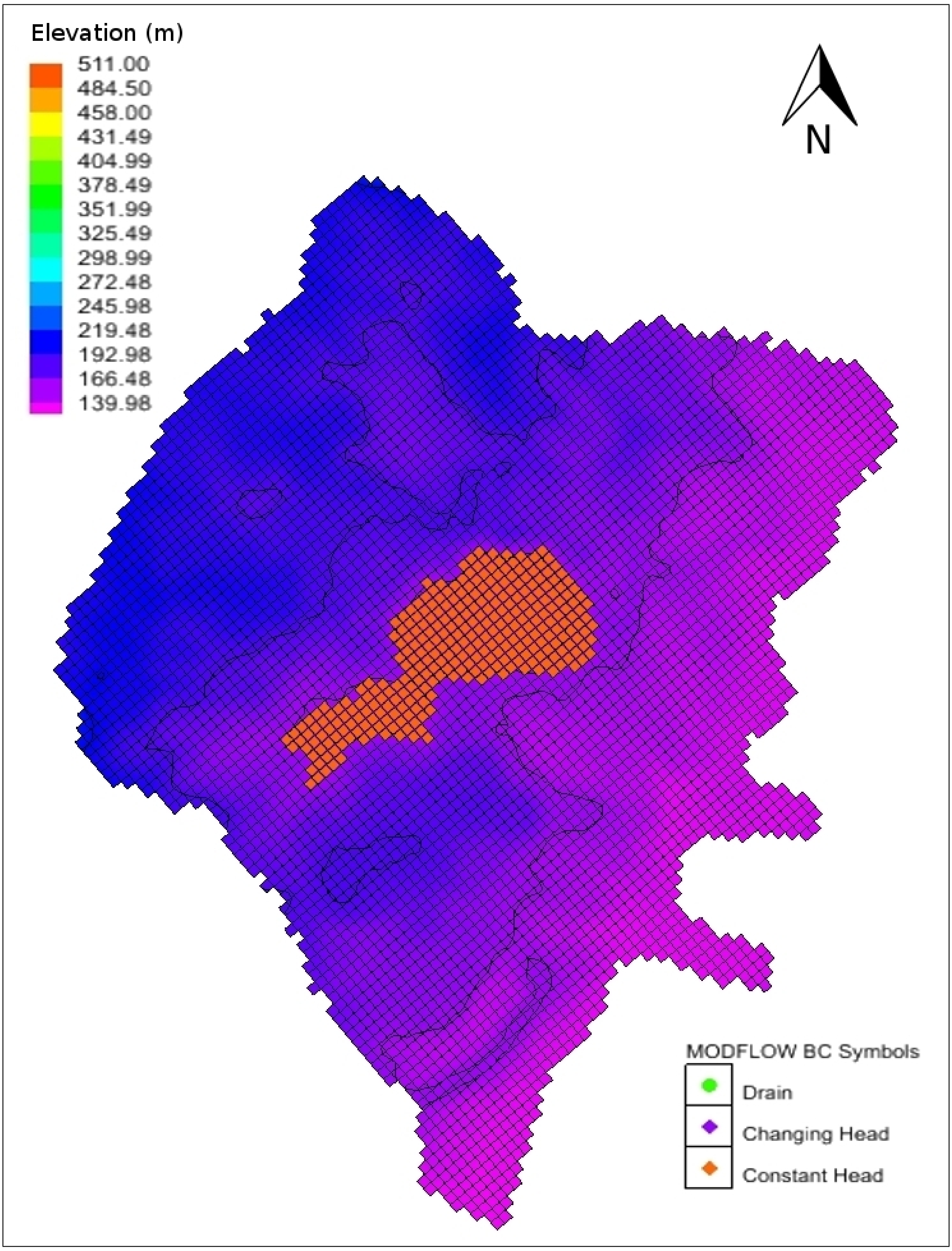


Figure 6.7: Top Elevation Contour Map and MODFLOW Boundary Conditions in Layer 3

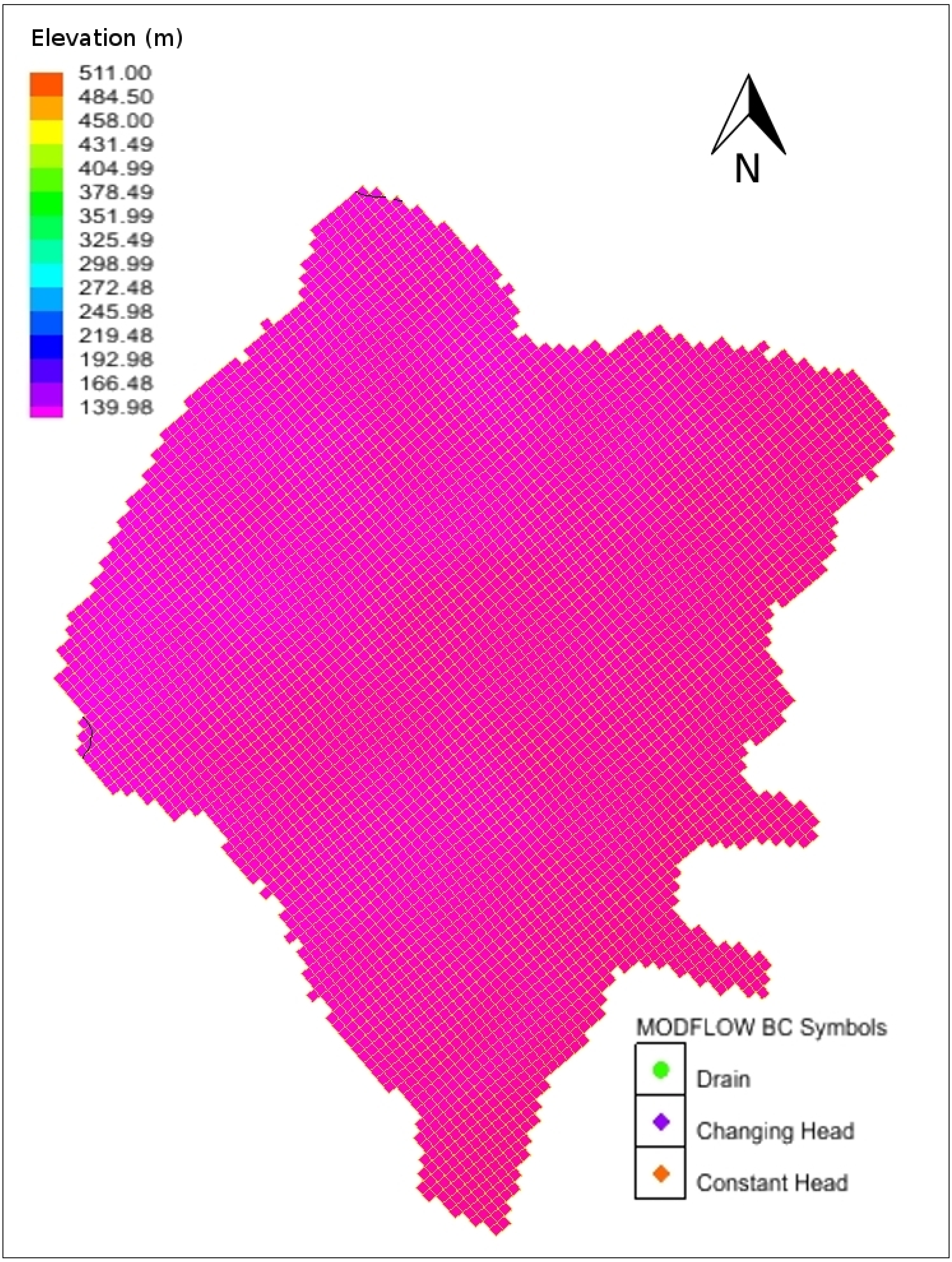


Figure 6.8: Top Elevation Contour Map and MODFLOW Boundary Conditions in Layer 4

Table 6.1: Parameters Used for Flow and Transport Model of the Study Area

Parameter	Value
Length of study area (m)	4000
Width of study area (m)	4000
Saturated thickness, b(m)	Variable
Grid spacing in x-direction, x (m)	40
Grid spacing in y-direction, y (m)	40
Number of layers in z-direction	4
Horizontal hydraulic conductivity (m/day)	
Layer 1	0.98673
Layer 2	0.163555
Layer 3	0.014369
Layer 4	0.009365
Vertical hydraulic conductivity (m/day)	
all layers	0.2
Effective porosity for all layers, θ	0.28
Longitudinal dispersivity, α_L (m)	20
Transverse dispersivity, α_T (m)	4
Initial pollutant concentration (mg/l)	0.00

using the automatic parameter estimation tool available in GMS 7.01 package which is based on PEST (Doherty and Hunt, 2010). In the calibration process, estimates of the values of recharge for each recharge zone and the representative hydraulic conductivity of each layer were calibrated. Results of calibration are shown in Table 6.2 and in Figures 6.9 and 6.10. The calibration target at each observation location was set to an interval of 10 m with a confidence level of 85%. Calibrated values of head at each monitoring location with respect to the targets are shown in Figure 6.10. Box plots in Figure 6.10 are colour-coded to show if the calibration target was met. Green boxes show that the calibration target has been met at that monitoring location, yellow boxes show head measurements slightly off calibration target, whereas the red ones show head measurements that are completely off calibration target. Groundwater heads estimated during the calibration process have

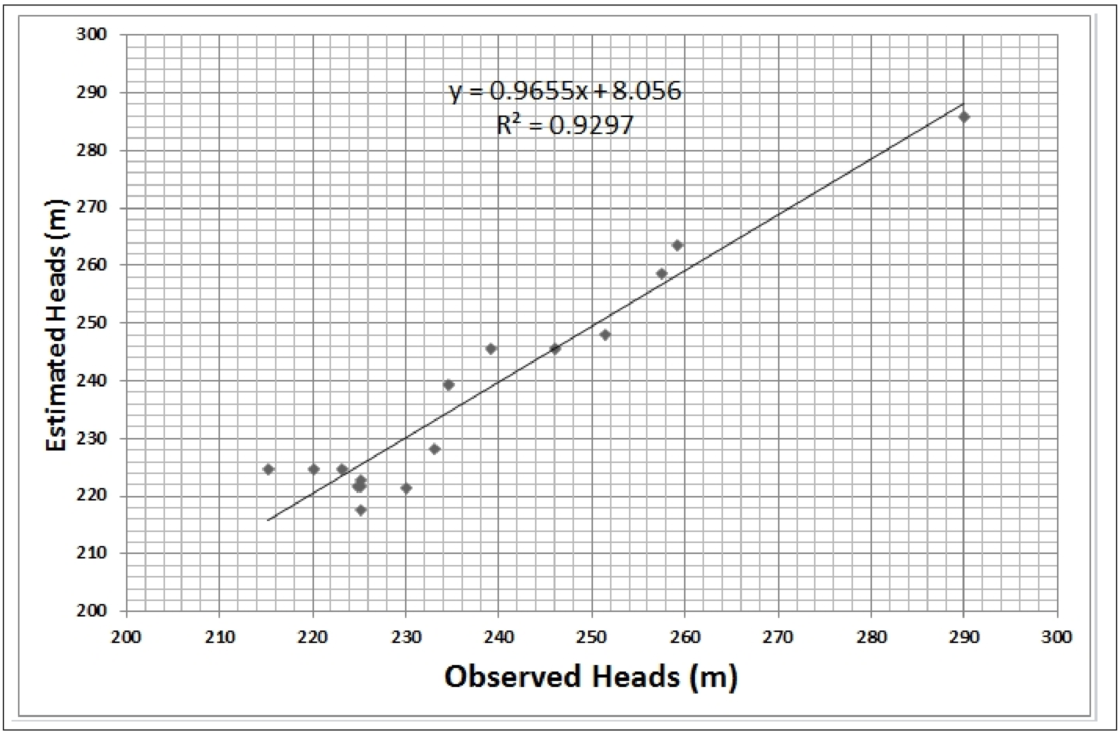


Figure 6.9: Estimated vs Observed Heads after Calibration

been plotted against the values measured at each monitoring location in Figure 6.9. It shows a linear correlation between observed and estimated groundwater head values with a correlation co-efficient of 0.93. Since only a limited amount of data was available to carry out the calibration, the calibrated model may not be very accurate. However, the aim of this study is to demonstrate the potential applicability of release history reconstruction methodology to distributed sources. Therefore, very accurate calibration of groundwater flow model may not be a necessary prerequisite to demonstrate the potential applicability.

The study area was divided into seven different recharge zones and the recharge rate for each was estimated in the calibration process. These estimated recharge rates and all the recharge zones are shown in Figure

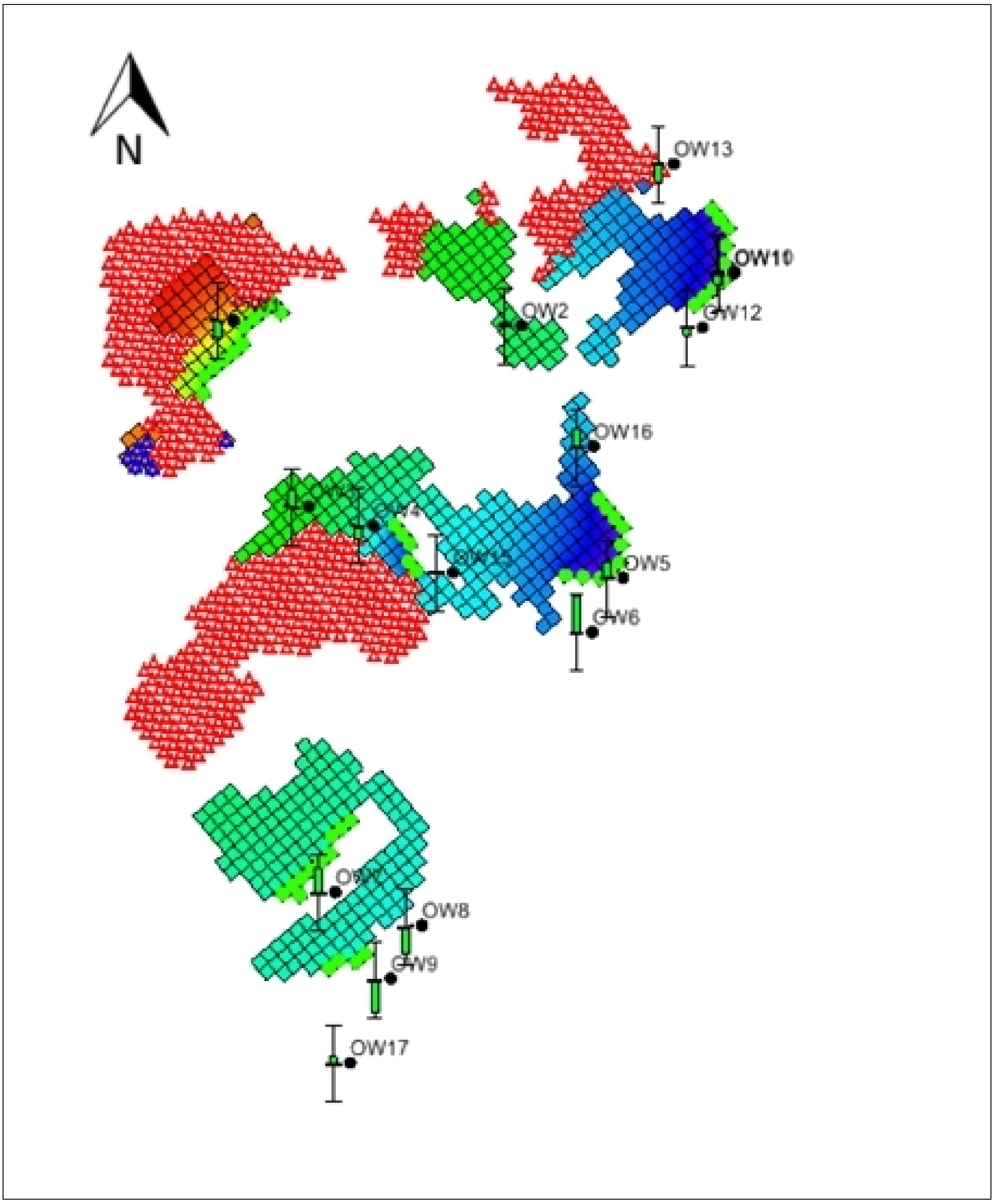


Figure 6.10: Calibrated Groundwater Model of the Study Area

Table 6.2: Observed and Estimated Values of Hydraulic Head at Various Monitoring Locations

Obs. Well	Obs. Well Location (m)			Observed Head	Observation Interval	Observation Confidence	Obs. Head Std. Dev.	Computed Head	Deviation
Name	X	Y	Z	(m)	(m)	(%)	(m)	(m)	(m)
OW1	231386.5	7383547	254.66	290	10	85	6.946705	285.8984	-4.1016
OW2	232315.9	7383531	257.46	257.46	10	85	6.946705	258.7181	1.2581
OW3	231629.1	7382948	259.11	259.11	10	85	6.946705	263.7027	4.5927
OW4	231838.7	7382886	251.342	251.342	10	85	6.946705	248.1169	-3.2251
OW5	232643.7	7382718	178.12	220	10	85	6.946705	224.5807	4.5807
OW6	232544.3	7382543	181.23	215.22	10	85	6.946705	225.0046	9.7846
OW7	231715.4	7381705	199.1	239.1	10	85	6.946705	245.7403	6.6403
OW8	231994.9	7381598	180.3	225	10	85	6.946705	217.9475	-7.0525
OW9	231893.6	7381426	186.22	230	10	85	6.946705	221.5887	-8.4113
OW10	233002.4	7383704	178.22	225.05	10	85	6.079568	221.8756	-3.1744
OW11	233001.9	7383699	185.43	224.8	10	85	6.946705	221.8963	-2.9037
OW12	232900.7	7383525	176.61	225.05	10	85	6.946705	223.0506	-1.9994
OW13	232807.1	7384053	179.13	233	10	85	6.946705	228.3411	-4.6589
OW15	232095	7382736	211.4	246.01	10	85	6.946705	245.7576	-0.2524
OW16	232548.6	7383142	201.2	234.49	10	85	6.946705	239.3924	4.9024
OW17	231762.9	7381154	184.32	223	10	85	6.946705	224.8584	1.8584

6.11. In this calibration process, distributed spatial recharges from each of the recharge zones were considered as calibration parameters. Groundwater potentiometric heads are calculated using the model parameters that are also mentioned in Table 6.1. Simulated heads from the calibrated model are plotted in Figures 6.12 to 6.15. It can be inferred from these figures that the general direction of groundwater flow is from north-west of the study area to the south-east towards the Dee River.

6.3 Transport Model

In order to demonstrate the applicability of the developed methodology for characterization of unknown distributed sources, an illustrative groundwater transport model was developed to simulate the transport of a conservative pollutant, present in the waste dumps as well as in the flooded open pit. A conservative pollutant was chosen for illustration purposes as it was not possible to incorporate some of the complex geochemical processes at this stage due to limited availability of field data.

Actual pollutant transport at the mine site involves a number of reactive chemical species and modelling the associated transport process requires a great deal of analytical and general information on the speciation and aquatic chemistry taking place in the groundwater. Due to lack of accessibility to the site and lack of resources for analytical studies on the aquatic chemistry of contaminated groundwater, it was not possible to simulate the actual pollutant transport processes occurring in this site. For performance evaluation of the methodology, it was therefore decided to use an illustrative contamination scenario, which can be verified for performance evaluation of

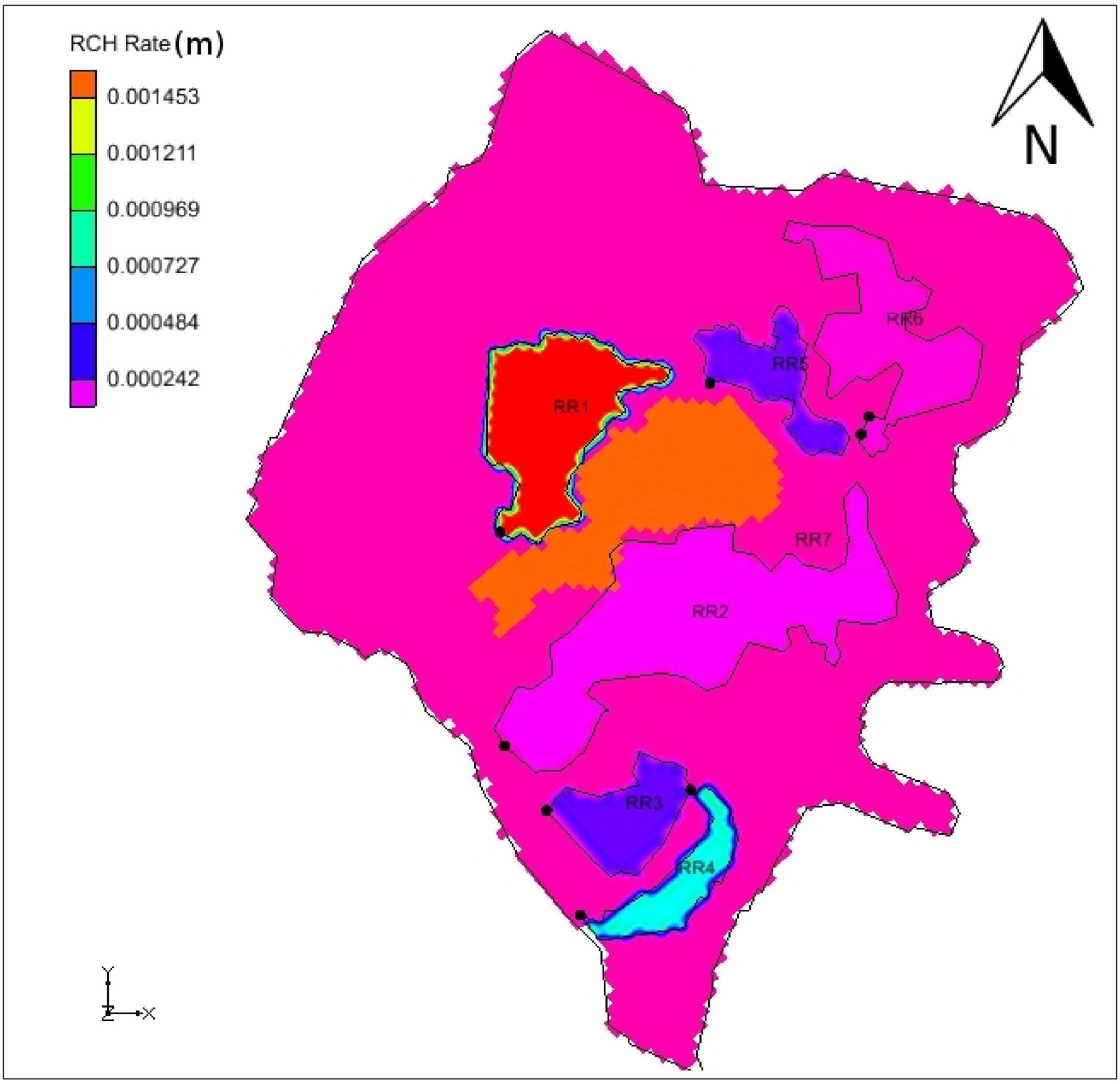


Figure 6.11: Recharge Rates for Various Recharge Zones in the Study Area

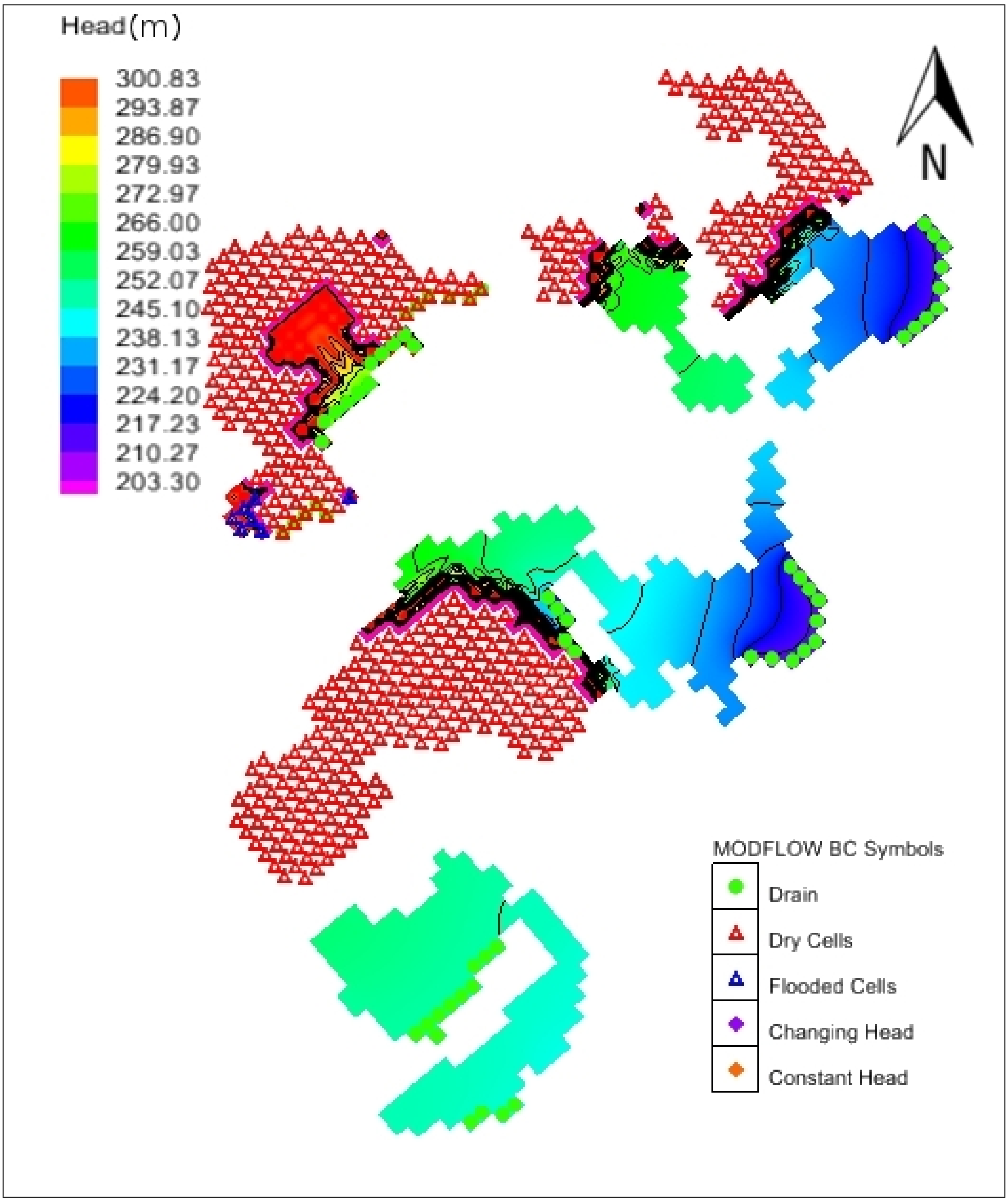


Figure 6.12: Simulated Heads in Layer 1

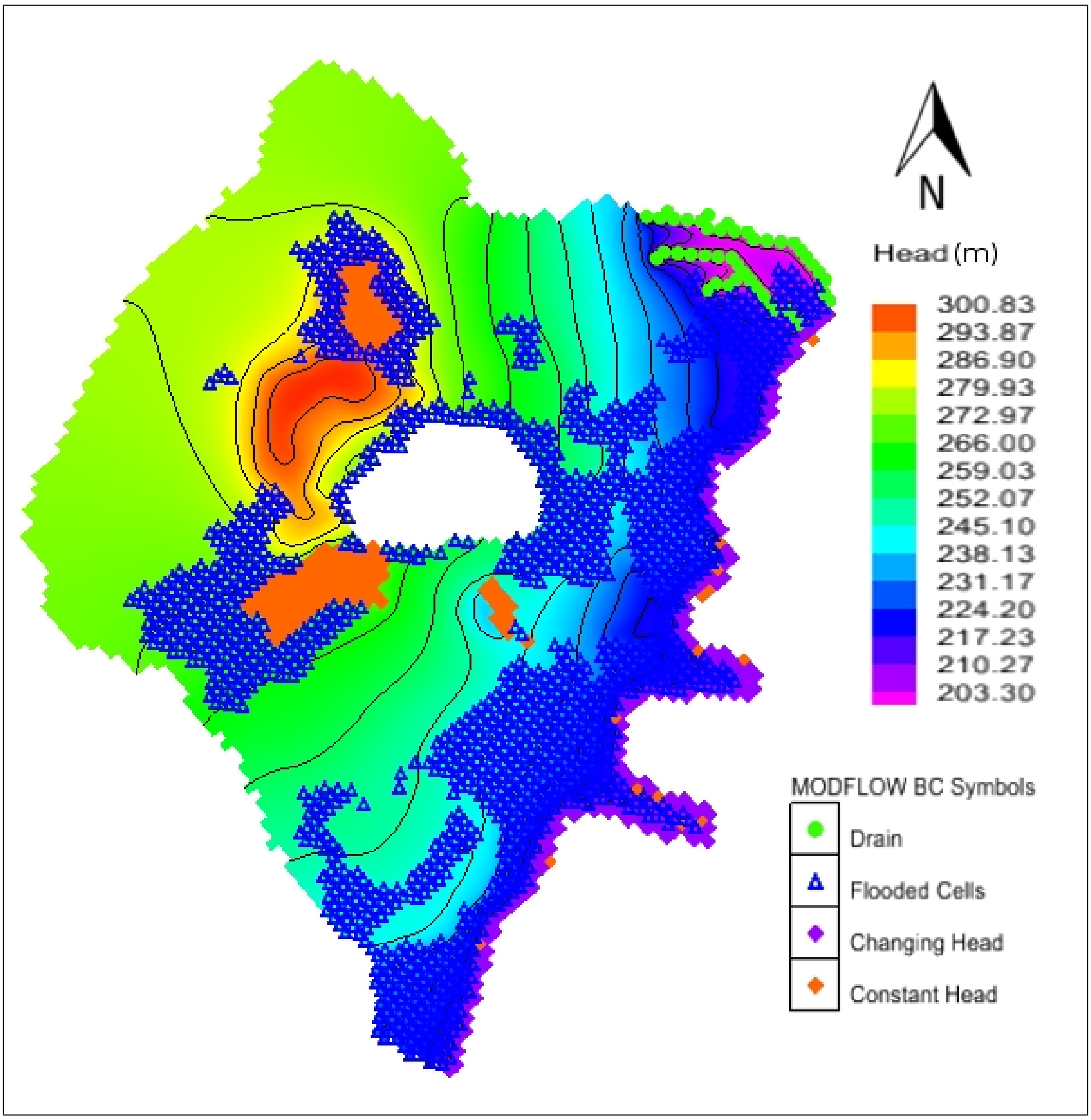


Figure 6.13: Simulated Heads in Layer 2

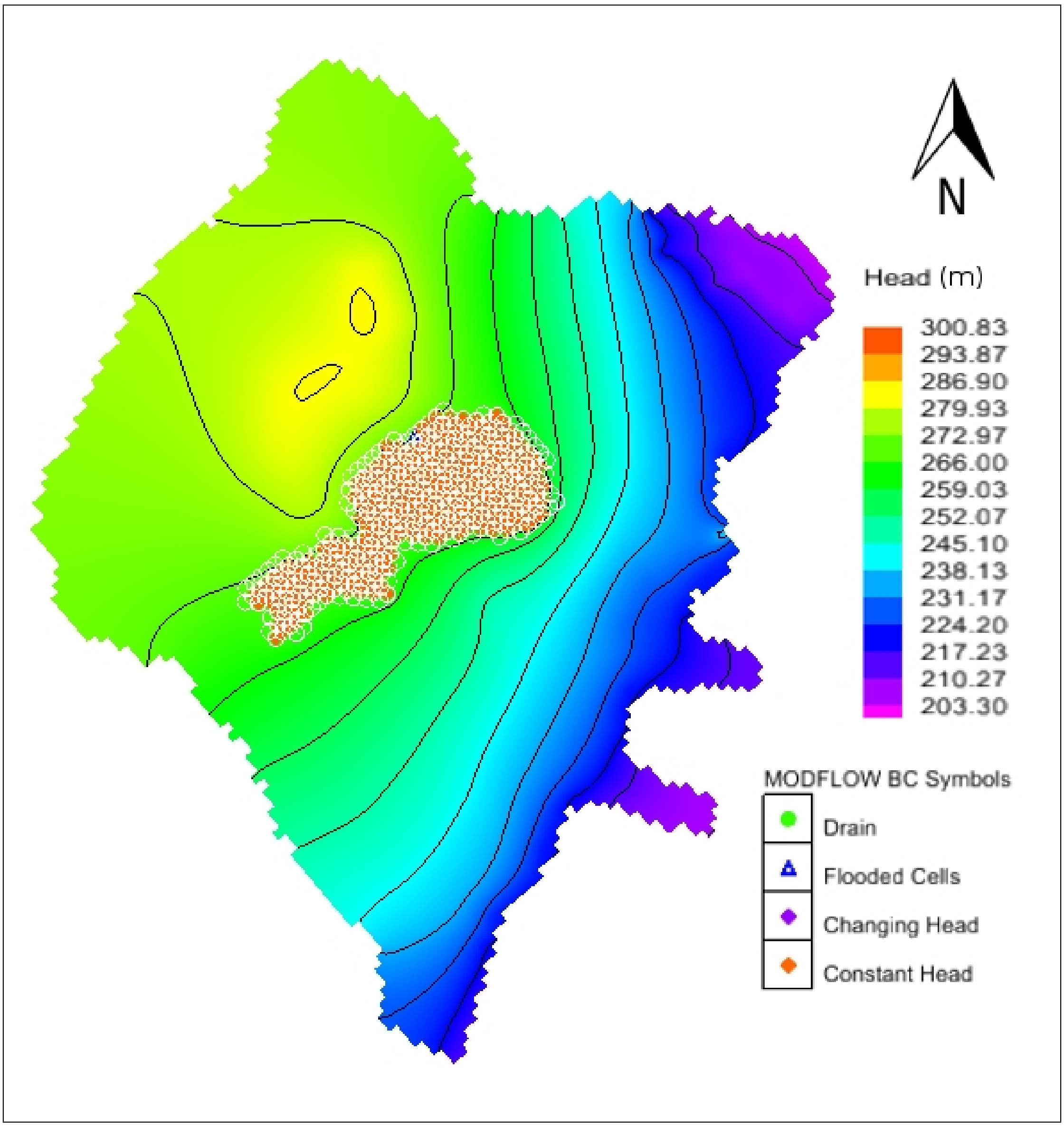


Figure 6.14: Simulated Heads in Layer 3

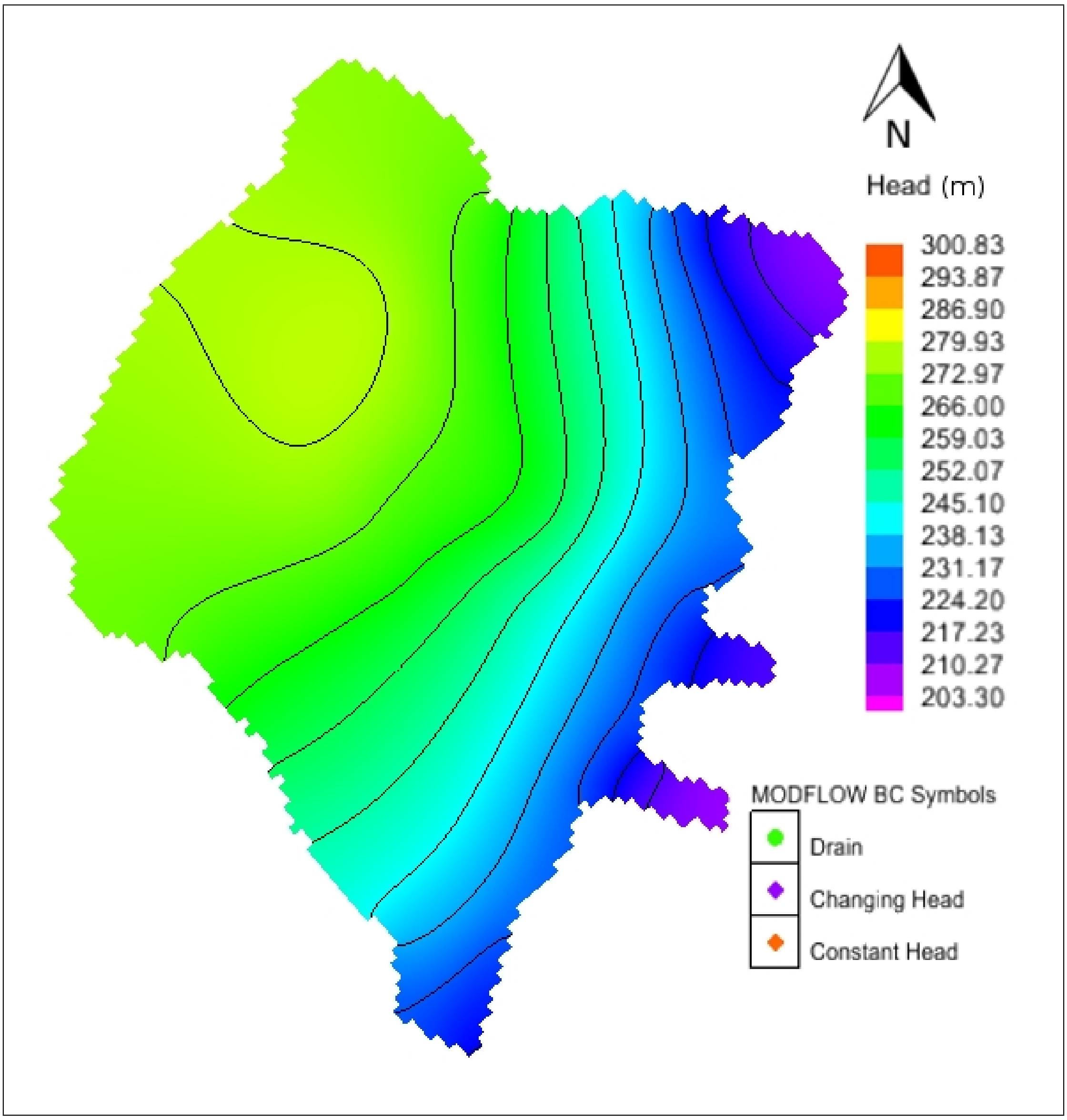


Figure 6.15: Simulated Heads in Layer 4

the methodology in general.

Observed pollutant concentrations were simulated using transport parameters as mentioned in Table 6.1. The transport model considered a time span of 15 years, which was divided into five stress periods of three years each. It was assumed that the sources are active only in the first three stress periods. Contaminant concentration measurements in the aquifer are assumed to be available only for the last stress period at an interval of 60 days. Sources of pollutant are assumed to be the flooded open pit and the recharge. Actual source concentrations are shown in Table 6.3.

Table 6.3: Actual Source Concentrations

Sources	Contaminant Concentration (mg/l)		
Open Pit	Stress Period 1	3 years	4.965
	Stress Period 2	3 years	5.528
	Stress Period 3	3 years	4.863
	Stress Period 4	3 years	0.000
	Stress Period 5	3 years	0.000
Recharge	Stress Period 1	3 years	0.0088
	Stress Period 2	3 years	0.0092
	Stress Period 3	3 years	0.011
	Stress Period 4	3 years	0.000
	Stress Period 5	3 years	0.000

In this performance evaluation, the flow regime is taken as steady state. Therefore, linkage of the source identification model using the flow simulation model was not necessary. The aquifer head distribution as obtained through calibration was specified as input for source identification purposes. The decision variable in the source identification process was the concentration of pollutant in recharge entering the groundwater system from the pit, which was considered as a constant head boundary, and the concentration of the pollutant in recharge coming from the mining waste dumps and vertical distributed recharge entering through the top layer.

The vertical recharge in this illustrative problem was assumed uniformly distributed spatially. However, it is possible to separately delineate these sources, i.e. different wastes can be considered separately. The concentration values shown in Table 6.3 are the actual concentrations specified for the pit and vertical recharge. These specified concentration values were utilized for generating the concentration measurement values at designated observation locations for this performance evaluation purpose.

6.4 Performance Evaluation of Release History Reconstruction Methodology

Potential applicability of the developed methodology was evaluated based on the ability of the developed methodology to accurately estimate the actual source concentrations as per the discussed scenario even with moderate levels of concentration measurement errors. Erroneous observation data are generated with the formulation described in Equation 3.7. In order to test the methodology for different levels of measurement errors, three different values of fraction 'a' (0.05, 0.1 and 0.15) are used. For each value of fraction 'a', five randomized realizations of the observed pollutant concentration data are generated. These erroneous observations are used to reconstruct the release histories of pollutant sources. Average values of objective functions and estimated source concentrations are shown in Figure 6.16.

The source identification solution given in Figure 6.16 shows that the estimated unknown concentration values matched fairly well with the actual specified concentrations. This inference is true even when concentration measurement errors are incorporated. It can be inferred

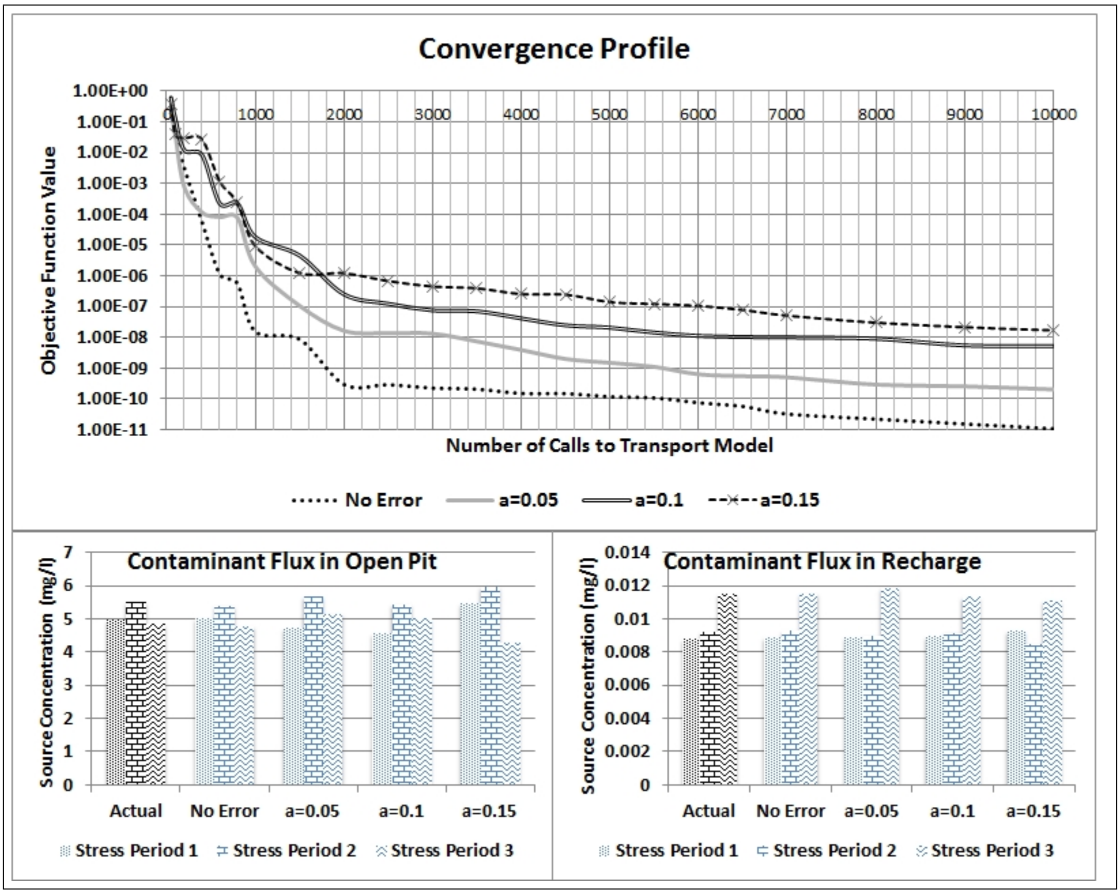


Figure 6.16: Estimated Source Concentrations and Convergence Profile for Various Error Levels

from Figure 6.16 that the optimization algorithm converges even with substantial levels of error in measurement data. It can be also noted from the estimated source concentrations that the pollutant concentration values for recharge are less sensitive to errors in pollutant concentration measurement. This is mainly because recharge has a much larger areal extent when compared with that of the open pit lake. Even for the open pit lake the values of source concentration do not fluctuate unreasonably with increasing errors in measured concentrations. This shows that the developed methodology is potentially applicable for characterization of a wide range of distributed pollutant sources by incorporating surface-groundwater interactions. Groundwater pollution from unknown pollutant sources at abandoned mine sites is a widespread problem in Australia, including North Queensland. This methodology can be applied to enhance the probability of reliable source characterization at these sites.

6.5 Conclusion

The developed methodology for unknown groundwater source flux characterization was utilized for a distributed sources of contamination scenario, incorporating the hydraulic interaction between the surface and subsurface flow system. The potential applicability of the methodology was shown by using an illustrative example of an abandoned mine site. While the study area represented topographic and geologic characteristics of the abandoned mine site, the contamination scenario was assumed to involve only a conservative pollutant. Solution results show that the methodology is potentially applicable to distributed pollutant sources as well. The next chapter summarises major conclusions derived in this thesis.

Chapter 7

Conclusions

This chapter summarises major conclusions derived in this thesis. It also highlights some of the limitations of the methodologies developed. The objective of this study was to develop a robust methodology for unknown groundwater pollutant source identification using linked simulation-optimization approach in scenarios where little prior information exists on the various characteristics of the pollutant sources. Also, the possibility of the contaminated aquifer being under management was incorporated. This study was aimed at addressing some of the major limitations in the present state-of-the-art in the use of linked simulation optimization approach for solving unknown groundwater pollutant sources. These include:

1. Sparsity of concentration measurement data.
2. Inefficient monitoring network for concentration measurements.
3. Difficulty in establishing the time of pollutant source activity initiation.
4. Applicability of optimal source characterization to distributed sources.
5. Problems associated with achieving a global optimal solution efficiently.

An attempt was made to address the problems associated with computational efficiency in achieving global optimal solution by using adaptive simulated annealing (ASA) as the optimization algorithm. It was shown that when ASA is adopted as the optimization algorithm, in general, computational efficiency in achieving global optimal solution is expected. ASA can also handle uncertainty in hydrogeologic parameters and errors in concentration measurement more efficiently as compared to genetic algorithm.

Exact characterisation of pollutant sources becomes very difficult in groundwater aquifer contamination scenarios where no prior information exists on any of the unknown pollutant source characteristics, and the pollutant concentration data is available only from one, or a very few, initial detection wells. In such situations, it is particularly important to have a reliable estimate of the time of first activity of the source. This is so because the solution to the problem of unknown pollutant source identification is non-unique. Because time and space are both variables, multiple combinations of source location and its time of first activity can result in the same impacts on pollutant concentrations at the location being monitored. This non-uniqueness is particularly prominent when monitoring information is very limited.

The problem of finding the time of initial activity of sources has been addressed in this study by using principles of time series matching. The observed concentration time series is treated as a test sequence that is matched to the entire estimated concentration breakthrough curve for that monitoring location.

Often, observed information is sparse. In practice, it is very common to either miss or delay one or more concentration observations at a given monitoring location. Existing state-of-the-art methodology incorporates the use of Euclidean distance as a measure of similarity between the observed and the estimated time sequences. This approach is not very efficient in addressing the problem of misaligned data series of missing/sparse data. In this study, it is proposed to use dynamic time warping distance as a similarity measure to address these issues.

Heterogeneity in hydrogeological parameters such as hydraulic conductivity and porosity has been considered for all illustrative as well as real-life examples. In the illustrative examples, layered heterogeneity was considered. It was assumed that the hydraulic conductivity and porosity are uniform in each individual geologic layer while the layers represented vertical heterogeneity. In real-life examples, different spatial zones of the hydrogeological parameters were considered in each layer based on the parameter values measured at a number of locations in the study area.

Applicability of the developed methodologies has been demonstrated by the use of illustrative or real-life contamination scenarios, and study areas. The developed methodology for release history reconstruction has also been extended to the characterization of distributed pollutant sources.

Results of performance evaluation of each of the methodologies indicate their potential for field application. However, there are some limitations to the methodologies developed in this study and these limitations need to be addressed in future studies. The extension of the methodology to include distributed pollutant sources and incorporating surface-groundwater interactions should be useful in applying the source characterization methodology to abandoned mine sites. Widespread pollution of subsurface and groundwater due to pollutants originating from unknown sources in such mine sites is a widespread problem in many parts of Australia, including North Queensland. This study enhances the probability of reliable source characterization in these critical sites. Some of the major limitations are:

1. The methodology developed incorporating DTW for preliminary estimation of actual pollutant sources, although capable of taking into account multiple potential sources, the inference is limited to one actual source.

2. The methodologies developed in this study are sensitive to uncertainties in hydrogeological parameters. The spatial random nature of the hydrogeologic parameters needs to be incorporated more rigorously. Although, such heterogeneities were incorporated in simulating the field conditions for performance evaluation purposes.

3. This study assumes that groundwater flow follows Darcy's Law. Fractures or cracks in the subsurface have not been incorporated. In some of the mine sites, fissures and fractures may be present. This aspect needs further consideration.

4. Some of the performance evaluations are based on the assumption that the calibrated model represents actual field conditions as closely as possible.

5. The developed methodologies are computationally intensive. It is possible to further explore computational efficiency.

References

Alapati, S. and Kabala, Z. J. Recovering the release history of a groundwater contaminant using a non-linear least-squares method. *Hydrological Processes*, 14(6):1003–1016, 2000. 12

Anderson, M. P. and Woessner, W. W. *Applied groundwater modeling: simulation of flow and advective transport.* Academic Press, San Diego, CA, 1992. 25, 42

Andricevic, R. and Kitanidis, P. K. Optimization of the pumping schedule in aquifer remediation under uncertainty. *Water Resources Research*, 26(5): 875–885, 1990. 27

Aral, M. M.; Guan, J., and Maslia, M. L. Identification of contaminant source location and release history in aquifers. *Journal of Hydrologic Engineering*, 6 (3):225–234, 2001. 14, 34

Atmadja, J. and Bagtzoglou, A. C. Pollution source identification in heterogeneous porous media. *Water Resources Research*, 37(8):2113–2125, 2001a. 12

Atmadja, J. and Bagtzoglou, A. C. State of the art report on mathematical methods for groundwater pollution source identification. *Environmental Forensics*, 2(3):205–214, 2001b. 6, 11

Bagtzoglou, A. and Atmadja, J. Mathematical methods for hydrologic inversion: The case of pollution source identification. In Kassim, Tarek, editor, *Water Pollution*, volume 3 of *The Handbook of Environmental Chemistry*, pages 65–96. Springer, Berlin, 2005. URL http://dx.doi.org/10.1007/b11442. 10.1007/b11442. 11, 12

Bagtzoglou, A. C.; Tompson, A. F. B., and Dougherty, D. E. Probabilistic simulation for reliable solute source identification in heterogeneous porous media. In Ganoulis, J., editor, *Water Resources Engineering Risk Assessment*, pages 189–201. Springer-Verlag, Heidelberg, 1991. 36

Bagtzoglou, A. C.; Dougherty, D. E., and Tompson, A. F. B. Application of particle methods to reliable identification of groundwater pollution sources. *Water Resources Management*, 6:15–23, 1992. ISSN 0920-4741. URL http://dx.doi.org/10.1007/BF00872184. 10.1007/BF00872184. 36

Bear, J. *Dynamics of fluids in porous media*. Dover Publications, 1988. 40

Bellman, R. and Kalaba, R. On adaptive control processes. *Automatic Control, IRE Transactions on*, 4(2):1–9, 1959. 31

Chadalavada, S. and Datta, B. Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers. *Water Resources Management*, 22(6):651–670, 2008. 22

Chadalavada, S.; Datta, B., and Naidu, R. Uncertainty based optimal monitoring network design for a chlorinated hydrocarbon contaminated site. *Environmental Monitoring and Assessment*, 173:929–940, 2011a. ISSN 0167-6369. URL http://dx.doi.org/10.1007/s10661-010-1435-2. 10.1007/s10661-010-1435-2. 6, 24, 36

Chadalavada, S.; Datta, B., and Naidu, R. Optimisation approach for pollution source identification in groundwater: An overview. *International Journal of Environment and Waste Management*, 8(1):40–61, 2011b. 6

Cieniawski, S. E.; Eheart, J. W., and Ranjithan, S. Using genetic algorithms to solve a multiobjective groundwater monitoring problem. *Water Resources Research*, 31(2):399–409, 1995. 20

Cressie, N. Spatial prediction and ordinary kriging. *Mathematical Geology*, 20:405–421, 1988. ISSN 0882-8121. URL `http://dx.doi.org/10.1007/BF00892986`. 52

Dagan, G. Stochastic modeling of groundwater flow by unconditional and. *Water Resources Research*, 18(4):835–848, 1982. 26

Datta, B. 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, 2002. 50

Datta, B. and Dhiman, S. D. Chance-constrained optimal monitoring network design for pollutants in ground water. *Journal of Water Resources Planning and Management*, 122:180, 1996. 20

Datta, B. and Purwar, D. K. Optimal design of groundwater quality monitoring network incorporating uncertainties. In *Proc., National Symp. on Environment*, pages 129–131. Bhabha Atomic Research Center, Bombay, India, 1992. 90

Datta, B.; Beegle, J. E.; Kavvas, M. L., and Orlob, G.T. Development of an expert system embedding pattern recognition techniques for pollution source identification. Technical report, National Technical Information Service, Springfield, VA, U.S.A., 1989. 30

Datta, B.; Chakrabarty, D., and Dhar, A. Optimal dynamic monitoring network design and identification of unknown groundwater pollution sources. *Water Resources Management*, 23(10):2031–2049, 2009a. 34, 63

Datta, B.; Chakrabarty, D., and Dhar, A. Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters. *Journal of Hydrology*, 376(1-2):48–57, 2009b. 34, 63

Datta, B.; Chakrabarty, D., and Dhar, A. Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters. *Journal of Hydrology*, 376(1-2):48–57, 2009c. doi: 10.1016/j.jhydrol.2009.07.014. 16

Datta, B.; D., Chakrabarty, and Dhar, A. Identification of unknown groundwater pollution sources using classical optimization with linked simulation. *Journal of Hydro-Environmental Research*, 5(1):25–36, 2011. 17, 34

Dhar, A. and Datta, B. Multiobjective design of dynamic monitoring networks for detection of groundwater pollution. *Journal of Water Resources Planning and Management*, 133:329, 2007. 22

Dhar, A. and Datta, B. Logic-based design of groundwater monitoring network for redundancy reduction. *Journal of Water Resources Planning and Management*, 13(1):88–94, 2010. 23, 36

Doherty, J. E. and Hunt, R. J. *Approaches to highly parameterized inversion: A guide to using PEST for groundwater-model calibration*. US Department of the Interior, US Geological Survey, 2010. 111, 141

Fetter, C. W. *Applied Hydrogeology*, volume 691. Prentice Hall, New Jersey, 1994. 24, 38, 40, 114

Foster, S. S. D. and Chilton, P. J. Groundwater: the processes and global significance of aquifer degradation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 358(1440):1957–1972, 2003. 2

Freeze, R. A. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. *Water Resources Research*, 11(5):725–741, 1975. 26, 51, 52, 63

Freeze, R. A. and Cherry, J. A. *Groundwater*. Prentice Hall, 1977. 40

Gaddis, E. B.; Glennie, P. R.; Huang, Y., and Rast, W. *GEO5: Global Environmental Outlook*, chapter Chapter 4: Water. United Nations Environment Program, 2012. 1

Gelhar, L. W. *Stochastic Subsurface Hydrology*. Prentice-Hall, Englewood Cliffs, NJ, 1993. 27, 52

Gorelick, S. M.; Evans, B., and Remson, I. Identifying sources of groundwater pollution: an optimization approach. *Water Resources Research*, 19(3): 779–790, 1983. 6, 12, 34

Government of Queensland, . Water resource (Fitzroy Basin) plan 2011, subordinate legislation 2011 no. 283 made under the Water Act 2000. Technical report, Government of Queensland, 2011. xvi, 130

Harbaugh, A. W. *MODFLOW-2005: The US Geological Survey Modular Ground-water Model–the Ground-water Flow Process*. US Geological Survey Reston, VA, 2005. 43, 129

Haykin, S. *Neural networks: A comprehensive foundation*. Prentice Hall, 1999. 29, 47

He, L.; Huang, G. H., and Lu, H. W. A coupled simulation-optimization approach for groundwater remediation design under uncertainty: An application to a petroleum-contaminated site. *Environmental Pollution*, 157 (8-9):2485–2492, 2009. 16

Herrera, G. S. and Pinder, G. F. Space-time optimization of groundwater quality sampling networks. *Water Resources Research*, 41(12):W12407, 2005. 22

Hilton, A. B. C. and Culver, T. B. Groundwater remediation design under uncertainty using genetic algorithms. *Journal of Water Resources Planning and Management*, 131:25, 2005. 35

Holland, J. H. Adaptation in natural and artificial systems. *University of Michigan Press, Ann Arbor, MI*, 31, 1975. 28, 47

Hudak, P. F. A method for designing configurations of nested monitoring wells near landfills. *Hydrogeology Journal*, 6(3):341–348, 1998. 21

Ingber, L. Very fast simulated re-annealing. *Mathematical and Computer Modelling*, 12(8):967–973, 1989. 35

Ingber, L. Adaptive simulated annealing (ASA). *Global optimization C-code, Caltech Alumni Association, Pasadena, CA*, 1993. 35

Ingber, L. Adaptive simulated annealing (ASA): Lessons learned. *Control and Cybernetics*, 25:33–54, 1996. 29, 35, 48

Ingber, L. and Rosen, B. Genetic algorithms and very fast simulated reannealing: A comparison. *Mathematical and Computer Modelling*, 16(11): 87–100, 1992. 28, 35, 46

Jarvis, A.; Reuter, H. I.; Nelson, A., and Guevara, E. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), 2008. 127

Javandel, I.; Doughty, C., and Tsang, C. F. *Groundwater transport: Handbook of mathematical models*. DOE/SF/00098-T8, American Geophysical Union, Washington, DC; Lawrence Berkeley Lab., CA (USA), 1984. 24, 40

Jesson, E. E. and Bamber, B. J. Mount Morgan geo-physical survey for underground water. Technical report, Commonwealth of Australia, Department of National Development, Bureau of Mineral Resources Geology and Geophysics, 1959. 129

Keidser, A. and Rosbjerg, D. A comparison of 4 inverse approaches to groundwater-flow and transport parameter-identification. *Water Resources Research*, 27(9):2219–2232, 1991. 46

Kirkpatrick, S. Optimization by simulated annealing - quantitative studies. *Journal of Statistical Physics*, 34(5-6):975–986, 1984. 28, 46, 47

Kitanidis, P. K. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research*, 22(4):499–507, 1986. 26

Knopman, D. S. and Voss, C. I. Multiobjective sampling design for parameter estimation and model discrimination in groundwater solute transport. *Water Resources Research*, 25(10):2245–2258, 1989. 19

Kollat, J. B. and Reed, P. M. A computational scaling analysis of multiobjective evolutionary algorithms in long-term groundwater monitoring applications. *Advances in Water Resources*, 30(3):408–419, 2007. 22

Kollat, J. B.; Reed, P. M., and Maxwel, R. M. Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics. *Water Resources Research*, 47(W02529), 2011. doi: 10.1029/2010WR009194. 18

Lee, S. I. and Kitanidis, P. K. Optimal estimation and scheduling in aquifer remediation with incomplete information. *Water Resources Research*, 27(9): 2203–2217, 1991. 27

Loaiciga, H. A. An optimization approach for groundwater quality monitoring network design. *Water Resources Research*, 25(8):1771–1782, 1989. 19

Loaiciga, H.A.; Charbeneau, R.J.; Everett, L.G.; Fogg, G.E.; Hobbs, B.F., and Rouhani, S. Review of groundwater quality monitoring network design. *Journal of Hydraulic Engineering - ASCE*, 118(1):11–37, 1992. 18

Lu, G.; Clement, T. P.; Zheng, C., and Wiedemeier, T. H. Natural attenuation of BTEX compounds: Model development and field-scale application. *Ground Water*, 37(5):707–717, 1999. ISSN 1745-6584. doi: 10.1111/j.1745-6584.1999.tb01163.x. 114

Mahar, P. S. and Datta, B. Optimal monitoring network and ground-water-pollution source identification. *Journal of Water Resources Planning and Management*, 123(4):199–207, 1997. 6, 13, 21, 34, 36, 90

Mahar, P. S. and Datta, B. Identification of pollution sources in transient groundwater systems. *Water Resources Management*, 14(3):209–227, 2000. 6, 14, 34

Mahar, P. S. and Datta, B. Optimal identification of ground-water pollution sources and parameter estimation. *Journal of Water Resources Planning and Management*, 127(1):20–29, 2001. 6, 13, 34

Mahinthakumar, G. K. and Sayeed, M. Hybrid genetic algorithm – local search methods for solving groundwater source identification inverse problems. *Journal of Water Resources Planning and Management*, 131:45, 2005. 6, 15, 34, 35

Massmann, J. and Freeze, R. A. Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 1. Methodology. *Water Resources Research*, 23(2): 351–367, 1987. 18

McDonald, M. G. and Harbaugh, A. W. A modular three-dimensional finite-difference groundwater flow model. techniques of water resources investigations 06-a 1., 1988. 43

McKinney, D. C. and Loucks, D. P. Network design for predicting groundwater contamination. *Water Resources Research*, 28(1):133–147, 1992. 19

Meyer, P. D. and Brill, E. D. A method for locating wells in a groundwater monitoring network under conditions of uncertainty. *Water Resources Research*, 24(8):1277–1282, 1988. 19, 20

Meyer, P. D.; Valocchi, A. J., and Eheart, J. W. Monitoring network design to provide initial detection of groundwater contamination. *Water Resources Research*, 30(9):2647–2659, 1994. 19

Michalak, A. M. and Kitanidis, P. K. Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. *Water Resources Research*, 40(8):W08302, 2004. 11

Minsker, B. Long-term groundwater monitoring-the state of the art. *American Society of Civil Engineers, stock*, (40678), 2003. 18

MLA, Meat & Livestock Australia. Mount Morgan climate history, 2008. 127

Morris, B. L.; Lawrence, A. R.; Chilton, P. J.; Adams, B.; Caylow, R. C., and Klinck, B. A. Groundwater and its susceptibility to degradation: A global assessment of the problems and options for management. *UNEP Early Warning and Assessment Report*, 03(3), 2003. 2

Ogata, A. and Banks, R. B. *A solution of the differential equation of longitudinal dispersion in porous media: Fluid movement in earth materials*. United States Government Printing Office, Washington, 1961. 42

Pebesma, E. J. and Heuvelink, G. B. M. Latin hypercube sampling of Gaussian random fields. *Technometrics*, 41(4):303–312, 1999. ISSN 00401706. URL `http://www.jstor.org/stable/1271347`. 52

Pinder, G. Optimal search strategy for the definition of a DNAPL source. Technical report, Defense Technical Information Center, United States of America, 2009. 17, 18

Puech, V. Upper Macquarie groundwater model. Technical Report VW04680, Office of Water, NSW Government and National Water Commission, Australia, 2010. 107, 109, 110, 111, 114

Rabiner, L. and Juang, B. H. *Fundamentals of speech recognition*, volume 103. Prentice Hall, 1993. 31, 80

Reed, P. M. and Minsker, B. S. Striking the balance: Long-term groundwater monitoring design for conflicting objectives. *Journal of Water Resources Planning and Management*, 130:140, 2004. 21

Senin, P. Dynamic time warping algorithm review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 2008. 31

Sidauruk, P.; Cheng, A. H. D., and Ouazar, D. Ground water contaminant source and transport parameter identification by correlation coefficient optimization. *Ground Water*, 36(2):208–214, 1998. 12

Singh, R. M. and Datta, B. Groundwater pollution source identification and simultaneous parameter estimation using pattern matching by artificial neural network. *Environmental Forensics*, 5(3):143–153, 2004. 14, 34

Singh, R. M. and Datta, B. Identification of groundwater pollution sources using GA-based linked simulation optimization model. *Journal of Hydrologic Engineering*, 11:101, 2006. 6, 15, 34, 35

Singh, R. M. and Datta, B. Artificial neural network modeling for identification of unknown pollution sources in groundwater with partially missing concentration observation data. *Water Resources Management*, 21: 557–572, 2007. doi: 10.1007/s11269-006-9029-z. 34

Singh, R. M.; Datta, B., and Jain, A. Identification of unknown groundwater pollution sources using artificial neural networks. *Journal of Water Resources Planning and Management*, 130:506, 2004. 14, 34

Skaggs, T. H. and Kabala, Z. J. Recovering the history of a groundwater contaminant plume: Method of quasi-reversibility. *Water Resources Research*, 31(11):2669–2673, 1995. 12

Snodgrass, M. F. and Kitanidis, P. K. A geostatistical approach to contaminant source identification. *Water Resources Research*, 33(4):537–546, 1997. 12

Sun, A. Y.; Painter, S. L., and Wittmeyer, G. W. A constrained robust least squares approach for contaminant release history identification. *Water Resources Research*, 42(4):W04414, 2006a. 11, 34

Sun, A. Y.; Painter, S. L., and Wittmeyer, G. W. A robust approach for iterative contaminant source location and release history recovery. *Journal of Contaminant Hydrology*, 88(3-4):181–196, 2006b. 11, 34

Sun, N. Z. *Inverse Problems in Groundwater Modeling*, pages 12–37. Kluwer Academic Publishers, 1994. 5, 49

Taube, A. The Mount Morgan gold-copper mine and environment, queensland; a volcanogenic massive sulfide deposit associated with penecontemporaneous faulting. *Economic Geology*, 81(6):1322–1340, 1986. xvii, 132, 133, 134

Tikhonov, A. N. and Arsenin, V. Y. Solutions of ill-posed problems. *WH Winston, Washington, DC*, 1977. 4, 11

Unger, C.; Laurencont, T.; Keliher, L.; McCombe, C., and Bartley, G.

Rehabilitation plan for the Mount Morgan minesite central Queensland. Technical report, Queensland Government, Department of Natural Resources and Mines, 2003. xvii, 126, 131

Wada, Y.; Van Beek, L. P. H.; Van Kempen, C. M.; Reckman, J. W. T. M.; Vasak, S., and Bierkens, M. F. P. Global depletion of groundwater resources. *Geophysical Research Letters*, 37(L20402), 2010. 1

Wagner, B. J. Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling. *Journal of Hydrology*, 135(1-4):275–303, 1992. 13

Wagner, B. J. and Gorelick, S. M. A statistical methodology for estimating transport parameters: Theory and applications to one-dimensional advective-dispersive systems. *Water Resources Research*, 22(8):1303–1315, 1986. 13

Wels, C.; Findlater, L., and McCombe, C. Assessment of groundwater impacts at the historic Mount Morgan mine site, Queensland, Australia. In *Proceedings of the Seventh International Conference on Acid Rock Drainage*, 2006. xvi, 128, 129, 132, 135

Yeh, H. D.; Chang, T. H., and Lin, Y. C. Groundwater contaminant source identification by a hybrid heuristic approach. *Water Resources Research*, 43 (9):W09420, 2007. 6, 15, 28

Yeh, J. T. C. Stochastic modelling of groundwater flow and solute transport in aquifers. *Hydrological Processes*, 6(4):369–395, 1992. ISSN 1099-1085. doi: 10.1002/hyp.3360060402. 27, 52

Yenigül, N. B.; Elfeki, A. M. M.; Gehrels, J. C.; van den Akker, C.; Hensbergen, A. T., and Dekking, F. M. Reliability assessment of groundwater monitoring networks at landfill sites. *Journal of Hydrology*, 308(1):1–17, 2005. 20

Zheng, C. MT3D: A modular three-dimensional transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems. *Report to the US Environmental Protection Agency, Robert S. Kerr Environmental Research Laboratory, Ada, OK*, 1990. 25

Zheng, C. and Bennett, G. D. *Applied contaminant transport modeling: Theory and practice*. Van Nostrand Reinhold, New York, 1995. 25

Zheng, C. and Wang, P. P. MT3DMS: A modular three-dimensional multi-species transport model for simulation of advection, dispersion, and chemical reactions of contaminants in ground-water systems. Documentation and user's guide. In *Contract Report SERDP-99-1, U.S. Army Engineer Research and Development*, 1999. 25, 44