“What is important is to keep learning, to enjoy challenge and to tolerate ambiguity. In the end there are no certain answers”

Martina Horner

University of Alberta

Probabilistic Modeling of Natural Attenuation of Petroleum Hydrocarbons

by

Amir Hossein Hosseini

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Civil and Environmental Engineering

Fall 2009
Edmonton, Alberta

Examining Committee

Dr. Dave Chan, Department of Civil and Environmental Engineering

Dr. Jaime Gómez-Hernández, Universidad Politécnica de Valencia

Dr. Dave Sego, Department of Civil and Environmental Engineering

Dr. Kevin Biggar, Department of Civil and Environmental Engineering

Dr. Clayton Deutsch, Department of Civil and Environmental Engineering

Dr. Carl Mendoza, Department of Earth and Atmospheric Sciences

To my lovely wife Sara

Your enduring love, support and encouragement means the world to me

To my parents Saeed and Eli

For giving me the courage to set and reach goals

Abstract

Natural attenuation refers to the observed reduction in contaminant concentration via natural processes as contaminants migrate from the source into environmental media. Assessment of the dimensions of contaminant plumes and prediction of their fate requires predictions of the rate of dissolution of contaminants from residual non-aqueous-phase liquids (NAPLs) into the aquifer and the rate of contaminant removal through biodegradation. The available techniques to estimate these parameters do not characterize their confidence intervals by accounting for their relationships to uncertainty in source geometry and hydraulic conductivity distribution. The central idea in this thesis is to develop a flexible modeling approach for characterization of uncertainty in residual NAPL dissolution rate and first-order biodegradation rate by tailoring the estimation of these parameters to distributions of uncertainty in source size and hydraulic conductivity field.

The first development in this thesis is related to a distance function approach that characterizes the uncertainty in the areal limits of the source zones. Implementation of the approach for a given monitoring well arrangement results in a unique uncertainty band that meets the requirements of unbiasedness and fairness of the calibrated probabilities. The second development in this thesis is related to a probabilistic model for characterization of uncertainty in the 3D localized distribution of residual NAPL in a real site. A categorical variable is

defined based on the available CPT-UVIF data, while secondary data based on soil texture and groundwater table elevation are also incorporated into the model. A cross-validation study shows the importance of incorporation of secondary data in improving the prediction of contaminated and uncontaminated locations. The third development in this thesis is related to the implementation of a Monte Carlo type inverse modeling to develop a screening model used to characterize the confidence intervals in the NAPL dissolution rate and first-order biodegradation rate. The development of the model is based on sequential self-calibration approach, distance-function approach and a gradient-based optimization. It is shown that tailoring the estimation of the transport parameters to joint realizations of source geometry and transmissivity field can effectively reduce the uncertainties in the predicted state variables.

Acknowledgement

It is my pleasure to thank my co-supervisors Dr. Kevin Biggar, Dr. Clayton Deutsch, Dr. Carl Mendoza and Dr. Dave Sego for their support, encouragement and advice throughout my studies at the University of Alberta. In particular, I would like to thank Dr. Clayton Deutsch for his continued support, suggestions, help and advice and for opening the doors of the world of stochastic modeling and geostatistics to me.

The financial support for this research provided by the University of Alberta, Alberta Ingenuity Fund, Provost Doctoral Entrance Award, and WorleyParsons Komex graduate scholarship is greatly appreciated.

My family has been a continued source of support. I would like to acknowledge the encouragement of my parents Saeed and Eli. Most importantly, I would like to thank my wife Sara for her love, understanding and support along the way.

Table of contents:

List of Tables:

List of Figures:

Figure 7-2: The concentration profile for the reference homogenous case, as well as concentration profiles for four realizations with homogenous hydraulic conductivity and variable dissolution rate constants. The red arrows show the locations of the two point sources used in the subsequent estimation of first-order biodegradation rate constants – Page 267

Figure 7-3: The concentration profile for the reference heterogeneous case, as well as concentration profiles for four realizations with heterogeneous hydraulic conductivity and constant source concentration. The red arrows show the locations of the two point sources used in the subsequent estimation of first-order biodegradation rate constants – Page 268

Figure 7-4: The concentration profile for the reference heterogeneous case, as well as concentration profiles for four realizations with heterogeneous hydraulic conductivity and variable dissolution rate constants. The red arrows show the locations of the two point sources used in the subsequent estimation of first-order biodegradation rate constants – Page 269

Figure 7-5: The reference hydraulic conductivity field (the case with $\sigma_{lnK} = 1.0$), the location of the source zone and the types and values of the flow boundary conditions in the synthetic reactive contaminant transport problem – Page 272

Figure 7-6: The hydraulic head contours for the reference synthetic aquifers with log-normal hydraulic conductivity distributions with a mean equal to 4.1×10^{-5} m/s and standard deviations equal to (a) σ_{lnK} =1.0, (b) σ_{lnK} =1.5 and (c) σ_{lnK} =2.0 in natural logarithmic units. The centerline wells are depicted by black circles – Page 273

Figure 7-7: The development of the dissolved plumes after 7 years, when reaching to a semi-steady-state condition for the reference synthetic aquifers with log-normal hydraulic conductivity distributions with a mean equal to 4.1×10^{-5} m/s and standard deviations equal to (a) σ_{lnK} =1.0, (b) σ_{lnK} =1.5 and (c) σ_{lnK} =2.0 in natural logarithmic units – Page 274

List of Acronyms:

BTEX: Benzene-Toluene-Ethylbenzene-Xylenes

BU: Bayesian Updating

CDF: Cumulative Distribution Function

CG: Conjugate Gradients

CORONA: Consortium for Research on Natural Attenuation

CPT: Cone Penetration Testing

CPT-UVIF: Ultra-Violet Induced Fluorescence Cone Penetration Testing

CUB: Centerline of Uncertainty Band

DF: Distance Function

DNAPL: Dense Non-Aqueous Phase Liquid

DOS : Distance of Stabilisation

FDM: Finite Difference Method

FEM: Finite Element Method

GSLIB: Geostatistical Software Library

GW: distance to Ground Water

HMOC: Hybrid Method of Characteristics

HP: higher Permeability hydro-facies

IK: Indicator Kriging

KL: Karhunen Loeve expansion

KPCA: Kernel Principal Component Analysis

LNAPL: Light Nonaqueous Phase Liquid

LP: Lower Permeability hydro-facies

LUST: Leaking Underground Storage Tanks

LVM: Locally Varying Means

MAPS: Maximum A-Posteriori Selection

MCMC: Markov Chain Monte Carlo

MCS: Monte Carlo Simulation

MDS: Multi Dimensional Scaling

MMOC: Modified Method of Characteristics

MNA: Monitored Natural Attenuation

MOC: Method of Characteristics

MP: Medium Permeability hydro-facies

MTBE: Methyl Tertiary-Butyl Ether

NAPL: Nonaqueous Phase Liquid

OCR: Over Consolidation Ratio

PAH: Poly-cyclic Aromatic Hydrocarbons

PCBs: Polychlorinated Biphynels

PDF: Probability Density Function

PHC: Petroleum Hydrocarbons

PR: Permanence of Ratios

SBT: Soil Behavior Type

SCM: Site Conceptual Model

SGS: Sequential Gaussian Simulation

SIS: Sequential Indicator Simulation

SIS-LVM: Sequential Indicator Simulation with Locally Varying Means

SISI: Soil In-situ State Index

SRF: Stationary Random Function

SSC: Sequential Self-Calibration

SVOC: Semi-Volatile Organic Carbon

TOD: Time of Dissolution

TOR: Time of Remediation

TOS: Time of Stabilization

UB: Uncertainty Band

US EPA: United State Environmental Protection Agency

UVIF: Ultra-Violet Induced Fluorescence

VOC: Volatile Organic Carbon

WUB: Width of Uncertainty Band

List of Nomenclature:

A^{rel} (Chapter 3): measure of accuracy

[**A**] (Chapter 4): band matrix in DKR factorization

[**A**] (Chapter 5): matrix of descritized transmissivities

α (Chapter 3): scaling factor

α_L (Chapter 4): longitudinal dispersivity

α_T (Chapter 4): transverse dispersivity

α_x (Chapter 2): longitudinal dispersivity

α_y (Chapter 2): horizontal transverse dispersivity

α_z (Chapter 2): vertical transverse dispersivity

B (Chapter 4): aquifer thickness

$\mathbf{b_r}$ (Chapter 5): vector of estimated parameters in iteration r

β (Chapter 3): separation factor

C (Chapters 2, 4, 5, 7): dissolved concentration

$\hat{C}$ (Chapter 7): dissolved concentration of a tracer

C (Chapter 3): normalization constant in Bayesian updating

C (Chapter 5): diagonal scaling matrix

C_0 (Chapter 7): point source concentration

$\hat{C}_0$ (Chapter 7): point source concentration of a tracer

c (Chapter 3): constant controlling short-scale variations in interpolation

cv_i (Chapter 5): coefficient of variation for the observed concentration

cv_{nc} (Chapter 6): coefficient of variation for concentrations with Gaussian noise

C_{Γ_1} (Chapter 2): specified concentration along the boundary Γ_1

C_I (Chapter 3): stationary indicator covariance function

$C_m^{n^*}$ (Chapter 4): Average concentration of particles in each cell

C_r (Chapter 4): courant number

C_s (Chapters 2, 5): dissolved concentration of species *s*

C_s (Chapter 4): concentration of solute in the influx

C_s^0 (Chapter 2): dissolved concentration of species *s* at time *t = 0*

C_s^{eq} (Chapters 2, 4, 5, 6, 7): equilibrium concentration for species *s*

C_s^{sol} (Chapters 2, 4, 5, 6, 7): solubility limit for species *s*

C_{sr} (Chapter 2): source concentration

$C_{T_{UVIF}}$ (Chapter 3): measure of closeness

$C_{T_{UVIF}}^{rel}$ (Chapter 3): measure of relative closeness

D (Chapter 4): dispersion coefficient

D_{ij} (Chapters 2, 4, 5, 6, 7): directional dispersion coefficient

DF_{ID}^* (Chapter 3) interpolated distance function by inverse distance technique

D_m (Chapter 6): diffusion coefficient of NAPL species

$\mathbf{d_r}$ (Chapter 5): updating vector for the estimated parameters

d_{50} (Chapter 6): median grain size

erf(*x*) (Chapter 2): error function

F (Chapter 5): generic term for objective function

F_C (Chapter 5): least square objective function for concentrations

$F'_{C,j}$ (Chapter 5): modified objective function accounting for source size quantile

F_h (Chapter 5): weighted squared difference in simulated, observed state variables

F_T (Chapter 5): plausibility term

f^{DC} (Chapter 3): data conditioning factor

f_i (Chapter 2): function representing dispersive flux normal to the boundary Γ_2

f_{oc} (Chapter 2): organic fraction of soil

f_s (Chapters 2, 4, 5, 6, 7): is the mole fraction of the species *s*

f_s (Chapter 3): sleeve friction

g_i (Chapter 2): function for advective/dispersive flux normal to the boundary Γ_3

$\gamma(h)$ (Chapter 3): variogram function

{**h**} (Chapter 5): vector of unknown nodal piezometric head values

I (Chapter 5): identity matrix

i (Chapter 6): hydraulic gradient

$i(\mathbf{u}_\alpha; z_k)$ (Chapter 3): stationary categorical variable at location α for category *k*

J_{tol} (Chapter 5): tolerance value for minimization of the object function

[**K**] (Chapter 4): preconditioning matrix

K_G (Chapters 6, 7): geometric average of hydraulic conductivity measurements

k (Chapter 7): lumped decay coefficient

k_i (Chapters 2, 4, 5, 6, 7): hydraulic conductivity in the direction of flow

k^{NAPL} (Chapter 2): mass transfer rate coefficient for NAPL

k_{dis} (Chapters 5, 6, 7): dissolution rate coefficient for NAPL

k_{dis}^{PR} (Chapters 5, 6, 7): prior value for dissolution rate coefficient for NAPL

L (Chapter 4): Characteristic length

λ (Chapters 2, 4, 5, 6, 7): first-order decay (biodegradation) coefficient

λ_{α} (Chapter 3): kriging weight

λ_i^k (Chapter 5): ordinary kriging weights

λ_{ID} (Chapter 3): inverse distance weight

λ^{PR} (Chapters 5, 6, 7): prior value for first-order biodegradation coefficient

M_0 (Chapter 1): mass of initial contaminant

$M_{threshold}$ (Chapter 1): regulatory threshold

m_r (Chapter 5): Marquardt parameter

n (Chapter 5): number of observations

ND (Chapter 5): number of observed concentration

$p(k)$ (Chapter 3): stationary prior probability of soil type k

P_e (Chapter 3): Peclet number

Pe' (Chapter 6): modified Peclet number

P_j (Chapter 3): probability associated with each quantile map

P_j^* (Chapter 3): calculated probability of falling inside each quantile map

Q (Chapter 5): sink/source term in 2D flow equation

q_c (Chapter 3): tip resistance

q_j (Chapter 5): source size quantile

q_s (Chapter 4): volumetric rate at which water is added to or removed from system

q_{sr} (Chapter 2): volumetric flow rate representing fluid sources (or sinks)

θ (Chapters 2, 4, 5, 6, 7): porosity

θ (Chapter 3): size of the search angle

θ (Chapter 5): angle between the updating vector and steepest descent vector

$\{\theta\}$ (Chapter 4): new search direction in CG algorithm

R_f (Chapter 3): friction ratio

R_n (Chapter 2): chemical reaction term

R_{NA} (Chapter 1): natural attenuation rate

R_{source}^{NAPL} (Chapter 2): dissolved NAPL influx rate

ρ_b (Chapters 2, 4, 5, 6, 7): bulk density of the subsurface medium

ρ_r (Chapter 5): damping parameter

s (Chapter 5): measure of fit

$S_{k_{dis}}$ (Chapter 5): sensitivity of concentrations with respect to dissolution rate

S_λ (Chapter 5): sensitivity of concentrations with respect to biodegradation rate

S_{v_x} (Chapter 5): sensitivity of concentrations with respect to groundwater velocity

S_s (Chapter 2): specific storage

S_s^{NAPL} (Chapter 2): NAPL mass of substrate s per unit mass of dry soil

σ^2 (Chapter 5): variance of head measurement error based on Gaussian noise

$\sigma_{k_{dis}}$ (Chapter 7): standard deviation of dissolution rate constant

σ_{krig} (Chapter 5): kriging standard deviation

σ_{lnK} (Chapters 5, 6, 7): standard deviation of hydraulic conductivity

$\sigma_{\ln k}$ (Chapter 5): standard deviations of prior values for dissolution rate

$\sigma_{\ln\lambda}$ (Chapter 5): standard deviations of prior values for biodegradation rate

σ_{nH} (Chapter 6): standard deviation of head measurement error by Gaussian noise

T (Chapter 5): transmissivity

t (Chapter 5): proportionality parameter

T_t^{NAPL} (Chapter 4): equivalent mass of all inert and non-biodegradable materials

T_{UVIF} (Chapter 3): categorical stationary random variable for presence of NAPL

v_{avg} (Chapter 6): average groundwater velocity

v_c (Chapter 7): contaminant velocity

v_i (Chapter 2): seepage velocity

v_{ij} (Chapter 5): coefficients of the inverse of the kriging estimation matrix

W_S (Chapter 7): source width perpendicular to average flow direction

w_i (Chapter 5): diagonal elements of the weight matrix

w_{ij} (Chapter 5): weights equal to inverse of head error covariance matrix

ω (Chapter 3): interpolation exponent

ω (Chapter 4): time-weighting factor in MOC approach

ω^n (Chapter 4): a scalar to minimize the error in CG algorithm

ω_s (Chapter 4): molecular weight of substrate s

ω_t (Chapter 4): equivalent molecular weight of mixture of inert materials

$\mathbf{X}_r$ (Chapter 5): matrix of sensitivities

$\mathbf{y}$ (Chapter 5): vector of observed concentrations

$\mathbf{y}(\mathbf{b_r})$ (Chapter 5): vector of simulated concentrations

$Y_{k,\text{krig}}$ (Chapter 5): ordinary kriging estimate at any master point location

y_{GW} (Chapter 3): continuous random variable representing distance to water table

y_{SCI} (Chapter 3): continuous random variable representing soil texture

Z_{GW} (Chapter 3): elevation of groundwater table at the data location

Z_{rel} (Chapter 3): the measure of relative elevation

Z_{UVIF} (Chapter 3): elevation of the data point in the global coordinate system

CHAPTER 1

INTRODUCTION

Groundwater comprises about 98% of the available fresh water in the world. Therefore, protection and restoration of groundwater quality is of great global importance. The vulnerability of groundwater resources to soil and aquifer contamination caused by petroleum hydrocarbons (PHC) and organic chemicals is a widespread problem that is associated with significant technical and economical challenges. These are exacerbated by the difficulties associated with locating and removing the underground contamination by traditional extraction and excavation methods. Thus, application of cost-effective in situ remediation approaches such as Monitored Natural Attenuation (MNA) that utilizes the natural attenuation capacity of the subsurface has become widespread.

Reduction in concentrations of the organic contaminants in the subsurface is due to combination of physical, geochemical and biochemical processes. MNA is defined as deliberate use of naturally occurring biodegradation, sorption, and dispersion processes to remediate contaminated groundwater systems. According to US EPA (1999a), MNA can be considered as an appropriate remediation strategy, providing that it meets site remediation objectives within a timeframe that is reasonable compared to that offered by other methods.

Conceptually, estimating the length of time required for natural processes to remove a particular contaminant from a groundwater system is a simple mass-balance problem, termed time-of-remediation (TOR). According to Chapelle et al.

(2003), a TOR can be defined as the time required for lowering contaminant mass below a given threshold ($M_{threshold}$) and can be quantified by:

$$TOR = [M_0 - M_{threshold}]/R_{NA} \quad [1.1]$$

where, M_0 is the mass of initial contaminant to be lowered below a regulatory threshold by the rate of natural attenuation processes (R_{NA}) in a groundwater system. According to Equation [1.1], one needs an estimate of the mass of the contaminants present as well as the rate of ongoing natural attenuation processes acting on the contaminants. As pointed out by Chapelle et al. (2003), the reliability of any remediation time estimates is directly affected by the reliability of these parameters.

In real contamination scenarios, however, the estimation of the TOR is far more complex than what is indicated by Equation [1.1]. This is due to complexities and uncertainties associated with the rates of natural attenuation, which can be variable in space and time, as well as uncertainties associated with the load of existing contaminants in the aquifer, and uncertain source characteristics.

The focus of this research is on the development of a simple and flexible screening level model that can be used to predict different components of a TOR problem with their uncertainty for groundwater contamination scenarios associated with petroleum hydrocarbons.

1.1 Statement of the problem

In the context of groundwater contamination with petroleum hydrocarbons, a TOR problem can be conceptualized as three interactive components (Chapelle et al. 2003): (1) estimation of the length of the contaminant plume under semi-steady-state condition, (2) estimation of the time needed for the non-aqueous-phase-liquid (NAPL) source to dissolve, disperse and biodegrade, and (3) estimation of the time required for the plume to stabilize to a smaller size, upon reduction of the source. There is significant uncertainty associated with the solution of a TOR problem due to uncertainties associated with source properties and the uncertainties associated with contribution and efficiency of concentration-reducing mechanisms.

The important source properties are the source geometry and the dissolution rate of contaminant species into groundwater. The concentration-reducing mechanisms can be grouped into transport mechanisms, phase transfers and transformation mechanisms. The important transport mechanisms are advection, dispersion and dilution (recharge). The important phase transfer mechanisms are sorption and volatilization, and the important transformation mechanisms are aerobic and anaerobic biodegradation and chemical transformations.

Aerobic and anaerobic biodegradation are the only mechanisms that are responsible for destructive removal of petroleum hydrocarbon contaminants (PHCs) in groundwater. Many biodegradation models that simulate rather complex kinetics and multi-component reactions have been developed. Examples of these techniques are instantaneous reaction kinetics (Borden and Bedient 1986) and Monod kinetics (Monod 1949 and Rifai et al. 2000). It is evident that many of the required kinetic parameters for these complex models can not be measured or estimated by routine natural attenuation protocols. Thus, utility of these models is often limited (Rifai and Rittaler 2005). Simpler approaches with limited number

of parameters are often preferred as they can be supported by the available data (Essaid et al. 2003).

Application of first-order reaction models is quite common in natural attenuation studies, particularly at the screening level. Based on the concentrations measured at monitoring locations, the field-scale first-order rates are estimated by trial and error calibration (Borden et al. 1986, Chaing et al. 1989 and Lu et al. 1999), by inverse modeling techniques (Carrera and Medina 1996 among others) or by field approaches such as mass-flux (Borden et al. 1997, King et al. 1999, Peter et al. 2004) and concentration-distance relationships with centerline and off-centerline measurements (Buscheck and Alcantar 1995, Chapelle et al. 1996, Wiedemeier et al. 1996, Zhang and Heathcoat 2003, Stenback et al. 2004). The parameter estimates by trial and error calibration techniques are modeler-dependent and a measure of uncertainty is not often available. Among the inverse modeling approaches (Reviewed in Chapter 2), to the author's knowledge, none of them quantifies the non-linear confidence intervals in the estimated first-order rates under uncertainty of source properties (size, geometry and dissolution rate) and hydraulic conductivity distribution through generating multiple realizations and conditioning them to the observed concentrations. In the case of field estimation techniques (mass-flux and concentration-distance techniques), the estimated first-order rates are affected by heterogeneity and uncertainty in hydrogeological and mass transport properties as well as the size of the source zone and may over/underestimate the true rate constant up to two-orders of magnitude (Bauer et al. 2006, Kubert and Finkel 2006, Beyer et al. 2007). In addition to the effects of heterogeneity and uncertainty in groundwater velocity and dispersivities, the field estimation approaches are also limited in that they do not account for rate-limited mass transfer between NAPL source and groundwater (Chapter 7).

Understanding of the NAPL source dissolution rate is another important factor when investigating different aspects of a TOR problem. A number of numerical (Dillard and Blunt 2000, and Dillard et al. 2001, Parker and Park 2004, Christ et al. 2006) and experimental (Imhoff et al. 1994, Powers et al. 1994, Nambi and Powers 2003) works have studied the rate of dissolution of NAPL into groundwater at pore and field scales. In these works, the estimated dissolution rate constants were determined either under controlled laboratory conditions or under simplified field conditions with local mass flux/concentration measurements. Many of the experimental studies in this area resulted in formulation of mass transfer correlations which relate dimensionless forms of the mass transfer coefficient (e.g. Sherwood number) to dimensionless forms of system properties (e.g. Reynolds number and Schmidt number) at a pore-scale. Dillard et al. (2001) applied the pore-scale model of Powers et al. (1994) to field-scale problems and derived a number of correlations between Peclet number (a function of groundwater velocity and particle size distribution) and the rate of mass transfer between NAPL and water. According to the results by Dillard et al. (2001), one-order of magnitude variability exists in the value of dissolution rate for every given value of Peclet number. This variability is attributed to different degrees of saturation of NAPL. Given the uncertainties associated with groundwater velocity, residual NAPL saturation and particle-size distribution, few orders of magnitude uncertainty exists in the estimated dissolution rate constants based on the proposed correlations. Later studies (Essaid et al. 2003, Parker and Park 2004) showed that the dissolution rate constant estimates by the pore-scale correlations over-estimate the field-scale values for up to three orders of magnitude. In addition to pore-scale models, a number of upscaled screening models have also been proposed to incorporate the effects of spatial variations in NAPL saturations and flow by-passing to approximate field-scale dissolution (Parker and Park 2004, Zhu and Sykes 2004, among others). These upscaled models have been developed using explicit descriptions of pore networks and NAPL saturation distribution or

by fitting solutions of a simplified one-dimensional component mass balance equations to experimentally or numerically generated results, and they typically assume that the flow field is under steady-state condition. According to Christ et al. (2006), the parameterization of the upscaled models tend be more site-specific and depending on the specific source-zone scenario used for calibration of the upscaled models, the flux-weighted concentration predictions may over-estimate the true values by more than one order of magnitude. As a matter of fact, the rate of dissolution of NAPL into groundwater depends on a number of spatially and temporally variable factors: the interfacial area between the NAPL and water and wettability of porous media (Imhoff et al. 1994, Bradford et al. 1998), aquifer heterogeneity (Kueper and Frind 1991, Mayer and Miller 1996), the size and shape of NAPL blobs (Powers et al. 1994) and the groundwater velocity (Pfannkuch 1984). It is obvious that many of these factors as well as other chemical and biological parameters that affect the dissolution of NAPL (Chu et al. 2007) can not be easily quantified or estimated by monitoring networks to be used by pore-scale or upscaled models in contaminant transport modeling studies. Thus, NAPL dissolution is often mathematically simulated with a lumped mass transfer rate coefficient that controls the rate-limited dissolution (Chapter 2). An option is to calibrate such mass transfer rate (or mass flux rate) to site-specific observations through trial and error calibration or by inverse modeling.

In the hydrogeology literature, a number of inverse modeling techniques have been reported to delineate the sources of contaminants. Gorelick et al. (1983) formulated a simulation-optimization approach using linear programming and multiple regressions. Their work was developed for non-reactive contaminants, assuming no uncertainty in the aquifer properties. They used their models to identify pipe leak locations and discrete point sources. Wagner (1992) presented a maximum likelihood approach for a single species non-reactive contaminant, in which some potential disposal zones were initially considered and their associated

disposal fluxes were estimated simultaneously with aquifer properties. In a probabilistic framework, Wilson and Liu (1994) solved the advection-dispersion equation backward in time by keeping the dispersion part positive and reversing the advection part. They generated maps of 'travel time probability' and the 'location probability'. Snodgrass and Kitanidis (1997) presented a stochastic approach by combining Bayesian theory and geostatistical techniques. Their method, which was developed in one-dimension, incorporates uncertainty in contaminant source concentration, but the location of potential sources must be known a priori. An inverse method based on correlation coefficient optimization was presented by Sidauruk et al. (1998), where they estimated dispersion coefficients, flow velocities as well as the location and time origin of the pollutant based on analytical solutions. An inverse procedure was developed by Sciortino et al. (2000) based on a three-dimensional analytical transport model and a gradient based optimization approach to identify the source location and the geometry of a dense-non-aqueous-phase-liquid (DNAPL) pool. Aral et al. (2001) used an optimization approach based on genetic algorithms to infer the release history and source location of a contaminant. Mahar and Datta (2001) developed a staged approach to identify the source location and designed an optimized monitoring network under steady and transient conditions. First, they used an optimization approach to initially identify an unknown pollution source based on observation data. Then, they simulated different realizations of contaminant plume using perturbed sources. Ultimately, the optimized locations of monitoring wells and an improved estimation of the sources was obtained. More recently, Neupauer and Lin (2006) presented an approach to condition backward probability density functions (PDF) of source location to concentration data. Their work was an extension to the work of Neupauer and Wilson (2001) and was presented to characterize the uncertainty in location or release time of an instantaneous point source of contamination. Sun et al. (2006) formulated a constrained robust least squares estimator to characterize the uncertainty in source location and release

time of a non-reactive contaminant in an uncertain but homogenous transmissivity field. For a NAPL source zone, Newman et al. (2006) formulated an algorithm based on simulated annealing and minimum relative entropy to estimate non-decaying flux through a vertical flux plane. Yeh et al. (2007) also have proposed an approach based on combination of simulated annealing and three-dimensional solute transport modeling. They considered a suspected source area and estimated the source location, release concentration and release period for homogenous and heterogeneous aquifers.

None of the above inverse modeling techniques deals with characterization of source properties when the reaction rates of contaminants are unknown or uncertain. Therefore, the application of these approaches would be limited for the purpose of this study. Also, similar to the case of first-order biodegradation rate, to the author's knowledge, none of the previous works gives a measure of non-linear confidence interval for the dissolution rate constant through implementation of Monte Carlo type inverse modeling approaches under uncertainty of the source size and hydraulic conductivity.

Simultaneous characterization of uncertainty in rate-limited dissolution and field-scale biodegradation is important for development of an advanced screening tool for management of the TOR problems; but it is subject to potential numerical instabilities. For a real site with crude oil (BTEX) contamination and a simple representation of the source zone, Essiad et al. (2003) implemented inverse modeling in an 'optimal' sense to estimate NAPL dissolution rate and individual first-order biodegradation rates for BTEX compounds as well as other parameters such as the recharge rate, hydraulic conductivity, and transverse dispersivity. They only achieved convergence when they estimated a single dissolution rate for all BTEX compounds and coupled the simulation of BTEX compounds through simultaneous use of oxygen during aerobic biodegradation (crossover effect). In

other words, they failed to estimate individual dissolution rate and first-order biodegradation constants for each BTEX component due to high correlation between these parameters that results in parameter non-uniqueness. As pointed out by Carrera and Neuman (1986) and reviewed by Friedel (2005) and Carrera et al. (2005), the four primary reasons for parameter non-uniqueness are precision of numerical solution (e.g. round-off errors in calculation of sensitivities), numerical dispersion, local minima in parameter space and correlation among parameters. In the problem of simultaneous estimation of dissolution rate and biodegradation rate, non-uniqueness (uncertainty) of the parameter estimates is not only due to above-mentioned numerical instabilities (high correlation among parameters and existence of local minima), but also due to uncertainty in model structure and/or values of other hydrogeological or mass transport parameters. As discussed before, to the author's knowledge, no previous work is reported to characterize the uncertainties in dissolution rate and first-order biodegradation rate due to uncertainties in model source geometry (size) and other parameters (hydraulic conductivity distribution) through implementation of Monte Carlo type inverse modeling. It is observed that (Chapter 5) the uncertainties (non-uniqueness problems) associated with numerical instability problems can be avoided by designing an appropriate monitoring network and defining the objective function in such a way that it adequately preserves the information contained in the zones well away from the source (edge of the plume).

In addition to uncertainties associated with the location, size and rate of dissolution of residual NAPL sources, uncertainties exist in their localized distribution as well. A number of related analytical, numerical (reviewed in Chapter 3), and inverse modeling approaches (reviewed in the previous paragraphs) have been reported in the literature to delineate the distribution of NAPL sources and investigate the downstream impacts. Because the mechanisms that govern the distribution/re-distribution of residual NAPLs are quite complex

and site-specific (Chapter 3), development of probabilistic models using site-specific data is deemed appropriate. A number of field investigation approaches have been proposed for collection of site-specific data in NAPL source zones. The ultra-violet-induced-fluorescence cone penetration testing or CPT-UVIF (Pepper et al. 2002, Kram et al. 2004) is an advanced technology that is commonly used for this purpose. The work of D'Affonseca et al. (2008) is the only work in the literature that uses the data collected by CPT-UVIF technology for characterization of a dense-non-aqueous-phase-liquid (DNAPL) for the purpose of subsequent fate and transport modeling. They developed their model by deterministically calibrating the lateral and vertical extent of the DNAPL source zone with medium to coarse grained sand strata. Although the deterministic model of D'Affonseca et al. (2008) has made use of high-resolution CPT-UVIF data to improve characterization of a DNAPL source zone, their model is limited because the deterministic calibration of areal and vertical extent of NAPL source to soil horizons underestimates the existing uncertainty associated with short-scale variability of residual NAPL distribution, unresolved heterogeneities of soil properties and possible impacts of groundwater surface fluctuations on distribution/redistribution of immiscible contaminants.

1.2 Research objectives

The primary goal in this thesis is to develop a simple screening model that can be used to (1) simultaneously characterize the non-linear confidence intervals in the values of dissolution rate constant and first-order biodegradation rate under uncertain source geometry and hydraulic conductivity distribution and, (2) reduce the uncertainty in the TOR state variables being mass loaded into the aquifer and plume dimensions by stochastic inverse modeling. To achieve this goal, a number of objectives have been defined in the following paragraphs.

The first objective in this thesis is to develop a boundary modeling approach to characterize the uncertainty in areal limits of the source zones based on the site-specific well arrangement. These non-stationarity limits cannot be modeled by traditional geostatistical approaches such as trend-modeling, indicator simulation or object-based modeling. The approach should be simple and flexible, and based on the available arrangement of monitoring wells. The most important requirements for such boundary modeling approach are unbiasedness and fairness of the calibrated probabilities and the associated uncertainty band. The source zone geometry and its dimensions are important in virtually all contaminant transport modeling studies; however, to date no explicit modeling approach for delineation of uncertainty in the areal extent of the source based on the arrangement of monitoring wells has been presented.

The second objective of this thesis is development of a probabilistic model for characterization of uncertainty in the 3D localized distribution of residual NAPL in a real PHC contaminated site, while accounting for the effects of soil heterogeneities and the distance to water table. The CPT-UVIF technology is a valuable tool in high-resolution data acquisition for mechanical properties of the soil and NAPL contamination; however, it has not been used in a geostatistical

inference framework for development of probabilistic models for (residual) NAPL contamination.

The third objective of this thesis is to develop a mass transport simulator that is simple and flexible and can be used in subsequent parameter estimation studies. The primary reasons to develop the transport code rather than using commercial/open-source software are flexibility and ease of implementation and customization. A number of flexibilities include (1) simple GSLIB-like parameter files and Geo-EAS file formatting that can be executed in LINUX scripts for automating the processes; (2) flexible mass transfer mechanisms (e.g. rate-limited dissolution), and (3) simultaneous calculation of sensitivity coefficients as well as optimization for subsequent parameter estimation.

The fourth objective of this thesis is to perform a simple Monte Carlo type inverse modeling to characterize the non-linear confidence intervals in important mass transport parameters and state variables under uncertainty of hydraulic conductivity field and source geometry, to study the possibility of reducing the uncertainties in the source size and to study the effects of observation errors on the outcomes. In multi-state nonlinear inverse problems associated with subsurface mass transport, three different approaches can be considered: (1) decoupled inverse process involving two separate inverse problems (flow and transport) to be solved (Strecker and Chu (1986), Mishra et al. 1989), (2) the flow and transport equations are loosely coupled by sequentially solving both inverse problems in every non-linear parameter estimation step (Medina and Carrera 1996, Essiad et al. 2003), and (3) fully coupled inverse problems (Sun and Yeh, 1990), where the governing equations of the forward problem are coupled by construction through their dependent variables. The primary motivation for coupling the flow and transport inverse problems is to take advantage of the crossover effect (calibrating the model parameters against multiple types of

observation data). In this work, however, a decoupled approach is adapted. The motivations for adapting a decoupled approach are (1) to avoid the computational burden associated with the calculation of sensitivity coefficients in a multi-state system (Sahuquillo et al. 1999) while the significance of incremental value of such coupling in improving the estimations of dissolution rate and first-order biodegradation rate is unknown; and (2) to avoid the potential adverse effects that perturbations of hydraulic conductivity field may have on potentially unstable problem of simultaneous estimation of the dissolution rate and biodegradation rate (as observed by Essaid et al. 2003 and in Chapter 5). As a secondary objective, the instability problems associated with simultaneous estimation of dissolution rate constant and first-order biodegradation rate constant are also investigated.

The fifth and the last objective of this thesis is related to characterizing the uncertainty in the field-estimated first-order biodegradation rate constant under more realistic source conditions and due to uncertainty in the observed hydraulic heads, through a series of Monte Carlo simulations with simple 1D and 2D problem settings. It is well-known that the field estimated first-order rates by concentration-distance approaches are prone to uncertainty due to uncertainty in the estimated seepage velocity and longitudinal dispersivity, as well as missing the centerline of the plume in the heterogeneous aquifers (Bauer et al. 2006). The centerline approaches proposed in the literature are based on the assumption of constant concentration in the source area that affects the estimation of the true first-order biodegradation rate constant. The other unresolved issue associated with the centerline approaches is the effects of measurement errors in hydraulic heads on the layout of the centerline wells and estimation of first-order biodegradation rate by 1D concentration-distance relationships.

1.3 Thesis outline

Chapter 2 reviews natural attenuation principles and modeling. Section 2.1 reviews the dominant natural attenuation mechanisms including transport, phase transfer and chemical and biological transformation mechanisms. Section 2.2 reviews the efficiency of natural attenuation in destructive removal of different organic contaminants. In Section 2.3, the guidelines and principles associated with application of monitored natural attenuation (MNA) as a remediation technique are explained. Section 2.4 reviews the analytical and numerical models that are typically used to simulate the natural attenuation of organic compounds. In Section 2.5, uncertainty management in natural attenuation is considered and a review of literature on stochastic modeling of natural attenuation and on inverse modeling and parameter estimation in groundwater flow and mass transport are implemented. The parameterization for modeling natural attenuation of PHCs that is used in all subsequent modeling work is presented in Section 2.6.

In Chapter 3, geostatistical modeling with CPT-UVIF data is considered. Section 3.1 presents a cases study where categorical and continuous variables are defined based on raw CPT data and geostatistical techniques are used to construct a probabilistic model for the geology of a real contaminated site. In section 3.2, an explicit boundary modeling approach (distance function approach) is developed and used together with geostatistical techniques to develop a probabilistic model for 3D distribution of residual NAPL characterized by high-resolution data-acquisition instruments (CPT-UVIF). The effects of soil texture and distance to groundwater table (smear zone effect) are incorporated in the model through geostatistical data integration techniques. Finally, a cross-validation exercise is implemented to evaluate the performance of the geostatistical techniques and to assess the value of secondary data sources in improving the predictive ability for spatial distribution of residual NAPL.

Chapter 4 develops and verifies the groundwater flow and mass transport simulators that subsequently act as a transfer function for parameter estimation. In Section 4.1, development of a code for steady-state groundwater flow based on the finite volume approach with a conjugate gradient solver is presented and is verified by comparing its outcomes for to those of MODFLOW (McDonald and Harbaugh 1988). In Section 4.2, a particle tracking code based on semi-analytical method of Pollock (1988) is developed and is verified against the particle tracking package MODPATH. In Section 4.3, a reactive contaminant transport simulator is developed based on the Method of Characteristics (MOC) (Konikow and Bredehoef 1978). Rate-limited NAPL dissolution, advection, dispersion and first-order biodegradation are the mechanisms that are incorporated in the governing equation and the subsequent numerical discretization. The code is verified against MT3DMS (Zheng and Wang 1999) for homogeneous and heterogeneous conditions.

Chapter 5 implements stochastic parameter estimation to characterize the uncertainty in the first-order biodegradation rate constant and dissolution rate constant, under uncertain source geometry and hydraulic conductivity field. In order to avoid computational burden (while the value of coupling for estimation of the parameters is unknown) and potential instabilities, a decoupled inverse problem is adapted. As the first step of the decoupled inverse problem, the sequential self-calibration approach (Gomez-Hernandez et al. 1997) is reviewed and the relationship between the errors in hydraulic head observations and uncertainty in hydraulic conductivity realizations are studied (Section 5.1). A simple gradient-based inverse modeling approach is then developed in Section 5.2 to simultaneously estimate the values of the parameters for joint realizations of source zone geometry and hydraulic conductivity field and a sensitivity analysis is performed. In Section 5.3, the stability of the inverse problem and the issues

associated with high correlations between the parameters and existence of local minima are studied.

Chapter 6 presents a synthetic example to investigate the performance of the uncoupled Monte Carlo type inverse modeling approach in characterization of uncertainty in the dissolution rate and first-order biodegradation rate and state variables. The results of the decoupled approach are compared to those of a set of Monte Carlo simulations with uncertain parameter values derived from the literature (based on field estimation techniques). The effect of ranking and screening the calibrated realizations (based on the value of a modified form of objective function) in reducing the uncertainty in the source size and the state variables is also studied. Last, the effects of error in head and concentration observations and uncertainty about the time of release are investigated.

In Chapter 7, the uncertainty in field-estimated first-order biodegradation rate constants by centerline approaches is investigated. In Section 7.1, three widely-used field estimation techniques for first-order rate constants are reviewed. In Section 7.2, the effects of variability in the dissolution rate constant and error in head observations on the uncertainty of the estimated first-order rate is investigated in simple 1D and 2D examples.

In Chapter 8, conclusions and recommendations for future research are presented.

In Appendix A, the distance-based simulation (Scheidt et al. 2008 and Caers 2008) is implemented to investigate the possibility of expanding the set of accepted realizations (Section 5.1) by the newly-developed approach.

In Appendix B, the parameter files of the two codes that are developed for boundary modeling under uncertainty namely **mlimit2d** and **ubcalib2d**, as well as

the parameter estimation code **optkna** that is developed as an extension to the MOC mass transport simulator **snasim** are presented.

CHAPTER 2

NATURAL ATTENUATION: FUNDAMENTALS & MODELING

The extensive use of petroleum hydrocarbons (PHCs) and associated releases into the environment has resulted in widespread soil and groundwater contamination. Common sources of contamination are leaking underground storage tanks (LUST), pipelines, oil exploration activities, storage pits near production oil wells and refinery wastes. High costs of engineered clean-up systems and their disappointing performance has led to searches for alternative remediation strategies, and rethinking of remediation goals and time frames. The utilization of intrinsic attenuating capacity of soils to achieve clean-up goals has become a widely-accepted remediation scheme for PHC contaminants. This approach is quite often considered as a passive remediation scheme and is generally identified as Monitored Natural Attenuation (MNA).

In essence, the term Monitored Natural Attenuation refers to "*... reliance on natural processes to achieve site-specific remedial objectives. Where found to be a reliable remedy, monitored natural attenuation may be used within the context of carefully controlled and monitored site clean-up approach. To be considered an acceptable alternative, MNA would be expected to achieve site remedial objectives within a timeframe that is reasonable compared to that offered by other more active methods. MNA is always used in combination with 'source control' that is, removal of the source of the contamination as far as practicable ...*" (US EPA 1999a). Depending on hydrogeologic conditions, type of contaminants, and

extent and distribution of contaminants at a given site, MNA could be the preferred choice of remediation strategy.

In general, target contaminants for natural attenuation are volatile organic carbons (VOCs), semi-volatile organic carbons (SVOCs) and fuel hydrocarbons. Fuels and halogenated VOCs are commonly evaluated for natural attenuation. Pesticides also can naturally attenuate, but the process may be less effective and may be applicable to only some compounds within this group. Natural attenuation may also be appropriate for some heavy metals, where natural attenuation processes result in a change in their valence state and immobilization (FRTR 1999).

PHCs are among the most common contaminants in the environment and are present in both Non-Aqueous-Phase-Liquid (NAPL) and dissolved form. Under the right conditions, natural attenuation contributes considerably in the breakdown of the dissolved PHCs to other non-toxic end products, while the clean-up costs are much less comparing to implementation of active remediation schemes. However, natural attenuation is not expected to remediate NAPL (US EPA 1999b).

It should be noted that reliance on natural attenuation is not a 'no-action' approach to site clean-up. A careful study of the site conditions is necessary to estimate the effectiveness of natural processes in reducing contaminant concentrations over time. Data collection and analysis should be used in conjunction with risk assessment to verify the effectiveness of natural processes before it is perceived as the preferred remediation strategy (Ellis and Golder 1997). In fact, evaluation of the viability of this remediation technique in site clean-ups involves determination of ongoing natural attenuation processes and estimation of results of the processes in the future. Thus, development of a long-term monitoring plan is required for every contaminated site with MNA as its

remediation strategy. This monitoring plan is to be used to (1) decide on the effectiveness of the natural attenuation processes in achieving site clean-up goals, (2) record any changes in conditions affecting natural attenuation processes, and (3) trigger other active treatment technologies, if needed (US EPA 1999b).

In this chapter, first, the most important mechanisms contributing to natural attenuation of organic contaminants are briefly reviewed. Next, natural attenuation of different organic contaminants is studied and the applicability of MNA as a remediation strategy is evaluated. Advantages and limitations of the MNA as a remediation strategy are then briefly discussed. A large number of studies related to modeling fate and transport of organic contaminants have been reported in the literature. These may be grouped into two major categories namely analytical and numerical approaches and both are subsequently reviewed in this chapter. The works related to uncertainty management for natural attenuation as well as parameter estimation techniques in general hydrogeology literature are reviewed next in this chapter. As stated in Chapter 1, the main objective of this thesis is the development of a better understanding of existing uncertainty in natural attenuation modeling by performing stochastic parameter estimation. Therefore, the parameterization that is used in subsequent parameter estimation is presented in the last section of this chapter.

2.1 Review of dominant mechanisms

Reduction in concentration of organic contaminants in the subsurface is due to combination of transport mechanisms, phase transfers and transformation mechanisms. Figure 2-1 shows a conceptual illustration of some of the most important natural attenuation processes that affect the fate of PHCs in aquifers. Among transport mechanisms, advection, dilution, and hydrodynamic dispersion are the most important processes that affect PHC natural attenuation in nature. Sorption and volatilization are the most important phase transfer mechanisms. Chemical transformation and biodegradation are the important transformation mechanisms, which are the major processes in destructive removal of organic contaminants (Wiedemeier et al. 1999). The transport, phase transfer and transformation mechanisms are briefly reviewed in the subsequent paragraphs.

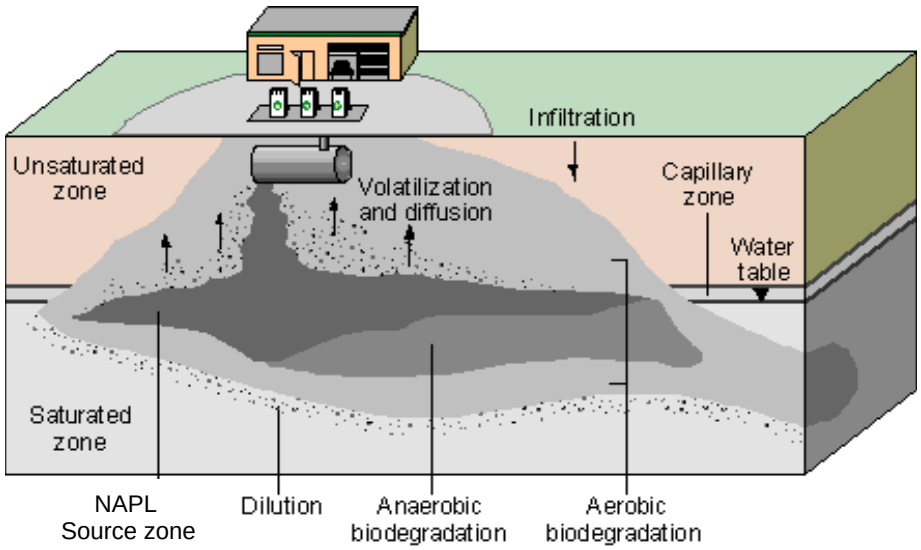


Figure 2-1: Conceptual illustration of the most important natural attenuation processes associated with PHCs (after Bekins and Rittmann 2001)

2.1.1. Transport mechanisms: Advection

Movement of solutes by bulk groundwater movement is defined as advection. Advective transport of solutes can be illustrated by considering a solute that moves with the same velocity as groundwater and does not react chemically or biologically in the subsurface. The rate of movement of solutes through the subsurface is quantified by seepage velocity and depends on aquifer properties including hydraulic conductivity, effective porosity and hydraulic head gradient (Wiedemeier et al. 1995).

2.1.2. Transport mechanisms: Dilution

Dilution (recharge) is defined as infiltration of water into groundwater system via discharge from surface water bodies or precipitation. This additional water will dilute the contaminants and also may provide an influx of electron acceptors to the system, which will affect the geochemical balance and may increase the rate of biodegradation (Wiedemeier et al. 1995). Infiltration of precipitated water through the unsaturated zone may introduce additional electron acceptors as well as dissolved oxygen to the system, and also dissolved organic carbon as an electron donor. In the case of surface water, a similar scenario may be observed, and the water entering the groundwater will dilute the plume of contaminant, and add electron acceptors and possible electron donors to the groundwater system (Suthersan 2002). The simplistic approach adopted for contaminant attenuation through dilution assumes that dilution of the original contaminant mass will ultimately deliver the total pollutant load over some unidentified time interval.

2.1.3. Transport mechanisms: Hydrodynamic dispersion

Lateral spreading of the contaminant mass in groundwater is mainly due to hydrodynamic dispersion, which results in reduced contaminant concentrations. This lateral spread also introduces contaminants into previously pristine areas of

groundwater system. There are two components of hydrodynamic dispersion: mechanical dispersion and molecular diffusion (Fetter 1993).

Mechanical dispersion occurs as a result of local variations in velocity and tortuous flow paths. Variations of rate and direction of transport velocities are due to aquifer characteristics, such as heterogeneities in small-scale hydraulic conductivity and porosity of surrounding soils. The component of hydrodynamic dispersion which is contributed by mechanical dispersion expressed as a function of seepage velocity. Molecular diffusion takes place as a result of the concentration gradients created within the zones of contamination (Suthersan 2002). It is significant only when the groundwater velocities are low. The diffusive flux of a dissolved contaminant, at steady-state, can be described by Fick's first law.

2.1.4. Phase transfer mechanisms: Sorption

The relevant contaminant phases in the subsurface include dissolved in groundwater, sorbed to soil grains, the immiscible phase, and the gas phase (for volatile compounds). Phase transfers can increase or decrease the contaminant concentration within the groundwater plume depending on the transfer mechanisms involved, the nature of contaminant and geochemistry of the aquifer system. Sorption is a result of attraction of an aqueous species to the surface of a solid. The underlying principle behind this attraction results from some form of bonding between the contaminant and adsorption receptor sites on the solid. The degree of sorption is dependent on the composition of the aquifer matrix including organic carbon content, clay mineral content, bulk density, specific surface area, and total porosity. Contaminant properties are also important, specifically solubility and the octanol-water partitioning coefficient (Wiedemeier et al. 1995).

Many contaminants, including chlorinated solvents, BTEX and dissolved metals, are removed from the solution by sorption onto the aquifer matrix, thus slowing the movement of contaminants. This slowing of contaminant transport is called retardation of the contaminant relative to the average seepage velocity of groundwater and results in a reduction in dissolved organic concentrations in groundwater (Bennett and Zheng 1995). Sorption can also influence the relative importance of volatilization and biodegradation. For example, as solute concentrations decrease due to other factors such as biodegradation and dilution, the amount of contaminant desorbing and reentering the solution will probably increase. However, for some compounds the rates of desorption may be so slow that the adsorbed mass may be considered as permanent residual within the time scale of interest (Suthersan 2002). If the sorption mechanisms do not permit easy desorption or removal of the sorbed pollutants, retention of the pollutants occurs (Yong and Mulligan 2004). If this is the case, decreasing in concentrations of contaminants occurs as transport of the pulse continues away from the source.

2.1.5. Phase transfer mechanisms: Volatilization

Volatilization is a non-reactive mechanism which removes contaminant mass from soil and groundwater. Volatilization of a contaminant into the gas phase depends on the contaminant vapor pressure and Henry's law constant (Wiedemeier et al. 1999). Other factors affecting the volatilization of contaminants from groundwater include contaminant concentration, the change in concentration with depth, diffusion coefficient of the compound, temperature, and sorption. Volatilization itself does not destroy contaminant or permanently immobilize it (Suthersan 2002). Volatilized contaminants can biodegrade in some circumstances but also can re-dissolve in infiltrating groundwater or be transported to the surface, where they may be broken down by sunlight. BTEX has the highest volatility of aromatic hydrocarbons. The high solubility and

volatility of BTEX account for the mobility of these compounds and their ability to dissolve in groundwater.

2.1.6. Transformation mechanisms: Chemical transformations

Chemical transformations degrade contaminants without requiring the presence of microorganisms. However, only halogenated compounds undergo these reactions in groundwater.

2.1.7. Transformation mechanisms: Biological transformations

Biological transformation (biodegradation) is a process in which naturally occurring microorganisms break down target substances, such as fuels and chlorinated solvents, into often less toxic or non-toxic substances. Certain microorganisms degrade components of fuels or chlorinated solvents found in the subsurface. The ability of microorganisms to metabolize PHCs, or use nutrients depends on the redox state in the groundwater, and different microorganisms have evolved to take advantage of varying conditions. In most organisms, including bacteria, the metabolic process requires the exchange of oxygen and carbon. Biodegradation can occur in the presence of oxygen, aerobic conditions, or without oxygen, anaerobic conditions.

Biodegradation is important because many significant components of petroleum hydrocarbons such as BTEX and some PAHs can be biodegraded by microorganisms indigenous to the subsurface. Commonly, PHC constituents mobile in the environment, except MTBE, are readily biodegradable. Under the right conditions, microorganisms can cause or assist chemical reactions that change the form of the contaminants so that little or no health risk remains. There are three biodegradation processes which change the form of contaminants. First, the contaminant is used by the microbe as the primary energy source. Second, the contaminant can be used to transfer energy in which case it is reduced; and third,

when biodegradation occurs in response to a secondary reaction between the contaminant and an enzyme produced during an unrelated action, which is termed cometabolism. For fuel hydrocarbons containing BTEX, the first process is dominant.

In the context of the first biodegradation process, hydrocarbons such as BTEX are used as a substrate (food source) and are oxidized as an electron donor, while in a sequence of reactions, one or more of a series of electron acceptors are reduced. In this sequence, oxygen is readily consumed as an electron acceptor (aerobic metabolism), which generates the highest energy yield. Once oxygen is depleted, other electron acceptors are used in the following order of preference (anaerobic metabolism): nitrate, manganese (IV), iron (III), sulfate and carbon dioxide (Downey et al. 1999; Langmuir 1997).

In the context of the second biodegradation process, the organic compound may also be used as an electron acceptor to aid respiration and transfer energy. Basically, all living organisms respire. That is, they use organic substances and other nutrients by breaking them down into simpler products. In the absence of oxygen, microorganisms often use chlorinated compounds to aid in respiration, not as a source of food. This is done through an electron transfer process (redox reaction). In the case where the carbon in the contaminant is the food source, the contaminant is an electron donor. In the case where the food is obtained from another source, the contaminant sometimes aids this transfer by accepting electrons that are released through respiration.

The third biological process which may result in degradation of mainly chlorinated solvents is cometabolism. When a chlorinated solvent is biodegraded through co-metabolism, it does not serve as primary food source or an electron acceptor. Cometabolism is the process in which biodegradation takes place as a

result of a secondary reaction, initiated for example, by enzymes produced by the metabolism of methane. Cometabolism has only been documented under aerobic conditions (Strauss 1998). In presence of organic substrate and dissolved oxygen, micro-organisms capable of aerobic metabolism will predominate over anaerobic forms. However, dissolved oxygen is rapidly consumed in the interior of contaminants plumes (Figure 2-2), converting these areas into anoxic (low oxygen) zones. Under these conditions anaerobic bacteria begin to utilize other electron acceptors to metabolize dissolved hydrocarbons. The main factors affecting the utilization of various electron acceptors include: (1) relative biochemical energy provided by the reaction, (2) availability of individual or specific electron acceptors at a particular site, and (3) kinetics of the microbial reaction associated with different electron acceptors (Rifai et al. 1997).

The transfer of electrons during the redox reaction releases energy that is utilized for cell maintenance and growth. The biochemical energy associated with alternative degradation pathways can be represented by the redox potential of the alternative electron acceptors: the more positive the redox potential, the more energetically favorable is the reaction utilizing that electron acceptor. Figure 2-3 presents the utilization sequence of aerobic and anaerobic electron acceptors during the biodegradation of fuel hydrocarbons. Based solely on thermodynamic considerations, the most energetically preferred reaction should proceed in the plume until all of the required electron acceptor is depleted. At that point, the next most-preferred reaction should begin and will continue until that electron acceptor is gone, leading to a pattern where preferred electron acceptors are consumed one at a time, in sequence (Langmuir 1997; Rifai et al. 1997; Rifai et al. 2000).

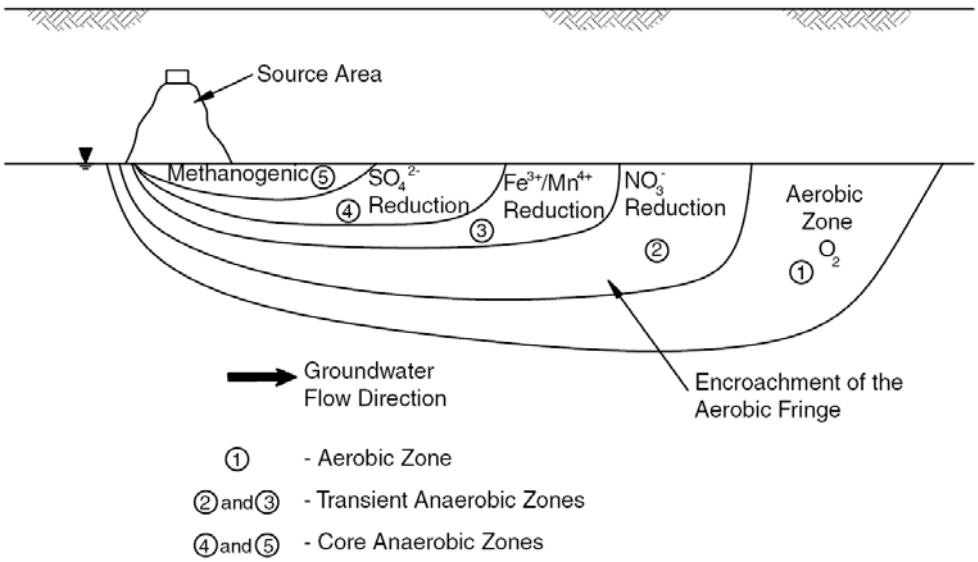


Figure 2-2: Schematic representation of terminal electron acceptors in advancing BTEX plume in a groundwater system (after Suthersan 2002)

Electron acceptor (1)	Type of reaction (2)	Metabolic by-products (3)	Redox potential (pH = 7, in mV) (4)	Reaction preference (5)
Oxygen	Aerobic	CO_2	+820	Most preferred
Nitrate	Anaerobic	N_2, CO_2	+740	⇓
Ferric iron	Anaerobic	Fe^{2+}	−50	⇓
Sulfate	Anaerobic	H_2S	−220	⇓
Carbon dioxide	Anaerobic	CH_4	−240	Least preferred

Figure 2-3: Sequential electron acceptors for BTEX (after Rifai et al. 2000)

Microorganisms are most effective at degrading low to moderate concentrations of contaminants. High concentrations and very low concentrations of contaminants may not be biodegradable. Contaminants in the NAPL phase are not effectively degraded by microorganisms (US EPA 1999a). As contaminants biodegrade, the products of the degradation process may or may not be less harmful than the original contaminants. Therefore, it is important to investigate

the site processes carefully to be sure that biodegradation is making the site safer. Fortunately, petroleum hydrocarbons appear to degrade to less harmful products in almost all cases. Also, under some conditions, the microbial activity involved in degrading the contaminants could cause mobilization of certain materials such as manganese or arsenic which could cause environmental problems. Monitoring for these potential problems is necessary (US EPA 1999b).

2.2 Natural Attenuation of organic contaminants

Application of natural attenuation as a remediation alternative for BTEX has recently increased dramatically. More recently, natural attenuation has been proposed for chlorinated solvents, heavy metals, radionuclides, and other contaminants for which the scientific understanding and field experience are much less robust (Bekins et al. 2001; NRC 2000). It is known that the components of PHCs are alkanes, cycloalkanes, aromatics, PAHs, asphaltenes and resins. In this section, applicability of MNA for various organic contaminants such as Aliphatics, BTEX, Chlorinated solvents, PAHs, MTBE, and Halogenated Aromatic Compounds (e.g. PCBs) is studied.

2.2.1. Aliphatic compounds: Alkanes, Alkenes and Cycloalkanes

Aliphatic compounds such as alkanes, alkenes and cycloalkanes are known by their branched straight chain molecular structure. Among these, low molecular-weight alkanes are most easily degraded by micro-organisms. For these compounds, as the chain length increases solubility as well as biodegradation rates decreases. Most alkanes present in PHCs are branched, which are not as biodegradable as unbranched ones. In transformation of alkanes, alcohols, aldehydes and fatty acids are sequentially formed (Higgins and Gilbert 1978).

Alkenes have not been extensively studied for biodegradation. For these compounds, it has been observed that having a double bond on the first carbon (namely 1-alkenes) may facilitate degradation rather than having the double bond at the other positions (Pitter and Chudoba 1990).

Due to their cyclic structure, cycloalkanes are not as biodegradable as alkanes. Their biodegradability also decreases as the number of rings increases in their

molecular structure. Oxidation of these compounds results in formation of cyclic alcohol with a ketone as final product (Pitter and Chudoba 1990).

2.2.2. Benzene, Toluene, Ethylbenzene and Xylenes (BTEX)

Natural attenuation processes, and in particular biological degradation, are known to be best documented for PHCs at petroleum fuel spill sites. Under proper site conditions, PHCs such as BTEX may degrade by biological transformations and produce non-toxic end products (US EPA 1999a). BTEX compounds are volatile, highly mobile and known to have adverse health impacts. If microbial activity is rapid enough, the dissolved BTEX contaminant plume will stop expanding, and regulatory standards for concentrations in groundwater may be met (Yong and Mulligan 2004). After degradation of a dissolved BTEX plume, a residue consisting of heavier petroleum hydrocarbons of relatively low solubility and volatility will typically be left behind in the original source area. Although this residual contamination may have relatively low potential for further migration, it still may pose a threat to human health or the environment either from direct contact with soils in the source area or by continuing to slowly leach contaminants to groundwater (US EPA 1999a). Thus, design and implementation of source control measures is always necessary to be in conjunction with MNA as the remediation scheme.

2.2.3. Chlorinated solvents

The potential for anaerobic biotransformation of chlorinated aliphatic hydrocarbons was first demonstrated during the early 1980s (Bouner et al. 1981). Later studies have shown that these compounds, which are usually denser than water and are referred to as DNAPLs, can biotransform under a variety of environmental conditions in the absence of oxygen. DNAPLs tend to sink through the groundwater column toward the bottom of the aquifer; and usually they are difficult to locate, delineate, and remediate even with active measures. In the

subsurface, these compounds represent source materials that can continue to contaminate groundwater for decades or longer (US EPA 1999a). As mentioned before, the most common anaerobic process for degrading chlorinated compounds is an electron transfer process called reductive dechlorination (Strauss 1998). Where excess organic material is available to serve as an electron donor and biogeochemical conditions support a reducing environment, biodegradation of these compounds is likely to occur (Suthersan 2002).

2.2.4. Polycyclic Aromatic Hydrocarbons (PAHs)

PAHs are compounds that have multiple rings in their molecular structure. Their molecular structure is complex and their water solubility is low. Petroleum refining activity, coke production process and wood preservation industry are the main sources of PAHs, some of which are known to be carcinogenic. PAHs biodegrade very slowly and as the number of rings increases, the compound becomes more difficult to degrade, because of decrease in solubility and volatility and increased sorption (National Research Council 2000; Yong and Mulligan 2004).

2.2.5. MTBE

MTBE (methyl tertiary-butyl ether) is a chemical compound that is manufactured by the chemical reaction of methanol and isobutylene. MTBE is produced in very large quantities and has been almost exclusively used as a fuel additive in motor gasoline. In the U.S., it is one of a group of chemicals commonly known as "oxygenates" because they raise the oxygen content of gasoline. At room temperature, MTBE is a volatile, flammable and colorless liquid that dissolves rather easily in water. MTBE leaking from Underground Storage Tanks may reach to and contaminate groundwater. MTBE contaminated groundwater is not drinkable, due to offensive taste and odor. MTBE is highly resistant to

biodegradation and there is also a dispute on the mechanism of its natural attenuation (Alvarez and Illman, 2006).

2.2.5. Halogenated Aromatic Compounds

PCBs, pesticides, plasticizers are examples of Halogenated Aromatic Compounds. These stable toxic compounds compose of benzene rings with halogen atoms attached to them and pose serious risks to the environment. The number and position of halogens affect the rate and mechanism of biodegradation. There are three major mechanisms for their transformation: hydrolysis, reductive dehalogenation, and oxidation (Yong and Mulligan 2004). Even though the use of Polychlorinated biphenyls (PCBs) is banned, in many cases these chemicals are present in sediments and aquatic environments. Similar to other halogenated aromatics, these compounds have high toxicity, stability and bioaccumulation potential. According to a number of studies, natural attenuation of PCBs in anaerobic sediments is catalyzed by bacteria in aquatic sediment and is taking place in high rates. This significant biodegradation rate is a result of dechlorination mechanisms, which converts the more highly chlorinated molecules to less chlorinated products, and results in formation of compounds with lower toxicity and lower accumulation potential (Suthersan 2002; Yong and Mulligan 2004).

2.3 MNA as a remediation scheme

In general, it is a prerequisite to show whether or not remediation by natural attenuation is appropriate for a particular site. In other words, natural attenuation should not be considered as a presumptive remedy. An appropriate site characterization includes assessment of potential risk, evaluation of the essence of source area control, and evaluation of potential effectiveness of the chosen remediation technique. Implementation of MNA in a particular site also needs demonstration of remedial progress through monitoring and assessment of samples taken. Monitoring should be conducted until it has been demonstrated that natural attenuation meets the remedial goals. A number of steps that have to be taken when the MNA is considered as a remedial alternative are briefly reviewed in this section.

2.3.1. Review of the available site-specific data and development of a site conceptual model (SCM)

For natural attenuation to be effective at a particular site, site-specific hydrogeology, geochemistry and contaminant properties must be verified. There are a number of requirements that must be met for a particular site for natural attenuation to be successful. These include minimal risk of contaminants coming into contact with receptors and presence of enough microorganisms, nutrients and terminal electron acceptors. According to Wiedemeier et al (1995), information to be obtained during data review includes nature, extent and magnitude of contamination, geologic and hydrogeologic data and locations of potential receptors. If it can be shown that intrinsic remediation is a potential remedial option, all future site characterization activities should include collecting the data necessary to support this remedial alternative. In fact, an extensive review of the site data results in the development of preliminary site conceptual model (SCM). The preliminary SCM demonstrates direction and velocity of groundwater flow,

location and distribution of contaminants at the source zone as well as in the dissolved plume, natural processes which theoretically affect the contaminants, and potential receptors. It also helps to determine any data shortcomings and to allocate additional data collection points. Based on the preliminary SCM, it can be determined whether natural attenuation is a major component in clean-up or other remedial measures should be considered.

Development of an appropriate SCM is an important step in verifying the viability of natural attenuation as a remedial alternative. It supports both the next step in site characterization and the subsequent modeling. Given the importance of development of a good SCM, a full chapter in this work (Chapter 3) is dedicated to development of a SCM for a PHC impacted contaminated site. The focus of the chapter will be in particular on characterization of soil texture and contaminant source distribution.

2.3.2. Collecting additional data in support of MNA

In case intrinsic remediation is deemed to be an effective option, further site characterization activities must be implemented to support this remedial option. In fact, to confirm that natural attenuation will continue to work, monitoring data must be obtained and linked to the conceptual model, and "footprints" of the mechanisms involved must be studied. These footprints include study of variations in concentration of contaminants, terminal electron acceptors, and donors, as well as products of biodegradation processes (such as Cl^- or Fe^{2+}) (Suthersan 2002). All these additional site characterizations help to evaluate the potential efficiency of MNA as a remedial option. In this regard, the rate of attenuation processes and the required time frame to achieve remediation goals must be estimated. In order to do such an evaluation, a three-tiered approach is usually practiced. According to US EPA's Directive for use of MNA at Superfund

and Underground Storage Tank sites, these three are lines of evidence in support of natural attenuation for a particular site (US EPA 1999a).

The first line of evidence is provided by historical groundwater and/or soil chemistry data that show reduction or stabilization in plume geometry and contaminant mass or concentration over time at appropriate sampling points. The second line of evidence involves recording hydrogeologic and geochemical data that can be used to demonstrate indirectly the type(s) of natural attenuation processes active at the site. This includes measuring the depletion of electron acceptors (dissolved oxygen, nitrate and sulfate), production of metabolic by-products (Mn(II), Fe(II) and methane), and presence of biodegradation metabolites or intermediates. The third line of evidence is microbiological data from field or microcosm studies, which is optional and directly demonstrates the occurrence of a particular natural attenuation process at the site and its ability to degrade the contaminants of concern (US EPA 1999a).

During the detailed site characterizations, there are often two objectives. The first is collection of data in support of ongoing natural attenuation processes. The second is collection of sufficient site-specific data, to be used in fate and transport modeling, to estimate future distribution and concentration of contaminants. According to Wiedemeier et al. (1995), in the context of detailed site characterizations in support of natural attenuation, the following parameters need to be determined:

- Spatial distribution and type of contaminants in groundwater;
- Delineation of contaminant source;
- Geochemistry of the aquifer;
- Regional hydrogeology: drinking water aquifers and regional confining units

- Site-specific hydrogeology, including local drinking water aquifers, location of water wells, patterns of current and future aquifer use, stratifications and preferential flow paths, grain size distribution, hydraulic conductivities, local surface water features, and local groundwater recharges and discharges;
- Potential pathways and receptors;

As mentioned above, site-specific studies, which directly demonstrate biological activity, may also be implemented. These studies may include techniques such as field dehydorgenase test or microcosm studies (Wiedemeier et al. 1995).

2.3.3. Summarizing the available data and refining the SCM

After implementation of detailed site characterization, the collected data should be incorporated into the SCM and should be used in quantifying advection, mechanical dispersion, dilution, sorption and biodegradation at the site. These calculations can give direct measures of intrinsic capacity of the groundwater system to attenuate the contaminants and support the conclusion that natural attenuation is actually taking place. These measures are also used in modeling natural attenuation over time (Wiedemeier et al. 1998).

In the context of refinement of the SCM, newly collected data are used to update the conceptual model. As a result, the SCM will give an accurate demonstration of hydrogeology, and nature and extent of contamination. Conceptual model refinement includes providing geologic logs, hydrogeologic sections, contour maps of contamination, contour maps of concentrations of electron acceptors, etc (Wiedemeier et al. 1995; Wiedemeier et al. 1998).

2.3.4. Simulating natural attenuation using solute fate and transport models

Natural attenuation models are typically applied for three different reasons (Rifai and Rittaler 2005):

- To determine if a plume is going to get larger;
- To determine how long will a plume be there;
- To develop a better understanding of the important natural attenuation processes at a given site.

As will be discussed later in further detail, both analytical and numerical models can be used to simulate the conditions at a contaminated site. Analytical models can give the exact solutions for very simplified scenarios. On the other hand, numerical models give approximate solutions; but they are able to handle complicated site hydrogeology. A key point in utilizing these models is incorporating the natural attenuation site data into the model to calibrate and validate the model for site-specific conditions.

When using these models, there are a number of issues that should be taken into account (NRC 2000). The first issue is related to the site-specific data used in development of the model. In this regard, the present and future site conditions must be consistent with the parameters and boundary conditions used. The second issue arises from over-fitting the model to data collected from the site. Models should not be forced if they do not fit the data. Instead, modifications in site conceptual model must be made or other models with different underlying assumptions should be used. The third issue arises from overestimating ability of models and reliability of their end results. It should be noted that uncertainties in data collected together with uncertainties in assumptions made during the modeling process, methods of verification and validation, and complexity of the site will affect the end results.

Development of a well-established conceptual model, using high quality data collected from the site, is a key to natural attenuation modeling. There are various packages available for predictive modeling of natural attenuation. Some of the most recognized ones are BIOSCREEN (Newell et al. 1996), BIOCHLOR (Aziz et al. 1999), BIOPLUME (Rifai et al. 2000), SEAM3D (Waddill and Widdowson 1998) and NAS (Mendez et al. 2004). Selection of a model to simulate natural attenuation at a site depends on underlying assumptions, limitations and site conditions such as hydrology and hydrogeology of the site. An exposure pathway analysis should also be implemented after performing predictive modeling. This is performed to support natural attenuation at the site, and includes identification and study of potential receptors at points of exposure under current and future conditions. Results of modeling efforts are core to this type of analysis (Carey et al. 1998).

2.3.5. Long-term monitoring

Long-term monitoring is necessary to demonstrate that contaminant concentrations continue to decrease at a rate sufficient to ensure that they will not become a health threat or violate regulatory criteria. In fact, the need to collect biogeochemical and groundwater quality data of the highest quality to predict the natural attenuation capacity of the system is the most important aspect of monitored natural attenuation at a site.

A long-term monitoring plan consists of placement of monitoring wells and development of a program of sampling, testing and analyzing. It is developed based on the site characterization data, the results of solute fate and transport modeling, and the results of exposure pathway analysis (Wiedemeier et al. 1995). There are two types of monitoring wells in a long-term monitoring plan. Performance-monitoring wells (or long-term monitoring wells) are intended to monitor behavior of the plume. They are used to determine the effectiveness of

the natural attenuation processes. They also check the parameters affecting these processes. Point-of-compliance wells (or contingency wells) are intended to detect movements of plume outside its perimeter and to prompt an action to manage the risk associated with such expansion. Usually there are also numbers of ambient monitoring wells that are located upgradient of the plume. Final number and placement of performance monitoring wells are determined based on regulatory requirements, and are dependent on factors such as type of the pollutants and source location, size and features of existing plume, hydrogeology of the site and results of modeling (Yong and Mulligan 2004). Contingency wells should be located approximately 150 m downgradient of the edge of the plume or the distance traveled by groundwater in two years, whichever is greater.

The frequency of monitoring should be enough to detect in a timely manner any changes in plume degradation and migration pattern. At a minimum, the monitoring program should be able to determine the attenuation rates and variations of attenuation rates in time (US EPA 1999b). It may be frequent at the beginning and then decreasing in frequency as the progress of natural attenuation becomes slower. But, it should be flexible to adjust to any changes that may occur. The sampling plan should be continued until the goals of the natural attenuation process have been safely met (Surampalli and Banerji 2002). In conjunction with well placements, a groundwater sampling scheme must be planned. As discussed before, in support of progress of natural attenuation in addition to decreases in concentration of organic chemical pollutants, other lines of evidence should also be demonstrated. Variations in concentrations of electron acceptors and carbon dioxide and methane production are these lines of evidence. (Wiedemeier et al. 1998).

2.4 Modeling natural attenuation

MNA is appropriate to use as a remediation strategy when "... *it will meet site remediation objectives within a timeframe that is reasonable compared to that offered by other methods*" (U.S. EPA, 1999a). Thus, estimating the amount of time required for natural attenuation processes to lower contaminant concentrations to given regulatory goals is needed when assessing MNA as a remedial alternative. As pointed out by Chapelle et al. (2003) this 'time-of-remediation' (TOR) problem is formulated as three interactive components: (1) estimation of the length of the contaminant plume, once it has achieved a steady-state configuration from a source area of constant contaminant concentration (distance of stabilization or DOS), (2) estimating the time required for a plume to shrink to a smaller regulatory accepted configuration when source-area contaminant concentrations are lowered by engineered methods (time of stabilization or TOS), and (3) estimating the time needed for Non-Aqueous Phase Liquid (NAPL) contaminants to dissolve, disperse and biodegrade below predetermined levels in contaminant source areas (time of dissolution or TOD). Conceptualization of the TOR problem is shown in Figure 2-4. Estimation of DOS is also useful in determination of remediation goals in source area as well. In fact, concentrations in a source area that will preclude contaminant transport to nearby sensitive receptors can be estimated. The need for quantitative assessment of the TOR problem has necessitated the development of natural attenuation analytical and numerical models to simulate advection, dispersion, sorption, and biodegradation. Some of these models are briefly reviewed in this chapter.

Most fate and transport models are based on mass balances that incorporate processes such as advection, dispersion, chemical reactions and biodegradation of target contaminants as a function of time. Such an expression can be represented in a compact form by:

$$\frac{\partial C}{\partial t} = \frac{1}{R_f}\left[D\nabla^2 C - v\nabla C - \lambda C\right] \quad [2.1]$$

where, C is the concentration of a dissolved contaminant at a given point in the aquifer at time t, R_f is the retardation factor for instantaneous, linear sorption, D is the hydrodynamic dispersion coefficient tensor, v is the groundwater velocity, λ is the first-order decay coefficient , and ∇ is the divergence operator (Alvarez and Illman 2006).

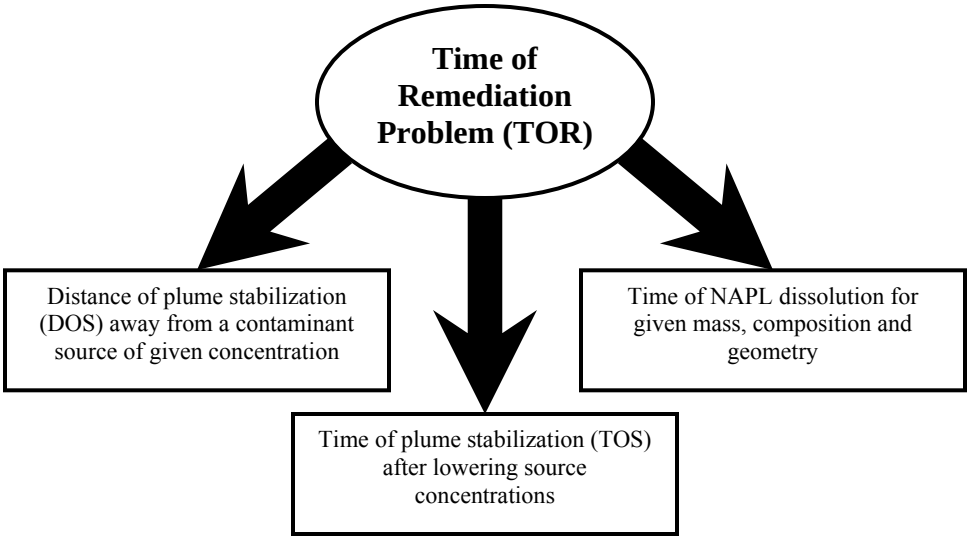


Figure 2-4: Conceptualization of the TOR problem (after Chapelle et al. 2003)

Physically, this expression means that the change in contaminant concentration with time is a function of the amount of plume spreading in all directions (dispersive transport term), the amount of contaminant that migrates with groundwater in bulk flow (advective transport term), and the amount of contaminant that is degraded, and the amount of contaminant that is sorbed to soil grains. The decay coefficient λ represents the processes such as hydrolysis, chemical redox reactions, volatilization and biodegradation. However, most studies have shown that the decay coefficient for BTEX compounds, which are the focus of this study, is primarily due to biodegradation (Chiang et al. 1989, Rifai et al. 1995).

The solution to Equation [2.1] can be obtained through a variety of analytical and numerical techniques. However, regardless of the solution technique used, the material properties of the domain and the forcing functions consisting of initial and boundary conditions and sink and source terms must be specified to obtain a unique solution to a flow and transport problem. Of course, in a steady-state problem, an initial condition is not required. For transient problems initial conditions must be specified as well as boundary conditions. The initial condition is used to specify the initial distribution of a contaminant within the model domain and can be represented by:

$$C(x, y, z, 0) = C_0(x, y, z) \qquad [2.2]$$

which states that at time zero, the concentration of contaminant throughout the modeling domain has an initial concentration distribution of $C_0(x,y,z)$.

Boundary conditions are used to specify fixed values of concentrations at boundary locations, the gradient at the physical boundaries or a mixed of the two. These three types of boundary conditions are formally known as Dirichlet, Neumann and Cauchy boundary conditions, respectively (Zheng and Bennett 1995). The Dirichlet boundary condition specifies the concentration along the boundary over a specific time period:

$$C(x, y, z, t) = C_{\Gamma_1}(x, y, z) \qquad \text{on } \Gamma_1 \quad \text{for } t > 0 \qquad [2.3]$$

where, $C_{\Gamma_1}(x, y, z)$ is the specified concentration along the boundary Γ_1. Neumann boundary condition specifies the concentration gradient normal to the boundary:

$$-D_{ij}\frac{\partial C}{\partial x_j} = f_i(x, y, z) \qquad \text{on } \Gamma_2 \quad \text{for } t > 0 \qquad [2.4]$$

where, $f_i(x,y,z)$ is a function representing the (dispersive) flux normal to the boundary Γ_2. For the Cauchy boundary condition, both the concentration along the boundary and concentration gradient across the boundary are specified, which

implies that both the dispersive and advective fluxes across that boundary are specified:

$$-D_{ij}\frac{\partial C}{\partial x_j}+v_i C = g_i(x,y,z) \qquad \text{on } \Gamma_3 \quad \text{for } t>0 \qquad [2.5]$$

where, $g_i(x,y,z)$ is a function representing the total flux (dispersive and advective) normal to the boundary Γ_3.

Source and sink terms appear in both the governing flow and transport equations. For the flow problem, the source/sink term represents a mechanism in which water is added or removed from the system. The source and sink terms are broadly classified as point or areal. These terms in the governing transport equation represent solute mass that is dissolved in groundwater and is added to or removed from the system through source and sink terms appearing in the governing flow equation. Examples of point sources and sinks include wells that are recharging or discharging, buried drains, and localized recharge resulting from contaminant spill events. Examples of physical features that are modeled as areal sources or sinks include recharge that takes place over a large area due to precipitation or irrigation, evapotranspiration, impoundments of contaminants such as sewage lagoons and mill tailings ponds (Alvarez and Illman 2006). It has been shown that the source term is an important element in modeling natural attenuation and must be addressed appropriately (McNab and Dooher 1998, Ling and Rifai 2007, Miles et al. 2008). The parameterization used in this work is presented in Section 2.6 and in Chapter 3 to address source term in terms of source zone geometry and loading of contaminants to groundwater and associated uncertainty.

The governing partial differential equation, the initial conditions and the boundary conditions form a “Boundary Value Problem”, which can be solved analytically or numerically. Analytical solutions are exact and are generally simple to

implement. Closed form solutions are usually available, which can be used readily to make preliminary predictions of fate and transport. However, they are generally limited to simple geometries and requires that aquifer and transport parameters be uniform throughout the modeled region. Numerical solutions on the other hand treat the boundary value problem as a system of algebraic equations or alternatively simulate transport by tracking a large number of particles in a known velocity field. The main advantage over analytical solutions is that they are flexible and can handle complex geometries. They can also incorporate spatial variation in the initial condition and both space and time variations in boundary conditions. However, numerical solutions can be prone to discretization and round off errors as well as numerical dispersion.

In the reminder of this chapter, the widely-used analytical and numerical models are briefly reviewed; and a proposed parameterization for stochastic modeling of natural attenuation is presented. A more detailed description of governing equations and numerical discretization of the mass conservation equations will be presented in Chapter 4 in the context of development of a natural attenuation simulation code.

2.4.1. Analytical models for natural attenuation

The analytical solutions are generally limited to steady, uniform flow and should not be used for groundwater flow or solute transport problems in strongly anisotropic or heterogeneous aquifers. These models additionally should not be applied under non-uniform flow conditions. Although analytical solutions can be obtained only under restrictive assumptions, the simplicity of analytical solutions makes them valuable as screening tools. In addition, analytical solutions are the primary means for testing and benchmarking numerical codes. A large number of analytical solutions for solute transport are available. The solution by Domenico (1987) which is widely used was implemented in the analytical model

BIOSCREEN (Newell et al. 1996). The solution by Domenico (1987) is designed for multi-dimensional transport of a decaying contaminant species and is given by:

$$C(x,y,z,t)=\left(\frac{C_0}{8}\right)\exp\left[\frac{x}{2\alpha_x}\left(1-\sqrt{\frac{1+4\lambda\alpha_x}{v_c}}\right)\right]\cdot\operatorname{erfc}\left[\frac{x-v_c t\sqrt{\frac{1+4\lambda\alpha_x}{v_c}}}{2\sqrt{\alpha_x v_c t}}\right] \qquad [2.6]$$
$$\cdot\left[\operatorname{erf}\left(\frac{y+Y/2}{2\sqrt{\alpha_y x}}\right)-\operatorname{erf}\left(\frac{y-Y/2}{2\sqrt{\alpha_y x}}\right)\right]\cdot\left[\operatorname{erf}\left(\frac{z+Z/2}{2\sqrt{\alpha_z x}}\right)-\operatorname{erf}\left(\frac{z-Z/2}{2\sqrt{\alpha_z x}}\right)\right]$$

where, C is contaminant concentration in space and time, C_0 is initial contaminant concentration at the source, x is distance downgradient of the source, y is distance from the centerline of the source, z is vertical distance from the groundwater surface to the measurement point, Y is the source width, Z is the source depth, α_x is longitudinal dispersivity, α_y is horizontal transverse dispersivity, α_z is vertical transverse dispersivity, λ is site-specific first-order decay coefficient, t is time, v_c is contaminant velocity in groundwater, erf(x) is error function and erfc(x) is complimentary error function. The key assumptions in the model are (Newell et al. 1996):

- The aquifer and the flow field are homogenous and isotropic;
- Groundwater velocity is fast enough to ignore the molecular diffusion;
- Adsorption is a reversible process represented by a linear isotherm.

As stated above, The Dominico solution is for multi-dimensional transport from a finite, planar, continuous source of contamination under transient condition. This solution can be easily modified to model other situations such as steady-state condition, decaying source condition, etc.

In BIOSCREEN, biodegradation is modeled by three different model types: solute transport without decay, solute transport with biodegradation modeled as a first-

order decay process (lumped-parameter approach), and solute transport with biodegradation modeled as an instantaneous biodegradation reaction with multiple electron acceptors. BIOSCREEN considers a very simple geometry for both modeling domain and the contaminant source.

BIOCHLOR (Aziz et al. 1999) is another analytical model that simulates remediation by natural attenuation of chlorinated solvents. Similar to BIOSCREEN, it has been programmed in Microsoft Excel spreadsheet environment and solves the Domenico analytical solute transport model. It can simulate 1D advection, 3D dispersion, linear sorption and biotransformation via reductive dechlorination (Section 2.2.3). Reductive dechlorination is assumed to occur under anaerobic conditions and dissolved solvent degradation is assumed to follow a sequential first-order decay process.

Analytical solutions can be very useful tools to evaluate contaminant behavior in groundwater and to characterize potential exposure pathways in risk assessment efforts. However, there are often deviations from restrictive modeling assumptions. These are commonly due to heterogeneity of the aquifer material which causes spatial variability in model parameters such as hydraulic conductivity, porosity, sorption capacity and even biodegradation kinetics. Moreover, complexities such seasonal fluctuations in groundwater table and local variations in groundwater velocity can not be incorporated into analytical models. Therefore, analytical models should be used as screening tools in initial site characterization studies and their results should be interpreted with caution.

2.4.2. Numerical models for natural attenuation

Most numerical models for solving Equation [2.1] can be classified as Eulerian, Lagrangian, or mixed Eulerian-Lagrangian (Neuman 1984, Baptista 1987).

Eulerian methods solve the transport equations on a fixed spatial grid. Primary Eulerian methods are Finite Difference Method (FDM) and Finite Element Method (FEM). They are commonly used to solve flow problems. They were also among the first methods applied to solute transport problems. They are both accurate and efficient in solving the dispersion-dominated problems (Zheng and Bennett 1995). However, they are prone to excessive numerical dispersion and artificial oscillation, when applied to advection-dominated problems. Although these types of errors may be mitigated when using fine grid spacing and small time-step sizes, the required computational effort may become prohibitive.

Lagrangian methods solve the transport equations in a deforming grid, or in deforming coordinates in a fixed grid. These methods are capable of solving transport problems with sharp concentration fronts both efficiently and accurately (Zheng and Bennett 1995). In contrast to Eulerian approaches, Lagrangian methods handle advection-dominated problems very well but have serious problems when dispersion must be solved together with advection. The random walk method (Ahlstrom et al. 1977, Prickett et al. 1981) is probably the most widely-used Lagrangian technique used in solute transport modeling. This approach handles the advection term by using a forward particle tracking technique, while associating a mass of the solute with each particle. The effect of dispersion is incorporated by adding a random displacement to the particle location after each advective movement. Sorption and decay are handled by adjusting the velocity of the particles and the mass carried by the particles (Zheng and Bennett 1995).

Mixed Eulerian-Lagrangian methods have emerged to combine the advantages of both Eulerian and Lagrangian approaches by solving the advection term with a Lagrangian method and dispersion and other terms with an Eulerian method (Neuman 1981). The forward-tracking method of characteristics (MOC), the

backward tracking modified method of characteristics (MMOC), and a hybrid of these two methods (HMOC) are among the Lagrangian-Eulerian approaches.

In the method of characteristics (MOC) (Garder et al. 1964, Konikow and Bredehoef 1978), a large number of particles are deterministically or randomly distributed across the whole modeling domain or in active areas of the modeling domain (when dynamic particles allocation is implemented). Unlike the random walk method, the particles in the MOC are not explicitly used to carry mass but used to represent the concentration field. Particles are tracked forward through the flow field using a particle tracking method to simulate the advection term. The concentrations carried by the particles are then redistributed over the FDM grid and the nodal values of cell concentrations (due to advection only) are calculated. The changes in cell concentrations due to dispersion, sink/source mixing and chemical reactions are then calculated by the standard FDM. The cell concentrations and subsequently particles concentrations are then updated to incorporate the concentration changes due to dispersion, sink/source mixing and chemical reactions, and the next step of the simulation is initiated. The MOC has been used in this work to develop a natural attenuation simulator. Implementation details of the approach are presented in Chapter 4 of this thesis. One of the most desirable features of the MOC technique is that it is virtually free of numerical dispersion caused by spatial truncation errors. The major drawback of this technique is that it can be slow and it requires a large amount of computer memory. The MOC can also lead to large mass balance discrepancies under certain situations because the discrete nature of particle tracking technique does not guarantee local mass conservation at a particular time-step. The computer memory requirement can be dramatically reduced by using a dynamic approach for particle distribution. The mass discrepancy problem can be mitigated by consistent velocity interpolation and using high-order particle tracking (Zheng and Wang 1999).

The modified method of characteristics (MMOC) (Russell and Wheeler 1983, Cheng et al. 1984) was originally developed to approximate the advection term accurately without sacrificing a great deal of computational efficiency. The MMOC is similar to MOC technique except in the treatment of the advection term. Unlike the MOC technique, which tracks a large number of moving particles forward in time and keeps track of the concentration and position of each particle, the MMOC technique places one fictitious particle at the nodal point of the fixed grid at each new time level. The particle is tracked backward to find its position at the old time level, the concentration associated with that position is used to approximate the concentration related to advection term at the new time level. The MMOC uses one particle for each finite difference cell. Therefore, the MMOC technique, used in conjunction with a simple lower-order interpolation scheme, is normally faster than the MOC technique and requires much less computer memory. The MMOC technique is also free of artificial oscillations if implemented with a lower-order interpolation scheme such as linear interpolation. However, with a lower-order interpolation scheme, the MMOC technique introduces considerable numerical dispersion, especially for advection-dominated problems. The intended use of the MMOC technique is where sharp fronts are not present. For advection-dominated problems with sharp fronts, the MOC approach is preferred (Zheng and Wang 1999).

The choice of the MOC and MMOC depends on the nature of the concentration field being modeled and computing resources available. The HMOC attempts to combine the strengths of the MOC and MMOC approaches by using and automatic adaptive scheme (Neuman 1984, Zhang and Wang 1999). The fundamental idea behind this scheme is automatic adaptation of the solution process to the nature of concentration field. When sharp concentration fronts are present, the advection term is solved by the MOC technique through the use of

moving particles distributed dynamically around each front. Away from such fronts, the advection term is solved by the MMOC technique with fictitious particles placed at the nodal points tracked directly backward in time. By selecting an appropriate criterion for controlling the switch between the MOC and MMOC schemes, the adaptive procedure can provide accurate solutions to transport problems over the entire range of Peclet numbers. The choice of the adaptive criterion, however, may not be obvious and the adaptive procedure may not lead to the optimal solution (Zheng and Wang 1999). In the next paragraphs, some of the numerical models that are commonly used to evaluate natural attenuation are briefly reviewed.

MT3D (Zheng 1990) is a modular 3D transport code used to simulate advection, dispersion and chemical reactions of one contaminant species at a time in groundwater. The flow-field for this model is explicitly generated by the widely used flow simulation code MODFLOW (McDonald and Harbaugh 1988). MT3D can simulate radioactive decay, biodegradation and linear and non-linear sorption. It contains four different solution methods: MOC, MMOC, HMOC, and FDM with forward difference scheme. MT3DMS (Zheng and Wang 1999) is an expanded version of MT3D that can simulate transport of multiple contaminant species at one time and has some other advancements such as an improved advection solver, a non-equilibrium sorption model, etc.

RT3D (Clement 1997) is a generalized multi-species version of the MT3D code. It comprises a much more detailed modeling of reactive contaminant transport as compared to MT3DMS. As with MT3D, RT3D also requires the groundwater flow code MODFLOW for computing spatial and temporal variations in groundwater head distribution. The RT3D code includes an implicit reaction solver (with seven pre-programmed reaction modules used to simulate different types of reactive contaminants) that makes the code sufficiently flexible for

simulating various types of chemical and microbial reaction kinetics. Using the pre-programmed reaction packages, instantaneous hydrocarbon biodegradation using multiple electron acceptors, kinetically limited hydrocarbon biodegradation by multiple electron acceptors, kinetically limited reaction with bacterial transport, non-equilibrium sorption, etc. can be modeled by RT3D.

The BIOPLUME III software package (Rifai et al. 1997) was developed based on the MOC approach. It models the sequential biodegradation beginning with oxygen as an electron acceptor and proceeding to nitrate, iron, sulfate and carbon dioxide based on the depletion of the preceding electron acceptor. In addition to aerobic and anaerobic biodegradation, other attenuating mechanisms such as advection, dispersion, sorption, and ion exchange are also modeled by BIOPLUME III. Three different kinetic expressions used to simulate biodegradation reactions are first-order decay, instantaneous reaction, and Monod kinetics.

The SEAM3D model (Waddill and Widdowson 1998) simulates 3D transport and sequential electron acceptor-based bioremediation in groundwater. Hydrocarbon contaminants are simulated as electron donors for microbial growth with available electron acceptors depleted sequentially. Each hydrocarbon substrate can produce a single daughter product. Biodegradation of each substrate follows Monod kinetics, modified to include the effects of electron acceptors and nutrient availability. Microbial biomass is simulated as scattered micorcollonies attached to the porous medium. One of the unique features of SEAM3D is that it treats NAPL mass as being entirely residual (or trapped) in porous medium and simulates the dissolution of NAPL into groundwater based on a concentration gradient drive between the nonaqueous phase and aqueous phase. As discussed later in this chapter and in Chapter 4, a similar approach has been implemented in natural attenuation simulator developed in this work.

More recently, natural attenuation software (NAS) (Mendez et al. 2004) has been developed which consists of a combination of analytical and numerical solute transport models. NAS is designed for application to groundwater systems consisting of porous, relatively homogenous media, and assumes that groundwater flow is uniform and unidirectional. It models advection, dispersion, sorption, NAPL dissolution and biodegradation. It also determines redox zonation, and estimates and applies varied biodegradation rates from one redox zone to the next. The interested reader is referred to Mendez et al. (2004) for capabilities and details of NAS.

2.5 Uncertainty management and parameter estimation

As discussed above, there are many studies on deterministic modeling of natural attenuation with analytical and numerical approaches; however, there are a number of works reported in the literature that deal with stochastic modeling and uncertainty management for natural attenuation of petroleum hydrocarbons. Uncertainty management related to natural attenuation has two aspects: (1) modeling uncertainty in response variables due to uncertain model parameters, and (2) characterizing the uncertainty in model parameters, using the observed response variables at a few locations. The first aspect is formally termed stochastic modeling and the second aspect is termed parameter estimation or inverse modeling. The uncertainty in the response variables due to uncertain model parameters can be quantified with the use of an analytical/numerical mass transport simulator and Monte Carlo Simulations, which is relatively straight forward. Parameter estimation (inverse modeling), however, is a challenging task that has been subject of much research in the recent years.

The main focus in this thesis is parameter uncertainty. Two different aspects of parameter uncertainty related to natural attenuation of petroleum hydrocarbons are considered in this work: (1) Chapters 5 and 6 involve the quantification of uncertainty (non-linear confidence intervals) in first-order biodegradation rate constant and LNAPL dissolution rate under uncertain source geometry and aquifer transmissivity; and (2) Chapter 7 involves the quantification of uncertainty in field-measured biodegradation rates in the heterogeneous aquifers. In Section 2.5.1, a few works related to stochastic modeling of natural attenuation are reviewed. In Section 2.5.2, the works related to inverse modeling for reactive/non-reactive mass transport in general hydrogeology literature are reviewed. In Section 2.5.3, the previous works related to the quantification of uncertainty in field-estimated biodegradation rate are presented.

2.5.1. Stochastic modeling of NA: Literature review

The work of McNab and Dooher (1998) is an early work dealing with characterizing the uncertainty in natural attenuation response variables. They develop a screening model to quantify uncertainties involved in the impact of biodegradation on hydrocarbon plume behavior. The approach is based on Monte Carlo simulation using an analytical solution to the advective-dispersive solute transport equation, including a first-order biodegradation term, coupled with mass balance constraints on electron acceptor use. The model was applied to an existing PHC contaminated site and the degree of uncertainty associated with model-predicted hydrocarbon concentrations and geochemical indicators as well as the roles of various parameter uncertainties (e.g. average hydraulic conductivity, first-order biodegradation rate, source term) in influencing forecasts were investigated. Lu et al. (1999) conducted a series of Monte Carlo Simulations to study the effects of uncertain hydraulic conductivity on the natural attenuation of BTEX compounds through aerobic degradation, denitrification, iron and sulfate reduction, and methanogenesis. They first quantified the uncertainty in hydraulic conductivity field by multiple realizations of spatially correlated random fields. Their simulated BTEX plumes were then analyzed for mass distributions and the relationship among various factors such as dissolved BTEX mass, plume spreading, and depletions of electron acceptors. They also investigated how the model responds to varying degrees of heterogeneity in the hydraulic conductivity field. Additional hydraulic conductivity realizations were created with the same mean but different variances and correlation lengths. In another work, Thornton et al. (2001) presented a quantitative methodology for field-scale performance assessment of natural attenuation using plume-scale electron and carbon balances and evaluation of the associated uncertainty. Their work focused on the calculation of global mass balances, using mass inputs from the plume source, background groundwater, and plume residuals in a simplified box model.

Uncertainty in the model predictions and sensitivity to different parameter values was assessed by Monte Carlo simulation. They concluded that the source area and infiltration primarily account for uncertainty in forecasts of the plume electron donor input; and estimates of electron acceptor inputs to the plume are primarily dependent on the transverse mixing zone width and less dependent on aquifer hydraulic conductivity and hydraulic gradient. More recently, Christensen et al. (2004) presented a statistical methodology for estimation of the relative efficiency of natural attenuation mechanisms, being oxidation using oxygen, denitrification, iron reduction and sulfate reduction, and the associated measures of uncertainty. They concluded that sulfate is the main cause of hydrocarbon removal in their study site; and oxygen is preferentially depleted at the upstream edge of the plume.

2.5.2. Inverse modeling and parameter estimation: Literature review

An important aspect of parameter uncertainty is related to using the state variables to estimate the model parameters and their uncertainty. An accurate description of physical and chemical properties of an aquifer is a prerequisite for accurate predictions of contaminant transport in groundwater. In practice, the field measurements of distributed hydrogeologic and geochemical parameters are only performed at limited sparse locations leaving a large portion of the aquifer untested. For some other geochemical parameters, they can be only reliably estimated under controlled laboratory experiments (e.g. dissolution rate constant, biodegradation rate constant). Inability to adequately estimate these distributed and single-valued parameters motivates us to apply/develop inverse modeling techniques that use the worth of piezometric head/concentration measurements for this purpose. Inverse modeling is implemented in this thesis to quantify the uncertainty in LNAPL dissolution rate and first-order biodegradation rate under uncertain source geometry and transmissivity field. The ultimate goal of such techniques is to build realistic aquifer models that are conditioned to all static data

(e.g. hydraulic conductivities) and are able to reproduce all dynamic data (e.g. piezometric heads, concentrations) within their margin of error. Such conditioned models are then used for more reliable future predictions.

Several inverse modeling techniques have been developed to incorporate piezometric head data into groundwater flow models. The linearized cokriging method (Kitanidis and Vomvoris 1983, Hoeksema and Kitanidis 1984 and Kitanidis and Lane 1985), the fast Fourier transform method (Gutjahr and Wilson 1989, Robin et al. 1993 and Gutjahr et al. 1994), the linearized semi-analytical method (Dagan 1985, Rubin and Dagan 1987, Dagan and Rubin 1988), the fractal simulation method (Grindrod and Impey 1991), the pilot point method (RamaRao et al. 1995), the maximum-likelihood method (Carrera and Neumann 1986) and the sequential self-calibration (SSC) method (Gomez-Hernandez et al. 1997) are the some important classic approaches. A thorough review and comparison of these approaches is presented in McLaughlin and Townley (1996) and Zimmerman et al. (1998). Some of the other important more recent techniques are proposed by Hu (2000), Caers and Hoffman (2006) and Fu and Gomez-Hernandez (2008). Hu (2000) adapted the gradual deformation approach for Gaussian fields with data conditioning, while preserving the correlation structure, Caers and Hoffman (2006) presented the probability perturbation approach for Gaussian and non-Gaussian fields, Fu and Gomez-Hernandez (2008) developed a new methodology based on the theory of blocking Markov-Chain Monte Carlo which uses a sampling algorithm for conditioning to piezometric heads (and travel times) instead of optimization. The SSC technique (Gomez-Hernandez et al. 1997) is applied extensively throughout this thesis. The details of the approach are reviewed in Chapter 5.

The integration of concentration data has been less widely studied. The works related to integration of concentration data can be grouped in three different

categories. The first group are the works that integrate the available information about the state variables (e.g. point measurements of concentration), and estimate the distribution of model parameters (e.g. transmissivity field, sorption) and state variables in an 'optimal' sense. An estimate of uncertainty may also be obtained as the conditional covariance of state variables (linear confidence intervals). The second group involves the works that do not update the transmissivity field as the main source of uncertainty in transport modeling. Instead, they parameterize the effects of unresolved heterogeneities through the estimation of first and second conditional moments of state variables. The third group involves the works that generate multiple realizations conditioned to point measurements of model parameters and the state variables. The generated realizations jointly represent the best estimate and uncertainty of model parameters and state variables in the context of Monte Carlo simulations.

The first group of inverse modeling techniques with concentration data involves the approaches that give a best estimate for each model parameter and state variable. The first group of researchers to integrate concentration data to estimate dispersivities was Umari et al. (1979). Their objective functions were based on L1 and L∞ norms in the residual space; and quasi-linearization was used to minimize the objective functions. Gorelick et al. (1983) estimated the location and magnitude of the source through linearizing the state equations. The coupled estimation of flow and transport parameters was performed by Strecker and Chu (1986). A two-step estimation procedure was developed, where, the transmissivity field was estimated using head data only in the first step, and in the dispersivities were estimated using concentrations and previously quantified transmissivities in the second step. A similar two-stage approach was adopted by Van Rooy et al. (1989) for an inhomogeneous aquifer. In the first step, they estimated transmissivity field by kriging. Then, they estimated the values of source and dispersivities using concentration data. The work of Keidser and Rosbjerg (1991)

was an extension to the work of Van Rooy et al. (1989) where they alternatively used head and concentration data to estimate transmissivities while treating the transport parameters as known. Then, they used the estimated transmissivities and the available concentration and head data to estimate the dispersivities. Using a least-square estimator and nonlinear regression, Wagner and Gorelick (1986) developed a coupled approach for estimation of flow and transport parameters and implemented that in a 1D example. Sun and Yeh (1990a) presented the adjoint-state method for calculation of sensitivity coefficients for coupled flow and transport problems. In their second paper (Sun and Yeh 1990b), they addressed the issue of identifiably for the coupled problems. Medina et al. (1990) and Medina and Carrera (1996) presented a fully coupled inverse modeling approach for flow and transport. They used finite element method for spatial descretization and finite difference method for temporal descretization to solve the flow and transport equations in steady-state and transient conditions. The spatial variability was parameterized by zonation method and maximum-likelihood estimation was used to estimate multiple flow and transport parameters. Optimization of maximum-likelihood estimator was achieved by the Marquardt method and confidence intervals were presented for the estimated parameters. Wagner (1992) used a very similar approach to that of Medina et al. (1990) but estimated the source location as well as the flow and transport parameters and their confidence intervals (reviewed in Chapter 1). In an extension to the work of Keidser and Rosbjerg (1991), Sonnenborg et al. (1996) also developed a coupled approach based on the maximum-likelihood method and the method of characteristics. They applied their model to a real site in Denmark and tested multiple parameterization of transmissivity field with zonation approach. Another application of maximum likelihood method was presented by Mayer and Huang (1999). They solved the coupled inverse problem by optimizing the maximum likelihood objective function by genetic algorithm for a transmissivity field that was parameterized by kriging instead of zonation. The variogram parameters were treated as unknown

and estimated in the optimization process. Inverse modeling has also been used for hydrocarbon and dense nonaqueous phase liquid (DNAPL) source identification (Parker and Islam 2000, Sciortino et al. 2000), estimating biodegradation rates in batch experiments (Schirmer et al. 1999), and estimation of BTEX biodegradation rates and LNAPL dissolution rate at a the Bemidji crude oil spill site (Essaid et al. 2003). Parker and Islam (2000) presented two methods for estimating release timing of LNAPL and its uncertainty. First, they developed a simple analytical model for LNAPL plume migration that predicts LNAPL plume velocity as a function of basic soil and fluid properties. Then, they used the first-order method and a Monte Carlo simulation approach to estimate the release time uncertainty. Sciortino et al. (2000) developed an inverse modeling technique to identify the location and the dimensions of a single-component DNAPL pool in a saturated porous medium under steady-state conditions. A 3D analytical model was used to quantify the transport of solute from a dissolving DNAPL pool, the inverse modeling was formulated as a least squares minimization problem, and it was solved by Levenberg-Marquardt method. They concluded that the inverse problem is non-unique and non-convex even in the absence of observation errors. Essaid et al. (2003) developed an inverse modeling technique and applied it to a real site with crude oil contamination. They used historical data over a course of 11 years and incorporated a multi-component transport and biodegradation model under steady-state condition for a homogenous aquifer. Their inverse modeling was successfully converged only when a single dissolution rate coefficient was estimated for all BTEX components. Assuming a stationary oil body with known geometry, they simulated the transport of dissolved oxygen as well as all BTEX components, and estimated a number of model parameters including the recharge rate, hydraulic conductivity, dissolution rate coefficient, individual first-order BTEX anaerobic biodegradation rates and transverse dispersivity. In this work, an estimation of parameter uncertainty was made through the definition of confidence intervals for each estimated parameter. The first group of inverse

modeling techniques gives optimal estimates of the model parameters. The parameterization (e.g. zonation) of the distributed parameters (e.g. transmissivity) is often too simple with an upper limit on the number of zones that can be estimated, while avoiding instability. Thus, the transmissivity realizations that are estimated using these approaches are too smooth and do not represent the true variability. It is also very important to note that the single valued parameters (e.g. single-species biodegradation rate constant) that are estimated by these approaches always correspond to a single optimal transmissivity field (and source geometry) and their true uncertainty can not be quantified.

The second group of inverse modeling techniques includes some early works that tried to parameterize the effects of unresolved heterogeneities through the estimation of first and second conditional moments of state variables. In their first paper, Graham and McLaughlin (1989a) presented a methodology to calculate the non-conditional ensemble concentration mean, macro-dispersive flux, and covariance of the concentration at any time and location. They used a numerical method to solve the moment propagation equations. In their approach, transmissivity field was considered the only source of uncertainty and the methodology was developed based on a small perturbation assumption. In their second paper Graham and McLaughlin (1989b) presented an improved methodology by conditioning the ensemble moments on field observations of hydraulic conductivity, piezometric head and solute concentration. The conditional moments were updated using a combination of propagation equations and a Kalman filter approach at each time step. In a Lagrangian framework, Rubin (1991a) presented an approach for conditioning the concentration ensemble moments to the measurements of hydraulic conductivity, head and velocity. In this work, the steady-state flow equation was linearized and multi-Gaussian distributions for transmissivity, head, flow and concentration were assumed. In a second paper, Rubin (1991b) extended the previous methodology for conditioning

to concentration measurements. The spatial and temporal covariances of concentration field were obtained by tracking a large number of particles. The calculated covariances were then used in cokriging to interpolate the concentration residuals and obtain the concentration deviation at each location at space and time. The main limitation of the second group of inverse modeling techniques is due to the fact that they do not update the transmissivity field. Also, they do not account for large heterogeneity variances.

The third group of inverse modeling techniques (also referred to as Monte Carlo methods) involves approaches that generate multiple realizations that are conditioned to observations of model parameters and state variables. A great advantage of these techniques is that they account for non-uniqueness in inverse modeling by generating multiple realizations. The classic examples of this group of inverse modeling techniques are the pilot point method (RamaRao et al. 1995) and the SSC approach (Gomez-Hernandez et al. 1997). In the pilot point method (RamaRao et al. 1995), first, multiple unconditional realizations are generated by turning bands method. Each of these unconditional realizations are then iteratively updated by addition of the estimated residuals by kriging. The kriging estimation of residuals is performed using the transmissivity deviations at the observation locations as well as deviations calculated at the pilot points in the calibration process. A sensitivity analysis and kriging are used to locate pilot points optimally. A least-square objective function and a gradient-based optimization algorithm were used to find the optimal deviations at the pilot points. LaVenue and de Marsily (2001) and Alcolea et al. (2006) presented the extensions of the pilot-point method for transient condition in fractured media and for use of prior information. The details of the SSC approach (Gomez-Hernandez et al. 1997) are given in Chapter 5. This technique was originally developed for a 2D steady-state condition with multi-Gaussian transmissivity distribution. Capilla et al. (1999) presented the extension of the method to non-Gaussian fields. The extension of

the SSC technique to transient flow condition by joint calibration of transmissivities and storativites was presented by Hendricks Franssen et al. (1999a). Hendricks Franssen et al. (1999b) developed the extension of the SSC to 3D flow in fractured media. Wen et al. (2003) presented the extension of the SSC to two-phase flow under transient conditions. Finally, Hendricks Franssen et al. (2003) and Hendricks Franssen et al. (2008) presented the extension of the SSC to coupled inverse modeling of groundwater flow and mass transport and integration of remote sensing data in stochastic groundwater flow models, respectively. The coupled inversion of groundwater flow and mass transport was developed using a Eulerian flow and transport simulator and based on calculation of sensitivity coefficients by adjoint state method (Sahuquillo et al 1999), which is computationally intensive. According to the results of the synthetic study reported in Gomez-Hernandez et al. (2003) and Hendricks Franssen et al. (2003), conditioning to both head and concentration data improves the prediction of and reduces the uncertainty in the predicted concentration field. Also, conditioning to piezometric head data substantially improves the predictions of transmissivity field and head distribution and reduces their uncertainty. Based on the results of their study, however, incremental value of incorporation of concentration data (in addition to head data) into characterization of transmissivity and head fields was minimal.

In addition to previous works developed to use the worth of concentration data into inverse modeling of groundwater flow and mass transport problems, there is a number of works that deal with incorporation of travel time data in inverse modeling. Harvey and Gorelick (1995) developed a cokriging method based on combining a linear estimator with numerical flow and transport simulations. Through first-order approximations of the governing flow and transport equations, the conditional covariances between the parameters (e.g. transmissivity) and the state variables (e.g. head and travel time) were updated sequentially. Their

methodology resulted in a transmissivity field that was conditioned to transmissivity, head and travel time data. Also, they concluded that the incorporation of solute travel time is crucial for identification of flow paths and flow barriers. Li and Yeh (1999) and Cirpka and Kitanidis (2000) presented similar cokriging approaches capable of incorporating travel time data. In both works, it was concluded that the head data are the most important observations that can improve the estimation of transmissivity field. Datta-Gupta et al. (1998) and Wen et al. (2002) presented Monte Carlo type approaches capable of incorporating travel time data, through the application of simulated annealing and semi-analytical quantification of sensitivity equations, respectively. Both works concluded that travel time data are important information in recognition of flow barriers and flow paths. Another recent work in this area includes the work of Huang et al. (2004). Huang et al. (2004) used travel time data in the context of the SSC approach to condition both transmissivity distribution and sorption distribution to head and travel time data.

2.5.3. Uncertainty in field estimated first-order rate: Literature review

Another aspect of parameter uncertainty (for natural attenuation) has been mainly studied in the context of characterizing the uncertainty in the biodegradation rate constant determined by some sort of field measurement methodology.

As discussed in Chapter 1, commonly-used field estimation techniques for first-order biodegradation rate constant include 1D and 2D concentration-distance approaches with centerline and off-centerline measurements (Buscheck and Alcantar 1995, Chapelle et al. 1996, Wiedemeier et al. 1996, Zhang and Heathcoat 2003, Stenback et al. 2003) and mass-flux techniques (Borden et al. 1997, King et al. 1999, Peter et al. 2004). The major shortcomings associated with these approaches are (1) they give estimated values that over/underestimate the true rate constant up to two orders of magnitude and (2) they do not provide a

measure for uncertainty in the estimated rates under site-specific conditions. The potential uncertainties associated with these field estimation techniques are studied in a number of works presented below.

Odencrantz et al. (2003) were the first group to point out the problems involving the estimation of biodegradation rate constant by calibrating the analytical models to field data and/or field measurement techniques. Using a 1D analytical model, they showed that uncertainty in field-scale flow velocity and dispersivity values translates into uncertainty in estimated first-order biodegradation rates. The requirement of sampling along the plume centerline in the majority of the existing field measurement techniques also may result in significant uncertainty (error) in the estimated first-order biodegradation rate constants. Bauer et al. (2006) evaluated measurement uncertainty of first-order degradation rates (estimated using centerline approaches) in heterogeneous aquifers. They first generated a number of synthetic heterogeneous aquifers with different heterogeneity levels. For each heterogeneity level and for each synthetic aquifer, they calculated the biodegradation rate constant with different centerline approaches and compared the results to the original values used in simulations. They showed that with increasing heterogeneity, the measured degradation rate constants become uncertain with a high variability (up to two orders of magnitude) around the true constant. They also conducted a series of sensitivity analyses to investigate the influences of source width, choice of dispersivity values and transport velocity. Beyer et al. (2006) performed a similar study to that of Bauer et al. (2006) but they adapted a slightly different degradation kinetics model, namely Michaels-Menten (MM) to their model and kept the original first-order biodegradation model as well. In addition to uncertainty assessment for measured biodegradation, they also studied the resulted uncertainty in plume lengths and concluded that the calibrated MM model performs better in estimating the plume lengths in heterogeneous aquifers. Beyer et al. (2007) performed similar study to that of

Bauer et al (2006) but this time to take into account the 'human factor', the centerline of the synthetic plumes were estimated by a large number of professionals who were given a number of point head and concentration measurements and the approximate location of the source zone. They also utilized off-centerline concentration measurements by using the analytical model of Stenback (2004) and least-square estimation to find the value of first-order biodegradation rate constant. In both cases they showed that the true rate constant is overestimated by a few orders of magnitude. For wider plumes, however, fitting the analytical model can reduce the degree of overestimation. Recently, Ukankus and Unlu (2008) investigated the effects of the range of correlation structure on the relationship between the heterogeneity of the aquifer system and error in estimated first-order biodegradation rate.

The uncertainties associated with mass-flux techniques have been investigated in a number of works including the work of Kubert and Finkel (2006), where they showed that the outcomes of the mass-flux techniques can be readily affected by heterogeneities in aquifer material and well spacing.

2.6 Parameterization for modeling NA of PHCs

Natural attenuation response variables should be stated in terms of a number of model parameters in such a way that (1) all the important mechanisms contributing in natural attenuation of PHCs are simulated; (2) the model parameters and their uncertainty can be estimated using the available direct measurements of the parameters and observations of the response variables and (3) the model is not over-parameterized.

Important natural attenuation mechanisms are explained in Section 2.1. If the concentration of the dissolved contaminants is the primary dependent variable, the most important transport mechanisms of dissolved contaminants can be represented by an expanded version of the advection-dispersion-reaction equation (Equation [2.1]) as given by Zheng and Wang (1999):

$$\frac{\partial(\theta C_s)}{\partial t} = \frac{\partial}{\partial x_i}\left(\theta D_{ij}\frac{\partial C_s}{\partial x_j}\right) - \frac{\partial}{\partial x_i}(\theta v_i C_s) + q_{sr}C_{sr} + \sum R_n \qquad [2.7]$$

where, C_s represents the dissolved concentration of species s, θ is the porosity of the subsurface medium, t is time, x_i is the distance along the respective Cartesian coordinate axis, D_{ij} is the hydrodynamic dispersion coefficient tensor, v_i is the seepage or pore water velocity, q_{sr} is the volumetric flow rate per unit volume of aquifer representing fluid sources (or sinks), C_{sr} is the concentration of the source (or sink) flux, and $\sum R_n$ is the chemical reaction term:

$$\sum R_n = -\rho_b \frac{\partial \overline{C_s}}{\partial t} - \lambda\theta C_s \qquad [2.8]$$

in which, ρ_b is the bulk density of the subsurface medium, $\overline{C}_s$ is the concentration of the contaminant species s sorbed on the subsurface solids, and λ is the first-order reaction rate for the dissolved phase. The first term on the right hand side of Equation [2.7] represents the hydrodynamic dispersion of the solutes (including

molecular diffusion and mechanical dispersion). The second term represents advection of the solutes due to the bulk movement of groundwater in the pore spaces. The third and forth terms on the right hand side of Equation [2.7] represent the source/sink of dissolved contaminants with a pre-specified concentration and influx (or outflux), and chemical reactions, respectively. In the chemical reaction term, aqueous-solid surface reaction (sorption) and first-order rate reaction are represented by Equation [2.8].

The governing transport equation is linked to the governing flow equation by the Darcy's law:

$$v_i = -\frac{k_i}{\theta} \cdot \frac{\partial h}{\partial x_i} \qquad [2.9]$$

where, k_i represents the hydraulic conductivity in the direction of flow, and h is the hydraulic head that is obtained from the solution of governing equation for groundwater flow:

$$\frac{\partial}{\partial x_i}\left(\theta k_i \frac{\partial h}{\partial x_i}\right) + q_{sr} = S_s \frac{\partial h}{\partial t} \qquad [2.10]$$

where, S_s is the specific storage coefficient of the porous medium, and the groundwater is assumed to be of uniform density (Zheng and Bennett 1995).

The Equations [2.7] through [2.10] represent advection, dispersion, sorption and degradation of dissolved contaminants in a transient groundwater flow regime. The timeframe for occurrence of natural attenuation processes is large compared to seasonal stresses that bring the groundwater into a transient condition. Thus, in the context of parameterizing a model for natural attenuation and for simplification purposes, the groundwater is assumed to be in a steady-state condition. As a result, the simplified version of Equation [2.10] can be written by:

$$\frac{\partial}{\partial x_i}\left(\theta k_i \frac{\partial h}{\partial x_i}\right) = -q_{sr} \qquad [2.11]$$

which represents the governing equation for groundwater flow in a steady-state condition. The source (sink) flux (q_{sr}) in Equations [2.7] and [2.11] may represent addition or removal of water to or from the system by areal recharge or pumping activities. Although the effect of areal recharge on groundwater flow velocity and dilution and advection of contaminants can be pronounced (Anderson and Woessner 1992), the uncertainty in areal recharge has not been modeled in this work. The reason is that estimation of areal recharge is very site-specific, as it is a function of precipitation and evapotranspiration. Characterizing the uncertainty in site-specific precipitation and evapotranspiration is outside the scope of this work. The model and associated codes, however, are general and can deterministically simulate the effect of areal recharge for field applications.

Sorption of dissolved PHCs to solid particles is a function of the organic fraction of soil (f_{oc}). Using an analytical model, Alvarez and Illman (2006) conducted a sensitivity analysis to evaluate to relative importance of different natural attenuation mechanisms (parameters). They varied different model parameters over the ranges prescribed in the literature. They concluded that uncertainty in biodegradation rate constant (λ) and groundwater velocity (v_i) are the most important parameters that affect the length of contaminant plumes. They also observed that the variability in the organic fraction of soil (f_{oc}) has a minor impact on the plume length. Their results were in good agreement with the results previously obtained by McNab and Dooher (1998). In addition to numerical observations reported in the literature, there are some other reasons for excluding the sorption in the model developed in this work: (1) estimation of organic fraction and its uncertainty is very site-specific, and the choice of prior values for f_{oc} is quite subjective, which in turn may affect the overall quality of calibration and estimation of other more important model parameters (e.g. NAPL dissolution rate, Biodegradation rate constant); (2) as mentioned before, the time frame for natural attenuation of PHCs in the environment is often in the order of decades. In

this situation, the adsorption sites are already occupied by contaminants and the effect of variability in soil organic matter on the overall process of natural attenuation is considered small.

The importance of the source term (source geometry and dissolution rate) in stochastic modeling and parameter uncertainty associated with natural attenuation has been discussed by previous researchers (McNab and Dooher 1998, Bauer et al. 2004, Ling and Rifai 2007, Miles et al. 2008). As pointed out by Waddill and Widdowson (1998) and Mendez et al. (2004), the source term can significantly affect the time of remediation (TOR) problem. It is common practice to set the contaminant concentrations to pre-specified values in the source areas. However, the rate of dissolution of NAPL into groundwater depends on the interfacial area between the NAPL and water, aquifer heterogeneity, the size and shape of NAPL blobs and the groundwater velocity (Pfannkuch 1984, Powers et al. 1994, Imhoff et al. 1994, Mayer and Miller 1996). Many of these factors can not be easily estimated under field condition; and as a result, the rate of dissolution of NAPL into groundwater usually involves large uncertainties under field conditions (Chapter 1).

If transport processes occur at a high rate relative to the NAPL dissolution rate, the aqueous phase concentration (C_s) may remain lower than the equilibrium concentration (C_s^{eq}) of the contaminant species s. This effect may be described mathematically (Parker et al. 1991, Imhoff et al. 1994) by a mass transfer rate coefficient (k^{NAPL}), such that the NAPL dissolution term for species s becomes:

$$R_{source}^{NAPL} = \max\left[0, k^{NAPL}\left(C_s^{\text{eq}} - C_s\right)\right] \qquad [2.12]$$

As stated by Parker et al. (1991), the equilibrium concentration (C_s^{eq}) of every contaminant species s can be written as:

$$C_s^{\text{eq}} = \text{f}_s \cdot C_s^{\text{sol}} \qquad [2.13]$$

where, f_s is the mole fraction of the species s in the mixture of PHC contaminants,

and C_s^{sol} is solubility of pure substrate in water. The NAPL mass of substrate *s* per unit mass of dry soil (S_s^{NAPL}) decreases as the dissolution occurs. This process can be represented by:

$$\frac{dS_s^{NAPL}}{dt} = -\frac{\theta}{\rho_b} R_s^{NAPL} \qquad [2.14]$$

In other words, dissolution causes the NAPL concentration of species *s* to decrease as the aqueous phase concentration increases.

As reviewed in Sections 2.1.7 and 2.4.2, many models that simulate rather complex kinetics and multi-component reactions have been developed. It is evident that many of the requisite kinetic parameters for these models can not be measured or estimated by routine natural attenuation protocols. Thus, utility of these models is generally limited to research (Rifai and Rittaler 2005). In a modeling context, simpler approaches with limited number of parameters are often preferred as they can be supported by site data and avoid over-parameterization. In this work, biodegradation of PHC contaminants is represented by a first-order reaction model, which can be expressed by:

$$C_s(t) = C_s^0 e^{-\lambda t} \qquad [2.15]$$

where, $C_s(t)$ is concentration at time t, C_s^0 is concentration at time t = 0, and λ is first-order degradation rate. Incorporating first-order biodegradation and NAPL dissolution models and excluding the sorption model in Equation [2.7] results in the partial differential equation representing dissolution-advection-dispersion-biodegradation of petroleum hydrocarbon contaminants in groundwater:

$$\frac{\partial(\theta C_s)}{\partial t} = \frac{\partial}{\partial x_i}\left(\theta D_{ij} \frac{\partial C_s}{\partial x_j}\right) - \frac{\partial}{\partial x_i}(\theta v_i C_s) + \theta \cdot \max\left[0, k^{NAPL}\left(C_s^{eq} - C_s\right)\right] - \theta \lambda C \qquad [2.16]$$

Equations [2.11] and [2.16] form the basis for all subsequent modeling work. According to these equations, a model of uncertainty for natural attenuation of PHCs in groundwater can be parameterized in terms of (1) exhaustive distribution

of hydraulic conductivities (transmissivities) over the modeling domain and its uncertainty, (2) site-specific first-order biodegradation rate constant and its uncertainty, (3) areal extent of residual NAPL source and its uncertainty, and (4) dissolution rate of NAPL into groundwater and its uncertainty. Uncertainty in the value of effective porosity is not modeled in this thesis, because variability in porosity and its impacts on the response variables are insignificant (Alvarez and Illman, 2006). The calculation of hydrodynamic dispersion tensor (D_{ij}) in Equation [2.16]) will be affected by variability in velocity field. The local dispersivity values, however, are set to small known values. This is because the cell sizes are always smaller than the range of the variograms used in geostatistical simulation of transmissivity fields.

CHAPTER 3

GEOSTATISTICAL MODELING WITH CPT-UVIF DATA

In every groundwater modeling and remediation project, uncertainty exists at three different levels (Sun 1994): (1) uncertainty in the site conceptual model (SCM), (2) uncertainty in model structure, and (3) uncertainty in the values of the model parameters. The primary objective of this thesis is to focus on parameter uncertainty for natural attenuation of petroleum hydrocarbons; however, if the SCM and model structure are not representative, subsequent characterization of parameter uncertainty will be of little value. Thus, it is important to develop a plausible SCM in natural attenuation studies for contaminated sites.

Geostatistical techniques can be applied to develop a conceptual model for geology, hydrogeology and contaminant source distribution for hydrocarbon impacted sites, using data collected by Ultra Violet Induced Fluorescence Cone Penetration Testing (CPT-UVIF) and piezometer readings. A few previous works have dealt with geostatistical modeling with cone penetration testing (CPT) data. In an early work, Yoon and O'Neill (1999) performed variogram analysis and applied ordinary kriging to map raw CPT data across the site. They used the mapped properties to compute ultimate pile capacities based on three different CPT design methods. Elkateb et al. (2001) performed geostatistical simulation using the CPT data (tip resistance) to quantify the effect of soil spatial variability on liquefaction susceptibility. They performed an explicit trend analysis and calculated and fitted the variograms for the residuals. Recently, Lenz and Baise (2007) defined an index for liquefaction potential based on raw CPT data and mapped the index to determine the spatial distribution of liquefaction potential. To the author's knowledge, the work of D'Affonseca et al. (2008) is the only

work reported in the literature that makes use of CPT-UVIF data for characterization of spatial distribution of a NAPL source zone and its relationship with soil properties. In their work, D'Affonseca et al. (2008) deterministically calibrated the areal and vertical extent of NAPL source to soil horizons previously characterized by CPT data. The majority of the previous works related to geostatistical modeling with CPT data include the application of raw CPT attributes (e.g. cone resistance) within the context of studying the liquefaction potential. Also, no previous work related to geostatistical characterization of NAPL source by CPT-UVIF data has been reported. A number of soil behavior type (SBT) charts have been proposed to correlate the raw CPT attributes to soil types defined based on their mechanical properties. In this work, instead of using the raw CPT attributes, continuous and categorical variables based on the SBTs are defined and used in subsequent geostatistical modeling. The primary motivation for application of SBT-derived variables instead of CPT raw attributes is due to an existing relationship between the SBTs and hydraulic properties of the soil through definition of hydro-facies that can be mapped across the site. From the CORONA[1] project, CPT-UVIF data and piezometer readings are available for a former flare pit site at west-central Alberta. In the first section of this chapter, a brief overview of the CPT device is presented and the CPT readings are used to define discrete (hydro-facies) and continuous variables correlated to the hydraulic properties of in-situ soil. In a case study, available geostatistical techniques are applied to develop a probabilistic model for ground conditions at the site. In the second section of this chapter, a distance function algorithm is developed to model uncertainty in areal limits. Then, the distance function approach is applied together with a number of widely used geostatistical techniques to develop a source zone modeling approach that accounts for important secondary sources of information such as distance to the point of release, soil texture and distance to water table elevation.

[1] Consortium for Research on Natural Attenuation at the University of Alberta

3.1 Geostatistical Modeling of Ground Conditions

3.1.1. CPT technology and soil profiling

CPT technology is a method of providing 'real time' information about soil type and associated design parameters at a high resolution. It was initially developed in the 1950s at the Dutch Laboratory for Soil Mechanics in Delft to investigate soft soils. It consists of a steel cone that is hydraulically pushed into the ground at up to 40,000 pounds of pressure. There are sensors on the tip of the cone that collect data. Standard cone penetrometers collect data using sensors the measure cone-tip pressure and sleeve friction (Figure 3-1). It is generally applied to depths up to 45 m, but in some occasions, it has been used as deep as 90 m. According to Robertson (1986), the CPT has three main applications in the site investigation process:

- To determine sub-surface stratigraphy and identify materials present;
- To estimate geotechnical parameters; and,
- To provide results for direct geotechnical design.

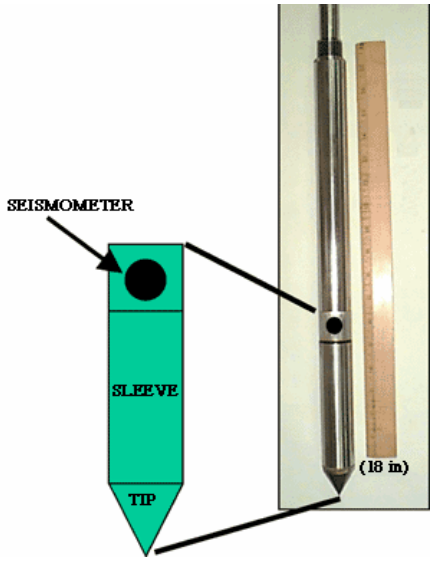
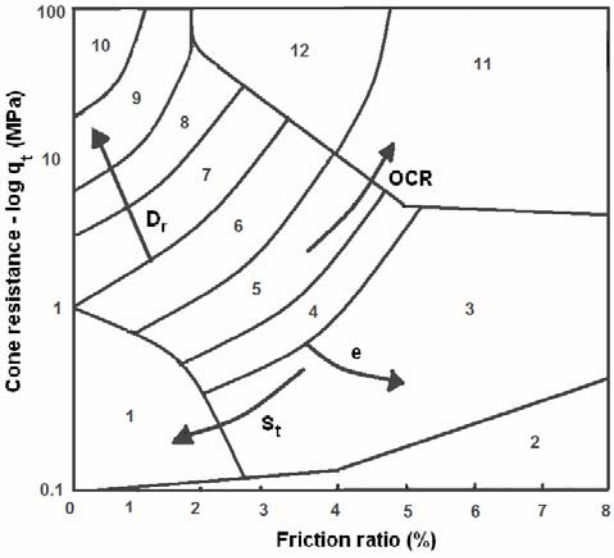


Figure 3-1: Schematic representation showing the position of the tip, sleeve and seismometer on a CPT cone (USGS website – accessed 13/01/09)

One of the major applications of the CPT is for soil profiling and classification. Extensive experience exists that relates CPT results to soil type. Experience has shown that typically the cone penetration resistance, or tip resistance is high in sands and low in clays, and the friction ratio, the ratio of the sleeve friction to the tip resistance (R_f=f_s / q_c) is low in sands and high in clays. This observation has been incorporated into several soil classification charts such as charts by Robertson et al (1986) and Douglas and Olsen (1981) among others (Figures 3-2 and 3-3).



Zone	*Soil Behavior Type*
1	*Sensitive fine grained*
2	*Organic material*
3	*Clay*
4	*Silty clay to clay*
5	*Clayey silt to silty clay*
6	*Sandy silt to clayey silt*
7	*Silty sand to sandy silt*
8	*Sand to silty sand*
9	*Sand*
10	*Gravely sand to sand*
11	*Very stiff fine grained (OC)*
12	*Sand to clayey sand (OC)*

Figure 3-2: Soil behavior type classification chart (after Robertson et al. 1986)

Douglas and Olsen (1981) claim that CPT classification charts can not be expected to provide accurate predictions of soil type based on grain size distribution but provide a guide to soil behavior type (SBT). The chart by Robertson et al. (1986) is one of the most commonly used CPT soil behavior type charts. It uses the basic parameters of cone (tip) resistance and friction ratio. The chart is global in nature and can provide reasonable predictions of soil behavior type for CPT soundings up to about 20 m in depth. The chart identifies general

trends in ground response such as increasing relative density (D_r) for sandy soils, increasing stress history (OCR), soil sensitivity (S_t) and void ratio (e) for cohesive soils (Figure 3-2).

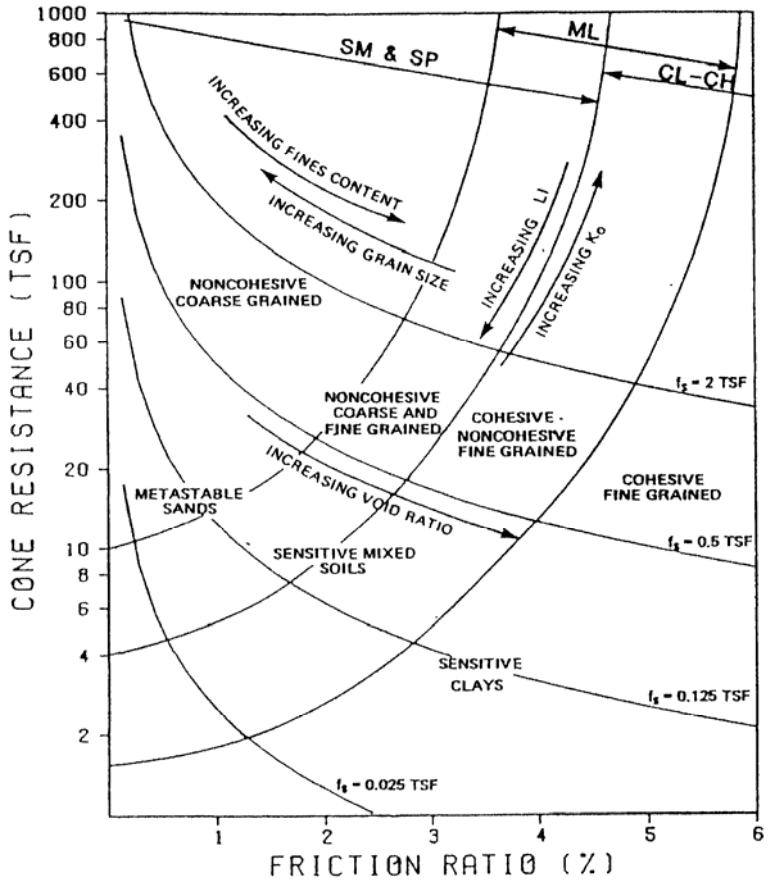


Figure 3-3: Soil behavior type classification chart (Douglas and Olsen 1981)

As presented by Robertson et al. (1986), an approximate estimate of soil hydraulic conductivity, or permeability, can be made from an estimate of soil behavior type using the CPT classification charts. Table 3-1 provides estimates based on the chart shown in Figure 3-2. The discrete nature of soil classification charts makes it difficult to perform a statistical analysis on CPT soil classification results. Also, as it can be observed in Table 3-1 there is some overlap between the hydraulic properties of different SBTs.

There are generally three elementary factors controlling tip resistance (q_c) and sleeve friction (f_s) in cone penetration testing (Douglas and Olsen 1981): (1) equipment and procedure, (2) soil composition, and (3) environmental components (soil in-situ state). In the case that rigorous standardization on test equipment and test procedure is followed, q_c and f_s would be affected only by soil composition and soil in-situ state. This is illustrated in a typical soil classification chart as two basic tendencies that are almost orthogonal to each other: soil type changes in one direction and in-situ soil state (*OCR*, soil sensitivity, etc.) changes in the other. This observation can be mathematically depicted, and two independent indices representing the two primary factors can be defined (Zhang and Tumay 2003). A curvilineal coordinate system (Figure 3-4) is empirically established along the tendencies in the soil classification charts of Figures (3-2 and 3-3). One of the curvilineal coordinates represents the soil type and the other corresponds to the soil-state. This curvilineal coordinate system is then transformed to a Cartesian coordinate system by conformal mapping as shown in Figure 3-4.

The conformal transformation is accomplished by the following equations:

$$U = -\frac{(a_1x - a_2y + b_1)(c_1x - c_2y + d_1) - (a_2x + a_1y + b_2)(c_2x + c_1y + d_2)}{(c_1x - c_2y + d_1)^2 + (c_2x + c_1y + d_2)^2} \qquad [3.1]$$

$$V = 10 + \frac{(c_1x - c_2y + d_1)(a_2x + a_1y + b_2) - (a_1x - a_2y + b_1)(c_2x + c_1y + d_2)}{(c_1x - c_2y + d_1)^2 + (c_2x + c_1y + d_2)^2}$$

$$x = 0.1539R_f + 0.8870\log q_c - 3.35 \quad , \quad y = -0.2957R_f + 0.4617\log q_c - 0.37$$

where, U and V represent Soil Classification Index (SCI) and Soil In-situ State Index (SISI), respectively. R_f is in % and q_c is in units of 100 kilo-Pascal. The coefficients in Equations [3.1] are a_1 = -11.345, a_2 = -3.795, b_1 = 15.202, b_2 = 5.085, c_1 = -0.269, c_2 = -0.759, d_1 = -2.960, and d_2 = 2.477.

Zone	*Soil Behavior Type*	*Range of permeability (m/s)*
1	*Sensitive fine grained*	3×10^{-9} to 3×10^{-8}
2	*Organic material*	1×10^{-8} to 1×10^{-6}
3	*Clay*	1×10^{-10} to 1×10^{-9}
4	*Silty clay to clay*	1×10^{-9} to 1×10^{-8}
5	*Clayey silt to silty clay*	1×10^{-8} to 1×10^{-7}
6	*Sandy silt to clayey silt*	1×10^{-7} to 1×10^{-6}
7	*Silty sand to sandy silt*	1×10^{-6} to 1×10^{-5}
8	*Sand to silty sand*	1×10^{-5} to 1×10^{-4}
9	*Sand*	1×10^{-4} to 1×10^{-3}
10	*Gravely sand to sand*	1×10^{-3} to 1
11	*Very stiff fine grained soil(OC)*	1×10^{-9} to 1×10^{-7}
12	*Very stiff sand to clayey sand (OC)*	3×10^{-8} to 3×10^{-6}

Table 3-1: Estimation of soil permeability using SBT chart by Robertson et al. (1986)

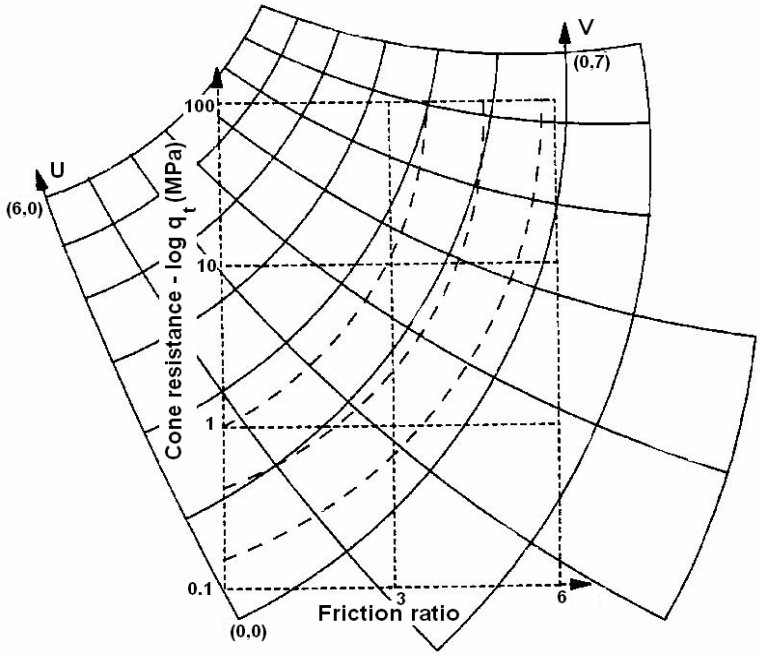


Figure 3-4: Curvilineal coordinate transformation by Zhang and Tumay (2003)

3.1.2. Geostatistical modeling of soil stratigraphy using CPT data

The study site is a former flare pit site (a CORONA site) located in west-central Alberta, Canada. Solid stem augering was used for initial drilling. Soil logs show a heterogeneous distribution of clay, silt and sandy units (Armstrong, et al. 2003). The location of the former flare pit is roughly known to be in the north of the site. The exact limits, however, are not known. The site slopes from north to south and the direction of groundwater is from northeast to southwest. According to initial soil sampling and groundwater analysis, small amount of free-phase hydrocarbon (mobile LNAPL) was suspected to remain at the site. The site is suspected to be primarily contaminated by residual hydrocarbon and it was characterized through logging and sampling 16 boreholes drilled using the solid stem auger method to approximately 5 m below ground surface. Based on these data, 18 CPT-UVIF cone holes were advanced in two phases and geo-mechanical properties of the soil were recorded in a high resolution fashion. The holes ranged in depth from 4 to 11 m below ground surface. Figure 3-5 shows the locations of the CPT-UVIF cone holes (used in the geostatistical simulation) for former flare pit site (with and without detection of any residual NAPL at any depth) and the geostatistical modeling domain. The modeling domain is 60 m in east-west direction, 80 m in north-south direction, and 16 m in depth with a Cartesian grid with 120 × 160 × 56 cells in each direction, respectively.

In geostatistics, the decision of which data should be pooled together for subsequent analysis is the "decision of stationarity". The decision of stationarity may be revised based on further data analysis. For instance, while observing a bimodal histogram for data, one might want to consider separating the data into two classes with distinct statistical and geological properties (Deutsch 2002). In fact, separating the data set into more homogenous geologic and hydrogeologic zones improves the accuracy of the estimates. Indicator kriging (simulation) offers a method for categorical data showing non-stationarity in its basic statistics.

Indicator kriging is used to calculate the probability of various 'soil types' or 'geological regions' with distinct statistical and geological features. This practice improves data homogeneity and makes the decision of stationarity more appropriate (Dagdelen and Turner 1996).

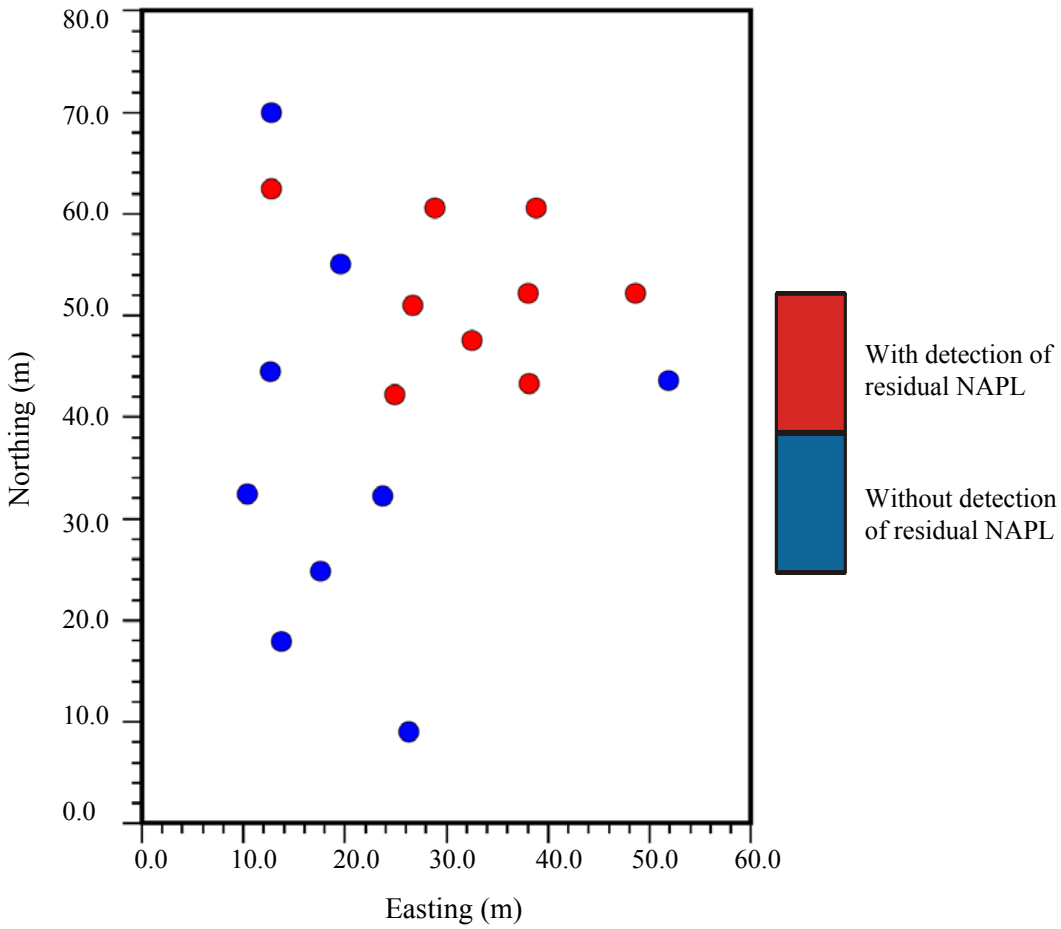


Figure 3-5: Locations of CPT-UVIF cone holes at the former flare pit site. The solid red circles show the data locations with detected residual NAPL and the solid blue circles show the data locations without any residual NAPL observed.

Figure 3-6-a shows the bimodal histogram of the soil classification index (SCI) for all data (after declustering) based on the data collected by 18 CPT holes across the site. By grouping SBTs together based on their hydraulic properties, we are able to define three hydro-facies with relatively distinct statistical (and hydraulic

properties). Figure 3-7 and Table 3-2 display the definition of three distinct facies, namely higher permeability unit (HP), medium permeability unit (MP) and lower permeability unit (LP) based on their hydraulic properties using CPT soil behavior type chart by Robertson et al. (1986) and their global proportions. The histograms of SCI attribute for the defined facies HP, MP and LP are shown in Figures 3-6-b, 3-6-c and 3-6-d, respectively.

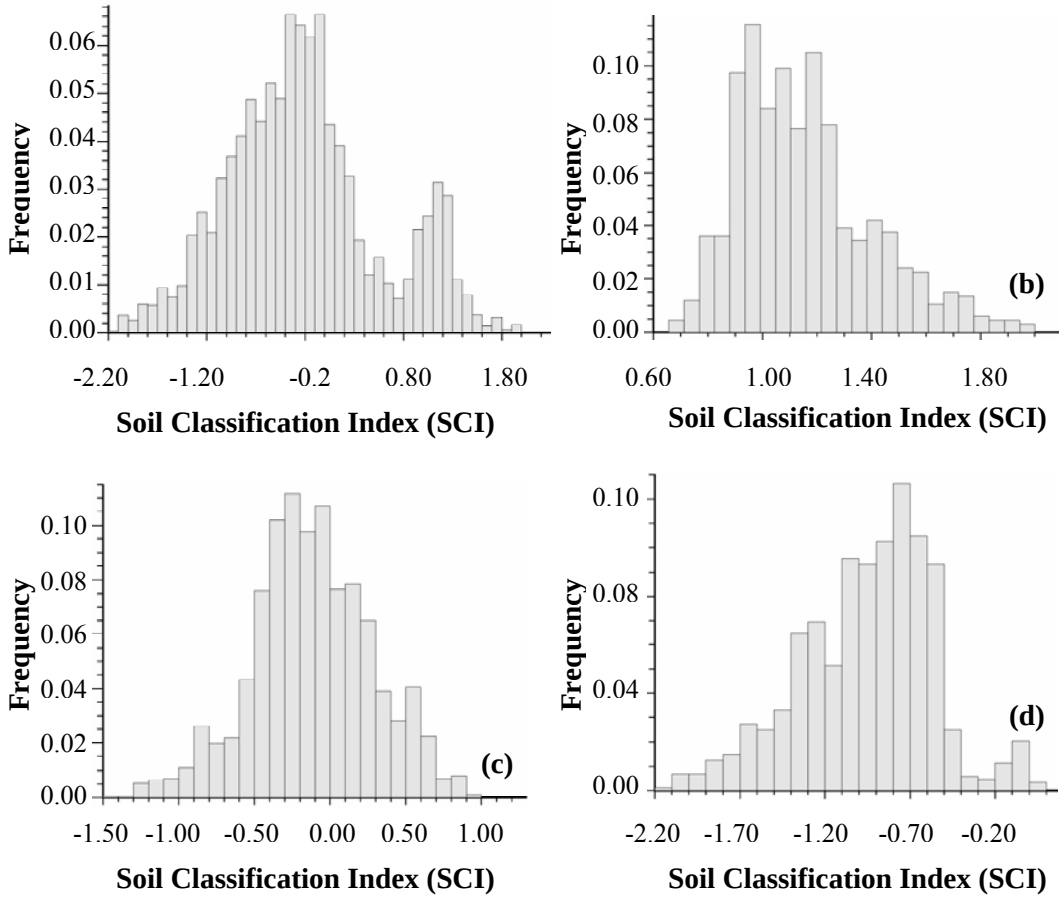


Figure 3-6: (a) The histogram of the SCI for all data – after declustering, and the histograms of the SCI for (b) HP facies, (c) MP facies and (d) LP facies.

It is still necessary to delineate the correct directions of geological continuity and detect any inclination in geological units. This requirement is primarily due to the

fact that in geostatistical analysis the model is constructed on a Cartesian grid (Deutsch 2002).

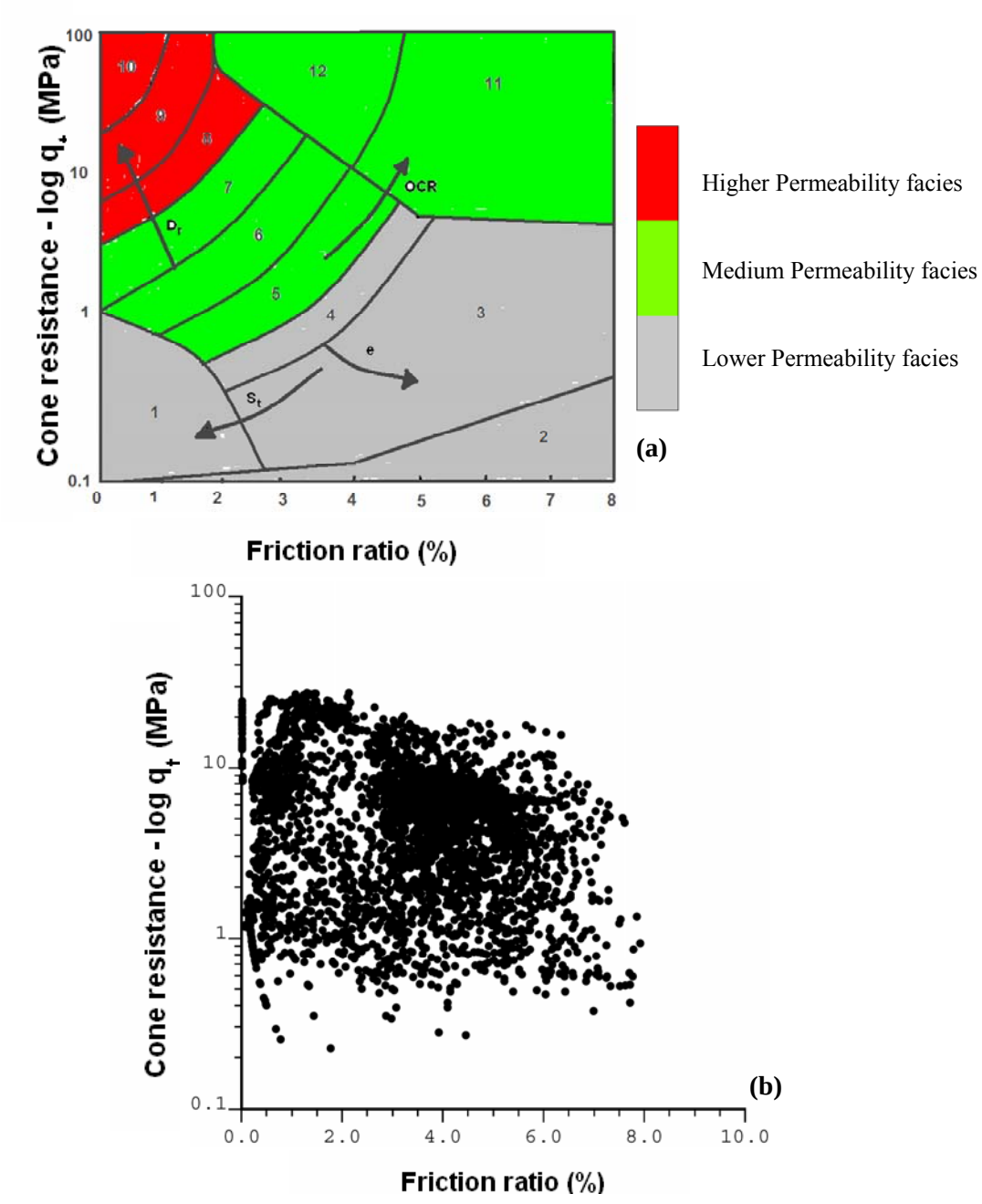


Figure 3-7: (a) Grouping SBTs into three categories (HP, MP and LP) based on their estimated permeability values, (b) the cross-plot of cone (tip) resistance versus friction ratio values obtained from the set of 18 CPT holes across the modeling domain.

Facies Type	*Global Proportion (%)*	*Range of permeability (m/s)*
LP	*29.5*	1×10^{-10} to 3×10^{-8}
MP	*55*	1×10^{-8} to 1×10^{-5}
HP	*15.5*	1×10^{-5} to 1×10^{-3}

Table 3-2: Global proportions and range of permeability values for HP, MP and LP facies; the ranges of permeability values are obtained based on the assumption of normally consolidated soil.

Thus, prior to calculation of directional variograms for different categories, a vertical coordinate transformation must be performed, considering various common deposition-erosion scenarios. As a standard practice, horizontal experimental variograms are calculated after coordinate transformation and the scenario showing highest degree of correlation is retained and the rest of the geostatistical modeling is done in the new coordinate system. The final results are back-transformed to the original coordinate system. Four different scenarios are schematically shown in Figure 3-8. As shown in Figure 3-9, the resulting stratigraphic variograms for different scenarios closely resemble each other for different categories. However, for MP facies (with the highest global proportion), slightly higher correlation is observed in the original vertical coordinate before stratigraphic transformation (elevation scenario). Thus, no stratigraphic transformation is implemented for this site and the original coordinate system is used in the subsequent calculation of experimental variograms.

Indicator kriging (IK) and simulation (Deutsch and Journel 1998) are used to directly estimate the distribution of uncertainty in categorical variables. The first step in indicator formalism is to code the data as indicator values:

$$i(\mathbf{u}_\alpha; z_k) = Prob\{\text{soil type } k \text{ present}\} = \begin{cases} 1, & \text{if soil type } k \text{ is present at } \mathbf{u}_\alpha \\ 0, & \text{otherwise} \end{cases} \qquad [3.2]$$

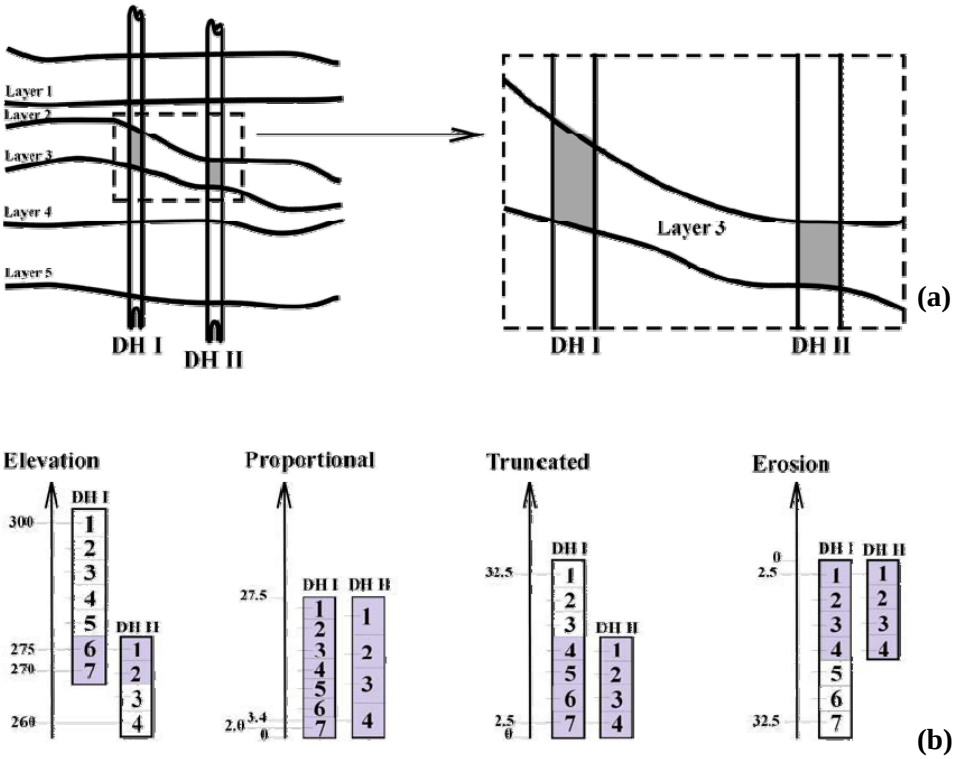


Figure 3-8: (a) A five layer geological structure with soil samples taken from Layer 3, (b) and four different stratigraphic transformation scenarios: elevation, proportional, truncated and erosion. The shaded composites represent horizontal variogram calculation pairs (after McLennan 2004).

The stationary prior probabilities of different soil types ($p(k),\ k=1,\ 2,\ 3$) have been determined using the histogram of declustered data (Table 3-2). According to [3.2], residuals can be calculated as:

$$Y(\mathbf{u}_\alpha; z_k) = i(\mathbf{u}_\alpha; k) - p(k) \qquad \text{with} \qquad \alpha = 1, 2, ..., \text{n} \quad \text{and} \quad k = 1, 2, 3 \qquad [3.3]$$

Kriging of the residuals is used to derive the probability of occurrence of each soil type at each unsampled location. Thus, the model of uncertainty at every unsampled location **u** will be (Deutsch 2002):

$$p_{IK}(\mathbf{u}; k) = \sum_{\alpha=1}^{n} \lambda_\alpha(k).Y(\mathbf{u}_\alpha; z_k) + p(k) \qquad \text{with} \qquad k = 1, 2, 3 \qquad [3.4]$$

where, the subscript *IK* denotes indicator kriging, λ_α's are kriging weights that are calculated by (simple) kriging equations and account for closeness to data

locations as well as overall uncertainty in data and redundancy in nearby data. The simple kriging equations are expressed by (Goovaerts 1997):

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}; z_k) C_I(\mathbf{u}_\alpha - \mathbf{u}_\beta; z_k) = C_I(\mathbf{u}_\alpha - \mathbf{u}; z_k) \quad \text{with} \quad \alpha = 1, ..., \mathrm{n}(\mathbf{u}) \qquad [3.5]$$

where, $C_I(\mathbf{h}; z_k)$ is the covariance function of the indicator random function $i(\mathbf{u}_\alpha; z_k)$. To solve the simple kriging system of equations, the covariance function must be calculated for different lags and in different directions. For this purpose, experimental indicator variogram calculation and indicator variogram modeling is performed. Figure 3-9 and Table 3-3 show the details of variogram calculation and modeling. As stated before, all variogram calculation and modeling is done in original coordinates system (elevation scenario). As shown in Table 3-3, spherical type variogram has been used in all the nested structures that are modeled for the directional indicator variograms. The spherical variogram is defined by:

$$\gamma(h) = c.\mathrm{Sph}\left(\frac{h}{a}\right) = \begin{cases} c.\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right], & \text{if } \mathrm{h} \le \mathrm{a} \\ c, & \text{if } \mathrm{h} \ge \mathrm{a} \end{cases} \qquad [3.6]$$

where, a is the range, and c represents a positive variance contribution or *sill* value. In the context of indicator kriging (or simulation) indicator variograms are calculated for each category separately, using the available site-specific data. The horizontal solid lines on the variograms show the *sill*, which represents maximum variability associated with each soil type in the modeling domain. It can be calculated by $p(1-p)$ where $p = F(z_k)$ is the global proportion of indicator variable before declustering. The range of a variogram is the distance where the the variogram reaches the *sill*. As the range becomes larger, smaller variability is observed in nearby data. Comparing the ranges of vertical and horizontal variograms in Figure 3-10, the effect of stratification can be clearly observed, since the vertical range is much smaller than the horizontal range.

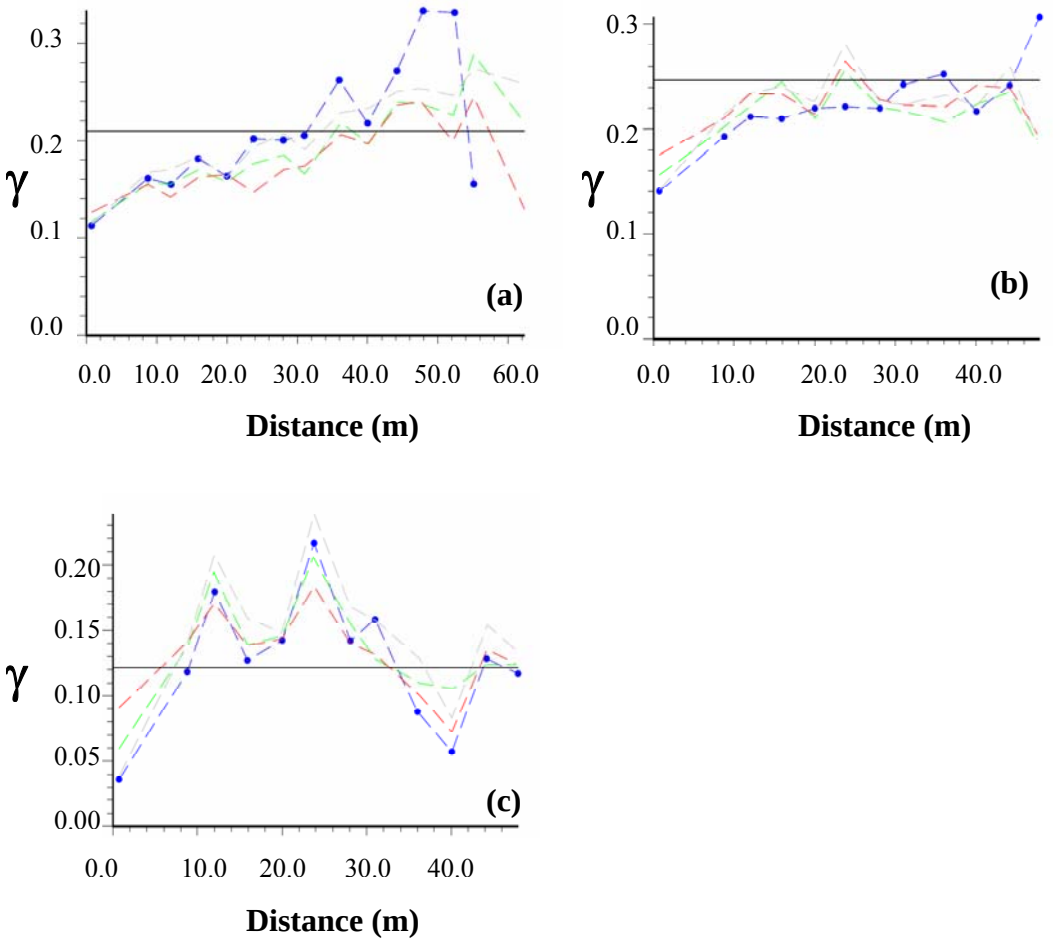


Figure 3-9: (a) Stratigraphic horizontal variograms for facies LP, (b) facies MP, and (c) facies HP. The blue, green, red and grey lines represent elevation, proportional, truncated and onlap scenarios.

Sequential Indicator Simulation (SIS) (Journel and Gomez-Hernandez 1993) is a Monte Carlo simulation technique built on Indicator Kriging (IK) explained above. In order to populate the whole modeling domain with simulated values, grid nodes are visited sequentially in a random path. At each grid node the following procedure is repeated: (1) searching for nearby data and previously simulated values, (2) performing IK to build a distribution of uncertainty, and (3) drawing a simulated value from the distribution of uncertainty.

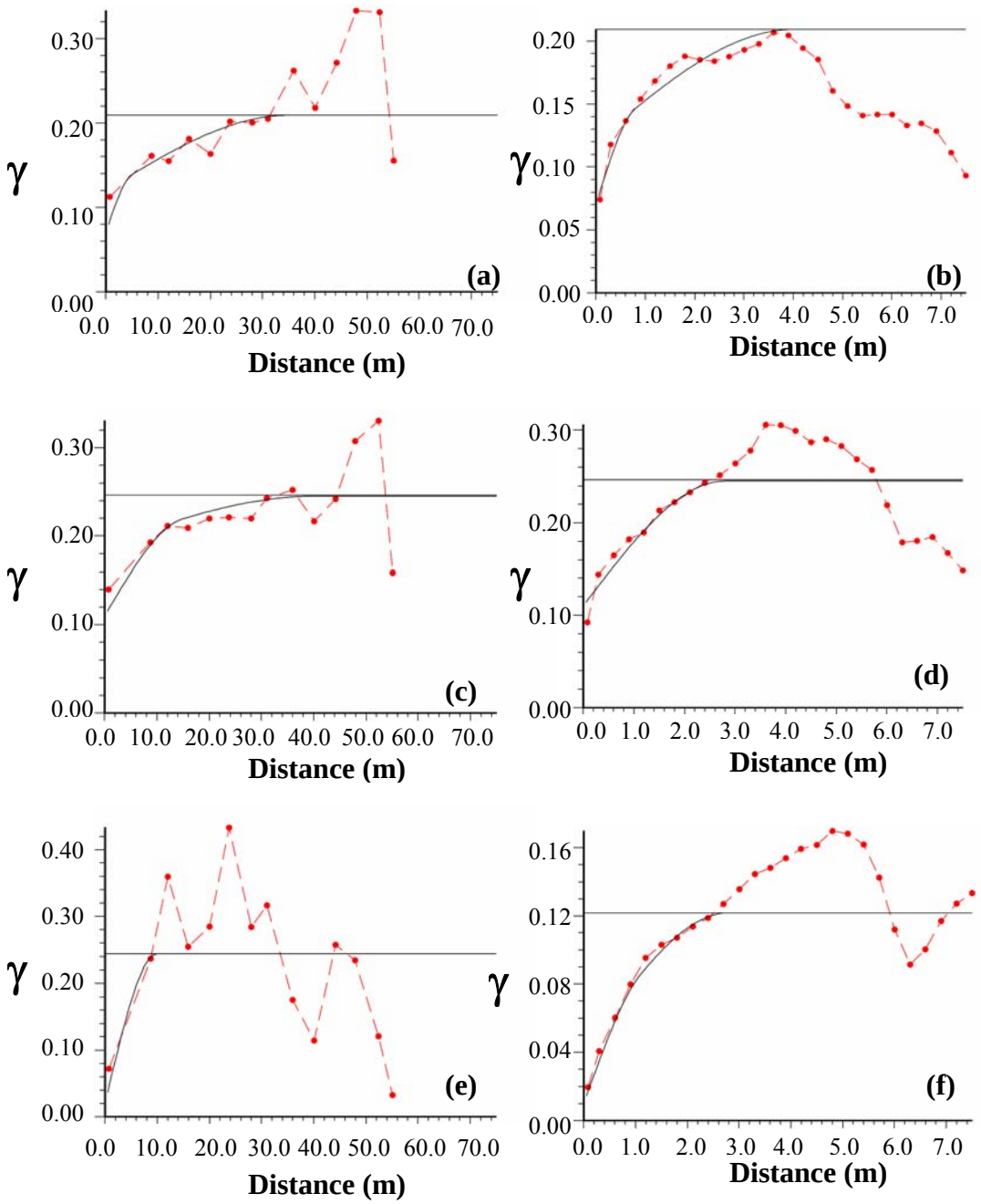


Figure 3-10: (a, b) Horizontal and vertical variograms for facies LP, (c, d) for facies MP and (e, f) for facies HP.

Category	Nugget	c_1	NS1	a_{h1}	a_{v1}	c_2	NS2	a_{h2}	a_{v2}
LP	0.07	0.05	*sph*	5.	0.8	0.0891	*sph*	35.	4.
MP	0.11	0.08	*sph*	15.	2.8	0.055	*sph*	40.	2.8
HP	0.01	0.03	*sph*	10.	1.1	0.0818	*sph*	10.	2.8

Table 3-3: Details of the variograms modeled for facies LP, MP and HP.

It should be noted that the series of indicator-derived probability values must be non-negative and sum to one. This is not guaranteed by indicator kriging and an a-posteriori correction is applied by re-standardizing the conditional probabilities to sum to one. A large number of equi-probabale realizations of 'soil-type' are generated. These realizations reproduce the input data equally well. A number of checks should be done to validate the geostatistical model. These checks include reproduction of input statistics such as histogram and variograms, honoring input data, consistency with the available information about geology of the site and closeness of estimated probabilities to the true soil types. Four indicator realizations of soil-type are shown in Figure 3-11. The realizations are clipped by ground topography that is modeled by (ordinary) kriging with surface elevation data available at boreholes and CPT holes. The SIS realizations often show unrealistic short-scale variations. The other concern is that the facies proportions often depart from their target input proportions. Especially, facies types with relatively small proportions may be poorly matched. The main source of this discrepancy is the order-relations correction algorithms. Therefore, a maximum a-posteriori selection (MAPS) technique (Deutsch and Journel 1998) was implemented to remove the unrealistic short-scale variations and to ensure the reproduction of input global proportions for all realizations.

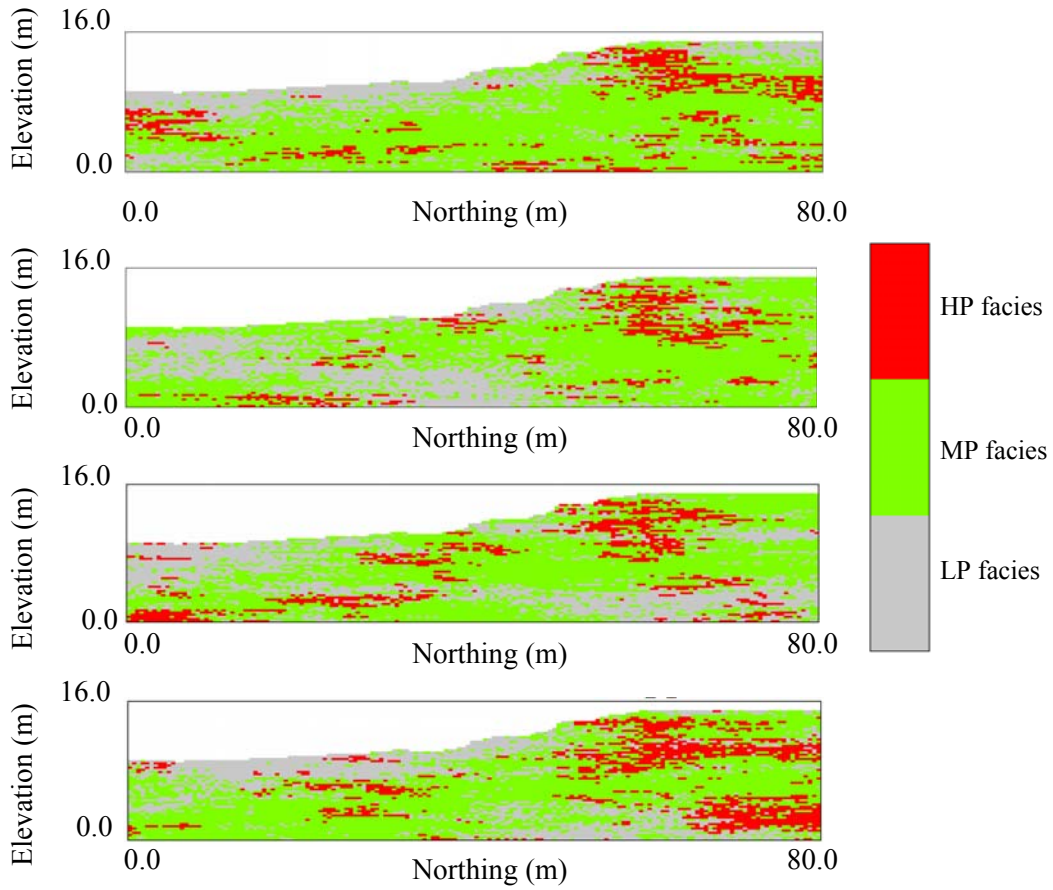


Figure 3-11: Cross-sectional view of four SIS realizations of soil-type. It is common in all realizations to see the higher permeability facies (HP) to be present at the northern part of the site.

The generated SIS realizations can be used for 3D stochastic contaminant transport simulation purposes by assigning average/calibrated hydraulic conductivity values (from Table 3-1) to each hydro-facies (as defined in Figure 3-6-a), while avoiding the non-stationarity problems. Application of CPT data in construction of 3D statistical models for hydraulic conductivity distribution can be considered useful and is preferred to traditional techniques such as slug tests, as CPT data are collected in a high-resolution fashion and can provide detailed information about layering in the aquifer. Rigorous calibration of hydraulic conductivities of the hydro-facies is a preferred approach as compared to taking

the average values from the Table 3-1. Such calibration can be implemented locally using the average values obtained by traditional point measurement approaches as well as in the field-scale by calibration to head data. It is, however, out of the scope of this research.

A posterior probability map (E-Type mean) can also be created to give a better understanding of the site conceptual model in terms of its hydraulic response. By combining the understanding from the site geology with the understanding of hydrogeology of the site (obtained by the studying the recorded groundwater table elevations at boreholes and cone holes as well as the geostatistical analysis), a simple conceptual hydrogeological model can be developed for the site. As shown in Figure 3-12, the hydrogeological system of the site is composed of an unconfined aquifer, with some perched water tables in the unsaturated zone. The site is mainly composed of glacial till; however, a relatively large sandy unit is embedded in the unsaturated zone at the northern part of the site. As will be explained in Section 3.2, the large sandy unit at the north of the site has an observed correlation with presence of residual NAPL contaminants.

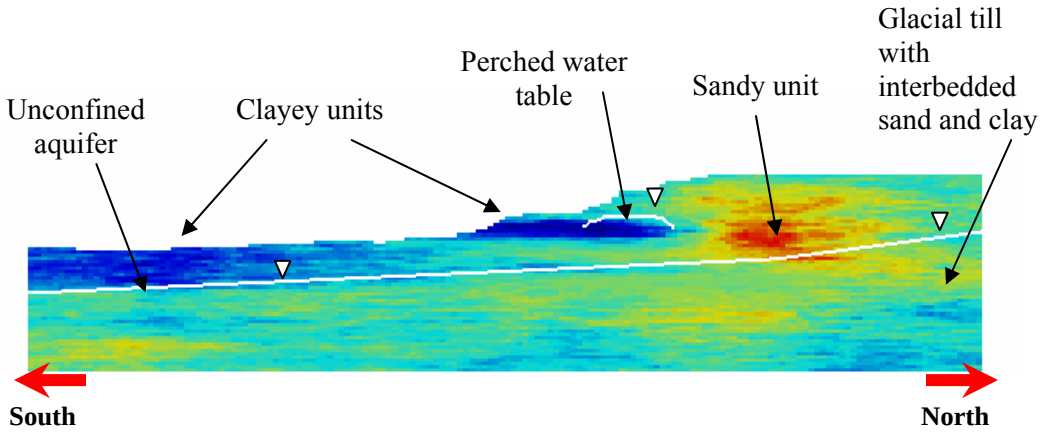


Figure 3-12: Development of a hydrogeologic SCM for the study site. The geology of the site is mainly composed of glacial till with a large sandy unit at the north of the site. The groundwater system is an unconfined aquifer with some perched water tables in unsaturated zone.

As a summary, in Section 3.1 of this thesis, a case study was presented where soil behavior type (SBT) charts were used directly to define categorical and continuous variables to construct a probabilistic model for hydrogeological properties of a real aquifer. The definition of categorical variables (hydro-facies) based on CPT data and construction of the associated probabilistic model can be used directly in 3D Monte Carlo simulations of contaminant transport or as a prior model for calibration of groundwater models (similar to the work of Poeter and McKenna 1995). The definition of continuous variables based on CPT data will subsequently be used (Section 3.2) in calibration and prediction of presence/absence of NAPL contamination given soil texture as a secondary variable.

3.2 Geostatistical Modeling of Residual NAPL Source

In contamination scenarios associated with petroleum hydrocarbons, a large portion of free phase hydrocarbon is recoverable using available technologies. It is very difficult, however, to recover residual NAPL with in-situ technologies and it usually becomes a long-term source of groundwater contamination. Despite several complex models proposed to quantify the rate of mass transfer between residual hydrocarbon and groundwater, less attention has been paid to the geometry of residual NAPL sources and their uncertainty in fate and transport modeling studies. In unconfined aquifers, distribution of residual LNAPL in the subsurface depends mostly on the distance to the point of release, soil texture and seasonal water table fluctuations.

When oil LNAPL is accidentally released, it migrates vertically and laterally under the gravity and capillary forces. When the volume of the release is sufficient, the LNAPL will migrate through the unsaturated zone to the capillary fringe and the water table (Figure 3-13-a). Due to capillary forces, some LNAPL is always retained in the soil pores as residual or immobile NAPL. In fact, LNAPL coexists with water (and air) in the soil pores. LNAPL saturations are always less than 100 percent but may range from as little as 5 percent to over 70 percent (Figure 3-13-b). As the remaining mobile LNAPL continues to migrate through the subsurface, the volume of mobile product decreases as NAPL becomes trapped as isolated droplets within the soil pore network. Thus, LNAPL plumes are spatially self-limiting, unless continually supplied from an ongoing release (API 2004). While migrating through the subsurface, LNAPL is significantly affected by the heterogeneous nature of the soil strata: slight differences in soil texture may promote preferential pathways within the aquifer horizontally and vertically (Essaid et al. 1993). Also, LNAPL is significantly influenced by vertical fluctuations in the water table. These fluctuations enhance

the development of the residual LNAPL in the smear zone. The residual NAPL is almost impossible to be removed and creates a long-term source of pollution as it partitions slowly into the aqueous and vapor phases.

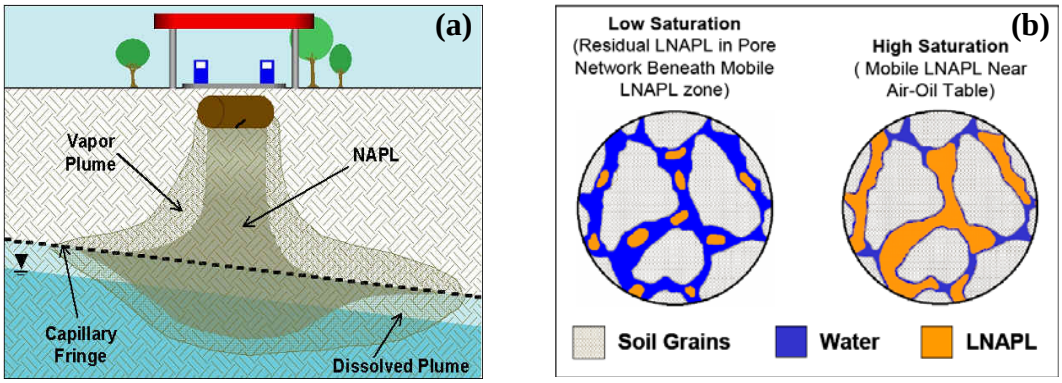


Figure 3-13: (a) A schematic representation of LNAPL release in the subsurface, and (b) Representation of LNAPL saturations in the pore space of saturated zone (API 2004)

A review of the literature shows a number of techniques are proposed to quantify the distribution of NAPL sources. They can be grouped into analytical techniques, numerical techniques and inverse modeling techniques (reviewed in Chapter 1).

An analytical solution has been presented by Huntley and Beckett (2002) based on soil capillary equations of van Genuchten (1980) to quantify the vertical saturation profiles for the LNAPL/water/air system across the capillary fringe. They then estimated the groundwater flux through the source zone based on the relative permeabilities influenced by phase saturations, and calculated the emissions to the aqueous phase. Although such analytical approaches are relatively simple and easy to implement, they are limited because they can only be applied to very simple geometries of NAPL source zones. As discussed above, the effect of soil texture heterogeneities and NAPL redistribution by temporal groundwater table fluctuations result in NAPL source geometries that are far more complex than what is modeled by analytical techniques. Thus, alternatively, a

number of numerical approaches have been proposed to simulate the LNAPL distribution and resulting contaminant emissions, whereby the geologic heterogeneity, transient flow conditions and dissolution factors can be considered. In multiphase flow modeling studies carried out at the site of a crude oil spill, Essaid et al. (1993) and Dillard et al. (1997) reproduced general large-scale features of the observed oil body. However, both studies had limited success in reproducing local oil saturations and highlighted the importance of accounting for the uncertainty in the hydrogeological parameter values. Essaid and Hess (1993) presented a stochastic modeling study of oil infiltration into a hypothetical glacial outwash aquifer by performing a series of Monte Carlo simulations with different spatial permeability distributions. They were able to reproduce the ensemble mean saturations by using the mean hydraulic properties. They, however, observed that localized oil saturations for individual realizations deviated considerably from those obtained using uniform mean properties as input. Similar observations have been reported by Kueper and Gerhard (1995). The works of Zhu (2001) and Chen et al. (2006) are among the other related works in this area.

Given the fact that mechanisms that govern the distribution/re-distribution of residual NAPLs are quite complex and site-specific, development of probabilistic models using site-specific data is required. A number of field investigation approaches have been proposed for collection of site-specific data in NAPL source zones. The ultra-violet-induced-fluorescence cone penetration testing or CPT-UVIF (Pepper et al. 2002, Kram et al. 2004) is an advanced technology that is commonly used for this purpose. To the author's knowledge, the work of D'Affonseca et al. (2008) is the only work that makes use of UVIF data for development of a model for DNAPL sources. Their work is deemed limited due to the fact that they perform a deterministic calibration between the UVIF response and soil properties.

As an alternative to previous approaches, a forward probabilistic model is proposed. The presented probabilistic model accounts for uncertainty in areal limits (through explicit boundary modeling) as well as the effects of soil texture and water table fluctuations (through geostatistical data integration). This work is particularly aimed at contamination by residual LNAPL, as the pools of residual LNAPL are quite consistent relative to the dynamics of groundwater flow and transport; and a static distribution can be reasonably assumed for residual LNAPL over a long period of time. It is also assumed that the approximate location of the source zone is roughly delineated by the observation wells. In other words, the proposed methodology gives a tool to evaluate the uncertainty in the areal extent and spatial distribution of a source zone whose existence is proven and its location is roughly known.

In this section, boundary modeling approaches are briefly reviewed; and the development and application of the distance function (DF) algorithm for modeling the areal limits of source zones are explained. In the next section of this chapter, a categorical stationary random variable is defined using CPT-UVIF data to represent presence/absence of residual LNAPL contamination. Two different geostatistical data integration techniques, namely sequential indicator simulation with locally varying means (LVM) and Bayesian updating (BU), are used to simulate the stationary random variable within the areal extent of the residual LNAPL plume with soil texture and the distance to water table as secondary variables. Assumptions of full data independence and conditional independence are adopted to combine secondary data sources. Subsequently, cross-validation is implemented to show the value of secondary data in improving the predictive ability and compare the performance of the two geostatistical techniques used.

3.2.1. The Distance Function (DF) Algorithm: Introduction

Geostatistical estimation and simulation techniques are applied with Stationary Random Functions (SRF) within geometric limits imposed by areal (or volumetric) bounding limits (or surfaces). There is often significant uncertainty in areal boundaries depending on the amount of data available. Thus, point measurements of soil NAPL concentrations may not be directly used in geostatistical modeling of contaminant source zone, unless the areal (volumetric) boundaries of the stationary domain is reasonably defined and its uncertainty is characterized. In fact, quantifying this uncertainty with probabilistic boundary models is an essential aspect of making a reasonable decision of stationarity.

A number of boundary modeling approaches have been proposed which range from entirely deterministic techniques to entirely stochastic techniques. Deterministic Digitization for mining applications (Houlding, 1994), volume function (Cowan et al. 2003), object-based modeling (Bridge and Leeder, 1979), surface-based modeling (Pyrcz et al. 2005), and stochastic pixel-based techniques such as sequential indicator simulation (Journel and Gomez-Hernandez 1993) are among the relevant techniques. The volume function approach represents boundary surfaces indirectly and results in a smooth volume function within which a constant valued surface representation of the desired boundary exists. It is simple and flexible. However, it needs a large amount of hard data and there is no access to global uncertainty. The distance function algorithm is similar to the volume function approach. But, it has been modified to satisfy the condition of unbiasedness and to have direct access to the global uncertainty. All other techniques (object-based modeling, surface-based modeling and stochastic pixel-based techniques) either require a large amount of hard and soft data that makes them limited when data is sparse, or require a-priori knowledge about the location of boundaries, or the possible shape of the object (source zone) being simulated, which are not readily available.

The definition of the DF is related to the notion of distance to an interface separating two distinct domains within which two different SRF will subsequently be developed for geostatistical modeling. Distance is measured to the nearest unlike data location. Distance can be positive or negative depending on direction from the inside or outside the source. Thus, a first guess for the bounding interface of interest would be the line corresponding to a constant value of DF=0. The DF smoothly varies between increasingly positive values inside and further away from the boundary interface to increasingly negative values outside and further away from the boundary interface.

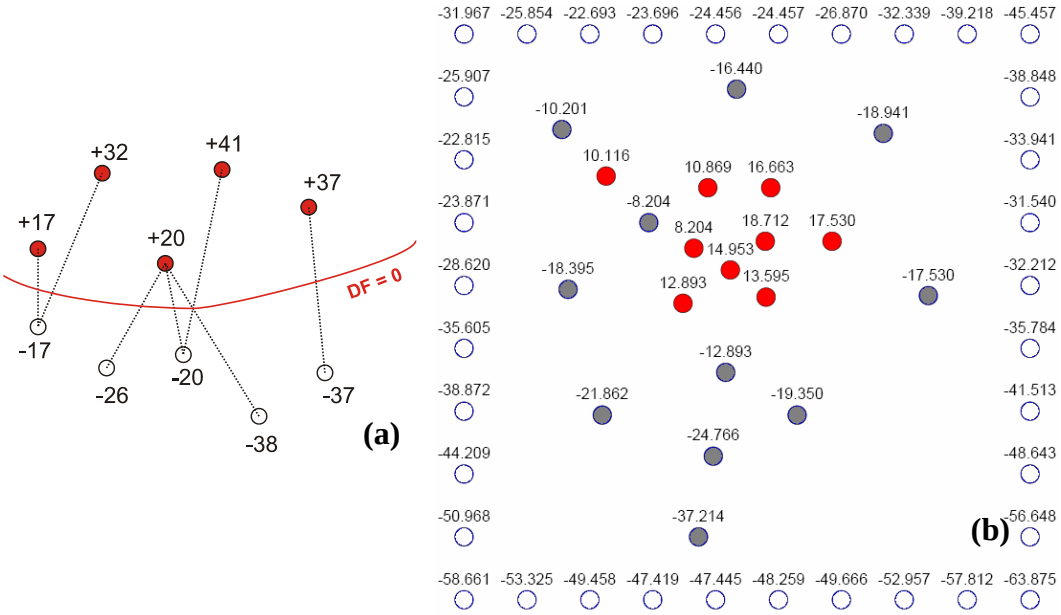


Figure 3-14: (a) Calculation of DF for sample points deemed inside and outside of the source zone and a first guess for the location of boundaries with tracing the line of DF = 0.0. (b) A boundary modeling example with an arbitrary number of control points (white circles) that are added at locations outside of the areal limits of interest. Red and grey circles show a layout of sample points deemed contaminated and uncontaminated, respectively.

An initial binary coding of the available sample data in terms of ‘inside’ and ‘outside’ the source zone is a prerequisite for constructing a DF. The DF is calculated for all sample points for subsequent interpolation. Figure 3-14 shows

(1) sampling locations and control points are coded for being inside and outside of the source zone, (2) DF calculated as the distance to the nearest unlike data location, and (3) an initial guess for the boundaries that traces the line corresponding to DF equal to zero.

If sampling locations are sparse, an interpolation technique must be employed to define the boundary interface. The algorithm used to interpolate the DF values should produce realistic variations in the DF across the domain. Kriging and inverse distance are flexible geostatistical estimation algorithms that can generate suitably smooth DF estimates for realistic boundary interfaces. In this work, inverse distance interpolation is employed, as negative weights are avoided and, by definition, the DF attribute is not a stationary variable. An inverse distance estimate at an unsampled location $\mathbf{u}_0$, is a weighted linear combination of N surrounding distance function data, $DF(\mathbf{u}_i)$, in a search neighborhood:

$$DF^*_{ID}(\mathbf{u}_0) = \sum_{i=1}^{N} \lambda_{ID}(\mathbf{u}_i) \cdot DF(\mathbf{u}_i) \qquad [3.7]$$

The weights, $\lambda_{ID}(\mathbf{u_i})$, are calculated by:

$$\lambda_{ID}(\mathbf{u}_i) = \frac{\dfrac{1}{\left(d(\mathbf{u}_i)\right)^{\omega} + c}}{\displaystyle\sum_{j=1}^{N} \frac{1}{\left(d(\mathbf{u}_j)\right)^{\omega} + c}} \qquad [3.8]$$

where, $d(\mathbf{u}_i)$ is the Euclidian distance between the estimation location $\mathbf{u}_0$ and the sample data at location $\mathbf{u}_i$, ω is the distance exponent and c is a constant controlling short-scale variations. The exponent ω controls the smoothness of the inverse distance estimates, which typically is between 0.5 and 2.0. The search neighborhood for inverse distance interpolation is often calibrated by cross-validation to limit the number of data used in interpolation (Rojas-Avellaneda and Silvan-Cardenas, 2006). In the context of the DF algorithm, larger search neighborhoods can ensure smooth estimates. DF conditioning data are reproduced exactly. The continuous DF is represented discreetly on a Cartesian grid. As the

grid resolution increases, a smoother boundary interface is obtained. The DF is not a stationary random function, and traditional probabilistic approaches cannot be used for uncertainty assessment. In this work, an alternative novel approach is proposed by introducing the concept of an uncertainty band and its associated parameterization.

3.2.2. The DF Algorithm: Quantification of global uncertainty – data conditioned estimation

Kriging and inverse distance weights are typically dependent on the geometrical arrangement of data, and not on the actual data values. McLennan (2007) presented an estimation methodology with data-dependent weights in kriging or inverse distance interpolation. The modified form of inverse distance interpolation can be represented by:

$$DF^*_{ID}(\mathbf{u}_0) = \sum_{i=1}^{N} f^{DC}(\mathbf{u}_i) \cdot \lambda_{ID}(\mathbf{u}_i) \cdot DF(\mathbf{u}_i) \qquad [3.9]$$

where, $f^{DC}(\mathbf{u}_i)$ is the data conditioning factor that depends on the actual numerical values of the data (not their geometrical arrangement). It is set equal to 1.0 in classical kriging and inverse distance. All other terms have been defined previously. The choice of $f^{DC}(\mathbf{u}_i)$ to DF data is crucial in this approach and corresponds to different positions of risk. As shown in Figure 3-15, dilated (larger) and eroded (smaller) boundaries can be generated with linear and quadratic parameterization. The negative sloping line gives gradually higher weight to DF samples that are more negative and gradually lower weights to DF samples that are more positive, producing a DF distribution from which eroded smaller boundary may be extracted. Virtually any amount of uncertainty can be tuned in with this approach. In the work of McLennan (2007), a linear parameterization was used and optimistic and pessimistic boundaries were identified by 'selection' of a number of different values for $f^{DC}_{\min}$.

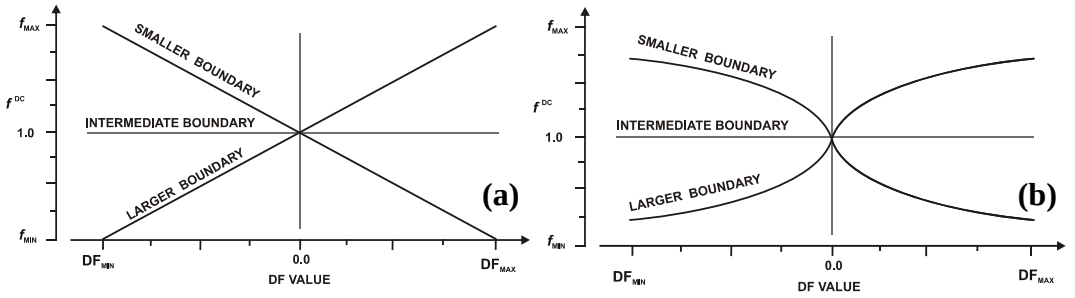


Figure 3-15: (a) Linear and (b) quadratic parameterization of data conditioning factors for dilated (larger), intermediate and eroded (smaller) boundaries (McLennan, 2007).

In this work, a more general parameterization scheme is used for data conditioning, which makes it possible to define a band of uncertainty and 'calibrate' the data conditioning factors for a given well arrangement. The modified inverse distance interpolation with the proposed data conditioning is given by:

$$DF^*_{ID}(\mathbf{u}_0) = \sum_{i=1}^{N} \alpha^{\frac{DF(\mathbf{u}_i)}{|DF(\mathbf{u}_i)|}} \cdot \lambda_{ID}(\mathbf{u}_i) \cdot \left[DF(\mathbf{u}_i) + \beta \cdot \frac{DF(\mathbf{u}_i)}{|DF(\mathbf{u}_i)|} \right] \qquad [3.10]$$

where, α and β are scaling and separation factors, respectively. All other terms have been defined previously. α can take any value greater than zero. Similar to $f^{DC}(\mathbf{u}_i)$, increasing or decreasing α results in dilated or eroded boundary domains. The effect of changes in α on the location of the boundaries (DF = 0.0) for the given well arrangement in Figure 3-13, is illustrated in Figure 3-16. The separation factor, β, is closely related to the notion of the uncertainty band and can take any value greater than or equal to zero. For a given well arrangement, the uncertainty band (UB) is defined as a probabilistic areal interval that, with a specific level of confidence includes the unknown actual boundary. Figure 3-17 shows an uncertainty band can be characterized by its width and centerline, and a uniform probability distribution. Other statistical distributions could be adopted.

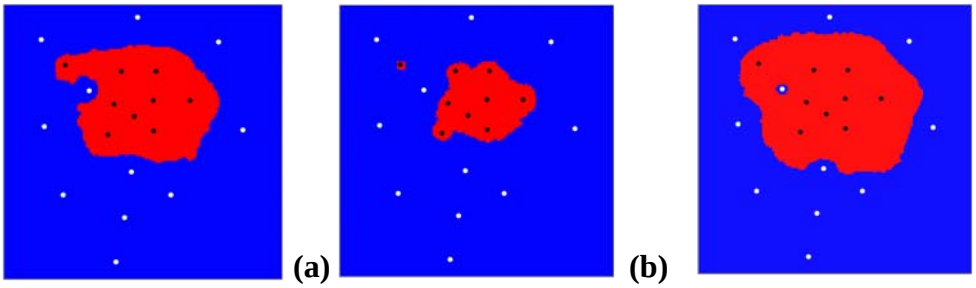


Figure 3-16: Location of boundaries corresponding to different scaling factors: (a) $\alpha = 1.00$, (b) $\alpha = 0.50$ and (c) $\alpha = 1.50$; with contaminated wells shown as black circles and uncontaminated wells shown as white circles.

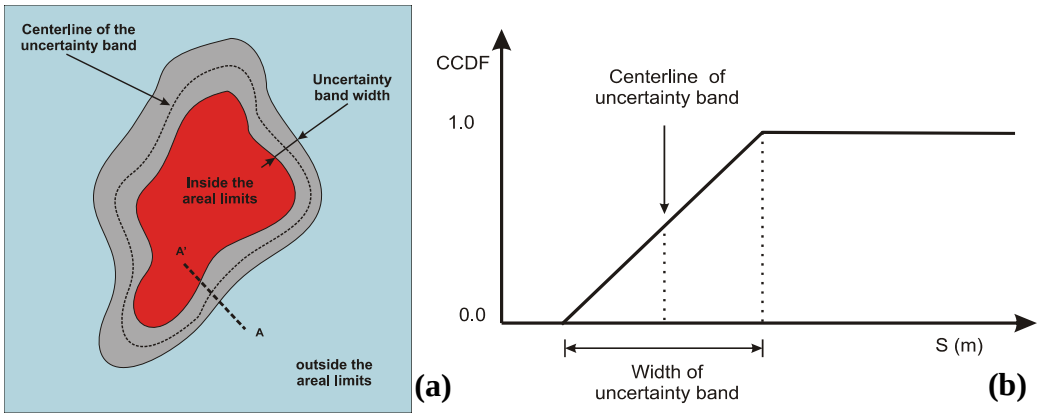


Figure 3-17: (a) Uncertainty band and (b) its conditional cumulative distribution function along A-A' cross-section.

As shown in Figure 3-18, α and β control the width (WUB) and the location of centerline (CUB) of the uncertainty band, and the uncertainty band is bounded by dilated and eroded boundaries. Figure 3-19 shows the uncertainty bands for different α and β values for the well arrangement shown Figure 3-13. As expected, the proposed data conditioning methodology results in wider uncertainty bands in areas where sample locations appear to be more distant. Control points can be added to locations which are known a-priori to be inside or outside of the contaminated area (Figure 3-14).

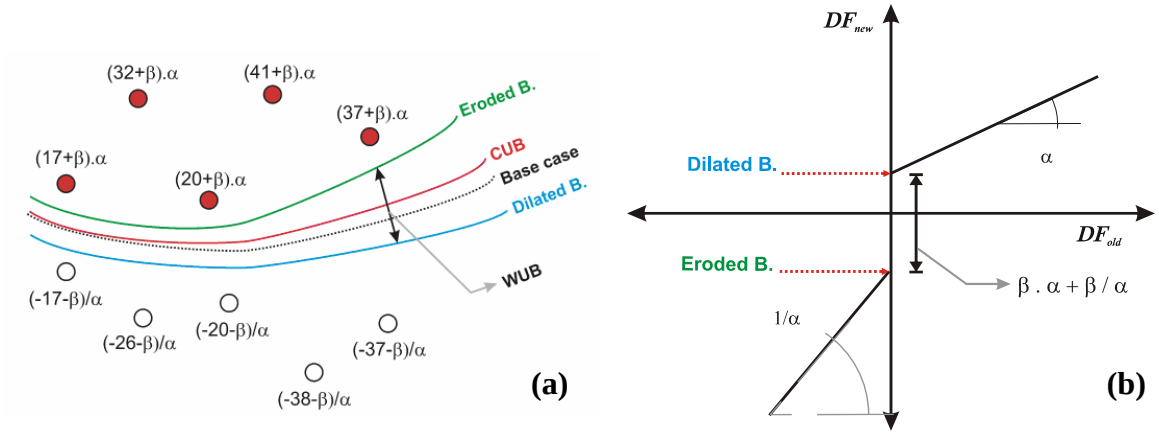


Figure 3-18: (a) The centerline and width of the uncertainty band, and data conditioning by α and β and (b) distance functions before and after data conditioning (α <1.0).

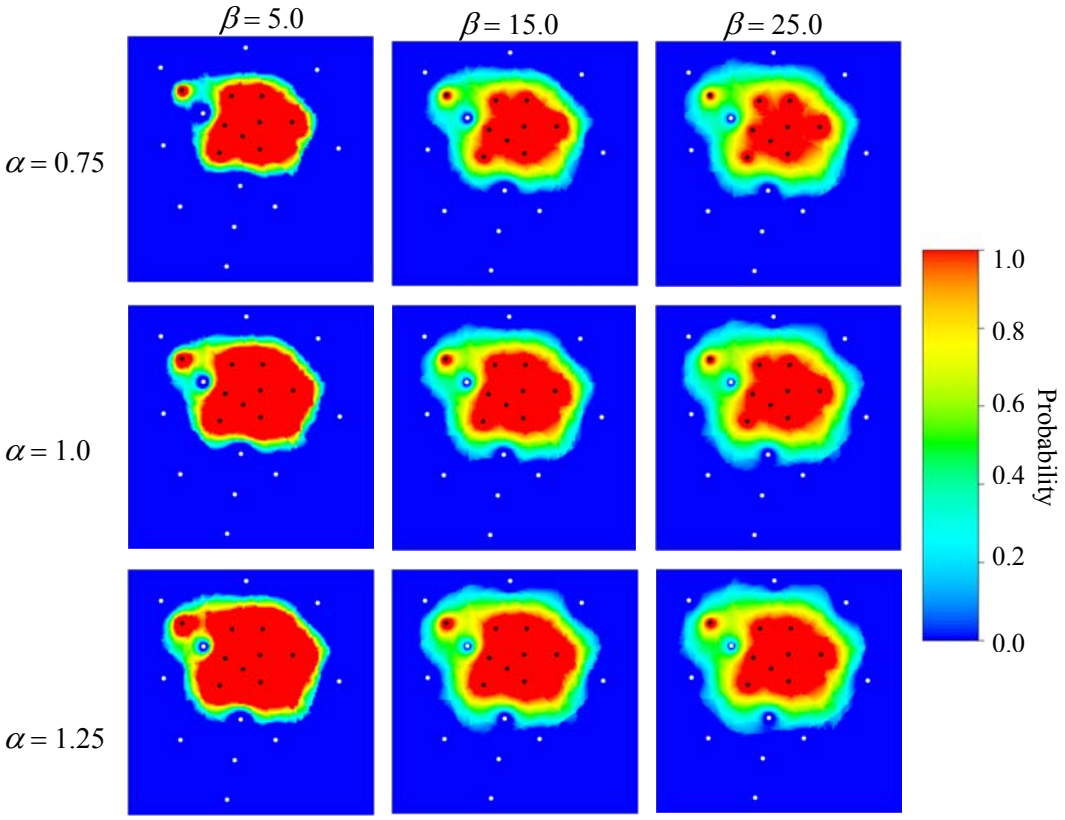


Figure 3-19: The centerline and width of uncertainty band, and data conditioning by α and β

3.2.3. The DF Algorithm: Quantification of global uncertainty - Calibration of the parameters

The space of uncertainty for areal limits is parameterized by scaling and separation (α and β) factors. An appropriate scaling factor, α, ensures the unbiasedness of the resulting uncertainty band and a proper separation factor, β, ensures the fairness of the probability distributions. In essence, α and β should be calibrated. In this work, a large number of synthetic limits are generated for the given well arrangement to calibrate scaling and separation factors. The generated synthetic boundaries are considered as ensemble of true boundaries for the given well configuration, and the α and β factors that lead to unbiased estimation of the plume size and fair uncertainty assessment are determined. More details are given below.

First, all observation wells and control points are coded as 1's (contaminated wells) or 0's (uncontaminated wells), and a directional search angle, θ, is specified. Then, for every contaminated well (coded as 1), a directional search is implemented and the directions that include a closest unlike (coded as 0) data location are identified as valid search directions and those include a closest like (coded as 1) data location are identified as null search directions (Figure 3-20). In order to generate a synthetic plume, for each contaminated well (coded as 1) a search direction is randomly selected. The selected search direction can be either a null search direction or a valid search direction. If a valid search direction is selected, a new imaginary well is added to the setting. This imaginary well is randomly located on a line that connects the original contaminated well to the closest unlike (uncontaminated) well location. The new imaginary well is randomly coded as either contaminated or uncontaminated. The DFs are then recalculated for the new setting and mapped by inverse distance interpolation. The line of DF = 0.0 is traced to create a new realization. The size of the search angle θ, its starting orientation and the size of the search neighborhood for interpolation

are important parameters in this process. Particularly, as θ decreases more short-scale features appear in the generated synthetic plumes. In this work, a wide range of directional search angle and search radii have been used to create a total of 400 realizations that represent the space of uncertainty associated with areal limits. Figure 3-21 shows nine synthetic realizations created for the given well configuration. The synthetic plumes resemble different possibilities (sizes and geometries) for the distribution of contaminants for the given well arrangement. This simple methodology generates realizations that represent different scenarios for the source zone distribution. Any other reasonable technique could be adapted and used to calibrate the α and β factors.

The α and β factors should be simultaneously calibrated. The centerline of the uncertainty band is calibrated to ensure unbiasedness; and the width of uncertainty band is calibrated to ensure fair probability distribution. The α and β values are determined through minimization of the following objective-function:

$$S(\alpha,\beta,R)=\sum_{j=q_1}^{q_M}\left[P_j-P_j^*(\alpha,\beta,R)\right]^2 \qquad [3.11]$$

where, P_j ($j = q_1, \ldots, q_M$) are the true probabilities corresponding to quantiles $q_1, \ldots, q_M$ used in optimization. The calculated probabilities $P_j^*(\alpha,\beta,R)$ are quantified as the proportion of synthetic plumes (R) that fall inside the $q_1, \ldots, q_M$ quantile (probability) maps. These quantile maps are derived from the uncertainty band calculated for a combination of α and β values. The larger the number of realizations used in calibration, the smoother the objective-function will be. However, increasing the number of quantiles and realizations will significantly increases the calculation time of the objective-function.

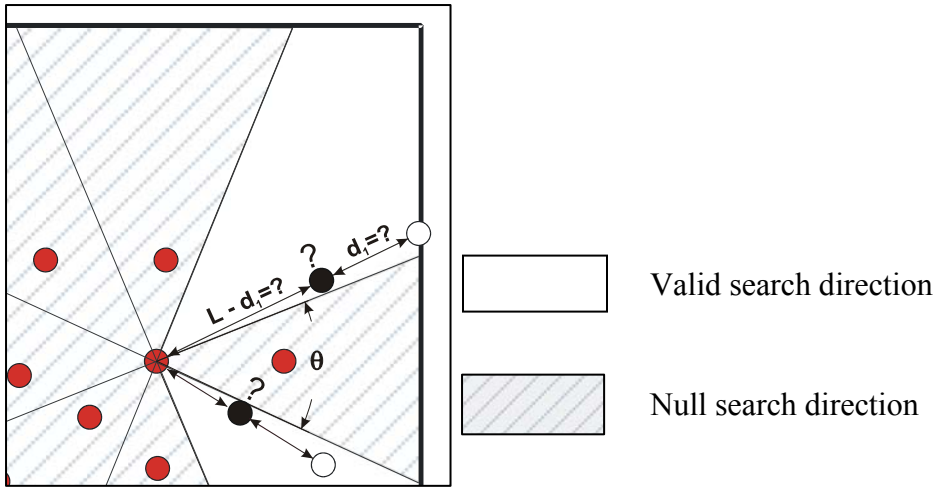


Figure 3-20: Generating multiple synthetic plumes by performing a directional search with a search angle (θ) and adding random wells (solid black circles) to the setting. (Red circles represent contaminated wells and white circles are uncontaminated wells)

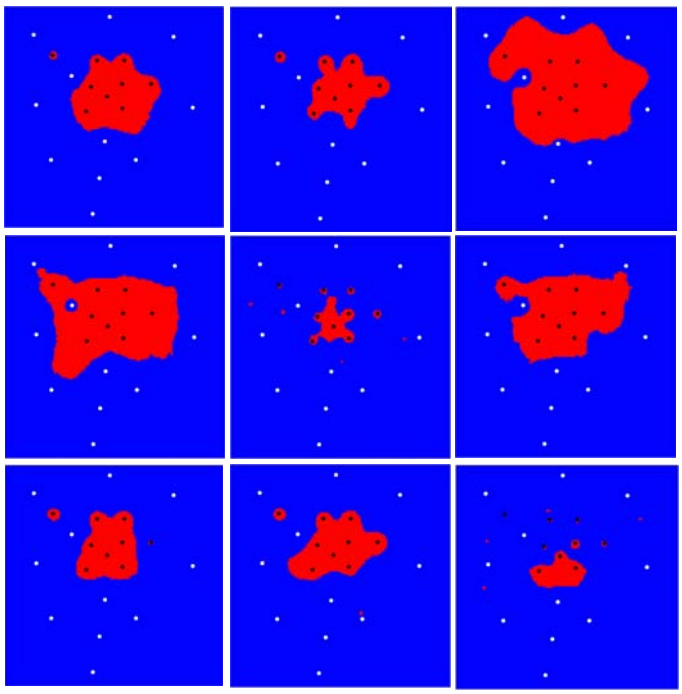


Figure 3-21: A few synthetic realizations created for the given well configuration. The hard coded data are honored in all realizations (Black circles represent contaminated wells and white circles are uncontaminated wells).

A downhill simplex algorithm (Nelder and Mead 1967) has been used to effectively reduce the required CPU time. The optimization algorithm results in calibrated values of $\alpha = 1.44$ and $\beta = 21.10$ for the given well configuration. 400 realizations of synthetic limits and q_{10}, q_{30}, q_{50}, q_{70}, q_{90} quantiles were used in the calibration process. Table 3-4 presents the calculated proportions corresponding to each quantile. There is a close match to the true probabilities. For illustration purposes, an exhaustive search on the values of α and β has been also implemented and the resulting objective-function surface is plotted in Figure 3-22. The exhaustive search results in calibrated α and β values, close to those obtained from downhill simplex algorithm. Figure 3-23 shows the uncertainty band and the p_{10}, p_{50} and p_{90} maps associated with calibrated values of α and β. Realizations can be drawn from the calibrated uncertainty band and used in Monte Carlo simulations to quantify the uncertainty in areal limits of the binary stationary random variable representing absence or presence of contaminants.

	$\alpha = 1.441$, $\beta = 21.1$					
	P_{10}	P_{30}	P_{50}	P_{70}	P_{90}	$S(\alpha,\beta)$
True	0.1	0.3	0.5	0.7	0.9	
Calculated	0.125	0.276	0.497	0.727	0.888	0.0021

Table 3-4: Comparing calculated proportions with true probabilities for the calibrated uncertainty band with $\alpha = 1.44$ and $\beta = 21.1$.

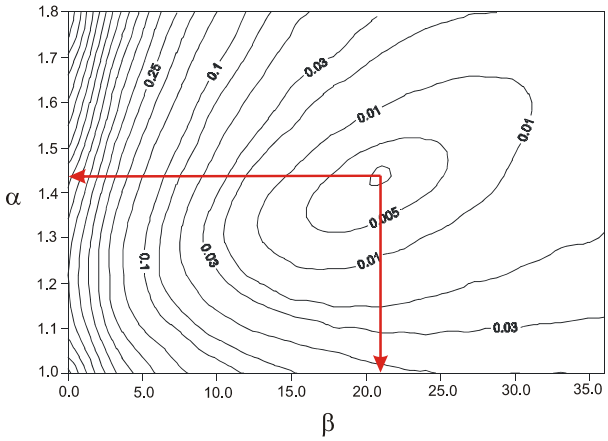


Figure 3-22: The objective-function surface for the given well configuration.

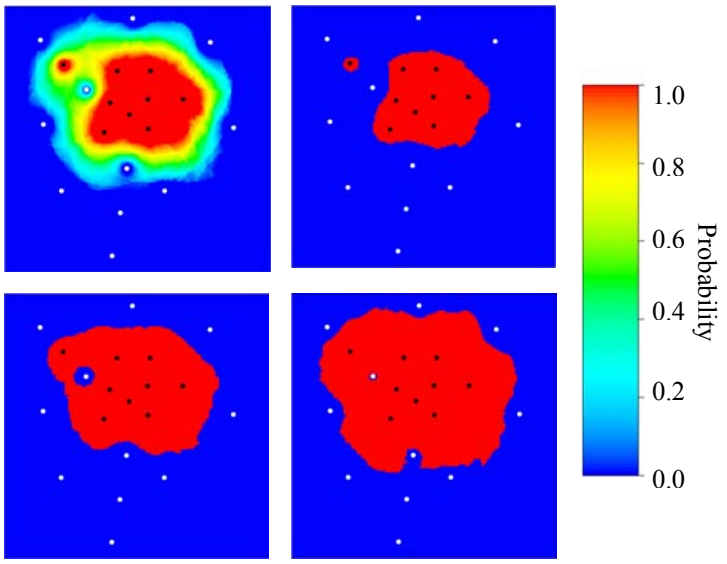


Figure 3-23: (a) The calibrated uncertainty band and associated (b) p90, (c) p50, and (d) p10 maps for the boundary modeling example.

3.2.4. Data Integration for characterization of distribution of residual NAPL

The boundary modeling approach presented above can be used to quantify uncertainty in the areal limits of a residual NAPL plume for the former flare pit site shown in Figure 3-5 (based on the layout of the 18 CPT-UVIF wells). Figure 3-24 shows the calibrated uncertainty band for the former flare pit site, which has been determined by assigning 56 control points at the boundaries of the modeling domain (one control point every 5 m). The associated calibrated scaling and separation factors are 1.19 and 38, respectively.

In order to model small-scale variability within the areal limits, UVIF data collected at a hydrocarbon impacted site are used to build the model of uncertainty in T_{UVIF} (the indicator variable that is defined based on UVIF readings and represents presence or absence of residual NAPL) across the modeling domain. Multiple realizations are then drawn from both the DF model and the

geostatistical model constructed for T_{UVIF}. These realizations are combined to form the final model of uncertainty for distribution of residual NAPL at the hydrocarbon impacted site.

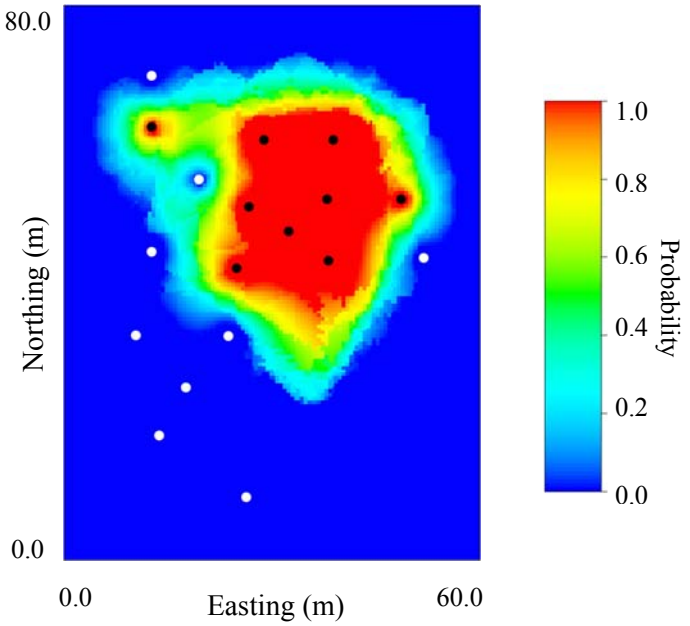


Figure 3-24: The calibrated uncertainty band for the given layout of CPT-UVIF cone holes. The associated scaling and separation factors are 1.19 and 38, respectively.

CPT-UVIF has been frequently used in environmental site characterization to delineate soil texture as well as lateral and vertical hydrocarbon distribution. Commercially available CPT-UVIF is a standard CPT cone coupled with a module to detect ultra-violet-induced-fluorescence generated by aromatic hydrocarbons. The UVIF module consists of a high intensity UV light projected through a sapphire window into the surrounding soil, and a photo multiplier tube sensor to record fluorescence. UVIF response may be affected by the size of pores and pore throats (Alostaz et al. 2008). This effect can introduce an artifact in prediction of distribution of NAPL, if the UVIF response is directly related to NAPL concentration or saturation. This artifact can be avoided by considering a

threshold value on UVIF voltage and defining the binary stationary random variable (T_{UVIF}) for presence/absence of NAPL contamination.

As stated before, the available CPT-UVIF data used in this study are from a former flare pit site located at west-central Alberta, Canada. Figures 3-5 and 3-24 show the locations of sampling points and the calibrated uncertainty band for the site, based on the layout of CPT-UVIF data. Site-specific data from multiple data sources such as soil texture and the distance to water table can be combined to create a 3D prior probability map for presence or absence of contamination. The conditional distribution of the binary random variable T_{UVIF} is then combined with the prior probability map to build a 3D updated posterior probability distribution.

3.2.5. Data Integration: Primary hard data

The recorded UVIF data may be affected by the size of pores and pore throats. This could introduce a bias in the distribution of residual NAPL, if we assume there is a direct relationship between the UVIF readings and NAPL saturation/concentration. To avoid the bias, a threshold value is calibrated based on the UVIF response voltage and site-specific residual NAPL concentrations. The categorical variable T_{UVIF} is defined as:

$$T_{UVIF}(u_\alpha) = \begin{cases} 1, & \text{if LNAPL is present at location } u_\alpha \\ 0, & \text{otherwise} \end{cases} \qquad [3.12]$$

The global proportions for presence and absence of contamination within areal limits are 0.267 and 0.733, respectively. Table 3-5 shows the details of the modeled variogram for the categorical variable T_{UVIF}. Based on the experimental variogram calculated for T_{UVIF}, a vertical trend seemed to exist. This vertical trend is not explicitly modeled since the vertical distribution of residual NAPL (T_{UVIF}) is closely related to and controlled by the distance to the groundwater table, which is considered as a secondary variable in this study.

3.2.6. Data Integration: Secondary soft data – soil texture

One of the most important aquifer properties affecting residual NAPL distribution is heterogeneity in soil texture. Short range variability in soil properties may result in preferential flow paths. In a heterogeneous setting, the distribution of NAPL is generally correlated to the distribution of effective (connected) porosity. In this study, cone penetration testing (CPT) data have been used to model the geological structure. The CPT data logger records the mechanical responses of the soil (cone resistance and sleeve friction) at a high resolution. As explained earlier, a curvilineal coordinate system is empirically established along the tendencies in soil classification charts (Figure 3-6). Figure 3-25 shows the cumulative distribution function (CDF) of a continuous random variable y_{SCI} defined based on the SCI data. The data for y_{SCI} shows that an areal trend is present with higher y_{SCI} values at the north-west of the site. Thus, to create a prior probability map for presence/absence of contamination based on SCI data, a six-step approach is followed: (1) the areal trend is explicitly modeled by (ordinary) kriging with a large search on a coarse grid (Figure 3-26-a), (2) the SCI data are de-trended and the residuals are determined (Figure 3-26-b), (3) sequential Gaussian simulation (SGS) (Deutsch and Journel 1998) is performed to simulate the residuals, (4) the areal trend is added back to each realization to create realizations of y_{SCI} over the entire grid, (5) the calibrated probabilities of absence/presence of contamination is assigned to all grid nodes based on the nodal values of y_{SCI} and the probabilities listed in Table 3-6, and (6) the E-Type mean of ensemble of realizations is computed and the conditional probabilities are standardized to give a prior probability map for presence/absence of residual NAPL based on the SCI data. The calibration in Table 3-6 is implemented based on the CDF of y_{SCI} and ten classes defined by decile thresholds (Figure 3-25).

Variable	Nugget	c_1	NS1	a_{h1}	a_{v1}
T_{UVIF}	0.043	0.153	*sph*	16.0	6.0

Table 3-5: Details of the variograms modeled for T_{UVIF}.

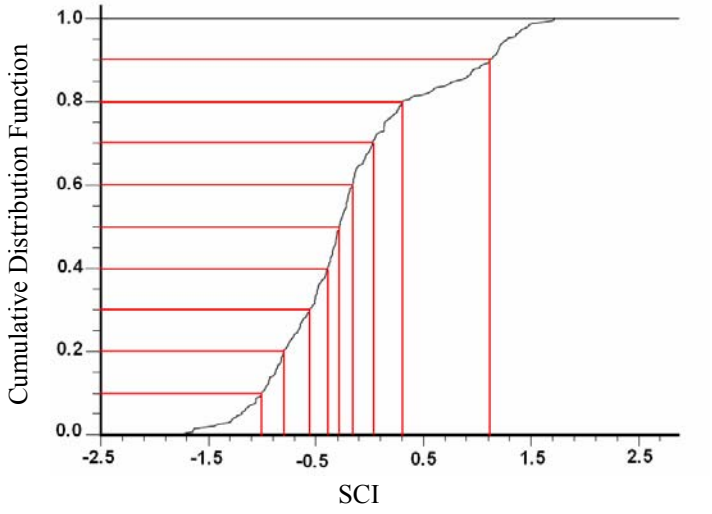


Figure 3-25: Cumulative histogram of SCI data with 10 classes defined by decile thresholds

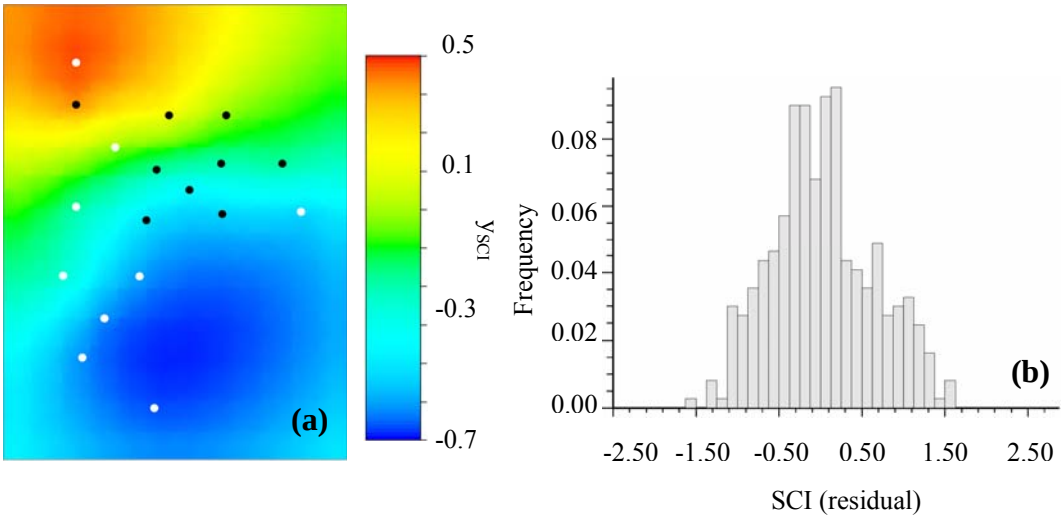


Figure 3-26: (a) The modeled areal trend and (b) the distribution of residuals of the SCI data after de-trending

3.2.7. Data Integration: Secondary soft data – distance to water table

The vertical movement of the groundwater table affects the volume of mobile and residual LNAPL. Given some mobile LNAPL sitting on the groundwater surface, a rise in groundwater table elevation causes the hydrocarbon to migrate upward as groundwater partially displaces it from the pore space. As water fills the pore network, LNAPL becomes trapped in the form of small droplets. These isolated droplets remain suspended in the network until the water table elevation drops. Lowering the water table enables the LNAPL to drain from the pore network. During drainage, droplets of LNAPL may remain entrapped within the pore interfaces, leaving residual LNAPL within the unsaturated zone. The resultant vertical movement of the water table produces a smear zone of residual LNAPL within the saturated and unsaturated zones. To account for the effects of groundwater table fluctuations, the relative elevation is defined as the distance of UVIF sampling point to groundwater table at the same location:

$$Z_{rel} = Z_{UVIF} - Z_{GW} \quad [3.13]$$

where, Z_{rel} is the 'relative elevation' at every data point, Z_{UVIF} is the elevation of the data point in the global coordinate system, and Z_{GW} is the elevation of groundwater table at the data location in global coordinate system. A continuous random variable y_{GW} is defined based on Z_{rel} data and mapped across the modeling domain by kriging. The presence/absence of contamination is calibrated against y_{GW} data and conditional probabilities are calculated. For this purpose, ten classes are defined using decile thresholds on the cumulative distribution function of y_{GW}, as shown in Figure 3-27. The resulting calibrated probabilities are presented in Table 3-7. The calibrated probabilities together with average groundwater surface mapped across the site by kriging give a three–dimensional map of prior probabilities for presence/absence of contamination based on relative elevations.

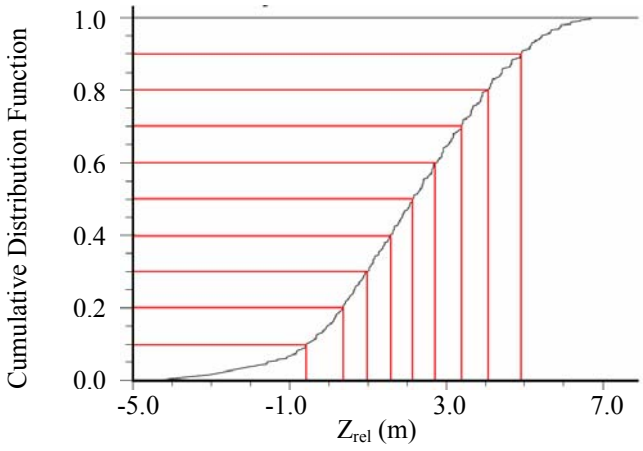


Figure 3-27: The CDF of Z_{rel} data with 10 classes defined by decile thresholds

SCI – class (y_{SCI})	$p(k=1\|y_{SCI})$	$p(k=0\|y_{SCI})$
[-2.14,-1.01]	0.093	0.907
[-1.01,-0.8)	0.151	0.849
[-0.8,-0.56)	0.200	0.800
[-0.56,-0.39)	0.302	0.698
[-0.39,-0.29)	0.213	0.787
[-0.29,-0.16)	0.314	0.686
[-0.16,0.04)	0.243	0.757
[0.04,0.31)	0.333	0.667
[0.31,1.11)	0.390	0.610
[1.11,1.77)	0.421	0.579

Table 3-6: Calibrated probabilities of absence/presence of contamination with respect to y_{SCI} data.

Z_{rel} – class (y_{GW})	$p(k=1\|y_{GW})$	$p(k=0\|y_{GW})$
[-4.3m,-0.57m]	0.212	0.788
[-0.57m , 0.38m)	0.294	0.706
[0.38m ,0.987m)	0.326	0.674
[0.987m ,1.63m)	0.461	0.539
[1.63m ,2.19m)	0.384	0.616
[2.19m ,2.77m)	0.338	0.662
[2.77m ,3.47m)	0.336	0.664
[3.47m ,4.14m)	0.083	0.917
[4.14m ,4.93m)	0.151	0.849
[4.93m ,6.68m)	0.029	0.971

Table 3-7: Calibrated probabilities of absence/presence of contamination with respect to y_{GW} data.

3.2.8. Data Integration: Combining the secondary data sources

Bayes law permits the calculation of the conditional probability $p(T_{UVIF}|y_{SCI}, y_{GW})$:

$$p(T_{UVIF}|y_{SCI}, y_{GW}) = \frac{p(T_{UVIF}) \times p(y_{SCI}|T_{UVIF}) \times p(y_{GW}|T_{UVIF}, y_{SCI})}{p(y_{SCI}, y_{GW})} \quad [3.14]$$

An easy way to combine the single event probabilities is to assume independence of the two data events. The assumption of data independence states that y_{SCI} and y_{GW} are independent $p(y_{SCI}, y_{GW}) = p(y_{SCI}) \times p(y_{GW})$. An additional assumption to simplify equation [3.14] is based on the assumption of conditional independence of events y_{SCI} and y_{GW} given the primary data, that is $p(y_{GW}|T_{UVIF}, y_{SCI}) = p(y_{GW}|T_{UVIF})$ and $p(y_{SCI}|T_{UVIF}, y_{GW}) = p(y_{SCI}|T_{UVIF})$. The equation for data integration with the assumption of full data independence can be obtained:

$$p(T_{UVIF}|y_{SCI}, y_{GW}) = \frac{p(T_{UVIF}|y_{SCI}) \times p(T_{UVIF}|y_{GW})}{p(T_{UVIF})} \quad [3.15]$$

Assumption of full data independence is not robust in presence of significant spatial correlation. A more robust approach is to assume the data are conditionally independent given the primary data event. The expression for conditional probability of the primary data event T_{UVIF} given the secondary data events y_{SCI} and y_{GW} is:

$$p(T_{UVIF}|y_{SCI}, y_{GW}) = \frac{p(T_{UVIF}) \times p(y_{SCI}|T_{UVIF}) \times p(y_{GW}|T_{UVIF})}{p(y_{SCI}, y_{GW})} \quad [3.16]$$

where the joint probability $p(y_{SCI}, y_{GW})$ is needed. According to Journel (2002) Bayesian analysis gets around this problem by considering ratios of updated probabilities of the type. This results in the expression for the permanence of ratios assumption:

$$\frac{\frac{p\left(\widetilde{T}_{UVIF}\middle|y_{SCI},y_{GW}\right)}{p\left(T_{UVIF}\middle|y_{SCI},y_{GW}\right)}}{\frac{p\left(\widetilde{T}_{UVIF}\middle|y_{SCI}\right)}{p\left(T_{UVIF}\middle|y_{SCI}\right)}}=\frac{\frac{p\left(\widetilde{T}_{UVIF}\middle|y_{GW}\right)}{p\left(T_{UVIF}\middle|y_{GW}\right)}}{\frac{p\left(\widetilde{T}_{UVIF}\right)}{p\left(T_{UVIF}\right)}} \qquad [3.17]$$

where, the event $\widetilde{T}_{UVIF}$ represents the compliment of the primary data event T_{UVIF}. Equation [3.17] results in the expression for the conditional probability based on the assumption of permanence of ratios (conditional independence):

$$p\left(T_{UVIF}\middle|y_{SCI},y_{GW}\right)=\frac{\frac{p\left(\widetilde{T}_{UVIF}\right)}{p\left(T_{UVIF}\right)}}{\frac{p\left(\widetilde{T}_{UVIF}\right)}{p\left(T_{UVIF}\right)}+\frac{p\left(\widetilde{T}_{UVIF}\middle|y_{SCI}\right)}{p\left(T_{UVIF}\middle|y_{SCI}\right)}\cdot\frac{p\left(\widetilde{T}_{UVIF}\middle|y_{GW}\right)}{p\left(T_{UVIF}\middle|y_{GW}\right)}} \qquad [3.18]$$

Two assumptions of full data independence and conditional independence are used in this study, and the associated conditional probabilities for presence/absence of contaminants are obtained. However, it is observed that the resulting conditional probabilities are very close due to a weak spatial correlation between the secondary data sources. Thus, the conditional probabilities obtained by the assumption of conditional independence (permanence of ratios) are only presented in this thesis.

3.2.9. Data Integration: Combining the prior probability maps with conditional probabilities

The calculated conditional probabilities are then combined with the prior probability map that is conditioned to indicator hard T_{UVIF} data only. For this purpose, sequential indicator simulation with locally varying means (LVM) and Bayesian updating (BU) techniques were used (Deutsch 2002).

The conditional probabilities can be incorporated as the locally varying means for kriging. The expression for probability of presence or absence of contaminant can be written by (Deutsch 2006):

$$i^*_{LVM}(\mathbf{u};T_{UVIF}) = \sum_{\alpha=1}^{n} \lambda_\alpha \cdot i(\mathbf{u}_\alpha;T_{UVIF}) + \left[1 - \sum_{\alpha=1}^{n} \lambda_\alpha\right] \cdot p(T_{UVIF}|y_{SCI}, y_{GW}) \quad [3.18]$$

where, $i^*_{LVM}(\mathbf{u};T_{UVIF})$ are the estimated local probabilities of presence/absence of contamination, n is the number of local data, λ_α, α = 1, …, n are the kriging weights, $i(\mathbf{u}_\alpha;T_{UVIF})$ is the local indicator data, and $p(T_{UVIF}|y_{SCI}, y_{GW})$ is the conditional probability obtained using secondary data y_{SCI} and y_{GW}.

Bayesian updating is one of the simplest forms of indicator cokriging: at each location along the random path, indicator kriging is used to estimate the probability of presence/absence of contamination conditioned to local hard data alone. Bayesian updating then modifies or updates the probabilities by (Deutsch 2006):

$$i^*_{BU}(\mathbf{u};T_{UVIF}) = i^*_{SK}(\mathbf{u};T_{UVIF}) \cdot \frac{p(T_{UVIF}|y_{SCI}, y_{GW})}{p_{T_{UVIF}}} \cdot C \quad [3.18]$$

where, $i^*_{BU}(\mathbf{u};T_{UVIF})$ are the estimated local probabilities of presence/absence of contamination, $p_{T_{UVIF}}$ is the global probability of absence/presence of contamination, and C is the normalization constant to ensure that the sum of the final probabilities is 1.0.

Expressions [3.17] and [3.18] are used to simulate the primary variable T_{UVIF} over a stationary domain. As discussed previously, the areal boundaries of the stationary domain are uncertain, when modeling the distribution of residual NAPL. To account for uncertainty in the areal boundaries, geostatistical realizations of the stationary random variable T_{UVIF} are explicitly clipped by independently drawn realizations of areal limits which are simulated using DF

approach explained earlier. Figure 3-28 shows planar slices of 3D binary realizations of presence/absence of contamination obtained by SIS-LVM and BU approaches clipped by independently drawn DF realizations.

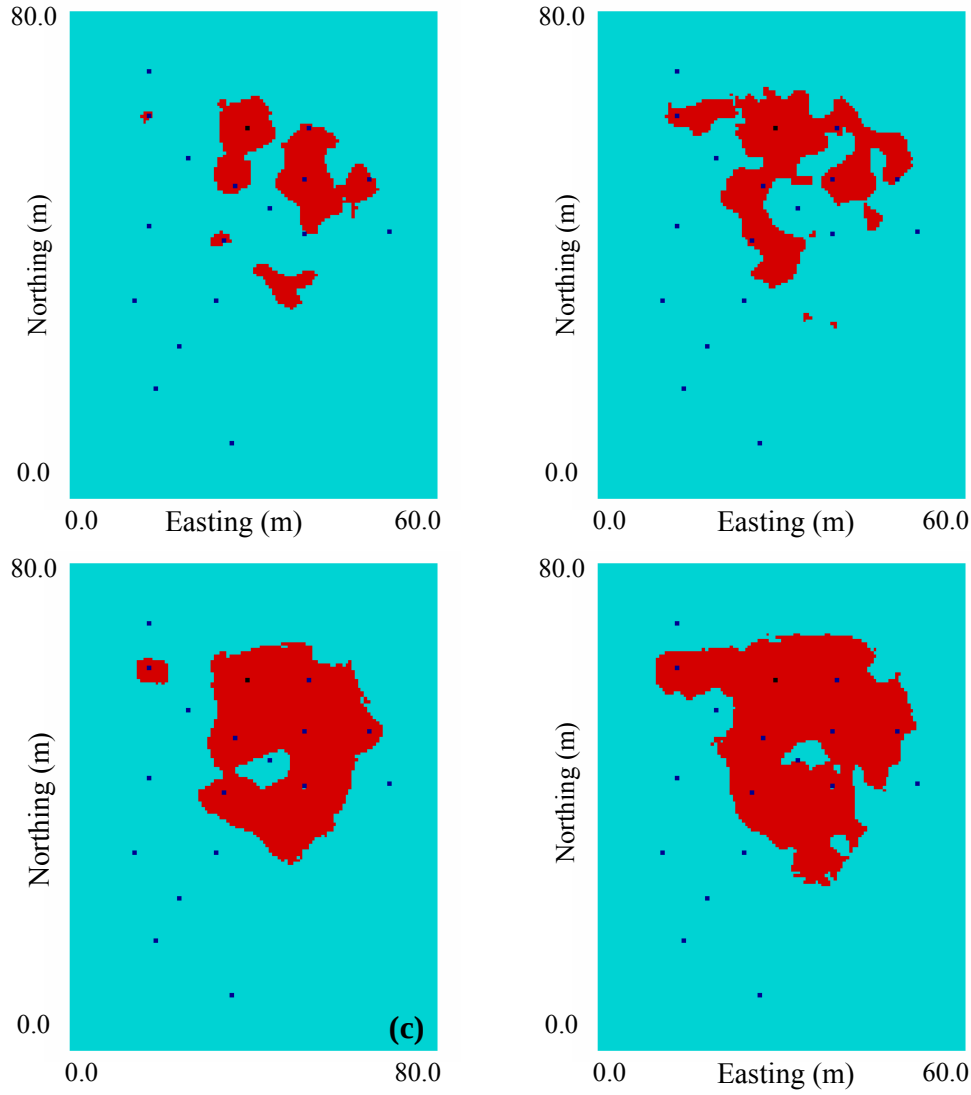


Figure 3-28: Planar slices (N_{XY}=30) of two 3D realizations of presence/absence of contamination obtained using SIS-LVM (a and b), and BU (c and d) techniques with soil texture and distance to water table as secondary information. The uncertainty in areal limits is explicitly accounted for by clipping the geostatistical realizations of stationary random function with equi-probabale realizations of areal limits simulated by boundary modeling approach presented in Section 3.2.3.

3.2.10. Cross-validation

Cross-validation methods are used to check the probabilistic prediction of the geostatistical modeling approach used in this work, and to investigate the value of secondary data sources in improving the predictive ability. Cross-validation is performed whereby each well is removed one at a time. A quantitative measure of 'closeness' to true categories can be summarized by:

$$C_{T_{UVIF}} = E\left\{p\left(\mathbf{u}_\alpha; T_{UVIF}\right) \middle| \text{true} = T_{UVIF}\right\} \quad [3.19]$$

which may be interpreted as the average predicted probability of the true categories.

The closeness measure can be easily interpreted relative to the global proportions. With no primary or secondary data the closeness measures will equal the global proportions. Thus, the measure of 'percent improvement' over the no data case can be expressed by:

$$C^{rel}_{T_{UVIF}} = \frac{C_{T_{UVIF}} - p_{T_{UVIF}}}{p_{T_{UVIF}}} \quad [3.20]$$

The third measure used for cross-validation is the measure of 'accuracy'. As the primary variable T_{UVIF} is a binary variable, at every cross-validation location four cases can be considered in terms of prediction of the true categories: (1) the location is truly contaminated and correctly predicted, (2) the location is contaminated but predicted to be clean, (3) the location is uncontaminated and correctly predicated, and (4) the location is clean but predicted to be contaminated. Cases (1) and (3) are plausible and cases (2) and (4) are not. A measure of 'accuracy' of predictions can be defined as:

$$A^{rel} = \frac{M - M_R}{1 - M_R} \quad [3.20]$$

where M and M_R are given by:

$$M = \frac{\sum_{i=1}^{N} \left(p_i^{11} + p_i^{00} - p_i^{10} - p_i^{01} \right)}{N} \quad , \quad M_R = p^1 \cdot p^1 + p^0 \cdot p^0 - 2p^1 \cdot p^0$$

where, N is the number of wells removed and replaced in the cross-validation, p_i^{11}, p_i^{10}, p_i^{00} and p_i^{01} are proportions corresponding to the cases 1 to 4, respectively. p^1 and p^0 are global proportions associated with T_{UVIF} = 1 and T_{UVIF} = 0, respectively. M is the global measure of plausibility. Its upper bound is 1.0, in the ideal case of correct prediction at all cross-validation locations. Its lower bound is M_R, which corresponds to the no-data case. Table 3-8 shows the measures of closeness and percentage improvement over the global probabilities, using secondary data only. Considering secondary data (particularly distance to groundwater table data) considerably improves the prediction of contaminated locations, even before incorporating the primary T_{UVIF} data. It is also observed that using both sources of secondary data (y_{SCI} and y_{GW}) improves the prediction ability and gives a higher degree of confidence. As mentioned earlier, the data integration schemes used in the study (with the assumptions of full data independence or conditional independence) seemed to give similar results.

Tables 3-9, 3-10 and 3-11 show the cross-validation results, while the indicator hard data (T_{UVIF} data) and secondary soft information (SCI data, groundwater elevation data and their combination with the assumption of permanence of ratios) are used. The results of indicator kriging (IK) with no secondary information show slight improvement in predictions over the global proportions. It is observed that inclusion of both sources of secondary data improves the predictive ability (Table 3-11) in prediction of contaminated locations. It is also observed that both SIS-LVM and Bayesian updating techniques overestimate the contaminated locations (conservative prediction). This overestimation is more serious in the case of Bayesian updating. In the case of SIS-LVM, this overestimation is only

observed when the distance to groundwater table is used as secondary information.

	T_{UVIF} = 0		T_{UVIF} = 1	
	$p_{T_{UVIF}} = 0.733$		$p_{T_{UVIF}} = 0.267$	
	$C_{T_{UVIF}}$	$C^{rel}_{T_{UVIF}}$ (%)	$C_{T_{UVIF}}$	$C^{rel}_{T_{UVIF}}$ (%)
SCI	0.7368	0.52	0.2828	5.92
GW	0.7568	3.25	0.3256	21.94
PR	0.7589	3.54	0.3441	28.88

Table 3-8: Measures of closeness $C_{T_{UVIF}}$ and percentage improvement over the global proportions $C^{rel}_{T_{UVIF}}$, using secondary data only (no primary data used). Secondary sources of information including soil classification index (SCI), distance to groundwater table (GW) and their combination with the assumption of permanence of ratios (PR) are used in the analysis.

$C_{T_{UVIF}}$	T_{UVIF} = 0			T_{UVIF} = 1		
	IK	**LVM**	**BU**	**IK**	**LVM**	**BU**
no secondary data	0.7532	-	-	0.2801	-	-
SCI	-	0.755	0.657	-	0.298	0.398
GW	-	0.734	0.679	-	0.338	0.663
PR	-	0.763	0.693	-	0.389	0.687

Table 3-9: Measure of closeness $C_{T_{UVIF}}$, while indicator hard data and secondary data from different data sources are accounted for. Secondary data including soil classification index (SCI), distance to groundwater table (GW) and their combination with the assumption of permanence of ratios (PR) are used in the analysis.

$C^{rel}_{T_{UVIF}}$ (%)	T_{UVIF} = 0			T_{UVIF} = 1		
	IK	**LVM**	**BU**	**IK**	**LVM**	**BU**
no secondary data	2.76	-	-	4.86	-	-
SCI	-	3.04	-10.35	-	11.81	49.40
GW	-	0.21	-7.32	-	26.90	148.45
PR	-	4.20	-5.42	-	45.99	157.60

Table 3-10: Measure of percentage improvement over global proportions $C^{rel}_{T_{UVIF}}$, while indicator hard data and secondary data from different data sources are accounted for. Secondary data including soil classification index (SCI), distance to groundwater table (GW) and their combination with the assumption of permanence of ratios (PR) are used in the analysis.

A^{rel} (%)	**IK**	**LVM**	**BU**
no secondary data	4.62	-	-
SCI	-	6.28	-5.28
GW	-	5.14	16.94
PR	-	14.01	21.23

Table 3-11: Measure of accuracy A^{rel}, while indicator hard data and secondary data from different data sources are accounted for.

The results of the LVM technique are more realistic in that it improves the predictive ability for prediction of both contaminated and uncontaminated locations. Similar results are obtained based on the measure of accuracy in Table 3-11. The assumption of conditional independence provides results that are very close to what is obtained by the assumption of full data independence. It should be noted that the improvements in predictive ability that is achieved by integration of SCI data as a secondary data source represents an upper bound to such improvements. This is due to the fact that all CPT-UVIF data are used in the construction of conditional probability distribution, which may or may not be available, when the technique is used for predictions. It is worth mentioning that the observed correlation between presence/absence of contamination and the SCI

data and the fact that soil properties usually give better (with less uncertainty) horizontal variograms are good motivations for integration of SCI data in prediction of contaminated locations with NAPL. The distance to water table is less affected by the number wells used in the analysis and as a result provides a more robust secondary source of information that can be used in prediction of contaminated locations.

In summary, in Section 3.2 of this thesis, first a boundary modeling technique was developed based on the concept of distance-function. The concept of uncertainty band was established and the calibration of uncertainty band against a large number of synthetic plumes was demonstrated. The calibrated uncertainty band resulted in a good match with the target quantiles showing the convergence of the proposed technique based on the calibration approach applied. The proposed distance-function approach can be very useful (1) to model (with uncertainty) the areal limits of non-stationary continuous or categorical random variables and (2) to provide a prior probability map for source zone sizes that can be used in Monte Carlo simulations of contaminant transport and/or Monte Carlo type inverse modeling studies (Chapter 5). Also in Section 3.2, data integration approaches were employed to create a probabilistic model for absence/presence of contamination within the areal limits of a NAPL source zone for a real aquifer. Secondary data sources such as the distance to water table and soil texture were used in the data integration and resulted in some improvement in predictive ability. Two different data integration techniques namely, Bayesian updating and sequential indicator simulation with locally varying means were employed. According to the results of cross-validation, it seemed that the Bayesian updating approach over-estimates the contaminated locations (with smaller global proportion) and underestimates the uncontaminated locations. The SIS-LVM approach, however, seemed to improve the prediction of both contaminated and uncontaminated locations over the global proportions. Finally, it was observed

that: integration of secondary data sources show an overall improvement in predictive ability (over the global proportions), the distance to groundwater provides a more robust source of secondary data as it is less dependent on the number of wells and their spatial locations, and the improvements achieved by the inclusion of soil texture data in the analysis provides an upper bound to what can be achieved by the incorporation of soil data, since all the collocated CPT data are used in construction of the 3D map for conditional probabilities. Despite this fact, incorporation of soil texture data is still useful as more reliable variograms can be calculated for soil data, comparing to hard T_{UVIF} data.

CHAPTER 4

NUMERICAL SIMULATION OF NATURAL ATTENUATION

An indispensable component of any inverse modeling study is the development of a transfer function such as a numerical simulator that relates the model parameters (e.g. porosity, hydraulic conductivity, biodegradation rate constant, etc) and their uncertainty to response variables (e.g. pressure head distribution, contaminant concentration, etc) and their associated uncertainty. A numerical simulator can be developed and used to obtain an approximate solution to the boundary value problems defined for groundwater flow and solute transport. This is usually associated with solving some form of advection-dispersion-reaction equation (Equation 2.1) by assuming a constant groundwater velocity or a coupled groundwater flow-and-transport model. When the solute concentration obtained by solution of the transport equation causes negligible variation in water density, porosity, etc, the flow equation and solute transport equation can be solved independently, that is, the flow equation is solved prior to transport simulation yielding the flow velocity distribution for all time periods and used as input to the transport code. Such a decoupled approach is applied in this work, as it has been implemented in common numerical models such as MOC (Konikow and Bredehoeft 1978 and Konikow et al. 1996), MT3D (Zheng 1990 and Zheng and Wang 1999), and RT3D (Clement 1997).

The main reason for constructing a numerical model for contaminant transport is to quantitatively analyze the natural attenuation processes at a given site and to predict contaminant concentrations at some locations and time. It is of great importance to recognize that all the transport and transformation mechanisms can

not be captured. The important aspect about correctly conceptualizing site processes is to make sure that the important mechanisms that affect contaminant transport in a significant way are captured and implemented correctly. Choosing the important processes depend heavily on the expert judgment. Nevertheless, previous studies (Chapter 2) have shown that source dissolution (and geometry), advection, dispersion, biodegradation and sorption can be considered as the most important mechanisms affecting natural attenuation of PHCs and can cause the largest uncertainties in the predicted concentrations. As discussed in Chapter 2, sorption of the dissolved PHCs to solid particles is a function of the organic fraction of soil (f_oc). Alvarez and Illman (2006) (among others) showed the variability in the organic fraction of soil has a minor impact on the plume length, when simulating natural attenuation of organic contaminants. Thus, for simplicity, the mechanism of sorption is not incorporated in the development of the mass transport simulator. Another simplifying assumption in this work is the assumption of steady-state groundwater flow. The timeframe for natural attenuation processes is large compared to that of seasonal stresses that bring the groundwater into transient condition, thus the assumption of steady-sate is reasonable.

In Section 4.1 of this chapter, the governing equations, the numerical discretization and the iterative matrix solver (C.A. Mendoza – personal communication – 2005) used in the 2D finite volume groundwater flow simulator (**flsim2d**) are presented and the flow simulator is verified by comparing its modeling outcomes for synthetic homogenous and heterogeneous aquifers to the results obtained by a MODFLOW. In Section 4.2, the technical details of the particle tracking code (**ptrack2d**) developed to model advective transport are explained and verified by comparing to a widely used particle tracking program in synthetic homogenous and heterogeneous aquifers. In Section 4.3, the mass transport equations that are numerically solved in this work are explained, their

numerical discretization is discussed in detail and the mass transport simulator (**snasim**) developed based on the Method-of-Characteristics (MOC) (Konikow and Bredehoef 1978) is presented. The verification of the **snasim** code is implemented by comparing its results to those of a widely used solute transport program MT3D.

4.1 Simulation of Steady-State Groundwater Flow

A mathematical groundwater model for steady-state (transient) conditions consists of a governing equation and boundary (and initial) conditions that simulates the flow of groundwater in a particular problem domain. The volumetric rate of flow per unit area (q=Q / A) is given by Darcy's law:

$$q_i = -k_i \frac{\partial h}{\partial x_i} \qquad [4.1]$$

where, k_i represents a principal component of the hydraulic conductivity tensor, and h is the hydraulic head that is obtained from the solution of continuity equation for groundwater flow:

$$\frac{\partial}{\partial x_i}\left(k_i \frac{\partial h}{\partial x_i} \right) - q_{sr} = S_s \frac{\partial h}{\partial t} \qquad [4.2]$$

where, S_s is the specific storage coefficient of the porous medium, q_{sr} represents the source/sink term (sink, if $q_{sr} > 0$.). The groundwater is assumed to be of uniform density. Equation [4.2] is valid when Darcy's law (Equation [4.1]) is considered to be valid, fluid (water) and the porous medium are assumed to be incompressible and the system is in saturated condition. Equations [4.1] and [4.2] also assume that the principal components of the hydraulic conductivity tensor are aligned with the global coordinate axes so that all non-principal components (cross terms) become zero. For steady-state groundwater flow, Equation [4.2] is simplified to:

$$\frac{\partial}{\partial x_i}\left(k_i \frac{\partial h}{\partial x_i} \right) = q_{sr} \qquad [4.3]$$

Except for very simple systems, analytical solutions to Equation [4.3] are not possible, and numerical methods are employed to obtain approximate solutions. In the context of the finite volume approximation, the continuous system described by Equation [4.3] is replaced by a set of discrete volumes in space and time, and

the partial derivatives are replaced by terms calculated from the differences in head values at these points. As shown in Figure 4-1, the value of the head at the cell represented by the indices (i, j) is h(i, j) and the values of heads at the adjacent cells are h(i-1, j), h(i+1, j), h(i, j-1), and h(i, j+1).

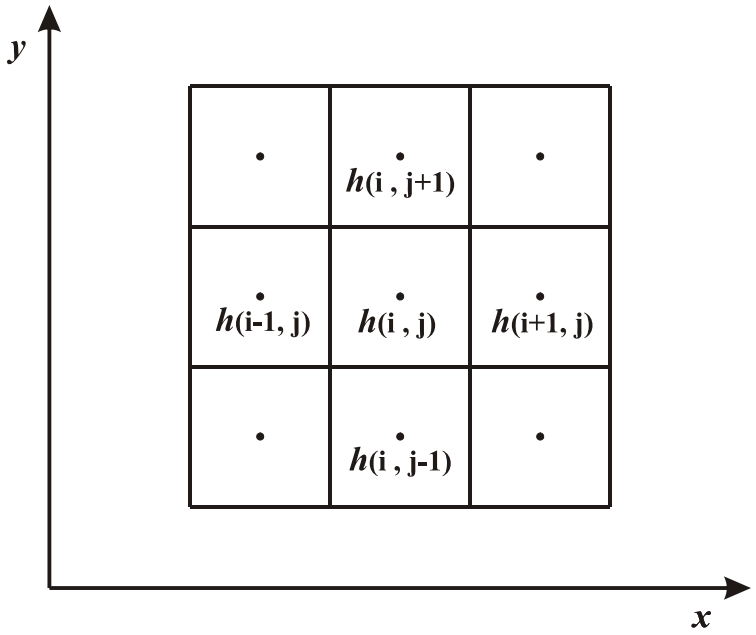


Figure 4-1: Finite volume grid showing the index numbering convention

4.1.1. Numerical discretization with finite volume method

Using a finite volume approximation in a two-dimensional setting, we may integrate Equation [4.3] over the volume (with depth of unity) and apply Green's Theorem, to obtain:

$$\oint_{\partial D}\left(k_x \frac{\partial h}{\partial x}\right)dy + \oint_{\partial D}\left(k_y \frac{\partial h}{\partial y}\right)dx = \int_D q_{sr}\,dA \qquad [4.4]$$

which is equivalent to:

$$\sum\left[\text{fluxes into cell (i, j)}\right] = q_{sr}\,.\text{Area of cell(i, j)} \qquad [4.5]$$

For a depth of unity, the flux from the cell (i-1, j) into cell (i, j) can be approximated by:

$$q_{i-1/2,j} = -k_{x(i-1/2,j)} \frac{(h_{i,j} - h_{i-1,j})}{\Delta x} . \Delta y \qquad [4.6]$$

where, $k_{x(i-1/2,j)}$ is the hydraulic conductivity at the cell interface. In this work, $k_{x(i-1/2,j)}$ is calculated as the geometric average of cell hydraulic conductivities:

$$k_{x(i-1/2,j)} = \frac{2k_{x(i-1)}k_{x(i)}}{k_{x(i-1)} + k_{x(i)}} \qquad [4.7]$$

Similar to Equation [4.7], all other flux terms can be approximated and Equation [4.5] can be re-written by:

$$\begin{aligned} \sum(\text{fluxes}) = & -k_{x(i-1/2,j)} \frac{(h_{i,j} - h_{i-1,j})}{\Delta x} . \Delta y - k_{x(i+1/2,j)} \frac{(h_{i+1,j} - h_{i,j})}{\Delta x} . \Delta y \\ & - k_{y(i,j-1/2)} \frac{(h_{i,j} - h_{i,j-1})}{\Delta y} . \Delta x - k_{y(i,j+1/2)} \frac{(h_{i,j+1} - h_{i,j})}{\Delta y} . \Delta x = q_{sr} . \Delta x \Delta y \end{aligned} \qquad [4.8]$$

Rearranging the terms in equation [4.8]:

$$\begin{aligned} & -k_{x(i-1/2,j)} \frac{\Delta y}{\Delta x} h_{i-1,j} - k_{y(i,j-1/2)} \frac{\Delta x}{\Delta y} h_{i,j-1} \\ & + \left[k_{x(i-1/2,j)} \frac{\Delta y}{\Delta x} + k_{y(i,j-1/2)} \frac{\Delta x}{\Delta y} + k_{x(i+1/2,j)} \frac{\Delta y}{\Delta x} + k_{y(i,j+1/2)} \frac{\Delta x}{\Delta y} \right] h_{i,j} \\ & - k_{x(i+1/2,j)} \frac{\Delta y}{\Delta x} h_{i+1,j} - k_{y(i,j+1/2)} \frac{\Delta x}{\Delta y} h_{i,j+1} = -q_{sr} . \Delta x \Delta y \end{aligned} \qquad [4.9]$$

The equation [4.9] can be similarly written for all cells resulting in a system of linear equations that can be represented in matrix form:

$$[\mathbf{A}]\{\mathbf{h}\} = \{\mathbf{q}\} \qquad [4.10]$$

where, $[\mathbf{A}]$ is a large $N \times N$ matrix of coefficients (N is the number of grid cells), $\{\mathbf{h}\}$ is the vector of unknown nodal pressure head values, and $\{\mathbf{q}\}$ is the right-hand side (flux) vector that represents loading and boundary conditions. Matrix $[\mathbf{A}]_{\mathrm{N \times N}}$ is a sparse matrix containing a large number of zero values that can be discarded by storing the components of the matrix using a band matrix.

It is important to manipulate matrices $[\mathbf{A}]$ and $\{\mathbf{b}\}$ in such a way that we can incorporate the boundary conditions. The different types of boundary conditions

are: (a) head is known for surfaces bounding the flow region (Dirichlet conditions); (b) flow is known across surfaces bounding the region (Neumann conditions); and (c) a head-dependent flux (or any other mix of (a) and (b)) is known across the surfaces bounding the region (Cauchy conditions) (Anderson and Woessner 1992).

In this work, Dirichlet and Neumann conditions (and their combination) are considered and coded in the groundwater flow simulator **flsim2d**. The Dirichlet and Neumann boundary conditions are represented by the following equations, respectively:

$$h(x_s, y_s, t) = h_0 \quad , \quad x_s, y_s \in S_D \qquad [4.11]$$

$$\frac{\partial h}{\partial n}(x_s, y_s, t) = f_0 \quad , \quad x_s, y_s \in S_N \qquad [4.12]$$

where, h_0 and f_0 are the fixed head and flux, respectively; and n represents the vector normal to a boundary. In cases where there are some non-zero fluxes across the boundaries (Neumann boundary conditions), a common practice is to consider some fictitious cells outside the boundary and to impose the fixed flux into the modeling domain by assigning some (fixed) head values at these fictitious cells. In this work, however, we treat all the Neumann boundary conditions as zero flux (no-flow) boundary condition (no reference to any fictitious cells) and include all the non-zero fluxes as source/sink flux in the flux vector $\{\mathbf{q}\}$. Fixed head (Dirichlet) boundary conditions are treated by a big number on the diagonal of the coefficients matrix $[\mathbf{A}]$ for the row corresponding to Dirichlet boundary and the same big number multiplied by the value of the fixed-head (Dirichlet) boundary condition h_0 in the flux vector $\{\mathbf{q}\}$. As a result, the solution value will be equal to h_0, as all other contributions in the row will be insignificant. The method is useful since the complex book keeping is avoided and the Dirichlet nodes may be easily changed.

4.1.2. Solution of system of linear equations

In order to compute the distribution of pressure heads and flow rates across the modeling domain, the system of linear equations (Equation [4.10]) must be solved. The solver adapted in this work is similar to the solver provided in a simple 2D homogeneous single-phase flow code (Dr. Carl Mendoza – Personal communication – 2005). The details of the solver are provided in the next paragraphs.

In choosing a solution method for the system of linear equations, there are a number of considerations: the memory requirement, the accuracy of the method, and the efficiency of the method. An iterative method consists of repeatedly performing steps to achieve better approximations to the solution of Equation [4.10]. Assuming $\{\mathbf{h}\}^n$ is a given approximate solution at step n, we are after a better approximation $\{\mathbf{h}\}^{n+1}$ at step $n + 1$. For this purpose, a 'preconditioning matrix' [**K**] is defined to approximate the matrix [**A**] ([**K**] = [**A**] + [**R**]) in such a way that [**K**] is non-singular and easy to invert. Equation [4.10] can be re-written in an iterative mode by:

$$[\mathbf{K}]\{\mathbf{h}\}^{n+1} = ([\mathbf{K}]-[\mathbf{A}])\{\mathbf{h}\}^n + \{\mathbf{q}\} \qquad [4.13]$$

that results in:

$$[\mathbf{K}]\{\mathbf{v}\}^{n+1} = \{\mathbf{r}\}^n \qquad [4.14]$$

where,

$$\{\mathbf{v}\}^{n+1} = \{\mathbf{h}\}^{n+1} - \{\mathbf{h}\}^n \quad \text{and} \quad \{\mathbf{r}\}^n = \{\mathbf{q}\} - [\mathbf{A}]\{\mathbf{h}\}^n \qquad [4.15]$$

The iterative algorithm in Equations [4.14] and [4.15] should be repeated until a convergence is achieved. For convergence, it is required that the spectral radius (the largest absolute value of eigenvalues) of matrix $[\mathbf{K}]^{-1}[\mathbf{R}]$ is less than one. This is achieved if [**K**] is non-singular and matrices $[\mathbf{K}]^{-1}$ and [**R**] are positive. Thus, the convergence of the algorithm is largely dependent on the choice of preconditioning matrix [**K**] and appropriate 'preconditioning algorithms' must be

employed. Later in this section, two different preconditioning algorithms are briefly explained and used in the flow simulator **flsim2d**. These methods are DKR incomplete LDU factorization (ILU) (Dupont et al. 1968) and modified incomplete factorization (MILU) (Gustafsson 1982).

The iterative algorithm presented by Equations [4.14] and [4.15] tends to require a large number of iterations, even after preconditioning. Thus, some sort of 'acceleration method' must be implemented (Letniowski 1989). The goal of an acceleration method is to decrease the number of iterations required in the basic iterative method presented above. The conjugate gradient (CG) algorithm is a widely used acceleration method that aims at finding a better search direction by replacing $\{\mathbf{h}\}^{n+1} = \{\mathbf{h}\}^{n} + \{\mathbf{v}\}^{n+1}$ with $\{\mathbf{h}\}^{n+1} = \{\mathbf{h}\}^{n} + \omega^{n}\{\boldsymbol{\theta}\}^{n}$, where $\{\boldsymbol{\theta}\}^{n}$ and ω^{n} are the new search direction and a scalar to minimize the error, respectively (Axelsson and Barker 1984). The vector $\{\boldsymbol{\theta}\}^{n}$ and scalar ω^{n} can be calculated as presented in Equation [4.16]. The strategy of choosing a search direction is that all search directions are conjugate, that is, if $\{\boldsymbol{\theta}\}^{i}$ and $\{\boldsymbol{\theta}\}^{j}$ are search directions at any two different iterations of the method then the inner product of $[\mathbf{A}]\{\boldsymbol{\theta}\}^{i}$ and $\{\boldsymbol{\theta}\}^{j}$ is equal to zero. It should be noted that the derivation of the CG method is well-established and the interested reader can refer to Shewchuk (1994) for the derivation and examples. The CG is an iterative algorithm and gives the solution at the n^{th} step by:

$$\{\mathbf{h}\}^{n+1} = \{\mathbf{h}\}^{n} + \omega^{n}\{\boldsymbol{\theta}\}^{n} \quad \text{with} \quad \omega^{n} = \frac{\left(\{\mathbf{r}\}^{n}, \{\mathbf{r}\}^{n}\right)}{\left(\{\boldsymbol{\theta}\}^{n}, [\mathbf{A}]\{\boldsymbol{\theta}\}^{n}\right)}$$

$$\{\boldsymbol{\theta}\}^{n} = \{\mathbf{r}\}^{n} + \zeta^{n}\{\boldsymbol{\theta}\}^{n-1} \quad \text{with} \quad \zeta^{n} = \frac{\left(\{\mathbf{r}\}^{n}, \{\mathbf{r}\}^{n}\right)}{\left(\{\mathbf{r}\}^{n-1}, \{\mathbf{r}\}^{n-1}\right)} \qquad [4.16]$$

$$\text{and} \quad \{\mathbf{r}\}^{n+1} = \{\mathbf{r}\}^{n} - \omega^{n}[\mathbf{A}]\{\boldsymbol{\theta}\}^{n}$$

The convergence of the CG algorithm is guaranteed for symmetric positive definite matrices, where the method converges to the correct solution much faster then the basic iterative solution (Shewchuk 1994).

The rate of convergence of CG type iterative approach can be accelerated by preconditioning or incomplete LU (ILU) decomposition. There are many ILU preconditioning methods available for general non-symmetric linear systems. Dupont, Kendall and Rachford's (DKR) LDU factorization (Dupont et al. 1968) is a widely used technique that works well for accelerating CG algorithm. In a sparse matrix data format, the matrix [**A**] consists of a diagonal band labeled γ_i, and off-diagonal bands labeled a_{ij}'s which are defined as the coefficient between the volumetric cell *i* and its adjacent cell *j*. The coefficients γ_i and a_{ij}'s can be calculated from Equation [4.9] (It should be noted that the *i, j* notation is slightly different in Equation [4.9]). An incomplete factorization LDU uses the idea of Gaussian elimination to find a reasonable approximation to the original sparse matrix [**A**]:

$$[\mathbf{L}][\mathbf{D}][\mathbf{U}] = [\mathbf{A}] + [\mathbf{E}] \qquad [4.17]$$

where, [**E**] is the error matrix, and [**L**], [**D**] and [**U**] are lower triangular, diagonal and upper triangular matrices, respectively. The sum of [**L**] + [**D**] + [**U**] should have a sparse banded structure close to the structure of [**A**] to minimize the work per iteration. If the bands of the original system are labeled as first degree, then higher degree bands are formed by fill-in resulting from elimination. Given a matrix [**A**], the elements of [**L**], [**D**] and [**U**] can be calculated: the original bands a_{ij}'s (matrix pattern) are retained in the factors (matrices [**L**] and [**U**]) and no more memory is required to store [**L**] and [**U**]. The diagonal bands of [**L**], [**D**] and [**U**], labeled d_i^{-1}, d_i, and d_i^{-1}, respectively, are constructed so that they combine to form the original diagonal band γ_i. Thus, only d_i needs to be solved:

$$d_i^{-1} = \gamma_i - \sum_{j=k(i)}^{j<i} a_{ij} d_j a_{ji} \qquad i = 1,2,3,...,\mathrm{n} \qquad [4.18]$$

where, index *j* is lower than *i*. Thus, Equation [4.18] can be solved by a marching procedure. It should be noted that in the ILU approach only d_i's are used and the higher degree bands (or shadow bands) are ignored. The error term [**E**] in

Equation [4.17] can be taken into account by using the modified ILU factorization (MILU), where the errors (or the values at the higher degree bands) are subtracted from the diagonal terms. This approach incorporates more information in the decomposition process; but at the same time it may cause problems. In fact, the problem may become less diagonally dominant as a larger number of higher degree bands are subtracted form the diagonal.

When implementing any of the preconditioning approaches jointly with CG method, the preconditioned matrix [**K**] replaces the original matrix [**A**], and calculation of ω^n and ζ^n in the CG method will be slightly changed. The complete preconditioned CG algorithm is then given by (Letniowski 1989):

$\{\mathbf{h}\}^0$ is given

Calculate $\{\mathbf{r}\}^0 = \{\mathbf{q}\} - [\mathbf{A}]\{\mathbf{h}\}^0$

Loop over $n = 1,2,3,...$ until convergence achieved

$$\text{Solve } [\mathbf{K}]\{\mathbf{v}\}^{n+1} = \{\mathbf{r}\}^n \qquad [4.19]$$

$$\{\mathbf{h}\}^{n+1} = \{\mathbf{h}\}^n + \omega^n \{\boldsymbol{\theta}\}^n \quad \text{with} \quad \omega^n = \frac{\left(\{\mathbf{r}\}^n, \{\mathbf{v}\}^n\right)}{\left(\{\boldsymbol{\theta}\}^n, [\mathbf{A}]\{\boldsymbol{\theta}\}^n\right)}$$

$$\{\boldsymbol{\theta}\}^n = \{\mathbf{r}\}^n + \zeta^n \{\boldsymbol{\theta}\}^{n-1} \quad \text{with} \quad \zeta^n = \frac{\left(\{\mathbf{r}\}^n, \{\mathbf{v}\}^n\right)}{\left(\{\mathbf{r}\}^{n-1}, \{\mathbf{v}\}^{n-1}\right)}$$

$$\{\mathbf{r}\}^{n+1} = \{\mathbf{r}\}^n - \omega^n [\mathbf{A}]\{\boldsymbol{\theta}\}^n$$

where, $\{\mathbf{v}\}^{n+1}$ and $\{\mathbf{r}\}^n$ are given by Equation [4.15]. In the procedure presented in this work, the convergence is achieved when the max norm of the residual vector $\{\mathbf{r}\}^n$ is equal to or less than or equal to a pre-specified convergence tolerance.

4.1.3. The single phase steady-state flow simulator code: flsim2d

A single-phase steady-state flow simulator for heterogeneous isotropic medium is programmed in the code **flsim2d**. Figure 4-2 shows the parameter file for this program.

```
                  Parameters for FLSIM2D
                  **********************

START OF PARAMETERS:
../data/Kxy.dat                  -file with transmissivity data
3                                -   columns for KXY
-1.0e21   1.0e21                 -   trimming limits
../data/BC.dat                   -file with specified head/flux
1   2                            -   columns for ID and head/flux
30     0.5     1.0               -nx,xmn,xsiz
30     0.5     1.0               -ny,ymn,ysiz
0                                -factorization ID - 0=ILU , 1=MILU
1.0E-06       200                -absolute conv tol, max no of iterations
flow2d.dbg                       -file for debugging output
flow2d.out                       -file for output hydraulic heads
```

Figure 4-2: The parameter file for **flsim2d** code

In Figure 4-2, the transmissivity data file (*../data/Kxy.dat*) contains isotropic transmissivity values ($T_{i,j} = k_{i,j} \times b_{i,j}$, where $b_{i,j}$ represents the saturated thickness of aquifer at every grid cell) for all grid cells (loops on *x* fastest, then *y*) in natural logarithm units in a standard Geo-EAS format. The boundary condition data file (*../data/BC.dat*) contains the type and the value of specified head/flux boundary conditions for all cells. The first column in the file represents the type (0=no-flow, 1=Dirichlet, and 2=Neumann) and the second column represents the value of the specified boundary condition. The discretization of the domain with a finite volume grid with rectangular but anisotropic cell sizes is specified by *nx*, *xmn*, *xsiz*, *ny*, *ymn*, and *ysiz*, which are the number of cells, the coordinate of the center of the first grid cell (in each direction) and the cell sizes in *x* and *y* directions, respectively. The choice of preconditioning is with *factorization ID* (0=ILU, 1=MILU). The *absolute convergence tolerance* and *maximum number of iterations* also must be specified.

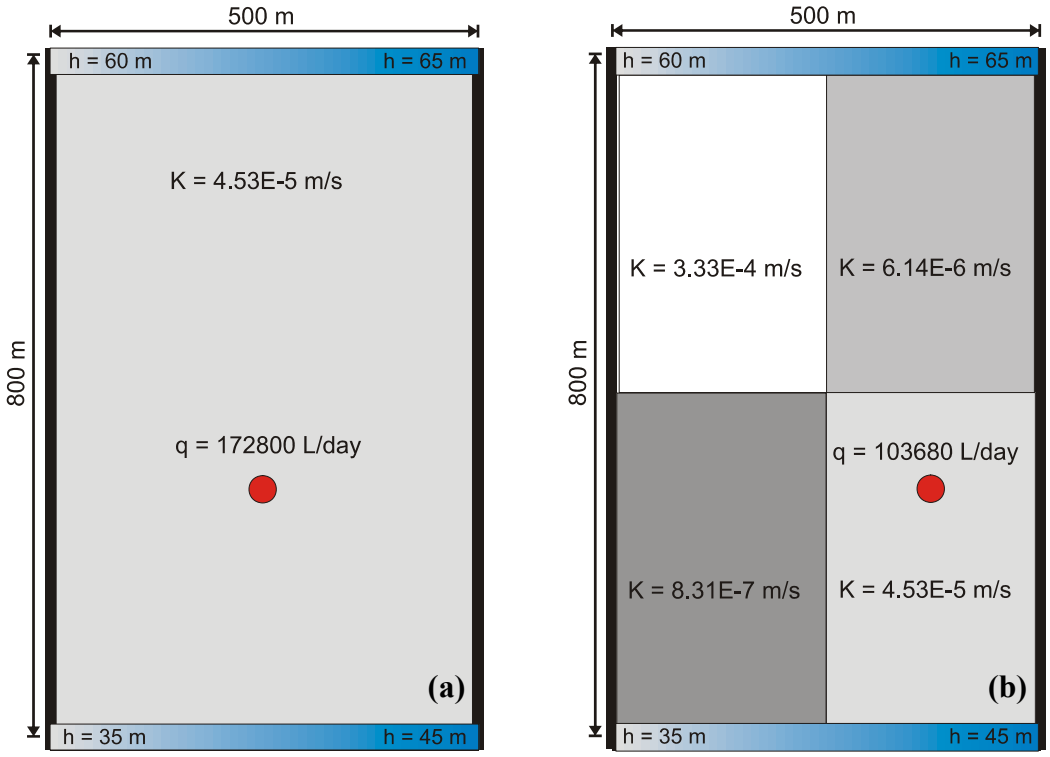


Figure 4-3: (a) The synthetic homogenous aquifer, and (b) the synthetic heterogeneous aquifer with fixed head boundary conditions at the north and south of the aquifers, no-flow (q_n = 0.0) boundary conditions at the east and west of the aquifers, and the given pumping rates of 172,800 L/day and 103,680 L/day.

4.1.4. Verification of the groundwater flow simulator code, flsim2d

Verification of the code flsim2d is implemented by comparing the modeling outcomes of the code to the results obtained by MODFLOW (McDonald and Harbaugh 1988). For this purpose, a synthetic homogenous site and a synthetic heterogeneous (with a simple geometry) site are considered, as shown in Figure 4-3. In Figure 4-3-a, the hydraulic conductivity of the homogenous aquifer is considered to be 4.53×10^{-5} m/s. In the case of the heterogeneous aquifer (Figure 4-3-b), the site has been equally divided in four sections with different hydraulic conductivities: 3.33×10^{-4} m/s, 6.14×10^{-6} m/s, 4.53×10^{-5} m/s, and 8.31×10^{-7} m/s. The pumping rate for the homogenous case is equal to 172,800 L/day, and for heterogeneous case is equal to 103,680 L/day. The boundary conditions for

both sites are identical. Dirichlet boundary conditions are assigned at the north and south of the site, and no-flow boundary conditions are assigned at the east and west of the site.

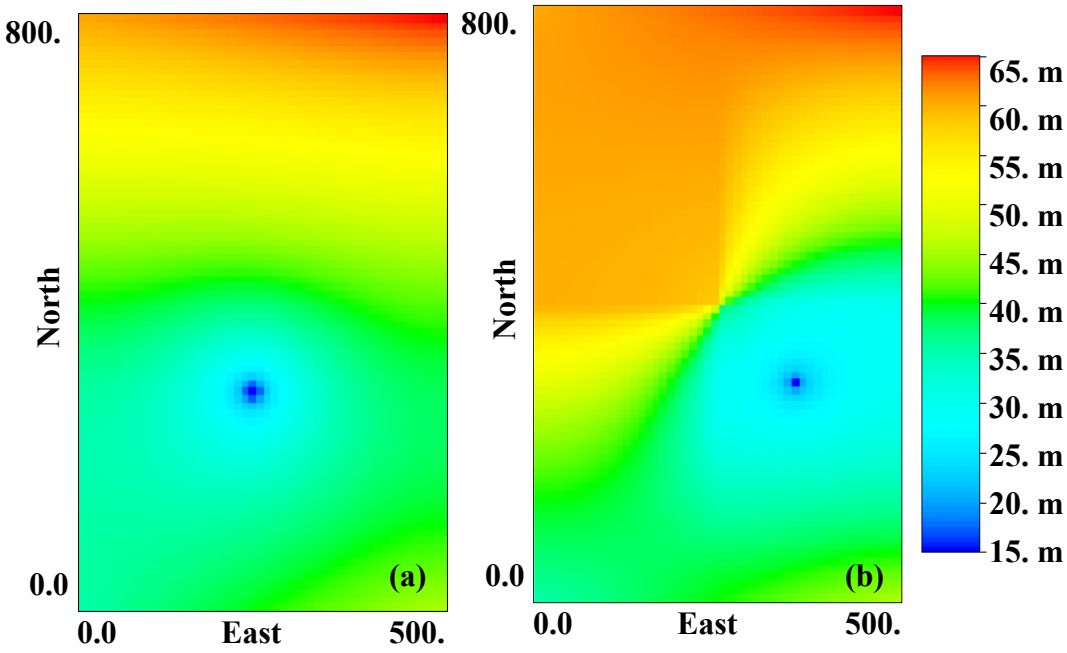


Figure 4-4: The calculated distribution of hydraulic head for (a) homogenous case and (b) heterogeneous case, using the **flsim2d** code.

Figure 4-4 shows the distribution of calculated heads for the given transmissivity distribution and boundary conditions using **flsim2d** code. Figures 4-5 and 4-6 show the pressure head contours for the homogenous and heterogeneous sites obtained using Visual MODFLOW.

Figures 4-7-a and 4-7-b show a direct comparison between the distribution of heads obtained by Visual MODFLOW, and the heads calculated for the same locations obtained by **flsim2d**. As observed in Figures 4-4, 4-5, 4-6 and 4-7, there is a very close agreement between the results of the **flsim2d** code and Visual MODFLOW.

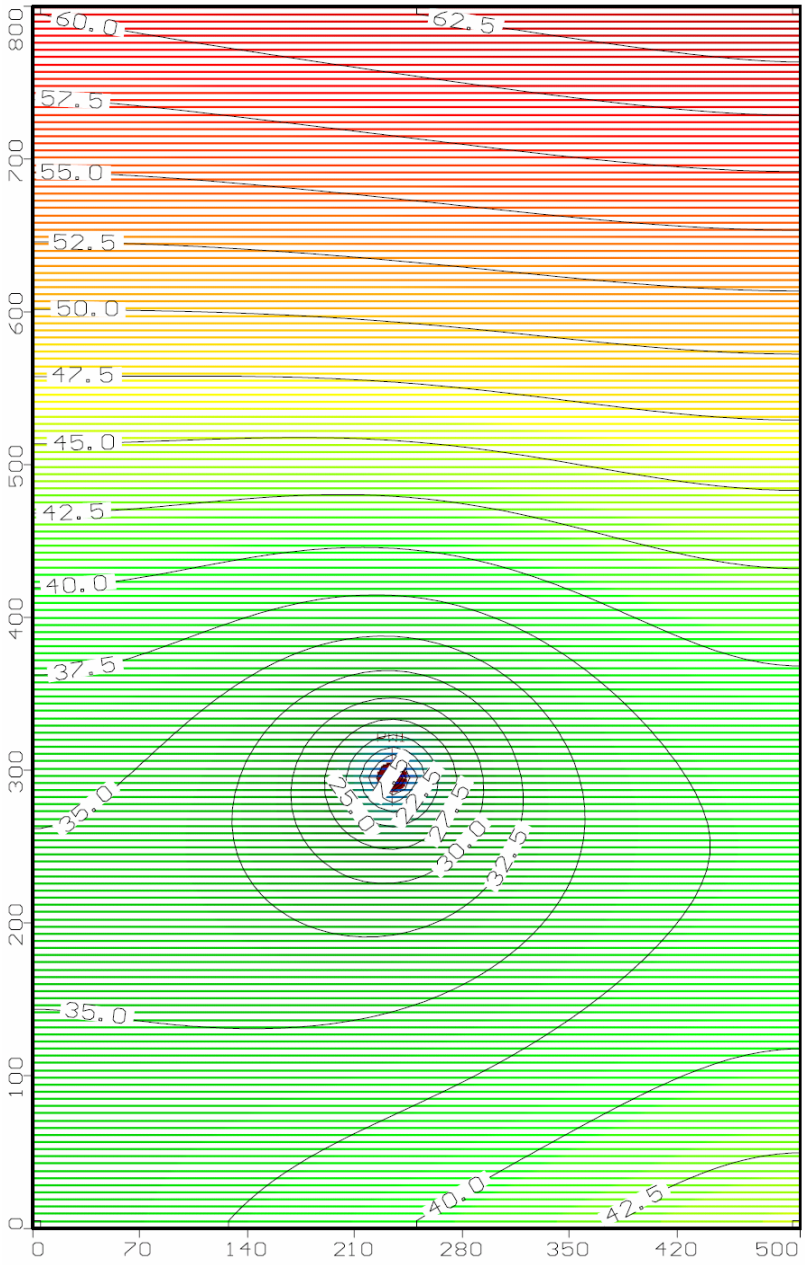


Figure 4-5: Hydraulic head contours for the homogeneous case obtained by MODFLOW

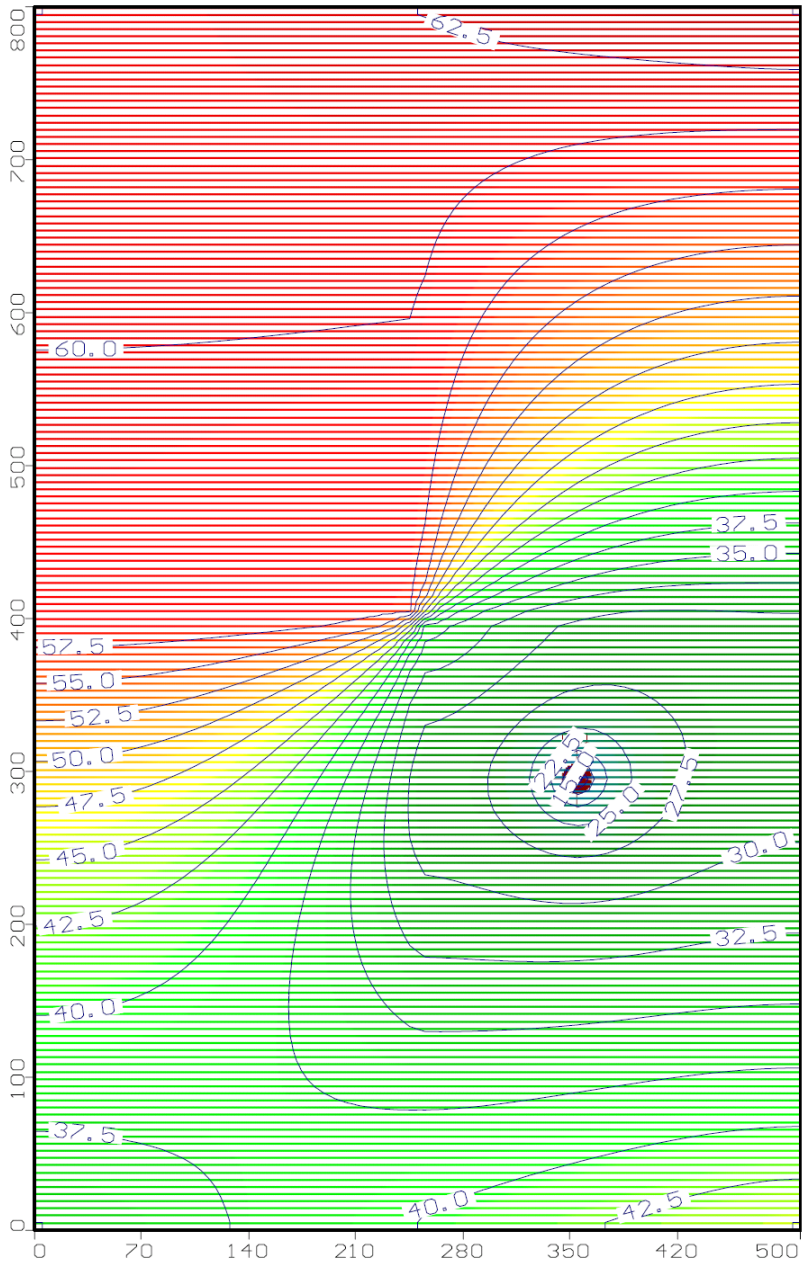
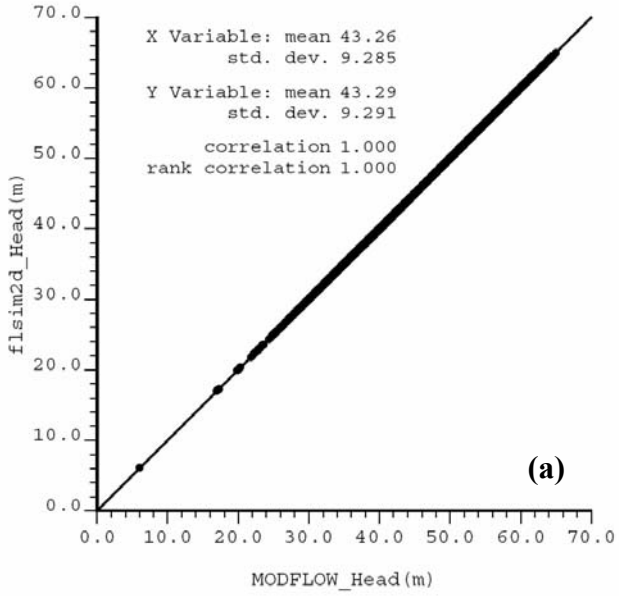


Figure 4-6: Hydraulic head contours for the heterogeneous case obtained by MODFLOW



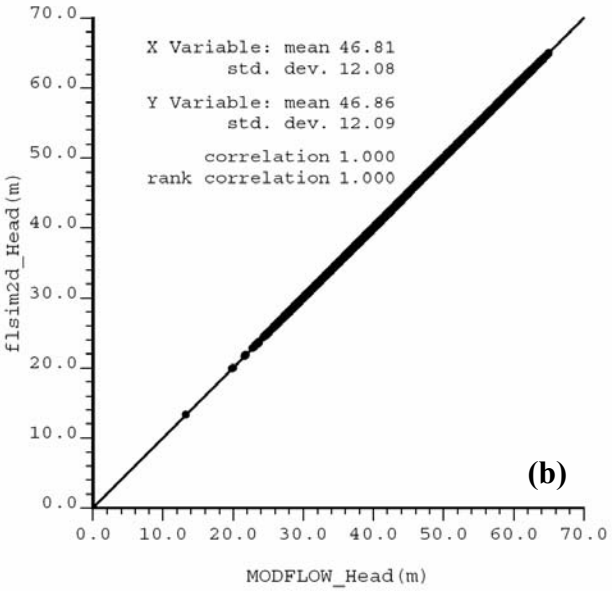


Figure 4-7: Cross-plots of heads obtained by MODFLOW and **flsim2d** for (a) homogeneous and (b) heterogeneous cases.

4.2 Simulation of Advective Transport

The amount of solute mass that is being transported through advection is a function of the solute concentration and the groundwater flow rate or seepage velocity. As discussed in Chapter 2, the governing transport equation is linked to the governing flow equation through the definition of seepage velocity:

$$v_i = -\frac{k_i}{\theta} \cdot \frac{\partial h}{\partial x_i} \qquad [4.20]$$

where, θ represents the effective porosity. In an Eulerian context, the mass conservation laws result in the following expression for advective mass transport (Zheng and Bennett 1995):

$$-\frac{\partial}{\partial x_i}(v_i C) + \frac{q_s}{\theta} C_s = \frac{\partial C}{\partial t} \qquad [4.21]$$

where, q_s represents the volumetric rate at which water is added to (or removed from) the system, and C_s represents the concentration of solute in the influx (or outflux). In this way, the term $q_s C_s$ represents the net rate at which solute is added to (or removed from) the system. The velocity term in Equation [4.21] is derived from Equation [4.20] using the distribution of hydraulic heads calculated previously. Equation [4.21] can be solved by standard numerical techniques such as finite difference method (FDM) or finite element method (FEM). However, in advection dominated problems, the Eulerian techniques suffer from some known numerical problems: artificial oscillation and numerical dispersion. Numerical dispersion is defined as unrealistic smoothing effect (smearing) at the sharp fronts as seen in the breakthrough curves of advection dominated problems. It can be quite significant when the physical dispersion is insignificant. Peclet number is defined as the measure of advection domination (Zheng and Wang 1999):

$$P_e = \frac{|v|L}{D} \qquad [4.22]$$

where, $|v|$ is the magnitude of the seepage velocity vector, L is a characteristic length (grid cell size in this case), and D is the dispersion coefficient.

4.2.1. Simulation of advective transport by particle tracking

To avoid numerical dispersion (and artificial oscillation), Lagrangian techniques can be alternatively used to solve for the advection term in solute transport problems. Zheng and Bennett (1995) showed that the rate of change in concentration (DC / Dt) identified with a particle or element of fluid that is moving along a pathline of the flow field can be derived using Equation [4.21] and calculated by:

$$\frac{DC}{Dt} = \frac{q_s}{\theta}\left(C_s - C\right) \qquad [4.23]$$

where, the rate of change in concentration (DC / Dt) is defined as:

$$\frac{DC}{Dt} = \frac{\partial C}{\partial t} + v_i \frac{\partial C}{\partial x_i} \qquad [4.24]$$

In purely advective mass transport, C_s and C are equal and the Equation [4.24] becomes:

$$\frac{DC}{Dt} = 0 \qquad [4.25]$$

which means that the concentration associated with a fluid element does not change with time, while the fluid element travels along the flow pathline. Thus, solving for the advection term with a Lagrangian approach is a matter of tracking the individual particle pathlines in the flow field. In advective transport the pathlines of contaminants coincide with the pathlines of groundwater flows, providing that the fluid density is uniform. The location of a particle at an arbitrary time t is given by:

$$\mathbf{p}(t) = \mathbf{p}(t_0) + \int_{t_0}^{t} \mathbf{v}(\mathbf{p}, t)dt \qquad [4.26]$$

where, $\mathbf{p}(t_0)$ is the particle location at time t_0, and **p** and **v** are the position and seepage velocity vectors, respectively. The numerical solution of Equation [4.26] involves starting from an initial position for each fluid particle (element) and finding its subsequent locations along the flow path in an infinite number of time steps. Numerical solution of Equation [4.26] requires the velocity field to be known continuously over the simulation domain, which is not the case when the head distribution is calculated previously over a grid with discrete cells. Thus, a velocity interpolation scheme is needed.

As discussed in Section 4.1, hydraulic heads can be calculated at the cell centers, using a block-centered finite volume (difference) method. As the first step in a velocity interpolation scheme, one can calculate the volumetric flow rates at the cell interfaces (Figure 4-8):

$$\begin{aligned} Q_{x(i-1/2,j)} &= -k_{x(i-1/2,j)} \frac{h_{i,j} - h_{i-1,j}}{\Delta x} \Delta y B \,,\; Q_{x(i+1/2,j)} = -k_{x(i+1/2,j)} \frac{h_{i+1,j} - h_{i,j}}{\Delta x} \Delta y B \\ Q_{y(i,j-1/2)} &= -k_{y(i,j-1/2)} \frac{h_{i,j} - h_{i,j-1}}{\Delta y} \Delta x B \,,\; Q_{y(i,j+1/2)} = -k_{y(i,j+1/2)} \frac{h_{i,j+1} - h_{i,j}}{\Delta y} \Delta x B \end{aligned} \qquad [4.27]$$

where, B is the aquifer thickness and all other terms are as defined previously. Using Equations [4.27], one can calculate the components of the seepage velocity as:

$$\begin{aligned} v_{x1} &= \frac{Q_{x(i-1/2,j)}}{\theta \Delta y B} \quad , \quad v_{x2} = \frac{Q_{x(i+1/2,j)}}{\theta \Delta y B} \\ v_{y1} &= \frac{Q_{y(i,j-1/2)}}{\theta \Delta x B} \quad , \quad v_{y2} = \frac{Q_{y(i,j+1/2)}}{\theta \Delta x B} \end{aligned} \qquad [4.28]$$

where, v_{x1}, v_{x2}, v_{y1} and v_{y2} are horizontal and vertical components of cell interface velocities in Figure 4-9. Given the cell interface velocities, one can calculate the seepage velocity components inside the cell by linear interpolation:

$$\begin{aligned} v_x &= A_x (x - x_1) + v_{x1} \quad \text{with} \quad A_x = \frac{v_{x2} - v_{x1}}{\Delta x} \\ v_y &= A_y (y - y_1) + v_{y1} \quad \text{with} \quad A_y = \frac{v_{y2} - v_{y1}}{\Delta y} \end{aligned} \qquad [4.29]$$

where, v_x and v_y are the horizontal and vertical components of velocity for a point inside the cell, and x_1 and y_1 are the coordinates of the lower left corner of the cell (Figure 4-9). Zheng and Bennett (1995) showed that calculation of velocity components by linear interpolation is consistent with the FDM solution of flow equation and mass conservation principles.

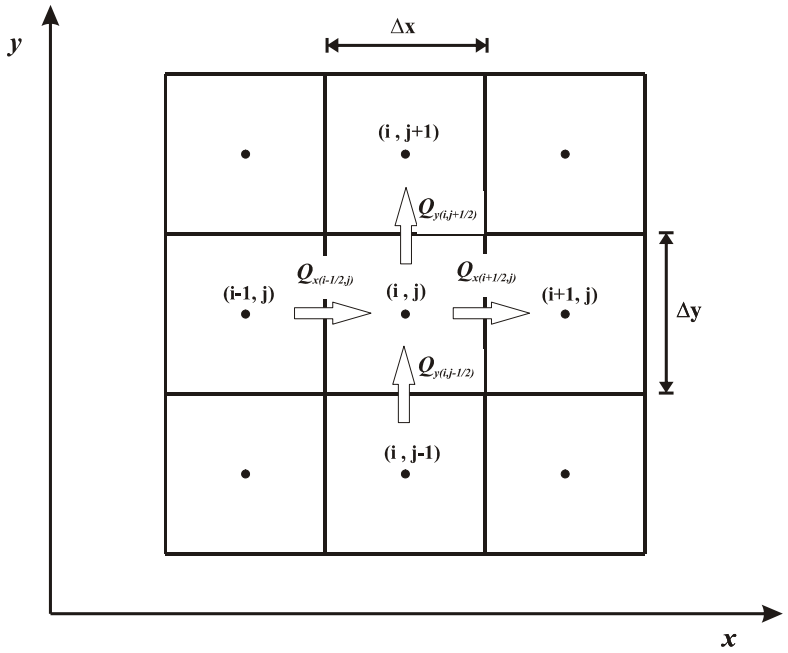


Figure 4-8: Schematic representation of volumetric flow rates on the cell interfaces (Modified from Zheng and Bennett (1995))

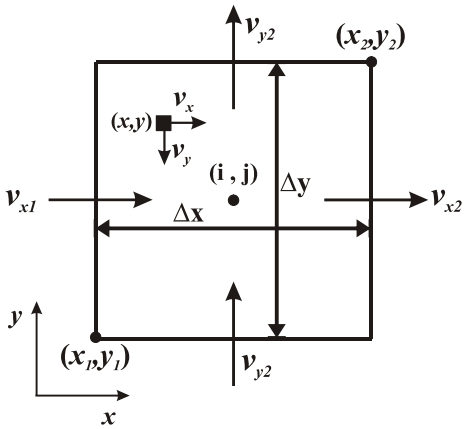


Figure 4-9: Schematic representation of calculation of seepage velocity components by interpolation (Modified from Zheng and Bennett 1995)

After the computation of seepage velocity components by interpolation, Equation [4.26] can be solved analytically. The 'semi-analytical' approach (Pollock 1988) to calculation of particle pathlines is based on numerical computation of velocities and analytical calculation of pathlines. In the context of the semi-analytical particle tracking approach, Pollock (1988) showed that the location of a fluid particle in a cell can be calculated at any arbitrary point in time (t_2) based on its velocity at any previous point in time (t_1) and the characteristics of the cell containing the particle:

$$\begin{aligned} x(t_2) &= x_1 + \frac{1}{A_x}\left[v_x(t_1)\exp(A_x\Delta t) - v_{x1}\right] \\ y(t_2) &= y_1 + \frac{1}{A_y}\left[v_y(t_1)\exp(A_y\Delta t) - v_{y1}\right] \end{aligned} \qquad [4.30]$$

In Equation [4.30], $v_x(t_1)$ and $v_y(t_1)$ are the velocity components at the point marking the particle location at time t_1 calculated using Equations [4.29]. When computing the particle pathlines, we are often interested in the location of moving particles at the cell interfaces, that is, the points of entrance and exit of each particle in any given cell. Rewriting Equation [4.30] gives the coordinates of exit points of a fluid particles in any cell:

$$\begin{aligned} x_e &= x_1 + \frac{1}{A_x}\left[v_{xp}\exp(A_x\Delta t_e) - v_{x1}\right] \\ y_e &= y_1 + \frac{1}{A_y}\left[v_{yp}\exp(A_y\Delta t_e) - v_{y1}\right] \end{aligned} \qquad [4.31]$$

where, v_{xp} and v_{yp} are the velocity components of a particle located at point p (in many cases the point of entrance), and Δt_e is the travel time to the exit point e, which is calculated by:

$$\Delta t_e = \min(\Delta t_x, \Delta t_y) \quad \text{with} \quad \Delta t_x = \frac{\ln(v_{x2}/v_{xp})}{A_x} \quad \text{and} \quad \Delta t_y = \frac{\ln(v_{y2}/v_{yp})}{A_y} \qquad [4.32]$$

Where there is no need to calculate the location of intermediate points within the cells, points p and e are treated as entrance and exit points at the cell interfaces and Equations [4.31] and [4.32] can be used repeatedly to find particle pathline.

The formulation of the semi-analytical method is slightly different when a groundwater divide is present, when the velocity components at two cell interfaces are identical, or when a weak sink is present. Presentation of applicable formulation for these exceptional cases is out of scope of this work and the interested reader is referred to Zheng and Bennett (1995) and Zheng and Wang (1999) for more details. It should be noted that these exceptional features are considered and programmed in the particle tracking code **ptrack2d** developed in this research.

4.2.2. The semi-analytical particle tracking code: ptrack2d

The semi-analytical particle tracking approach (Pollock 1988) presented by Equations [4.30], [4.31] and [4.32] has been programmed in the code **ptrack2d** for a heterogeneous aquifer under steady-state condition. Figure 4-10 shows the parameter file for **ptrack2d** code.

```
                  Parameters for PTRACK2D
                  ***********************

START OF PARAMETERS:
../data/ptcl.dat                -file with initial particle distribution
1   2                           -   columns for X and Y coordinates
-1.0e21   1.0e21                -   trimming limits
../data/Kxy.dat                 -file with transmissivity data
3   4                           -   columns for KX and KY
../data/BC.dat                  -file with specified head/flux
1   2                           -   columns for ID and head/flux
../data/heads.out               -file with calculated heads
1                               -   column for calculated head
30     0.5     1.0              -nx,xmn,xsiz
30     0.5     1.0              -ny,ymn,ysiz
0.3                             -average porosity
365    365                      -number of time steps, simulation time (days)
ptrack2d.dbg                    -file for debugging output
ptrack2d.out                    -file for output particle paths
```

Figure 4-10: The parameter file for **ptrack2d** code

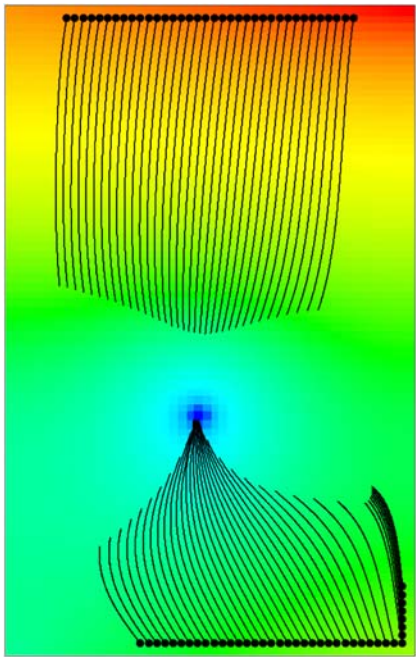
In Figure 4-10, the particle location data file (*../data/ptcl.dat*) includes the coordinates for the initial locations of the particles in two dimensions. The head distribution data file (*../data/heads.out*) includes the calculated heads for every grid cell (in a block-centered grid) and can be computed using the code **flsim2d**

presented previously. All other inputs are either self-explanatory or explained previously.

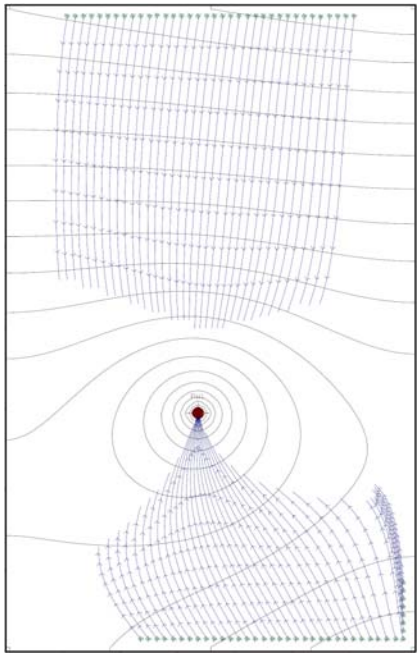
4.2.3. Verification of the particle tracking code, ptrack2d

Verification of the code is implemented by comparing its results to the modeling outcomes of MODPATH (Blandford and Huyakorn 1991), which uses a different particle tracking technique (first-order Euler algorithm and forth-order Runge-Kutta method). The model set-up is similar to the homogenous and/or heterogamous cases in Section 4.1.4 and Figure 4-3. Figures 4-11-a and 4-11-b show the particle paths for the homogeneous case (Figure 4-3-a) after 450 days obtained by **ptrack2d** and MODPATH, respectively. Figures 4-12-a and 4-12-b show the particle paths for the homogeneous case after 1850 days (when the system reaches steady-state) obtained by **ptrack2d** and MODPATH, respectively. Figures 4-13-a and 4-13-b show the particle paths for the heterogeneous case (Figure 4-3-b) after 500 days obtained by **ptrack2d** and MODPATH, respectively. Figures 4-13-a and 4-13-b show the particle paths for the heterogeneous case after 12500 days obtained by **ptrack2d** and MODPATH, respectively.

As observed in Figures 4-11 to 4-14, there is a very close agreement between the simulations results for homogenous and heterogeneous cases obtained by the code developed in this work, **ptrack2d**, and MODPATH.

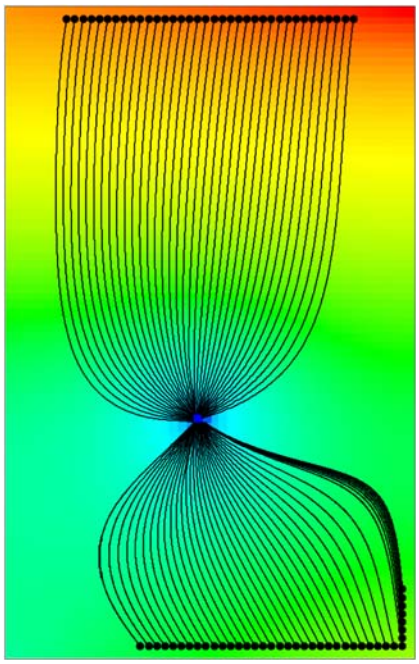


(a)

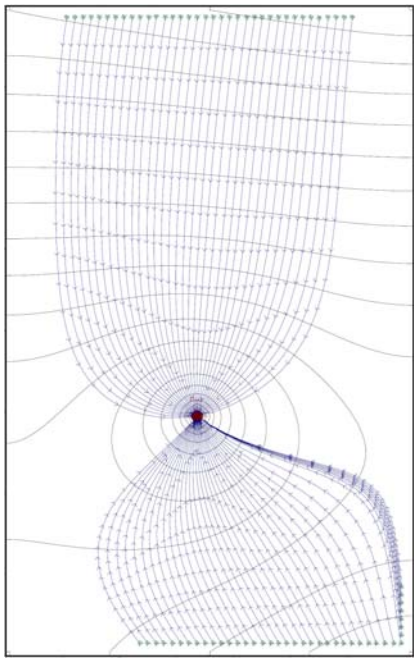


(b)

Figure 4-11: The particle paths in the homogenous case after 450 days calculated by (a) **ptrack2d** and (b) MODPATH.

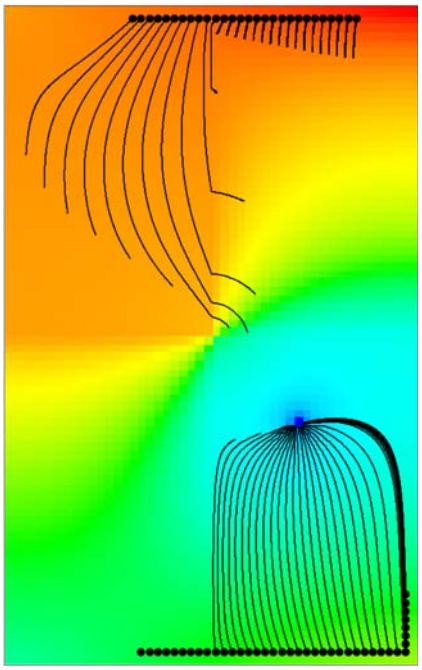


(a)

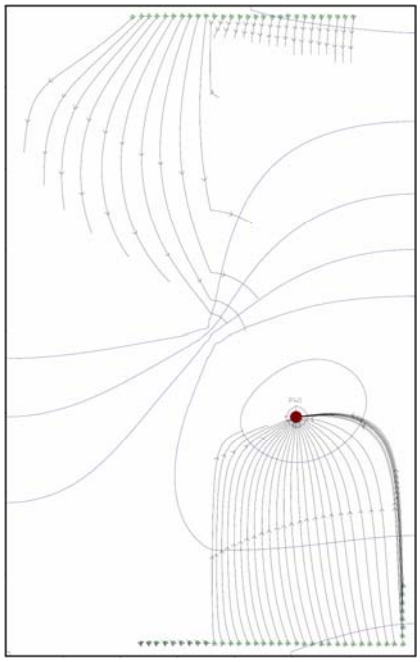


(b)

Figure 4-12: The particle paths in the homogenous case after 1850 days calculated by (a) **ptrack2d** and (b) MODPATH.

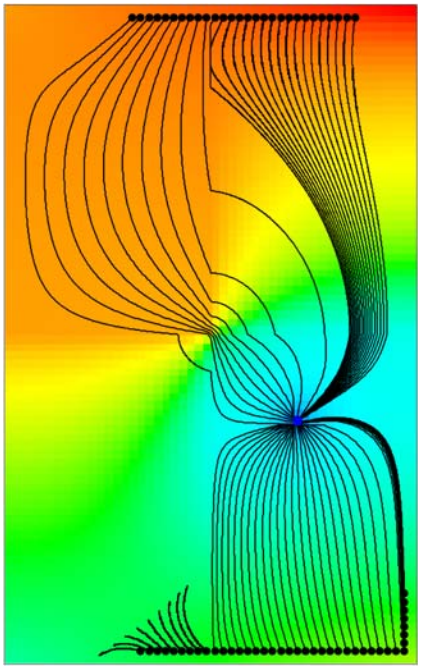


(a)

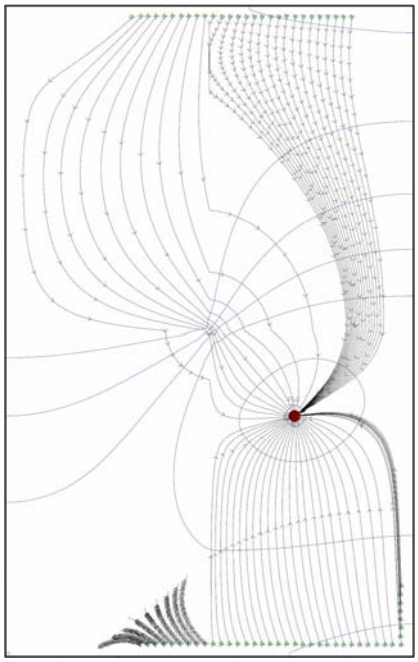


(b)

Figure 4-13: The particle paths in the heterogeneous case after 500 days calculated by (a) **ptrack2d** and (b) MODPATH.



(a)



(b)

Figure 4-14: The particle paths in the heterogeneous case after 12500 days calculated by (a) **ptrack2d** and (b) MODPATH.

4.3 Simulation of Reactive Contaminant Transport

As discussed in Chapter 2 (Section 2.6.3), the partial differential equation describing the dissolution, advection, dispersion and biodegradation of organic contaminants in groundwater is expressed by:

$$\frac{\partial(\theta C_s)}{\partial t} = \frac{\partial}{\partial x_i}\left(\theta D_{ij}\frac{\partial C_s}{\partial x_j}\right) - \frac{\partial}{\partial x_i}(\theta v_i C_s) + \theta \cdot \max\left[0, k^{NAPL}\left(C_s^{eq} - C_s\right)\right] - \theta\lambda C_s \quad [4.33]$$

where, C_s is the concentration of the contaminant species *s* (e.g. benzene), θ is the effective porosity, D_{ij} represents the dispersion tensor, λ is the first-order biodegradation rate constant, k^{NAPL} is the dissolution rate constant and C_s^{eq} is the equilibrium concentration of the contaminant species *s*, expressed by:

$$C_s^{eq} = f_s . C_s^{sol} \quad [4.34]$$

where, C_s^{sol} is the solubility limit for pure substrate *s* in water, and f_s is the mole fraction of the species *s* in the mixture of organic (and inert/non-biodegradable) materials and can be calculated by (Parker et al. 1991):

$$f_s = \frac{S_s^{NAPL}/\omega_s}{S_s^{NAPL}/\omega_s + T_t^{NAPL}/\omega_t} \quad [4.35]$$

where, S_s^{NAPL} is the mass of substrate *s* per unit mass of dry soil, and T_t^{NAPL} represents the equivalent mass of all inert and non-biodegradable materials *t* per unit mass of dry soil, ω_s is the molecular weight of substrate *s*, and ω_t is the equivalent molecular weight of mixture of all non-biodegradable and inert (insoluble) materials. The fraction T_t^{NAPL}/ω_t can be calculated by:

$$\frac{T_t^{NAPL}}{\omega_t} = \sum_{li=1}^{NI}\frac{I_{li}^{NAPL}}{\omega_{li}} + \sum_{lt=1}^{NT}\frac{TR_{lt}^{NAPL}}{\omega_{lt}} \quad [4.36]$$

where, I_{li}^{NAPL}, TR_{lt}^{NAPL}, ω_{li} and ω_{lt} represent each inert and tracer (non-biodegradable) material and their associated molecular weights, respectively. The NAPL mass of substrate *s* per unit mass of dry soil (S_s^{NAPL}) decreases as the

dissolution occurs. This process can be represented by (Waddill and Widowson 1998):

$$\frac{\mathrm{d}S_s^{NAPL}}{\mathrm{d}t} = -\frac{\theta}{\rho_b} R_s^{NAPL} \qquad [4.37]$$

where, ρ_b is the bulk density of the porous medium and R_s^{NAPL} represents the mass transfer rate given by:

$$R_s^{NAPL} = \max\left[0, k^{NAPL}\left(C_s^{eq} - C_s\right)\right] \qquad [4.38]$$

Thus, due to dissolution of NAPL into groundwater, soil NAPL concentration decreases and aqueous concentration increases. In Equation [4.33], the only term on the left hand side that accounts for change in concentration with time (a transient problem); the first term on the right hand side represents the hydromechanical dispersion; the second term represents advection; the third term represents dissolution; and the fourth term represents first-order biodegradation. This equation should be solved numerically to find the distribution of dissolved species concentration in space and time. Eulerian approaches such as the FDM and FEM, or Lagrangian approaches such as random walk method can be used for this purpose. As discussed in Section 4.2, Eulerian approaches are subject to some numerical difficulties such as numerical dispersion and artificial oscillation. On the other hand, Lagrangian approaches are significantly affected by discrete nature of their concentration fields, which makes their results sensitive to the number of particles used. Also, Lagrangian methods become less reliable when the dispersive component of transport process becomes more important. As a robust alternative, the MOC (a Lagrangian-Eulerian approach) has been used in this work and programmed into the code **snasim**. A dynamic particle allocation scheme is used in the code to reduce the required memory and speed up the simulation process. In the next section of this chapter, the numerical descretization of non-advective terms in Equation [4.33] by the FDM will be

presented in detail, and the connection between the solutions to the advective and non-advective terms will be discussed.

4.3.1. Numerical solution of the mass transport problem by the MOC

The MOC uses a particle tracking approach to handle the advection term in Equation [4.33], and applies the FDM to solve for all other non-advective terms. To solve the mass transport problem by the MOC, the transport equation is written in Lagrangian form. This can be achieved by combining Equations [4.21], [4.24] and [4.33] to give:

$$\frac{DC_s}{Dt} = \frac{1}{\theta}\frac{\partial}{\partial x_i}\left(\theta D_{ij}\frac{\partial C_s}{\partial x_j}\right) - \frac{q_s C_s}{\theta} + \max\left[0, k^{NAPL}\left(C_s^{eq} - C_s\right)\right] - \lambda C_s \qquad [4.39]$$

The second term on the right hand side of Equation [4.39] ($q_s C_s / \theta$) represents the effects of volumetric flow sources (or sinks) on the mass transport phenomena. The reactive mass transport simulator developed in this work does not account for external sink and sources (e.g. pumping wells) and influx of contaminants at the boundaries. These are not limiting assumptions in development of a stochastic model for natural attenuation of the PHCs. In fact, the goal is to study the behavior of non-engineered systems under naturally occurring attenuating mechanisms. Besides, the effect of influx (or outflux) of contaminants can be easily added to the stochastic natural attenuation simulator **snasim**. If these assumptions hold, and we assume that the effective porosity is constant across the modeling domain, Equation [4.40] can be re-written as:

$$\frac{DC_s}{Dt} = \frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C_s}{\partial x_j}\right) + \max\left[0, k^{NAPL}\left(C_s^{eq} - C_s\right)\right] - \lambda C_s \qquad [4.40]$$

The derivative on the left hand side of Equation [4.40] represents the rate of change in solute concentration along a characteristic curve of the velocity field. It can be approximated by finite difference as (Zheng and Wang 1999):

$$\frac{DC_s}{Dt} = \frac{C_m^{n+1} - C_m^{n^*}}{\Delta t} \qquad [4.41]$$

where, C_m^{n+1} is the solute aqueous concentration at cell *m* at time step *n+1*; $C_m^{n^*}$ is the solute aqueous concentration at cell *m* at the intermediate time step n^* due to advection alone; and Δt is the time increment from time step *n* to time step *n+1*. Combining Equations [4.40] and [4.41], one can calculate the solute concentration at the new time step by:

$$C_m^{n+1} = C_m^{n^*} + \Delta t \times RHS \qquad [4.42]$$

where, *RHS* represents the finite difference approximation to terms on the right hand side of Equation [4.40].

The solute concentration at the intermediate time step $C_m^{n^*}$ accounts for advection and is calculated by a Lagrangian approach. First, a number of moving particles are distributed across the flow domain at the beginning of the simulation. Depending on the modeler's choice, the particles may be distributed in areas with concentrations above a threshold (dynamic particle allocation), or everywhere across the domain. The may also be distributed (in each active cell) on a fixed pattern or on a random pattern (Zheng and Wang 1999). The program **snasim** distributes the particles in active cells on a random pattern and implements dynamic particle allocation. After allocation of particles, appropriate concentration value is assigned to each particle (based on concentration of the cell containing each particle), and the position and concentration of the particles are recorded. Particles are then tracked forward through the flow field with a small time increment. In a grid with uniform cell sizes, the average concentration at cell *m* due to advection alone can be calculated by averaging the concentrations of the particles that are within the cell at the end of each time increment, and written by:

$$C_m^{n^*} = \frac{1}{NP_m} \sum_{p=1}^{NP_m} C_p^n \qquad [4.43]$$

where, NP_m is the number of particles within the cell m (NP_m >0), and C_p^n is the concentration of the p^{th} particle at the previous time step n. In practice, however, the average concentration obtained by Equation [4.43] is not directly used in the finite difference approximation presented in Equation [4.42]. This is due to the fact that other transport mechanisms such as dispersion and biodegradation occur throughout the time increment not after finishing the advection. There is a standard approach to address this issue: a weighted concentration is defined based on $C_m^{n^*}$ and the concentration at the previous time step C_m^n. Thus, Equation [4.42] can be re-written by:

$$C_m^{n+1} = C_m^{n^*} + \Delta t \times RHS\left[\omega C_m^{n^*} + (1-\omega)C_m^n\right] \qquad [4.44]$$

where, ω is the weighting factor with a value between 0.5 and 1.0 (Zhang and Wang 1999). In the code **snasim**, the weighting factor ω is hard coded as 0.75.

4.3.2. Implementation details of the MOC as used in development of snasim

Equations [4.43] and [4.44] represent one step of the mass transport problem by the MOC. All mass transport simulators developed based on the MOC repeat these equations alternatively in successive transport steps until they reach the desired time. The implementation details of the approach as used in development of **snasim** are presented in this section. First, the details of dynamic particle allocation, used in the solution of advective transport, are presented. Then, numerical discretization of non-advective terms on a finite difference grid is presented to form the basis for solution of Equation [4.44]. Finally, the criteria for maximum time step size are discussed.

As discussed above, dynamic allocation of particles can help reduce the amount of memory required throughout the simulation. The particle allocation technique

used in **snasim** is similar to what is implemented in the widely used solute transport program MT3D-MS (Zheng and Wang 1999). However, different particle tracking approaches are used. In the MT3D family of mass transport simulators, numerical particle tracking techniques are used: first-order Euler algorithm (Konikow and Bredehoef 1978) and forth-order Runge-Kutta method (Zheng 1989). Similar to the particle tracking code **ptrack2d** (Section 4.2.2), a semi-analytical particle tracking approach is used in **snasim**. The semi-analytical approach provides the exact solution to the particle tracking problem presented in Equation [4.26]. However, it is constrained by the assumption of linear velocity interpolation.

In the context of dynamic particle allocation, the number of particles placed at each cell is normally set either at a high level (*NPH*) or at a low level (*NPL*). The criterion for switching between *NPH* and *NPL* is 'relative cell concentration gradient' ($DCCELL_{i,j}$) given by:

$$DCCELL_{i,j} = \frac{CMAX_{i,j} - CMIN_{i,j}}{CMAX - CMIN} \qquad [4.45]$$

where, $CMIN_{i,j}$ and $CMAX_{i,j}$ are the minimum and maximum cell concentrations in the cells adjacent to the cell (i , j) in the current time step; and *CMIN* and *CMAX* are the minimum and maximum concentrations in the entire field at the current time step. Depending on the value of $DCCELL_{i,j}$ being smaller or greater than a threshold, namely *DCEPS*, either *NPL* or *NPH* number of particles are placed at each cell. The threshold *DCEPS* is usually set to a small value close to zero. This methodology ensures that enough particles are placed in the areas with elevated concentration gradients. As the cell concentrations change with time and as particles leave (accumulate in) the cells in the source (sink) areas, it may become necessary to insert or remove the particles from the system. This is important in heterogeneous aquifers where heterogeneity may cause the flow paths to converge in some areas and diverge in other areas. In this work, insertion

and removal of particles are controlled by minimum and maximum allowed number of particles per cell, *NPMIN* and *NPMAX*, respectively. If the number of cells becomes smaller than the specified *NPMIN*, up to *NPL* or *NPH* numbers of particles are inserted into the cell without affecting the existing particles; and if the number of particles in each cell exceeds the specified *NPMAX*, all particles are removed and *NPH* numbers of particles are inserted into the cell to maintain the mass balance.

To solve Equation [4.42] at any desired time step, a finite difference approximation to the terms in the *RHS* is needed. The first term on the right hand side of Equation [4.40] represents the hydrodynamic dispersion. Its finite difference approximation can be written in an explicit (in time) form by:

$$\begin{aligned}\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial C_s}{\partial x_j}\right) \cong & \left(D_{xx}\right)_{i+\frac{1}{2},j}\frac{C^n_{i+1,j}-C^n_{i,j}}{\Delta x^2}-\left(D_{xx}\right)_{i-\frac{1}{2},j}\frac{C^n_{i,j}-C^n_{i-1,j}}{\Delta x^2} \\ & +\left(D_{xy}\right)_{i+\frac{1}{2},j}\frac{C^n_{i,j+1}+C^n_{i+1,j+1}-C^n_{i,j-1}-C^n_{i+1,j-1}}{4\Delta x\Delta y} \\ & -\left(D_{xy}\right)_{i-\frac{1}{2},j}\frac{C^n_{i-1,j+1}+C^n_{i,j+1}-C^n_{i-1,j-1}-C^n_{i,j-1}}{4\Delta x\Delta y} \\ & +\left(D_{yy}\right)_{i,j+\frac{1}{2}}\frac{C^n_{i,j+1}-C^n_{i,j}}{\Delta y^2}-\left(D_{yy}\right)_{i,j-\frac{1}{2}}\frac{C^n_{i,j}-C^n_{i,j-1}}{\Delta y^2} \\ & +\left(D_{yx}\right)_{i,j+\frac{1}{2}}\frac{C^n_{i+1,j}+C^n_{i+1,j+1}-C^n_{i-1,j}-C^n_{i-1,j+1}}{4\Delta x\Delta y} \\ & -\left(D_{yx}\right)_{i,j-\frac{1}{2}}\frac{C^n_{i+1,j-1}+C^n_{i+1,j}-C^n_{i-1,j-1}-C^n_{i-1,j}}{4\Delta x\Delta y}\end{aligned} \qquad [4.46]$$

where, D_{xx}, D_{yy} , D_{xy} and D_{yx} are the components of the dispersion tensor; and they may be approximated by:

$$\left(D_{xx}\right)_{i+\frac{1}{2},j}=\left[\left(\alpha_L\right)_{i+\frac{1}{2},j}\left(v_x\right)^2_{i+\frac{1}{2},j}+\left(\alpha_T\right)_{i+\frac{1}{2},j}\left(v_y\right)^2_{i+\frac{1}{2},j}\right]/\left(\hat{v}_{i+\frac{1}{2},j}\right)$$

$$\left(D_{xy}\right)_{i+\frac{1}{2},j}=\left[\left(\alpha_L\right)_{i+\frac{1}{2},j}-\left(\alpha_T\right)_{i+\frac{1}{2},j}\right]\left(v_x\right)_{i+\frac{1}{2},j}\left(v_y\right)_{i+\frac{1}{2},j}/\left(\hat{v}_{i+\frac{1}{2},j}\right)$$

$$
\begin{aligned}
\left(D_{xx}\right)_{i-\frac{1}{2},j} &= \left[\left(\alpha_L\right)_{i-\frac{1}{2},j}\left(v_x\right)^2_{i-\frac{1}{2},j} + \left(\alpha_T\right)_{i-\frac{1}{2},j}\left(v_y\right)^2_{i-\frac{1}{2},j}\right] / \left(\hat{v}_{i-\frac{1}{2},j}\right) \\
\left(D_{xy}\right)_{i-\frac{1}{2},j} &= \left[\left(\alpha_L\right)_{i-\frac{1}{2},j} - \left(\alpha_T\right)_{i-\frac{1}{2},j}\right]\left(v_x\right)_{i-\frac{1}{2},j}\left(v_y\right)_{i-\frac{1}{2},j} / \left(\hat{v}_{i-\frac{1}{2},j}\right) \\
\left(D_{yy}\right)_{i,j+\frac{1}{2}} &= \left[\left(\alpha_L\right)_{i,j+\frac{1}{2}}\left(v_y\right)^2_{i,j+\frac{1}{2}} + \left(\alpha_T\right)_{i,j+\frac{1}{2}}\left(v_x\right)^2_{i,j+\frac{1}{2}}\right] / \left(\hat{v}_{i,j+\frac{1}{2}}\right) \\
\left(D_{yx}\right)_{i,j+\frac{1}{2}} &= \left[\left(\alpha_L\right)_{i,j+\frac{1}{2}} - \left(\alpha_T\right)_{i,j+\frac{1}{2}}\right]\left(v_x\right)_{i,j+\frac{1}{2}}\left(v_y\right)_{i,j+\frac{1}{2}} / \left(\hat{v}_{i,j+\frac{1}{2}}\right) \\
\left(D_{yy}\right)_{i,j-\frac{1}{2}} &= \left[\left(\alpha_L\right)_{i,j-\frac{1}{2}}\left(v_y\right)^2_{i,j-\frac{1}{2}} + \left(\alpha_T\right)_{i,j-\frac{1}{2}}\left(v_y\right)^2_{i,j-\frac{1}{2}}\right] / \left(\hat{v}_{i,j-\frac{1}{2}}\right) \\
\left(D_{yx}\right)_{i,j-\frac{1}{2}} &= \left[\left(\alpha_L\right)_{i,j-\frac{1}{2}} - \left(\alpha_T\right)_{i,j-\frac{1}{2}}\right]\left(v_x\right)_{i,j-\frac{1}{2}}\left(v_y\right)_{i,j-\frac{1}{2}} / \left(\hat{v}_{i,j-\frac{1}{2}}\right)
\end{aligned}
\qquad [4.47]
$$

where, $\left(\alpha_L\right)_{i\pm\frac{1}{2},j\pm\frac{1}{2}}$ and $\left(\alpha_T\right)_{i\pm\frac{1}{2},j\pm\frac{1}{2}}$ are longitudinal and transverse dispersivities at the cell interfaces and can be calculated by:

$$
\begin{aligned}
\left(\alpha_L\right)_{i+\frac{1}{2},j} &= \frac{\left(\alpha_L\right)_{i,j} + \left(\alpha_L\right)_{i+1,j}}{2} \quad , \quad \left(\alpha_T\right)_{i+\frac{1}{2},j} = \frac{\left(\alpha_T\right)_{i,j} + \left(\alpha_T\right)_{i+1,j}}{2} \\
\left(\alpha_L\right)_{i-\frac{1}{2},j} &= \frac{\left(\alpha_L\right)_{i-1,j} + \left(\alpha_L\right)_{i,j}}{2} \quad , \quad \left(\alpha_T\right)_{i-\frac{1}{2},j} = \frac{\left(\alpha_T\right)_{i-1,j} + \left(\alpha_T\right)_{i,j}}{2} \\
\left(\alpha_L\right)_{i,j+\frac{1}{2}} &= \frac{\left(\alpha_L\right)_{i,j} + \left(\alpha_L\right)_{i,j+1}}{2} \quad , \quad \left(\alpha_T\right)_{i,j+\frac{1}{2}} = \frac{\left(\alpha_T\right)_{i,j} + \left(\alpha_T\right)_{i,j+1}}{2} \\
\left(\alpha_L\right)_{i,j-\frac{1}{2}} &= \frac{\left(\alpha_L\right)_{i,j-1} + \left(\alpha_L\right)_{i,j}}{2} \quad , \quad \left(\alpha_T\right)_{i,j-\frac{1}{2}} = \frac{\left(\alpha_T\right)_{i,j-1} + \left(\alpha_T\right)_{i,j}}{2}
\end{aligned}
\qquad [4.48]
$$

The second and third terms on the right hand side of Equation [4.40] represent the source dissolution and first-order decay. Their finite difference approximation can be written in an explicit (in time) form by:

$$
\max\left[0, k^{NAPL}\left(C_s^{eq} - C_s\right)\right] - \lambda C_s \cong \max\left[0, k^{NAPL}\left(C_s^{eq} - C_{i,j}^n\right)\right] - \lambda C_{i,j}^n \qquad [4.49]
$$

Substituting the expressions in Equations [4.43], [4.46], [4.47], [4.48] and [4.49] into Equation [4.44] gives the formulation for explicit time weighting in a finite difference approximation context.

In the context of solving the mass transport problem with the MOC, the size of the tracking time increment is controlled by an accuracy requirement:

$$\Delta t \le \min\left|\frac{C_r \Delta x}{v_x}, \frac{C_r \Delta y}{v_y}\right| \qquad [4.50]$$

where, C_r is the Courant number that is given by:

$$C_r = \frac{v \Delta t}{\Delta x} \qquad [4.51]$$

where, v represents the magnitude of the velocity vector at the cell interfaces, and Δx represents the (directional) grid discretization. As an explicit time-weighting scheme has been implemented in finite difference approximation of *RHS* terms in Equation [4.44], other stability constraints will also apply:

$$\text{Dispersion}: \quad \Delta t \le \frac{0.5}{\frac{D_{xx}}{\Delta x^2} + \frac{D_{yy}}{\Delta y^2}} \quad \text{with} \quad \begin{cases} D_{xx} = \dfrac{\alpha_L v_x^2 + \alpha_T v_y^2}{|v|} \\ D_{yy} = \dfrac{\alpha_L v_y^2 + \alpha_T v_x^2}{|v|} \end{cases}$$

$$\text{Dissolution}: \quad \Delta t \le \frac{1.0}{k_{NAPL}} \qquad [4.52]$$

$$\text{First-order decay}: \quad \Delta t \le \frac{1.0}{\lambda}$$

Similar to groundwater flow problems, three types of boundary conditions are possible in mass transport problems: (a) boundaries with given concentration (Dirichlet Condition), (b) boundaries with given concentration gradient (Neumann Condition), and (c) boundaries with given total (advective/dispersive) flux (Cauchy Condition). These boundary conditions are expressed by Equations [4.53] to [4.55], respectively:

$$C_s(x, y, t) = c(x, y, t) \qquad \text{on } \Gamma_1 \text{ with } t > 0.0 \qquad [4.53]$$

$$\theta D_{ij} \frac{\partial C}{\partial x_j} = f_i(x, y, t) \qquad \text{on } \Gamma_2 \text{ with t} > 0.0 \qquad [4.54]$$

$$\theta D_{ij} \frac{\partial C}{\partial x_j} - q_i C = g_i(x, y, t) \quad \text{on } \Gamma_3 \text{ with t} > 0.0 \qquad [4.55]$$

In impermeable boundaries both advective and dispersive fluxes are zero; and Equation [4.55] is set equal to zero. Assignment of zero dispersive flux boundary condition is commonly used (in Neumann and Cauchy boundaries), as the dispersive flux is often quite small and negligible, as compared to advective flux. An interesting case for boundary conditions is when an unknown advective flux exists at the boundaries, and its value is determined throughout the simulation. In this case, we may assign a Neumann boundary condition with zero dispersive flux and let the program calculate the advective flux internally.

4.3.3. The reactive mass transport simulator: snasim

A FORTRAN code, **snasim**, has been developed to simulate dissolution, advection, hydrodynamic dispersion, and first-order biodegradation of PHCs in groundwater, based on the MOC (presented in Sections 4.3.1 and 4.3.2). Figure 4-15 shows the parameter file for **snasim**. In the parameter file of **snasim**, the input data file containing the initial conditions (in milligrams per liter) and the appropriate column number are entered in lines 1 and 2. Lines 3 and 4 involve the input file and the data column containing soil NAPL concentrations (in the source zone) in grams of NAPL per grams of soil. The trimming limits that are applied on all source data are entered in the line 5 of the parameter file. Lines 6, 7 and 8 involve dry soil density (in grams per liter), total porosity, solubility limit of the solute (substrate), its mass fraction in the NAPL mixture, molecular weight of the solute (in grams per mole) and molecular weight of the non-biodegradable substances (in grams per liter). Line 9 contains the flag to specify the dissolution rate constant (k_{NAPL}) as a constant value or to draw it as a random variable (in Monte Carlo Simulations). The specified constant value is entered in line 10. Lines 11 and 12 involve the statistical distribution and the associated parameters

for the dissolution rate constant, when specified as a random variable. In lines 13 and 14, the input file for transmissivity data (in natural logarithmic units) and the associated column number are specified. The effective porosity is entered in line 15. Lines 16 and 17 contain the calculated steady-state heads (output of **flsim2d**) and its column number. The flow and transport boundary conditions data files and column numbers for type and value of the boundary conditions are entered in lines 18 to 21. The discretization of the simulation domain is specified in lines 22 and 23. The simulation mode is selected in line 24. As observed at the end of the parameter file, three simulation modes are available: (1) source dissolution, advection and biodegradation, (2) source dissolution, advection, dispersion and biodegradation, and (3) source dissolution, advection and dispersion. The dynamic particle allocation parameters, *NPL* and *NPH* are specified in line 25. The random number seed is specified in line 26. The values of *DCCELL*, *NPMIN* and *NPMAX* are specified in lines 27 and 28. Line 29 involves the value of the grid Courant number. Longitudinal and transverse dispersivities are specified in line 30. Similar to the dissolution rate constant, the biodegradation rate constant can be considered as a specified constant value or as a random variable with specified statistical distribution. The information about the biodegradation rate constant is entered in lines 31 to 34. Temporal discretization is specified in line 35. The program pauses if the size of time steps is larger than what is internally calculated as the maximum allowed time step size based on stability and accuracy constraints. The number of output events and the associated time steps are specified in lines 36 and 37. To speed up the simulation time and reduce the number of tracked particles, one can specify a minimum concentration threshold in milligrams per litre in line 38. All concentrations below this threshold are set equal to zero at the end of each time step. The file for debugging output and the debugging level are specified in lines 39 and 40.

```
                    Parameters for SNASIM
                    *********************

     START OF PARAMETERS:
1    ../data/initial.dat          -file with initial condition
2    3                            -   column for solute conc(mg/L)
3    ../data/source.dat           -file with source zone distribution
4    3                            -   column for NAPL conc(gr/gr)
5    -1.0e21   1.0e21             -   trimming limits
6    1.6       0.35               -dry soil density(gr/cm3),total porosity
7    0.00178   0.01               -substrate solubility(gr/cm3),mass fraction
8    78.1      101.56             -substrate,tracer molecular weight(gr/mole)
9    1                            -dissolution rate (0=specified,1=random)
10   0                            -   dissolution rate constant(specified)
11   0                            -   rate distribution (0=uniform,1=Gaussian)
12   0.01    0.05                 -   min,max (or mean and s.d. for Gaussian)
13   ../data/Kxy.dat              -file with transmissivity data
14   3                            -   column for Txy
15   0.3                          -effective porosity
16   ../data/heads.out            -file with calculated heads
17   1                            -   column for calculated head
18   ../data/FBC.dat              -file with flow boundary condition
19   1   2                        -   columns for ID and head/flux
20   ../data/TBC.dat              -file with transport boundary condition
21   1   2                        -   columns for ID and TBC
22   30     0.5     1.0           -nx,xmn,xsiz
23   30     0.5     1.0           -ny,ymn,ysiz
24   2                            -simulation mode
25   0      12                    -lower and higher no of random cell particles
26   67097                        -   random number seed
27   -1.0e5                       -relative cell concentration gradient
28   4      24                    -minimum and maximum no of particles
29   1.0                          -grid courant number
30   3.0    0.6                   -longitudinal and transverse dispersivity(m)
31   1                            -biodegradation rate (0=specified,1=random)
32   0.0025                       -   biodegradation rate constant(specified)
33   0                            -   rate distribution (0=uniform,1=lognormal)
34   0.001  0.005                 -   min,max (or mean and s.d. for lognormal)
35   365    3650                  -number of time steps, simulation time(days)
36   4                            -number of output events
37   37   73   183   365          -   time steps
38   0.001                        -minimum concentration
39   snasim.dbg                   -file for debugging output
40   1                            -   debugging level
41   snasim.out                   -file for output concentrations
42   sns-source.out               -output file for source concentrations



     simulation mode:
     1 = source dissolution + advection + biodegradation
     2 = source dissolution + advection + dispersion + biodegaradtion
     3 = source dissolution + advection + dispersion
```

Figure 4-15: The parameter file for **snasim** code

The names of output file for simulated concentrations and the output file for average source concentrations are specified in lines 41 and 42. The output file is replicated (and pre-named by the event number) by the program based on the number of output events and the simulated results for each event are written in separate output files.

4.3.4. Verification of the reactive mass transport code, snasim

Verification of the code **snasim** is performed by comparing modeling outcomes of the code with those of a widely used commercial code MT3DMS (Zheng and Wang 1999). Figure 4-16 shows the simulation domain and the input parameters for the sites with homogenous and heterogeneous hydraulic properties. For the homogenous case (Figure 4-16-a), transmissivity is equal to 4.53×10^{-5} m/s, longitudinal and transverse dispersivities are equal to 10.0 m and 1.0 m, and the first-order biodegradation rate constant is equal to 0.0075 day^{-1}. The source zone is considered to be 200.0 m by 100.0 m involving Dirichlet boundary condition with a fixed concentration of 50 mg/L. The spatial discretization of the simulation domain is conducted by defining a 50 × 80 uniform grid with cell sizes of 10 m × 10 m. The flow boundary conditions involve constant head boundary conditions at the north and south boundaries and no-flow boundary conditions at the west and east boundaries. The transport boundary conditions involve a Dirichlet boundary condition at the location of the source zone and a zero dispersive flux (Neumann) boundary condition at north, south, east and west boundaries. It should be noted that the assignment of a zero dispersive flux boundary condition at the borders result in calculation of unknown advective flux internally by the program, which ensures that the length of the plume is not restricted to length of the simulation domain. For the heterogamous case, the transmissivity distribution is shown in Figure 4-16-b, with four sections with hydraulic conductivities ranging from 8.31×10^{-7} m/s to 3.33×10^{-4} m/s. The dispersivities, first-order biodegradation rate constant, the size and concentration of the source zone and boundary conditions are the same as those of the homogenous case.

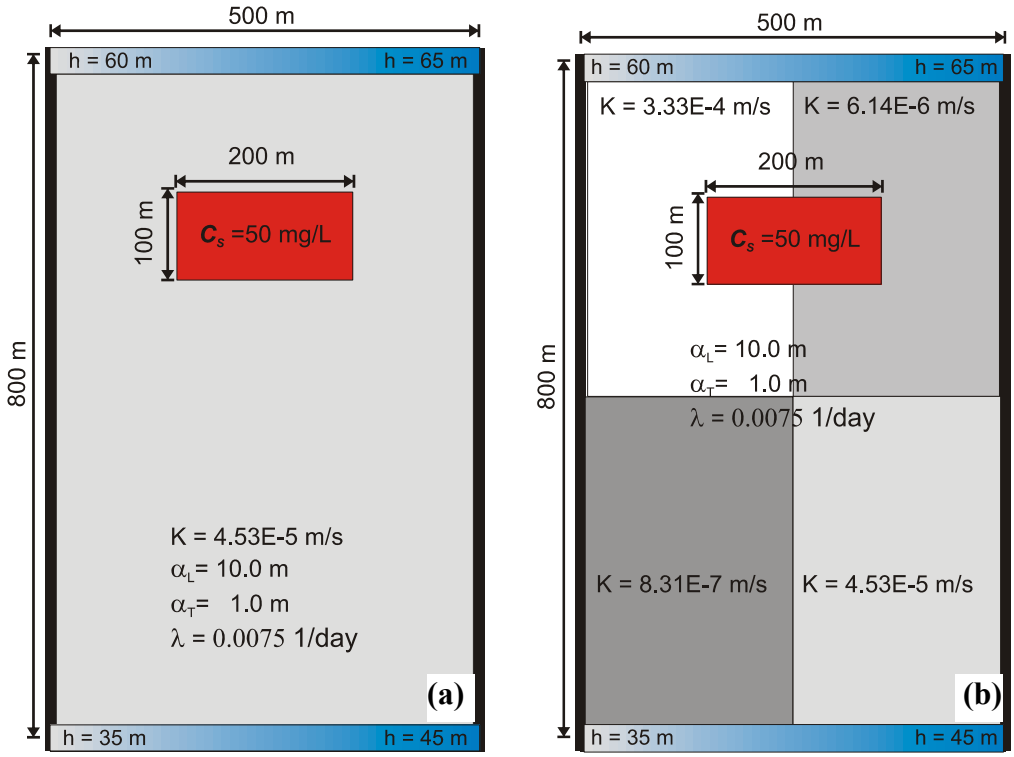


Figure 4-16: (a) The synthetic homogenous aquifer, and (b) the synthetic heterogeneous aquifer with fixed head boundary conditions at the north and south of the aquifers, no-flow (q_n = 0.0) boundary conditions at the east and west of the aquifers. The transport boundary condition involves a constant concentration boundary condition of 50 mg/L in the source area and zero-dispersive-flux ($D_{ij}\frac{\partial C}{\partial x_j} = 0.0$) boundary conditions at the north, south, east and west boundaries.

The simulations are conducted by **snasim** and MT3DMS (Zheng and Wang 1999) with similar dynamic particle allocation settings. The simulations results are recorded at time snapshots of one year, two years, five years and ten years.

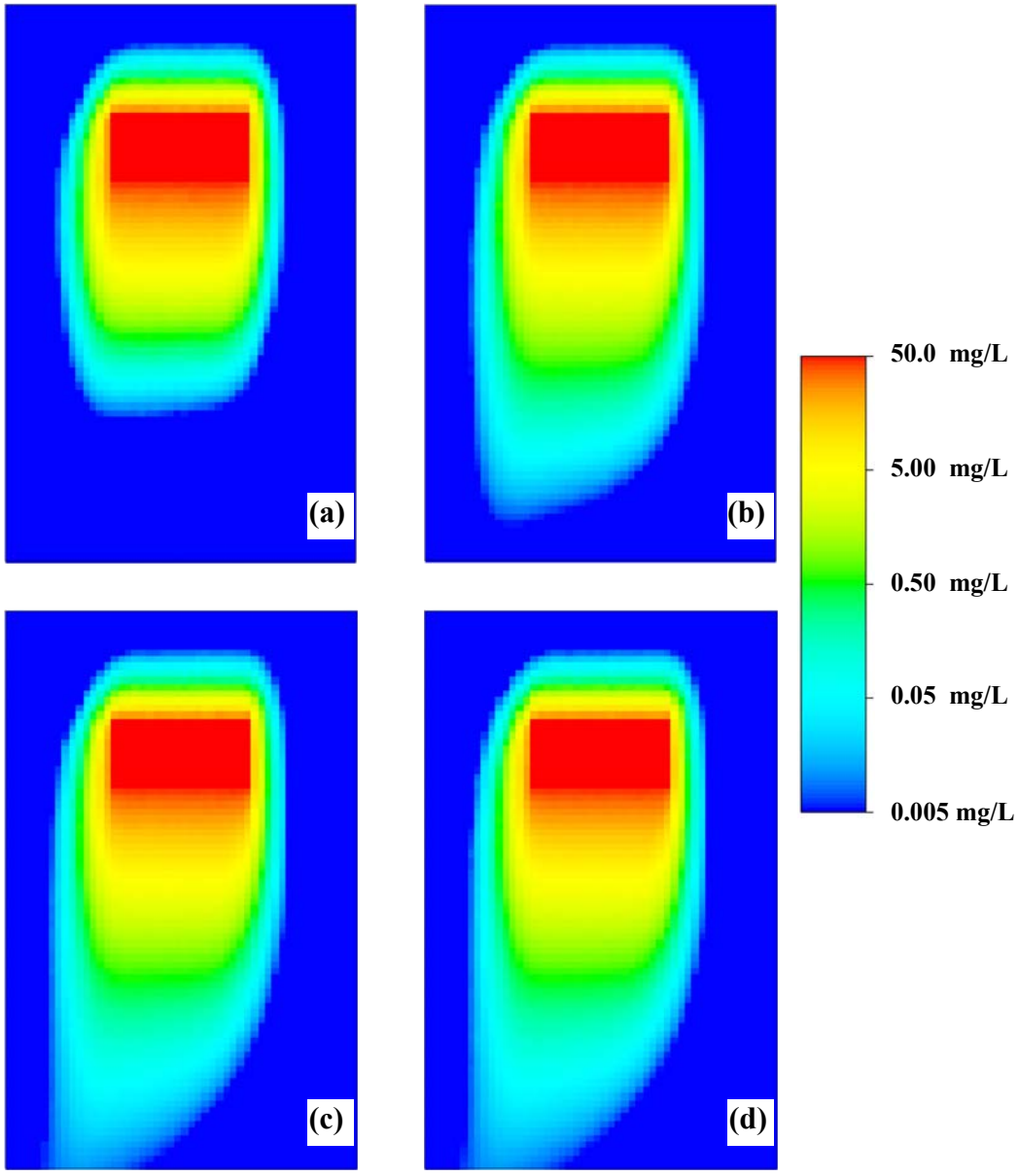


Figure 4-17: Development of the dissolved plume in the homogenous aquifer simulated by **snasim** after (a) 1 year, (b) 2 years, (c) 5 years and (d) 10 years.

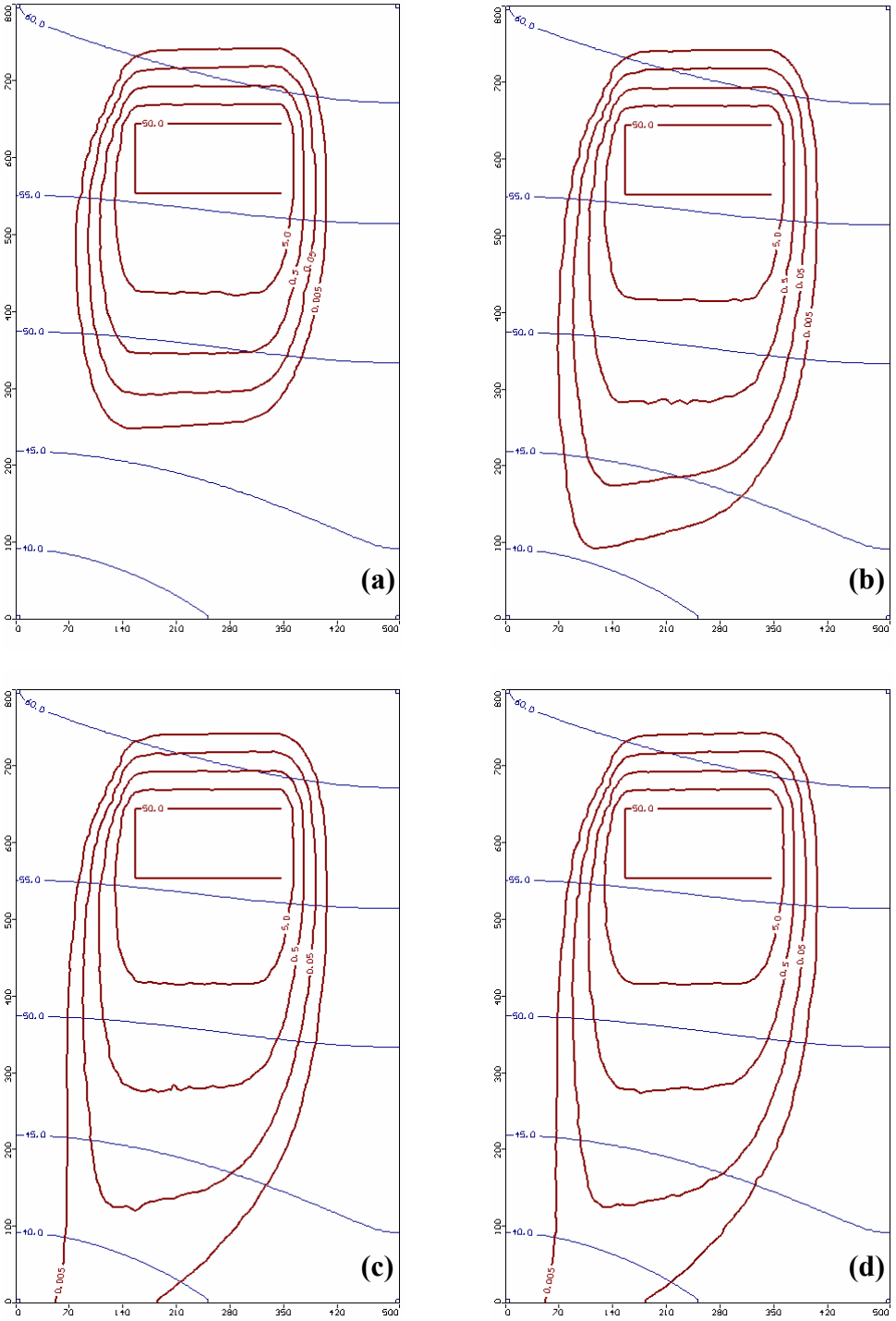


Figure 4-18: Development of the dissolved plume in the homogenous aquifer simulated by MT3DMS after (a) 1 year, (b) 2 years, (c) 5 years and (d) 10 years. The contour lines correspond to iso-concentration lines of 50, 5, 0.5, 0.05 and 0.005 mg/L.

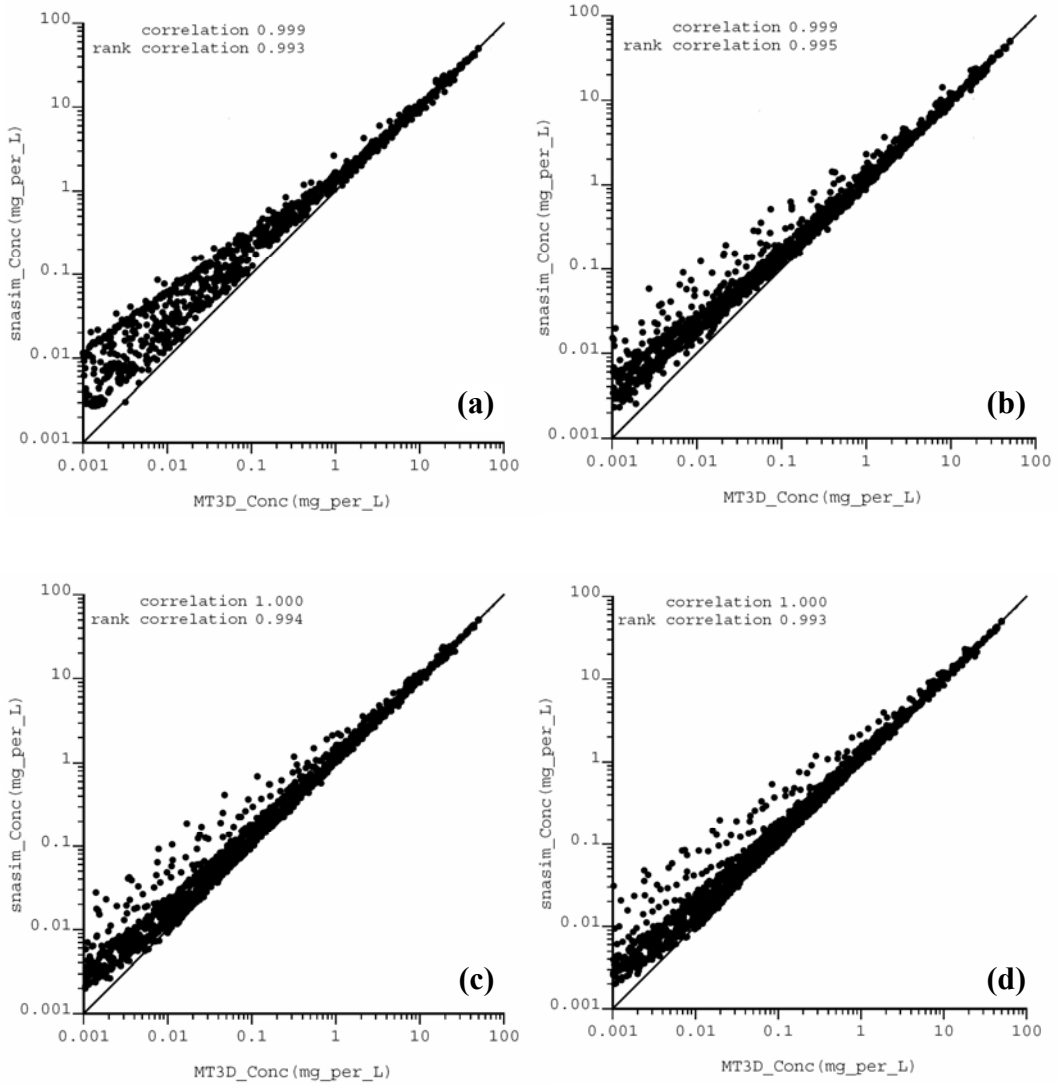


Figure 4-19: Cross-plots of concentrations obtained by MT3DMS and **snasim** after (a) 1 year, (b) 2 years, (c) 5 years and (d) 10 years.

Figures 4-17, 4-18 and 4-19 show the development of the reactive contaminant plume at different time steps as simulated by **snasim** and MT3DMS, and comparing the modeling outcomes for the homogenous case. Figures 4-20, 4-21 and 4-22 show the development of the reactive contaminant plume at different time steps, and comparing the results of **snasim** and MT3DMS for the heterogeneous case.

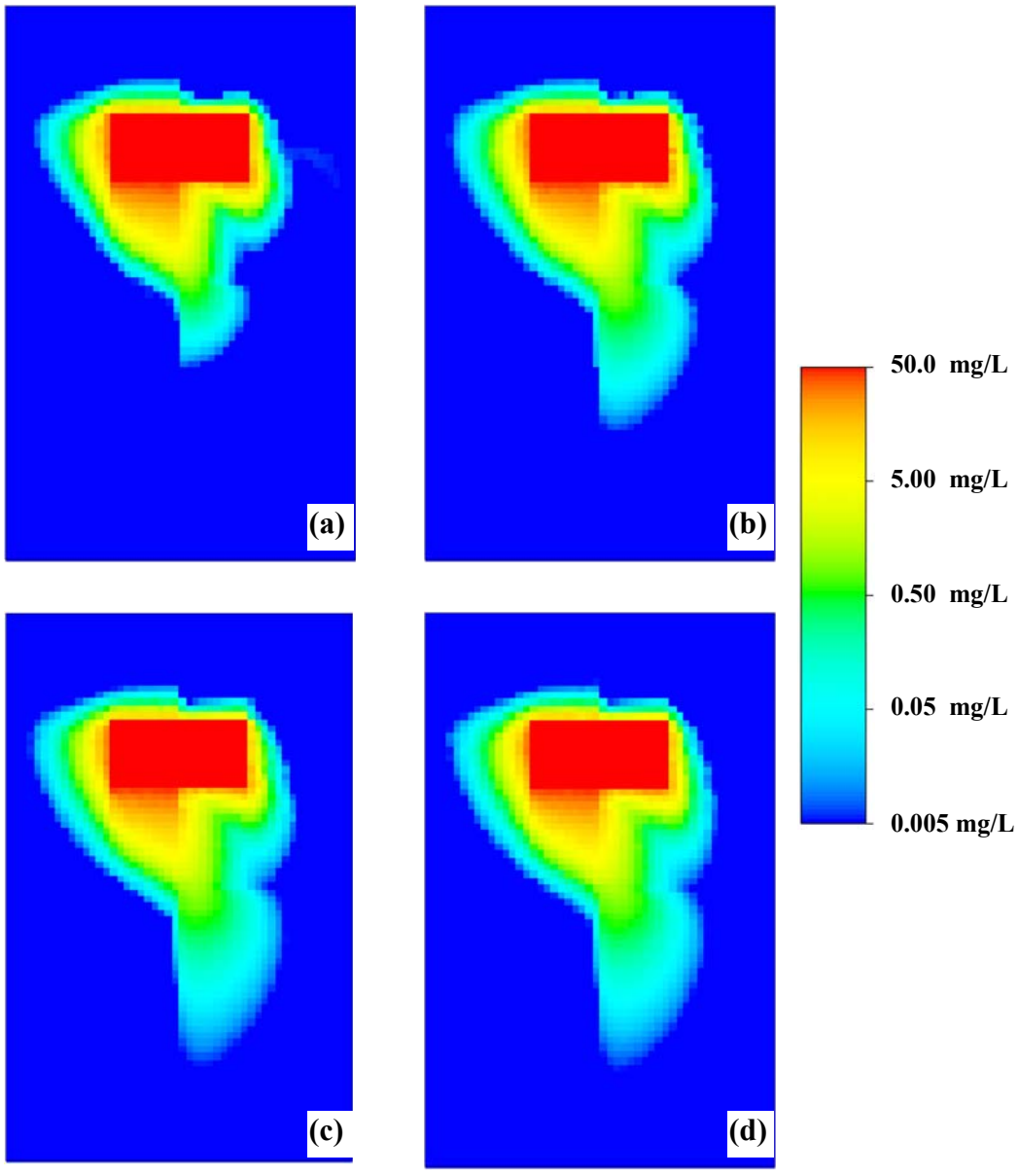


Figure 4-20: Development of the dissolved plume in the heterogeneous aquifer simulated by **snasim** after (a) 1 year, (b) 2 years, (c) 5 years and (d) 10 years.

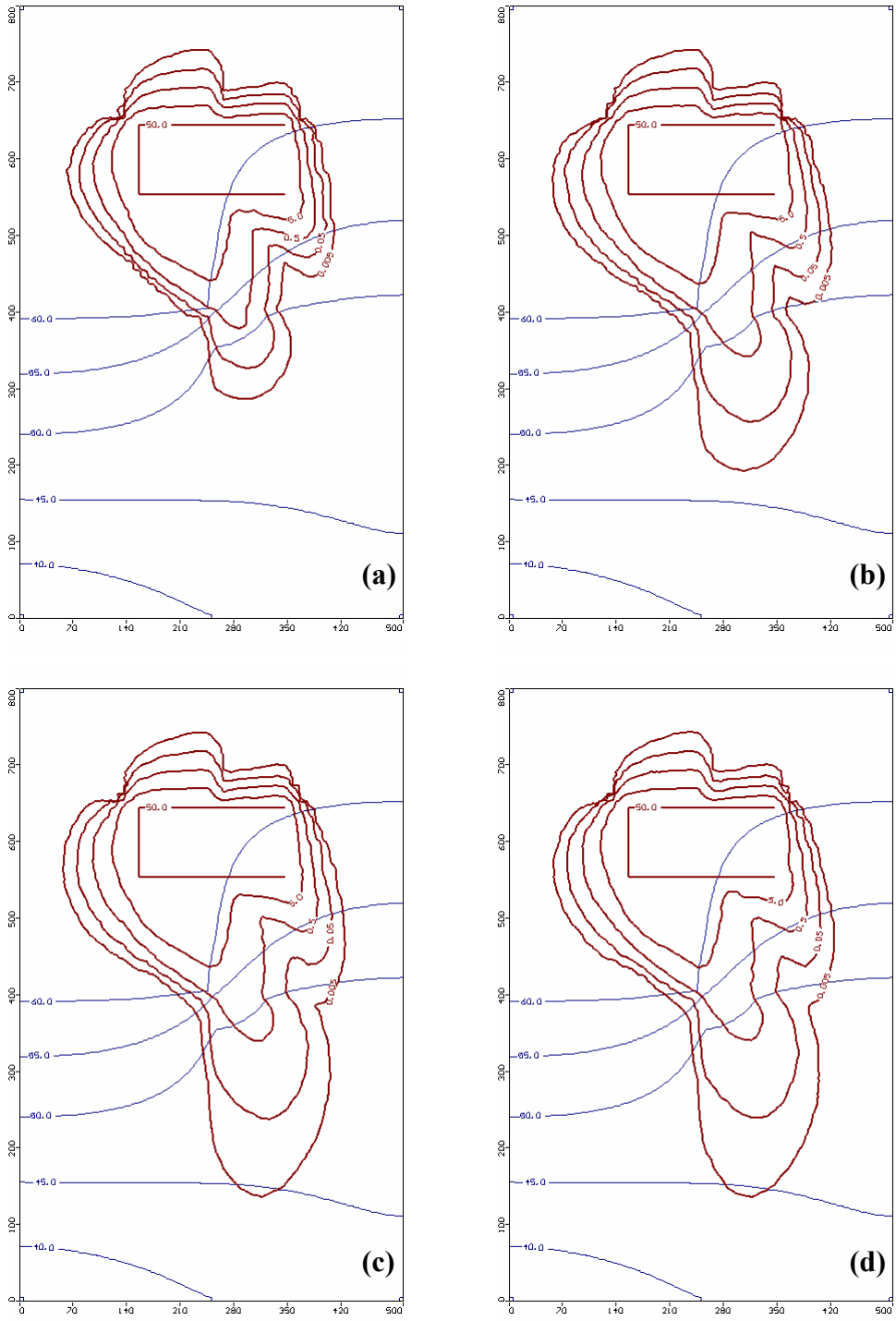


Figure 4-21: Development of the dissolved plume in the heterogeneous aquifer simulated by MT3DMS after (a) 1 year, (b) 2 years, (c) 5 years and (d) 10 years. The contour lines correspond to iso-concentration lines of 50, 5, 0.5, 0.05 and 0.005 mg/L.

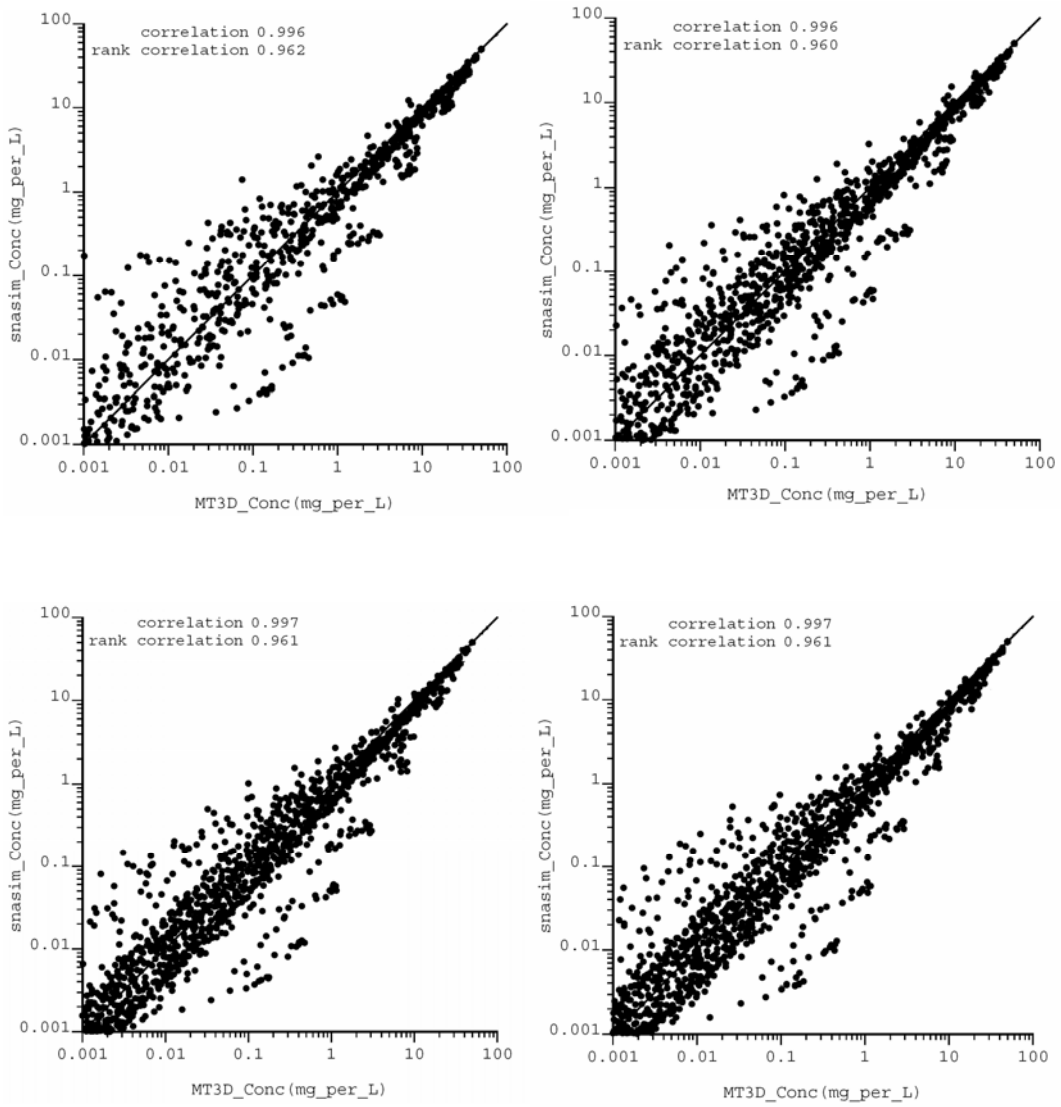


Figure 4-22: Cross-plots of concentrations obtained by MT3DMS and **snasim** after (a) 1 year, (b) 2 years, (c) 5 years and (d) 10 years.

Comparing Figure [4-17] to Figure [4-18], and Figure [4-20] to Figure [4-21], one can observe that the overall shape and size of the dissolved plumes simulated by **snasim** and MT3DMS are quite close for both homogenous and heterogeneous cases. According to the cross-plots presented in Figures [4-19] and [4-22], in general, there is a good agreement between the results of the mass transport

simulators. The discrepancy observed in the simulated concentrations is primarily observed at the edges of the plume, where concentrations are low. There are a few reasons for these relatively small inconsistencies: (1) The **snasim** program uses a semi-analytical particle tracking approach, while MT3DMS uses a numerical particle tracking scheme. As shown in Section 4.2.3, the inconsistency between the two particle tracking algorithms is negligible when the number of particles is small. In simulation with the MOC, however, there are a large number of particles (in the order of millions) and the difference between the outcomes of the two particle tracking approaches is more pronounced; (2) unlike **snasim**, MT3DMS uses an adaptive temporal discretization, that is, depending on the stability and accuracy criteria used, the time step size is adjusted throughout the simulation. In **snasim**, however, a constant temporal discretization scheme is used which satisfies the accuracy and stability criteria. This inconsistency between the temporal discretization in the two programs can be another reason for some inconsistency in the results; and (3) as explained in Section 4.3.2, there is an element of stochasticity in the dynamic particle allocation scheme used in the development of **snasim** program. The effect of selection of random number seed in execution of **snasim** is often small but it still may affect simulation results to some extent.

CHAPTER 5

STOCHASTIC PARAMETER ESTIMATION

An important aspect of uncertainty management for the contaminated sites time of remediation (TOR) problem is quantification of parameter uncertainty in the first-order biodegradation rate constant and dissolution rate constant. As discussed in Chapters 1 and 2 and 7, available estimation techniques for the first-order biodegradation rate constant do not quantify the non-linear confidence intervals under uncertainty of source properties (source size and dissolution rate) and hydraulic conductivity field, and in the case of field measurement techniques may result in estimates that are erroneous for up to a few orders of magnitude (Bauer et al. 2006, Beyer et al. 2007). The over/under-estimation of the rate constants may result in significant over/under-estimation of the state variables (e.g. plume length) as observed by Beyer et al. (2006). The dissolution of LNAPLs into groundwater has been extensively studied under laboratory conditions (Imhoff et al. 1994, Powers et al. 1994, Nambi and Powers 2003) and in numerical studies (Dillard et al. 2001, Parker and Park 2004, Zhu and Sykes 2004, Christ et al. 2006), where mass transfer correlations and upscaled dissolution models were proposed for field applications. It was observed that the upscaled models may be applied confidently only for the specific conditions under which the correlation parameters were developed and can over-predict or under-predict the flux-weighted concentrations by more than one order of magnitude (Christ et al. 2006). According to Christ et al. (2006), the local-scale (pore-scale) mass transfer correlations (for dissolution of residual NAPL ganglia) proposed by Powers et al. (1994) and Imhoff et al. (1994) have demonstrated applicability for a wider range of porous media, flow and entrapment conditions. Application of the local-scale

models under field conditions requires an estimation of effective dissolution rate that is associated with uncertainty. According to the results presented by Dillard et al. (2001), there is about one order of magnitude uncertainty in the estimated dissolution rates for a given modified Peclet number (that is directly correlated to the steady groundwater velocity) and different residual saturation conditions, as discussed in Chapter 6. This variability may be up to a few orders of magnitude when the groundwater velocity is uncertain or variable.

For groundwater management purposes at the field-scale, the value of monitoring data can be used to estimate these parameters through inverse modeling. These estimates, however, will be affected by uncertainty in model structure (source size) and other flow and transport parameters, such as distribution of hydraulic conductivity, dispersivities, etc. As pointed out in Chapter 1, Essiad et al. (2003) implemented inverse modeling to simultaneously estimate some of these parameters for a crude oil contaminated site. They simultaneously estimated the recharge rate, hydraulic conductivity value, dissolution rate constant (jointly for all BTEX compounds), first-order biodegradation rate (for each BTEX compound), and transverse dispersivity for a homogenous aquifer with a simple source geometry in an 'optimal' sense. They failed to estimate the individual dissolution rates for the BTEX compounds because of instability of the inverse problem that was stated to be due to a high correlation between the dissolution rate constant and biodegradation rate constant.

In this Chapter, the distributions (non-linear confidence intervals) of first-order biodegradation rate constant and dissolution rate constant are estimated under uncertain source geometry (size) and aquifer transmissivity through a simple inverse modeling approach. Tailoring the estimation of dissolution rate and first-order biodegradation rate to the distributions of uncertainty in the source geometry and transmissivity field through Monte Carlo type inverse modeling

helps to (1) characterize the inherent uncertainty in the values of these parameters, (2) reduce the uncertainty in the state variables and size and shape of the plume, and (3) possibly reduce the uncertainty in the source sizes by ranking the conditional realizations based on the values of the objective function.

As discussed in Chapter 1, in multi-state nonlinear inverse problems associated with subsurface mass transport, three different approaches can be considered, that is, decoupled flow and transport inverse problems, loosely (sequentially) coupled inverse problems and fully coupled inverse problems. In this work, an decoupled inverse problem is adapted. There are two reasons for decoupling the flow and transport inverse problems. The first and primary reason is to avoid the computational burden associated with the calculation of sensitivity coefficients in a multi-state system (Sahuquillo et al. 1999, Gomez-Hernandez et al. 2003) while the significance of such coupling in improving the characterization of uncertainty in dissolution rate and first-order biodegradation rate and the associated state variables is unknown. The second reason above is because simultaneous estimation of dissolution rate constant and biodegradation rate constant is prone to parameter non-uniqueness problems (local minima and high correlation between the parameters), as will be discussed in Section 5.3. Under this circumstance, addition of multiple degrees of freedom (local perturbations of hydraulic conductivities at multiple master locations) will likely have an adverse impact on the overall stability of the inverse problem. The second reason is subject to further research. However, for the purpose of investigating the potential problem, in a simple 1D example the correlation between sensitivities with respect to local velocity perturbations and sensitivities with respect to each of the two transport parameters are studied in Section 5.3.1.

In this Chapter, first, the sequential self calibration (SSC) approach (Gomez-Hernandez et al. 1997) used in conditioning multiple realizations of hydraulic

conductivity field to head data is reviewed. The model of uncertainty for hydraulic conductivity and head fields are combined with realizations of source zone geometry (obtained by the DF approach as in Chapter 3), and the rate constants are simultaneously estimated for each realization through inverse modeling with concentration data. In this work, it is observed that the potential instability of the decoupled inverse problem can be avoided through designing an appropriate layout for the monitoring network (Section 5.3).

In Section 5.1, the details of the SSC technique are presented and a post-processing step through ranking the realizations based on the measure of fit is implemented. This post-processing step is implemented to appropriately choose the conditional realizations that support a particular level of error in head measurements. This ranking and screening the realizations are to be used in an analysis of error in Chapters 6 and 7. In Section 5.2, the details of the parameter estimation step for the transport parameters are explained; and in Section 5.3 the stability of the inverse problem is investigated. In Chapter 6, a synthetic example is presented, where the results of the presented inverse modeling are compared to a set of Monte Carlo simulations to investigate the performance of the decoupled approach.

5.1 Sequential self-calibration approach

5.1.1. Some preliminary inverse modeling concepts

The basic governing equation for 2D steady-state groundwater flow in saturated porous media is given by (rewritten from Chapter 4):

$$\nabla.(T\nabla h(\mathbf{u})) = Q \qquad [5.1]$$

where, T, $h(\mathbf{u})$, and Q are transmissivity (hydraulic conductivity × saturated thickness), head and sink/source terms, respectively. Appropriate boundary (and initial) conditions are also defined for the system. Equation [5.1] represents the forward problem of finding the distribution of piezometric heads for a fully known transmissivity field and boundary conditions. Numerical discretization of Equation [5.1] by the finite volumes method results in the system of linear equations:

$$[\mathbf{A}]\{\mathbf{h}\} = \{\mathbf{q}\} \qquad [5.2]$$

where, $[\mathbf{A}]$ is the matrix of coefficients, $\{\mathbf{h}\}$ is the vector of unknown nodal piezometric head values, and $\{\mathbf{q}\}$ is the flux vector that represents loading and boundary conditions. According to Sun (1994), the forward problem in Equations [5.1] and [5.2] is 'well-posed', as it satisfies the requirements of well-posedness:

- **Existence:** There exists a function ($T \rightarrow h(\mathbf{u})$) which satisfies the governing equations and boundary conditions
- **Uniqueness:** There is no solution other than $h(\mathbf{u})$
- **Stability:** The changes in solution $h(\mathbf{u})$ are small for sufficiently small variations of input data.

Simultaneous or sequential adjustment of model structure as well as model parameters for the purpose of making input-output relation of the model fit to any observed excitation-response of the real system is termed 'model calibration'. If the model structure (e.g. zonation or spatial correlation) is assumed to be known,

the problem of determining model parameters (T in Equation [5.1]) from the observations of the model response ($h(\mathbf{u})$ in Equation [5.1]) and other available information is called parameter estimation or inverse modeling. Inverse problems are almost always 'ill-posed', that is, their solutions are often non-unique and prone to instability problems: small variations in observations results in unacceptably large variations in the estimated model parameters. In other words, for any given problem many different sets of property estimates may provide satisfactory and indistinguishable data fits. Some of these parameter estimates can be grossly in error with respect to the actual properties and may result in erroneous predictions. To reduce this statistical uncertainty one must decrease the number of unknowns through 'parameterization' and/or utilize additional information through 'regularization'.

Parameterization is the technique used to simplify the structure of an N-dimensional parameter space such that it can be represented approximately by an M-dimensional space (M<<N). According to Carrera et al. (2005), the most common parameterization approaches are zonation, point estimation and parameterization with pilot points (master points). In parameterization with zonation, the domain is partitioned in a set of zones. In point estimation, the model parameters are represented as random numbers with known distributions and parameterized by their statistical moments. In parameterization with pilot points, the aquifer properties are expressed as linear combination of some unknown model parameters and their direct measurements. The kriging weights are adopted as the weights in the linear combination.

Regularization refers to supplementing extra information that is independent of the measurement of the state variables. This extra information may be in form of designated parameter values at some points, or by limiting the admissible range of the model parameters at some locations. This supplementary information is called

'prior information'. As a classical approach, the inverse problem is transferred to an optimization problem by defining an objective function. The regularization terms (prior information) are either directly added to the objective function or considered as constrains in optimization of the objective function. In the context of groundwater flow problems, the regularized form of the objective function is expressed as:

$$F = F_h + \lambda F_T \qquad [5.3]$$

where, F_h represents the weighted squared difference in simulated and observed state variables (e.g. piezometric heads) which is expressed as:

$$F_h = \left(\mathbf{h} - \mathbf{h}^*\right)^T \mathbf{C}_h^{-1} \left(\mathbf{h} - \mathbf{h}^*\right) \qquad [5.4]$$

where, $\mathbf{h}, \mathbf{h}^*$ and $\mathbf{C}_h^{-1}$ are the vector of simulated heads, the vector of observed heads, and the matrix of weights. F_T in Equation [5.3] represents the plausibility term, and λ is the regularization coefficient. The plausibility term F_T is expressed by:

$$F_T = \left(\mathbf{Y} - \mathbf{Y}^*\right)^T \mathbf{C}_T^{-1} \left(\mathbf{Y} - \mathbf{Y}^*\right) \qquad [5.5]$$

where, $\mathbf{Y}, \mathbf{Y}^*$ are the vectors of point estimation and point measurements of the model parameters ($\mathbf{Y}=\{Y_i,\ i=1,\ \ldots,\ N_T\}=\{\ln T_i,\ i=1,\ \ldots,\ N_T\}$), N_T is number of prior values (point measurements of transmissivity T_i) and $\mathbf{C}_T^{-1}$ is the matrix of weights. As explained previously, alternatively, regularization may be implemented through constrained optimization. In this case, only the first term on the right hand side of Equation [5.3] is retained and some limits are imposed on the changes on the point estimations of the parameters throughout the optimization process.

Although appropriate parameterization of the model and regularization of the objective function crucially help reduce the instability and non-uniqueness problems, there still may not be a single set of parameters leading to a good representation of the reality. As a result, the inverse modeling techniques that

estimate the model parameters in an optimal sense, despite their computational efficiency, are not applied in this work. Instead, the Monte Carlo based (conditional simulation based) techniques such as SSC that generate multiple realizations conditioned to both piezometric head and hydraulic conductivity data are employed.

5.1.2. Sequential self-calibration approach: theory and implementation

The sequential self-calibration (SSC) approach was originally developed for a 2D steady-state condition with multi-Gaussian hydraulic conductivity distribution (Sahuquillo et al. 1999, Gomez-Hernandez et al. 1997). It was later extended for non-Gaussian distributions (through indicator simulation) (Capilla et al. 1999), for transient flow conditions (Hendricks Franssen et al. 1999a), for 3D flow in fractured media (Hendricks Franssen et al. 1999b), for two-phase fractional flow (Wen et al. 2003), for coupled flow and transport (Hendricks Franssen et al. 2003) and for integration of remote sensing data (Hendricks Franssen et al. 2008).

The SSC technique is an iterative technique coupling geostatistics and optimization. First, multiple conditional realizations of hydraulic conductivity field are generated using standard geostatistical techniques such as **sgsim** in GSLIB (Deutsch and Journel 1998). By construction, all the generated realizations honor the static data as well as input statistics such as histogram and variogram. Throughout the SSC process, the hydraulic conductivity field (and boundary conditions) is modified repeatedly without changing the spatial correlation structure of hydraulic conductivity until the head observations are also honored. Figure 5-1 shows the flowchart of the SSC technique. The unique aspects of SSC are designation of master points that significantly reduces the CPU time, a propagation of perturbations by kriging which preserves the spatial correlation of perturbations, and a computationally efficient algorithm for calculating the sensitivity coefficients within a single flow simulation run. The

major assumptions and simplifications of the SSC technique are (1) the support of point measurements of hydraulic conductivity is the same as that of grid blocks, and (2) a prior estimate of log-hydraulic conductivity variogram is available.

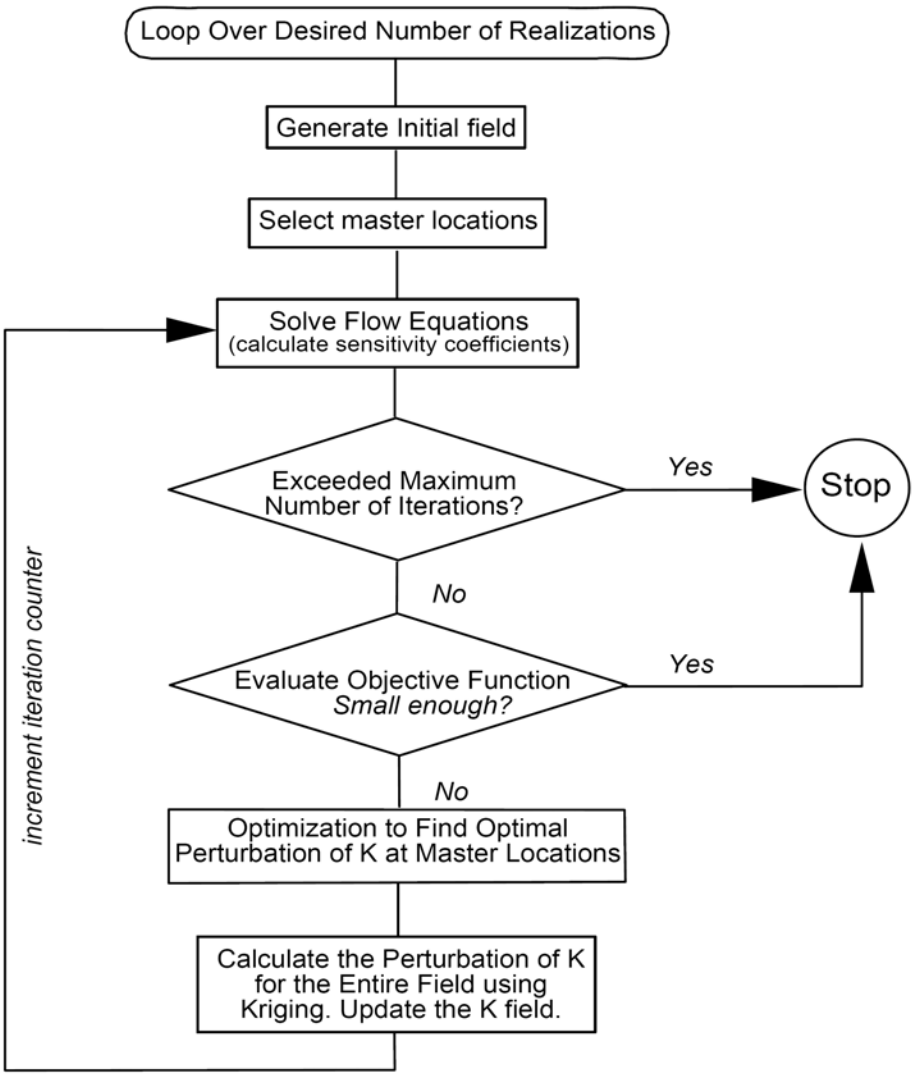


Figure 5-1: The flowchart of SSC technique (after Wen et al. 1999)

After constructing multiple conditional hydraulic conductivity realizations by standard geostatistical techniques, the flow equation (Equation [5.1]) is solved for every realization using standard finite difference method (Equation [5.2]) and the distribution of piezometric head is calculated. The hydraulic conductivity and

head distributions for each realization are represented in vector form by $\mathbf{Y} = \{Y_i, i=1, \ldots, N\} = \{\ln T_i, i=1, \ldots, N\}$, and $\mathbf{h} = \{h_i, i=1, \ldots, N\}$, respectively, where N is number of grid blocks. In the SSC approach, the reproduction of the measured piezometric heads is analyzed by evaluation of an objective function defined by (Gomez-Hernandez et al. 1997):

$$F = \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} w_{ij} \left(h_i^m - h_i\right)\left(h_j^m - h_j\right) \quad [5.6]$$

where, F is a generalized sum of squared differences between the measured heads $\left\{h_i^m, i = 1,..., n_h\right\}$ and calculated heads; and w_{ij} represent the weights that can be set to the inverse of the head error covariance matrix. It is quite common to assume that the errors (the sum of estimation and measurement errors) are spatially uncorrelated and Equation [5.6] can be written in a simplified form:

$$F = \sum_{i=1}^{n_h} w_i \left(h_i^m - h_i\right)^2 \quad [5.7]$$

In the SSC algorithm, regularization is implemented either by constraining the maximum departure of the updated field from the conditional expectation of log-hydraulic conductivity as given by the ordinary kriging estimate from the log-hydraulic conductivity data or by addition a plausibility term to the objective function (Gomez-Hernandez et al. 1997):

$$F = \sum_{i=1}^{n_h} w_i \left(h_i^m - h_i\right)^2 + \tau \sum_{i=1}^{m_Y} \sum_{j=1}^{m_Y} \nu_{ij} \left(Y_i^{(0)} - Y_i^{(k)}\right)\left(Y_j^{(0)} - Y_j^{(k)}\right) \quad [5.8]$$

where, τ is a trade-off coefficient, ν_{ij} are the coefficients of the inverse of the kriging estimation matrix, and $Y_i^{(0)}$ and $Y_i^{(k)}$ are the values of log-hydraulic conductivity fields at the master locations for the seed field and for the updated field after k iterations. According to Gomez-Hernandez et al. (1997) and Wen et al. (1999), conditioning to head data is considered to be achieved when the value of the objective function (Equation [5.7] or [5.8]) is less than a user-defined small tolerance.

As discussed previously, parameterization of hydraulic conductivity field in the SSC approach is through considering an adequate number of master locations and defining the perturbations in all grid blocks of the seed fields as a linear combination of perturbations at this limited number of master locations. Mathematically, this statement can be expressed by:

$$\Delta Y_i = \sum_{k=1}^{m_Y} \lambda_i^k \Delta Y_k \qquad [5.9]$$

where, ΔY_i represents the perturbation at every grid cell, ΔY_k represents the perturbation at master locations, and λ_i^k represents the ordinary kriging weights. Equation [5.9] ensures a smooth perturbation over entire field and minimal disturbance to the correlation structure of the seed fields. It should be noted that the set of master points include all hydraulic conductivity data locations as well as additional locations. Gomez-Hernandez et al. (1997) recommend that the master points should be located on a pseudo-regular grid with spacing in the order of one-third of the correlation range of the variogram used to construct the seed fields.

The perturbations at the master locations are calculated through optimization (minimization) of the objective function defined by Equations [5.7] or [5.8]. In the original presentation of SSC method in Gomez-Hernandez et al. (1997), constrained minimization of Equation [5.7] was implemented through a gradient based optimization algorithm (the method of feasible directions). The optimization process requires (1) the piezometric heads to be expressed as explicit functions of the perturbations at master locations, and (2) the derivative of piezometric heads with respect to log-hydraulic conductivity perturbation at all master locations (sensitivity coefficients) to be calculated. The first requirement is met through linearizing the Equation [5.2] by Taylor series expansion around the vector of piezometric heads associated with the seed field $\{h^{(0)}\}$. Thus, the linear approximation to the piezometric head field is given by:

$$\{h^l\} = \{h^{(0)}\} + \sum_{k=1}^{m_Y} \left(\frac{\partial}{\partial p_k} \{h\} \Big|_{\{h\}=\{h^{(0)}\}} p_k \right) \qquad [5.10]$$

where, the second term on the right hand side represents the approximate change in the head field due to small perturbations in hydraulic conductivity field (and boundary conditions).

The second requirement is met through the solution to the 'sensitivity equations'. Taking the derivative of Equation [5.2] and rearranging results in the set of sensitivity equations:

$$[A].\frac{\partial}{\partial p_k}\{h\} = \frac{\partial}{\partial p_k}\{Q\} - \frac{\partial}{\partial p_k}[A].\{h\} \qquad [5.11]$$

As pointed out by Gomez-Hernandez et al. (1997), the Equation [5.11] gives the full vector of sensitivity coefficients through solution of a set of *N* linear equations with the same matrix of coefficients as the original flow problem (Equation [5.2]). The derivatives of {Q} and [A] with respect to the parameters can be easily calculated. The reader is referred to Gomez-Hernandez et al. (1997) for the details on calculation of these derivatives as well as minimization of the object function defined in Equation [5.7].

Regularization of the objective function can be performed either by adding a plausibility term to the objective function or by constrained minimization of the objective function. As explained before, in the original presentation of SSC method, constrained minimization of Equation [5.7] was implemented. The constraints imposed are in the form of limiting the departure of final updated log hydraulic conductivity field from the conditional expectation of log-hydraulic conductivities given by the ordinary kriging from the log hydraulic conductivity data. According to Gomez-Hernandez et al. (1997), these constraints can be expressed by:

$$Y_{k,\min} - Y_k^{(l)} \le \Delta Y_k \le Y_{k,\max} - Y_k^{(l)} \qquad k = 1, ..., m_Y$$
where $Y_{k,\min}$ and $Y_{k,\max}$ are given by :
$$Y_{k,\min} = \min\left(Y_k^{(0)}, Y_{k,\text{krig}} - t\sigma_{\text{krig}}\right) \qquad k = 1, ..., m_Y$$
$$Y_{k,\max} = \max\left(Y_k^{(0)}, Y_{k,\text{krig}} + t\sigma_{\text{krig}}\right) \qquad k = 1, ..., m_Y \qquad [5.12]$$

where, $Y_{k,\text{krig}}$ is the ordinary kriging estimate at any master location and σ_{krig} is the corresponding kriging standard deviation, and *t* is the proportionality parameter which is set to 2 (Gomez-Hernandez et al. 1997) or 3 (Wen et al. 1999). The criteria for stopping the iterations on minimization of the objective function include (1) iterations continue until the objective function is sufficiently close to zero, or (2) improvements in the model fit (decrease in the value of the objective function) in consecutive iterations is insignificant, or (3) the maximum number of outer iterations has been exceeded.

As discussed before, a two-step decoupled approach has been applied in this thesis to characterize the uncertainty in dissolution and first-order biodegradation rates. In the first step of the proposed approach, the SSC technique is applied to generate hydraulic conductivity realizations conditioned to both hydraulic conductivity and head data. The SSC program used for this purpose is based on the source code presented by Wen et al. (1999). The original code was slightly modified and simplified for a steady-state groundwater application.

5.1.3. Screening the conditional realizations

As part of this research, the effect of error in head observations on subsequent estimation of dissolution rate and first order decay rate and predictions of the state variables will be investigated in Chapters 6 and 7. Thus, multiple realizations that support a given level of measurement errors are required. Some works in the literature investigate the relationship between the uncertainties in model parameters and measurement errors and/or predictive error variance. The work of

Vecchia and Cooley (1987) is one of the most important early works in this area. The works of Hendricks-Franssen and Gomez-Hernandez (2003), McKenna et al. (2003), Moore and Doherty (2005) and Tonkin et al. (2007) are among the recent works in this area for distributed systems. McKenna et al. (2003), Moore and Doherty (2005) and Tonkin et al. (2007) investigated the relationship between the predictive error variance (for transport) and uncertainty in the hydraulic conductivity fields. Moore and Doherty (2005) and Tonkin et al. (2007) use prediction error variance as an objective function for under-determined and over-determined systems in the context of linear and non-linear regression. In both works, they minimized the predictive error variance for a solute travel time. McKenna et al. (2003) added a strong regularization constraint to the objective function to address the relationship between the non-uniqueness in the T fields and resulting transport calculations. Hendricks-Franssen and Gomez-Hernandez (2003) investigated the impact of measurement error in stochastic inverse modeling by the SSC method. In their work, they set the tolerance value (J_{tol}) for minimization of the object function (Equation [5.7]) equal to:

$$J_{tol} = n\sigma^2 \qquad [5.13]$$

where, n is the number of head observations, and σ^2 is the variance of head measurement error based on Gaussian noise. Throughout data conditioning, they evaluated the objective function for each updated log-hydraulic conductivity field. The log-hydraulic conductivity field was considered as to be successfully conditioned to head data if the value of the objective function was below the pre-specified tolerance value (Equation [5.13]).

As discussed in Section 5.1.2, the SSC calibration stops when the value of the objective function is below a pre-specified value, or the improvements in the model fit are insignificant in consecutive iterations, or the maximum number of outer iterations exceeds the maximum allowed. These stopping criteria result in conditional realizations that do not show the same level of mismatch to head data,

as will be investigated through a numerical experimentation. A ranking and screening ensures that the level of mismatch (or fitting) of the conditional realizations with the head observations is consistent with the observations errors. One can define a measure of fit denoted by *s* as:

$$s^2 = \frac{F}{n} \qquad [5.14]$$

where, *F* is the value of the objective function while weights in Equation [5.7] are set to the inverse of the head error covariance matrix and *n* is the number of observations. Similar to Equation [5.13] and the work of McKenna et al. (2003) (among others), for every realization subject to conditioning to head data, the value of *s* should be as close to one as possible to (1) ensure the convergence of the conditioning algorithm and (2) avoid fitting the noise in observations. This concept can be used in ranking and screening the realizations conditioned to head data by the SSC approach. To investigate the distribution of *s* for realizations conditioned to head data by the SSC, a numerical experimentation is performed as follows.

Figure 5.2 shows a 2D reference log hydraulic conductivity field, the associated steady-state piezometric head response, and the locations of sampling points. The simulation domain is 250 m by 160 m and the squared shape grid blocks are 2.0 m by 2.0 m. The reference log hydraulic conductivity field has a mean of -10.1, standard deviation of 1.2 both in natural logarithm units ($\log_e$ m/s) and a spatial correlation defined by a spherical variogram with a nugget effect equal to 0.1 and a range of 36.0 m. The boundary conditions for the flow field include fixed head boundary conditions at the north and south of the site with constant head values of 4.5 m and 2.0 m, respectively; and no-flow boundary conditions at the west and east boundaries. There are 100 head observation locations (black circles) and 18 permeability measurement locations (white circles). Three data sets for piezometric heads have been created by introducing three different levels of Gaussian noise (σ = 0.10m, 0.15m, 0.25m) to the piezometric head values

sampled from the reference head field shown in Figure 5-2-b. Three sets of 1000 hydraulic conductivity realizations conditioned to hydraulic conductivity data only are generated by sequential Gaussian simulation program of GSLIB (Deutsch and Journel 1998). Figure 5-3 shows the histograms of the measure of fit *s* for the three sets of hydraulic conductivity realizations calculated based on different levels of error in the head observations, before conditioning to head data by the SSC. Figure 5-4 shows the histograms of the values of *s* for the three ensembles of realizations after conditioning to head data by the SSC. In the implementation of the SSC, a total of 126 master points are used (roughly two master points per correlation range in each direction), a maximum of 25 outer iterations are allowed, and the damping parameter and the minimum relative tolerance (normalized to the initial value of the objective function) are set to 0.2 and 0.01, respectively. Also, the minimum difference of objective function in two consecutive iterations, and the maximum number of times that the difference of objective function is smaller than the pre-specified value are set to 0.005 and 10, respectively. The minimum relative tolerance has been intentionally set to a small number (0.01) to ensure that the optimization process searches for the best possible fit to the observed heads to investigate the effects of different levels of noise on the histograms of *s*. In Figures 5-3 and 5-4, it can be observed that: (1) when the standard deviation of measurement error is large (noisy data set), there may be some realizations that have their measure of fit close to one (Figure 5-3-c) and are therefore acceptable even before conditioning to head data; (2) when the standard deviation of measurement error is small, there may only be a small number of conditioned realizations (by the SSC) that satisfy the requirement of *s* close to one (Figure 5-4-a); and (3) setting the target value of the objective function to a very small value when the observed head data is too noisy may result in conditional realizations that have a good fit (over-fitted) to noisy data set but their value of *s* deviates from one.

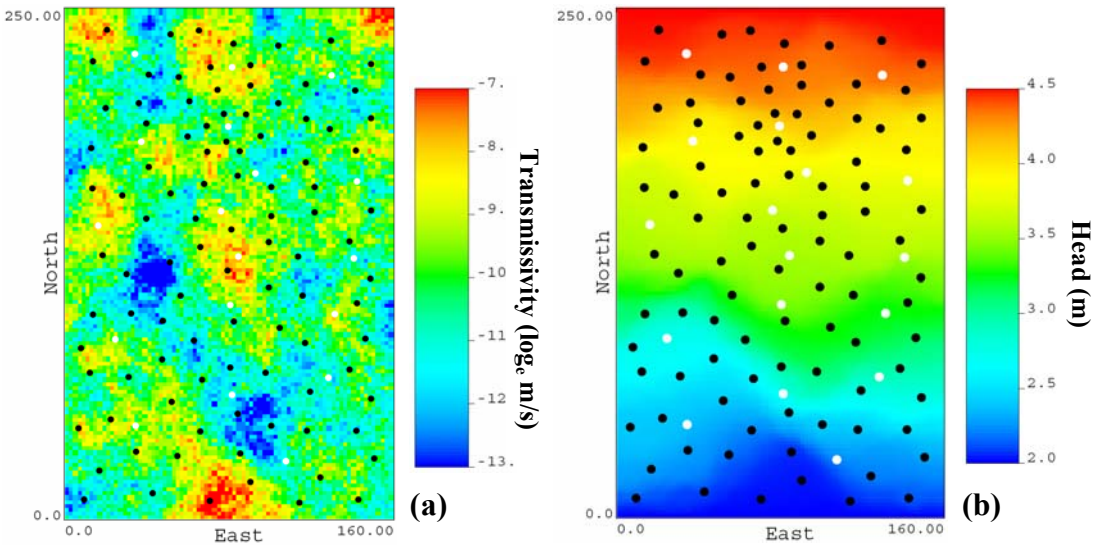


Figure 5-2: (a) Reference hydraulic conductivity field, and (b) the associated piezometric head response. There are 100 head observation locations (black circles) and 18 hydraulic conductivity measurement locations (white circles).

It should be noted that running the SSC program requires a large amount of CPU time. In the above example, the required CPU time for running the SSC algorithm for 1000 realizations under steady-state conditions (with known boundary conditions) was roughly equal to 62 hours on a Dell Precision PWS470 workstation with Intel Xeon ™ 3.00GHz CPU. Under transient flow condition, the required CPU time can be significantly larger. Therefore, generating a very large number of realizations, conditioning them to head data by the SSC, and accept/reject them based on the closeness of measure of fit *s* to one can be quite time-consuming.

In this work (Appendix A), an alternative way of generating multiple Gaussian realizations with their measure of fit *s* close to one has been examined. This alternative approach, termed distance-based simulation (Scheidt et al. 2008 and Caers 2008), builds on the concepts of multidimensional scaling (MDS), Karhunen-Loeve (KL) expansion, kernel principal component analysis (KPCA), and modeling and simulation in the metric and feature spaces.

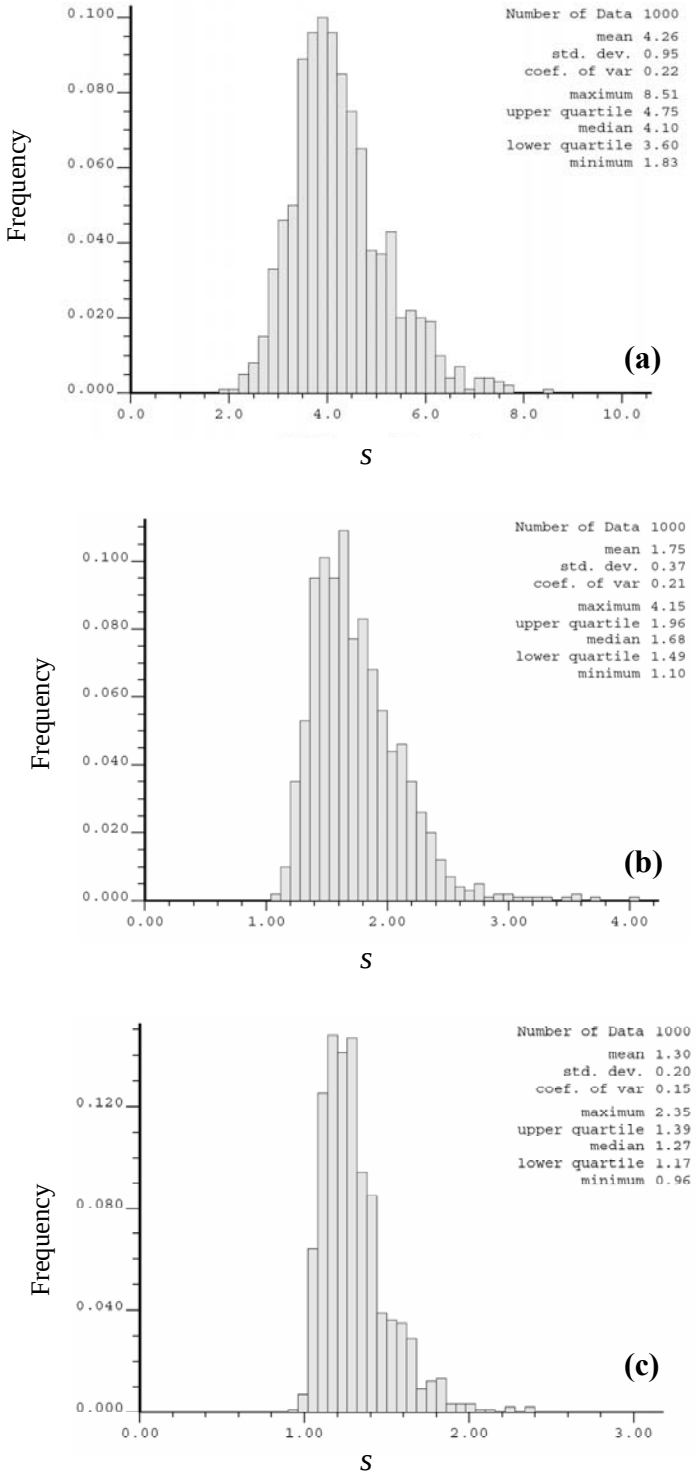


Figure 5-3: Histogram of the measure of fit s for the ensembles of realizations before conditioning to head data for measurement error standard deviations of (a) $\sigma = 0.10$ m, (b) $\sigma = 0.15$ and (c) $\sigma = 0.25$.

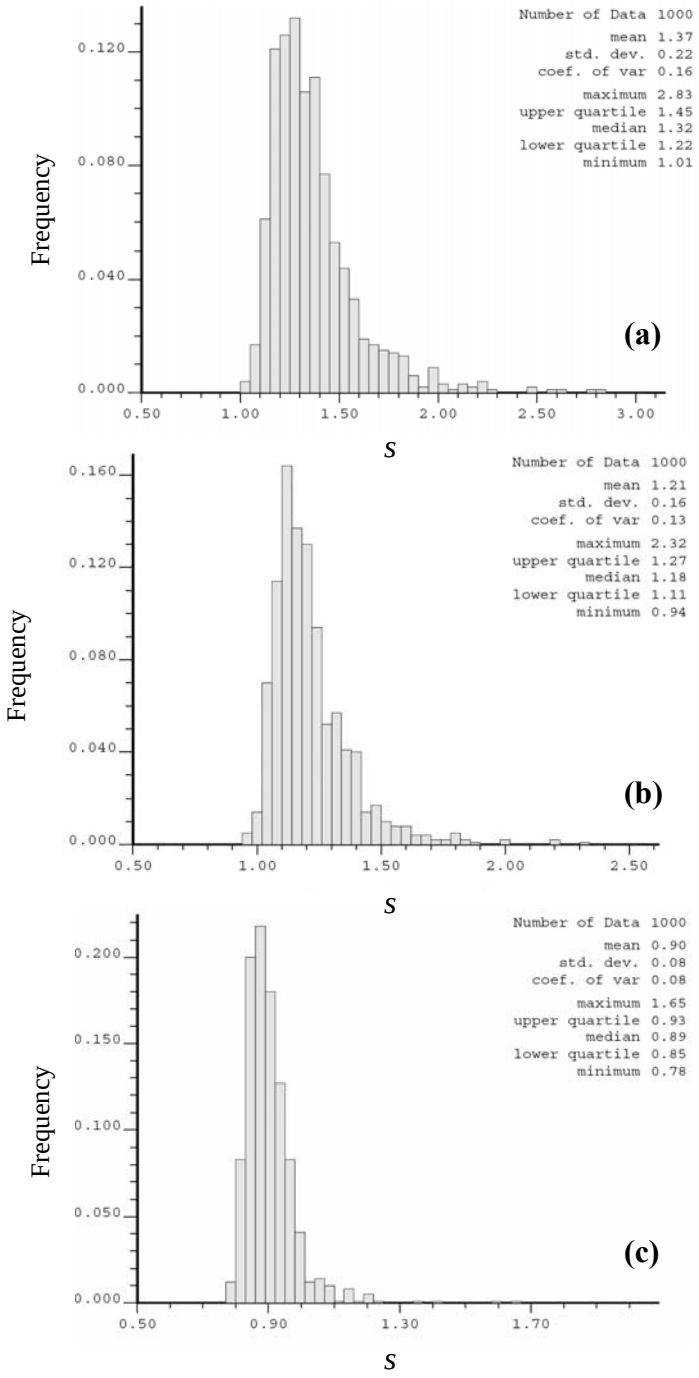


Figure 5-4: Histogram of the measure of fit *s* for the ensembles of realizations after conditioning to head data for measurement error standard deviations of (a) σ = 0.10 m, (b) σ = 0.15 and (c) σ = 0.25.

Using distance-based simulation, one may choose a limited number of realizations that have a desired response and may expand the set of realizations that have similar responses while being conditioned to static data and reproduce (some of) the input statistics. As discussed in Appendix A, this technique needs a very careful post-processing of the realizations to ensure the reproduction of input histogram (especially standard deviation) by the expanded set of realizations. It is also observed that many of the generated realizations may closely resemble the input realizations. Thus, despite being extremely fast technique in generating realizations, it has not been extensively used in this thesis.

As stated previously, ranking and screening the conditional realizations based on closeness of their measure of fit to one is used in Chapters 6 and 7 to create a subset of realizations that support a particular level of measurement error.

5.2 Stochastic parameter estimation for k_{dis} and λ

5.2.1. Formulation of the problem

As reviewed in Chapter 2, several analytical and numerical models have been developed to obtain a solution to the TOR problem through simulation of advection, dispersion, sorption, and biodegradation (decay) of solute petroleum hydrocarbon contaminants in groundwater. The input parameters to these models usually include flow and transport parameters such as areal recharge, hydraulic conductivity, effective porosity, dispersivities, retardation factor and solute decay rates (or biodegradation rate constants), and sometimes include dissolution rate constants and reaction kinetics. The modeling outputs (state variables) almost always involve the spatial and temporal distribution of heads and concentrations across the modeling domain. The length of the plume and the mass loaded into the aquifers are then estimated based the simulated concentrations.

In practice, estimates of the flow and transport parameters are obtained through field measurements, laboratory experiments, calibrating the models to site-specific data (inverse modeling); or as a last resource literature prescribed values are used. In any case, the estimated parameters are often prone to uncertainty. Inverse modeling techniques (Essaid et al. 2003, Hendricks Franssen et al. 2003 among others) have been used to estimate some of the flow and transport parameters. The distribution of hydraulic conductivity, the values of the decay rates (or biodegradation rate constants) of contaminant species and the parameters representing the source of contaminants (source geometry and dissolution rate) are the most important parameters with the most impact on the outcome of the TOR problem. It has been observed that simultaneous estimation of these parameters leads to potential instability problems (Essaid et al. 2003) in an inverse modeling context. The instability problems in joint estimation of first-

order biodegradation rate constant and dissolution rate constant can be avoided through weighting the mismatch between simulated and observed concentrations in the objective function inverse proportional to simulated concentrations (as used by Anderman and Hill 1999) and designing an appropriate monitoring layout for sampling concentrations. Figure 5-5 shows the workflow for the decoupled inverse problem.

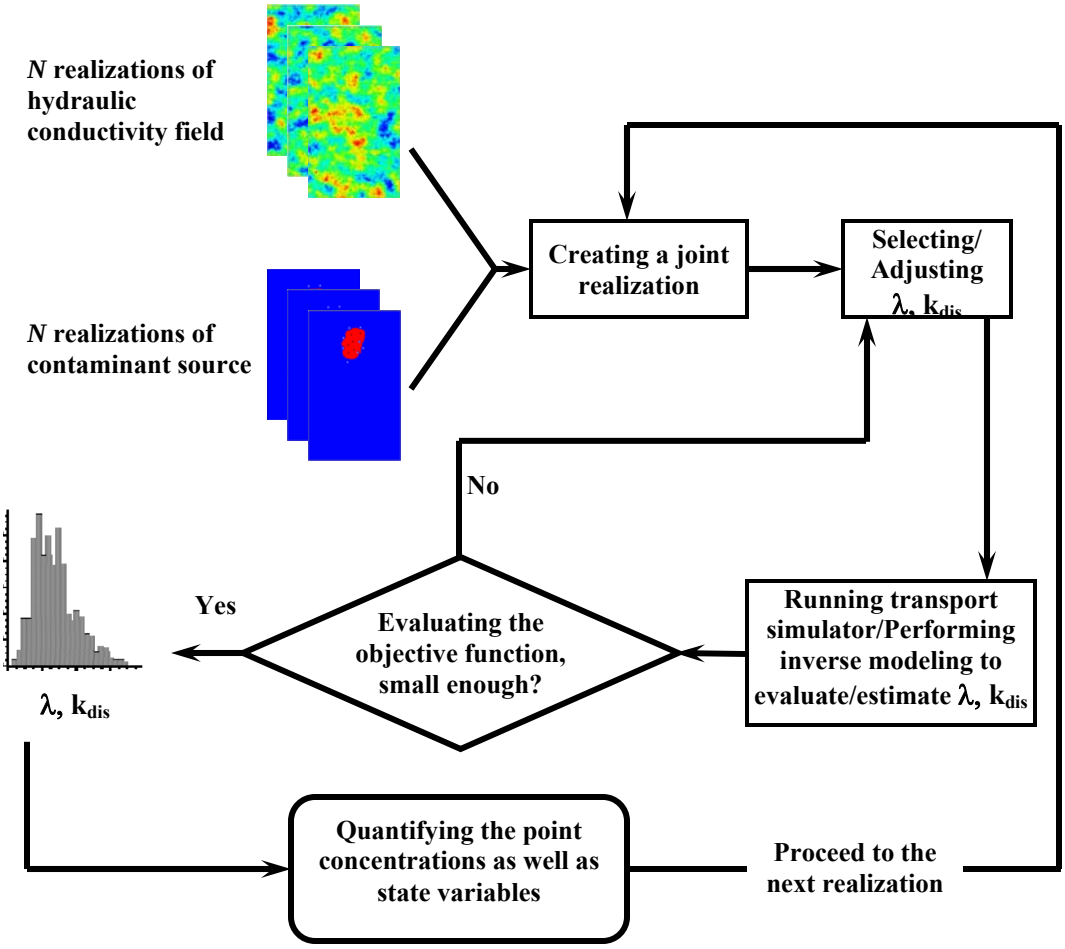


Figure 5-5: The workflow for the decoupled inverse problem.

First, joint realizations of source geometry and hydraulic conductivity field are constructed with the assumption of full statistical independence. The realizations of the hydraulic conductivity field are previously conditioned to head data by the SSC algorithm (Section 5.1). The values of the first-order biodegradation rate

constant, and dissolution rate constant are then estimated simultaneously for each of the joint realizations through inverse modeling. The mass transport simulator developed in Chapter 4 that numerically solves the Equations [4.33] to [4.38] as well as a modified Gauss-Newton optimization approach have been used to solve the inverse problem (Appendix B). The joint realizations with their optimized dissolution rates and first-order decay rates can then be used to make predictions with uncertainty. In Chapter 6, it is shown that tailoring the estimation of dissolution and first-order biodegradation rate constants to distributions of uncertainty in source geometry and transmissivity field characterizes the uncertainty in these parameters and effectively reduces the uncertainties in the prediction of state variables. The details for the implementation of inverse modeling for joint estimation of dissolution rate and first-order biodegradation rate are given in the next paragraphs.

The objective function that quantifies the mismatch between the observed and calculated concentrations is represented by:

$$F_C = \sum_{i=1}^{n_c} w_i \left(C_i^m - C_i\right)^2 \quad \text{with} \quad w_i = \frac{1}{cv_i^2 C_i^2} \qquad [5.15]$$

where, F_C is the sum of squared differences between the measured concentrations $\left\{C_i^m, i = 1,..., n_h\right\}$ and the calculated concentrations, w_i represent the weights that show the relative importance of different observations, and cv_i represent the coefficient of variation associated with the observations. Following the work of Anderman and Hill (1999) and as the low values of the concentrations at the locations far from the source of contaminants are important in improving the well posedness of the problem, the weights in Equation [5.15] are calculated as the square of the inverse of the product of coefficient of variation and simulated concentrations. Using this product instead of the standard deviation of the error in observations has two advantages: (1) the concentration values usually vary over a few orders of magnitude and the suggested weighting ensures that the small

concentrations (at the edges of the plume) get large enough weights to avoid having their information content vanished throughout the optimization; and (2) the author's experience with plotting the objective function surfaces showed that using simulated (rather than measured) concentrations and coefficient of variation leads into smoother objective function surfaces. This is also explained by Anderman and Hill (1999). In practice, there is often prior information available about the values of first-order biodegradation rate and the dissolution rate. As discussed before, these prior values are usually obtained through some sort of field/laboratory experiment or as literature prescribed values. The prior information can be used for regularization of the objective function (Equation [5.15]) or in constrained optimization. The regularized form of the objective function can be written by:

$$F_C = \sum_{i=1}^{n_c} w_i \left(C_i^m - C_i\right)^2 + w_k \left(\ln k_{dis}^{PR} - \ln k_{dis}\right)^2 + w_\lambda \left(\ln \lambda^{PR} - \ln \lambda\right)^2 \qquad [5.16]$$

where, k_{dis}^{PR}, k_{dis}, λ^{PR} and λ represent the prior value for the dissolution rate, the estimated value for the dissolution rate, the prior value for first-order biodegradation rate, and the estimated value for first-order biodegradation rate, respectively. The weights w_k and w_λ are calculated as the inverse of the variance of the prior values:

$$w_k = \frac{1}{\sigma_{\ln k}^2} \quad \text{and} \quad w_\lambda = \frac{1}{\sigma_{\ln \lambda}^2} \qquad [5.17]$$

where, $\sigma_{\ln k}$ and $\sigma_{\ln \lambda}$ are the standard deviations of prior values for dissolution rate and first-order biodegradation rate, respectively.

Although the weighting scheme used in the definition of the objective function defined by Equation [5.15] effectively avoids vanishing the information content at the low concentration locations, at the same time it introduces an artifact to the value of objective function that makes it inappropriate for ranking the realizations based on the value of the objective function. In other words, the weighting

scheme in Equation [5.15] always gives lower weights to joint realizations with larger source zones that have larger average concentrations across the modeling domain. In the example presented in this work (Chapter 6), it is observed that multiplication of the objective function by the value of the quantile associated with the source zone realization (drawn from the band of uncertainty defined by the distance function approach) can provide a more appropriate measure for ranking the realizations based on their objective function value. Thus, the modified objective function is defined by:

$$F'_{C,j} = \frac{1}{n}\sum_{i=1}^{n_c} q_j w_i \left(C_i^m - C_i\right)^2 \quad \text{with} \quad w_i = \frac{1}{cv_i^2 C_i^2} \qquad [5.18]$$

where, $F'_{C,j}$ and q_j represent the objective function and the source size quantile associated with the jth joint realization. In the synthetic example presented in Chapter 6, it is observed that by ranking based on the value of the modified objective function [Equation 5.18] and screening, one may reduce the uncertainty in the size of the source zone and reduce the bias in the predicted state variables.

To estimate the values of k_{dis} and λ, non-linear regression techniques (Cooley and Naff, 1990) are employed, where a gradient-based optimization algorithm is used iteratively to obtain an improved estimate of the set of parameters, until convergence achieved. In this work, the modified Gauss-Newton approach (Cooley and Naff 1990) is employed and the criteria for convergence are considered to be (1) the value of objective function (Equation [5.15]) is smaller than a pre-specified small value, or (2) the reduction in the value of the objective function is insignificant in a few consecutive iterations or (3) maximum number of iterations is reached (Appendix B). In each iteration of the Modified Gauss-Newton approach, the vector of estimated parameters is updated by addition of an updating vector $\mathbf{d_r}$ multiplied by a damping parameter ρ_r:

$$\mathbf{b_{r+1}} = \rho_r \mathbf{d_r} + \mathbf{b_r} \qquad [5.19]$$

where, $\mathbf{b_r}$ and $\mathbf{b_{r+1}}$ are the vectors of the estimated parameters in two consecutive iterations. The damping parameter ρ_r preserves the direction of $\mathbf{d_r}$ and ensures that the changes in the parameters remain less than the maximum allowed change specified by the user and has a damping effect on likely oscillations that may occur due to opposite directions in consecutive updating vectors ($\mathbf{d_r}$ and $\mathbf{d_{r-1}}$). In the modified Gauss-Newton method, the updating vector $\mathbf{d_r}$ is calculated by:

$$\left(\mathbf{C}^T\mathbf{X}_r^T\boldsymbol{\omega}\mathbf{X}_r\mathbf{C}+\mathbf{I}m_r\right)\mathbf{C}^{-1}\mathbf{d_r}=\mathbf{C}^T\mathbf{X}_r^T\boldsymbol{\omega}\left(\mathbf{y}-\mathbf{y}\left(\mathbf{b_r}\right)\right) \qquad [5.20]$$

where, $\mathbf{I}$ is the identity matrix, $\mathbf{C}$ is the diagonal scaling matrix, $\mathbf{X}_r$ is the matrix of sensitivities, $\boldsymbol{\omega}$ is the matrix of weights, $\mathbf{y}$ is the vector of observed concentrations, $\mathbf{y}(\mathbf{b_r})$ is the vector of simulated concentrations, and m_r is the Marquardt parameter. The expanded form of the vectors in Equation [5.20] can be presented by:

$$\mathbf{y}=\begin{bmatrix} C_1^m \\ C_2^m \\ \vdots \\ C_{ND}^m \\ \ln k_{dis}^{PR} \\ \ln \lambda^{PR} \end{bmatrix}_{N\times 1} \quad , \quad \mathbf{y}=\begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_{ND} \\ \ln k_{dis} \\ \ln \lambda \end{bmatrix}_{N\times 1} \quad , \quad \mathbf{X}_r=\begin{bmatrix} S_{k_{dis},1} & S_{\lambda,1} \\ S_{k_{dis},2} & S_{\lambda,2} \\ \vdots & \vdots \\ S_{k_{dis},ND} & S_{\lambda,ND} \\ 1 & 0 \\ 0 & 1 \end{bmatrix}_{N\times 2}$$

$$\boldsymbol{\omega}=\begin{bmatrix} \left(\frac{1}{cv_1C_1}\right)^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \left(\frac{1}{cv_2C_2}\right)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 0 & \left(\frac{1}{cv_{ND}C_{ND}}\right)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma^2_{\ln k_{dis}}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\sigma^2_{\ln \lambda}} \end{bmatrix}_{N\times N}$$

$$\mathbf{C}=\begin{bmatrix} \left(\mathbf{X}_r^T\boldsymbol{\omega}\mathbf{X}_r\right)_{11} & 0 \\ 0 & \left(\mathbf{X}_r^T\boldsymbol{\omega}\mathbf{X}_r\right)_{22} \end{bmatrix}^{-1/2}_{2\times 2} \qquad [5.21]$$

where, *ND* is the number of observations, and *N* equals *ND* + 2. The damping parameter always has a value less than one and can be calculated as the proportion of maximum allowed change to maximum calculated change. The maximum calculated change is obtained as the absolute value of the largest fractional change in the parameter values. The Marquardt parameter in Equation [5.20], which is used to change the direction of $\mathbf{d_r}$, has an initial value of zero and is updated in each step according to the following rule (Hill and Tiedeman, 2007):

$$\begin{cases} m_r^{new} = 1.5 m_r^{old} + 0.001 & , \text{ if } \cos(\theta) < 0.07 \\ m_r^{new} = m_r^{old} & , \text{ otherwise} \end{cases} \qquad [5.22]$$

where, θ is the angle between the Gauss-Newton updating vector $\mathbf{d_r}$ and the steepest descent vector which is calculated as the gradient of the objective function with respect to the estimated parameters. The components of the gradient vector can be calculated by:

$$\frac{\partial F_C}{\partial \alpha} = \sum_{i=1}^{ND} \left[\frac{2C_i^m}{cv_i^2 C_i^2} \left(1 - \frac{C_i^m}{C_i} \right) S_{\alpha,i} \alpha \right] + \frac{2}{\sigma_{\ln k}^2 \alpha} \left(\ln \alpha - \ln \alpha^m \right) \quad \text{with} \quad \alpha = k_{dis}, \lambda \qquad [5.23]$$

where, $S_{\alpha,i}$ represents the sensitivity coefficient defined as the derivative of concentration at the i^{th} observation location with respect to the estimated parameters, being either k_{dis} or λ. The calculation of the sensitivity coefficients in Equations [5.20], [5.21] and [5.23] is a crucial step in finding the optimum values for the model parameters. A number of different approaches have been reported in the literature for calculation of sensitivity coefficients, most importantly perturbation method, the method of sensitivity equations and the adjoint state method (Carrera et al 1990).

In this work, the method of sensitivity equations has been applied. Similar to the SSC algorithm, the sensitivity coefficients can be calculated simultaneously with calculation of simulated concentrations through solving the transport equations (Equations [4.33] to [4.38]). However, in the method of characteristics (used in Chapter 4 to develop the mass transport simulator code) the advection term is

handled through particle tracking and direct numerical differentiation of Equation [4.33] is not available. To overcome this problem, one may discretize the advection terms on the finite–difference grid and calculate the sensitivity coefficients at the same time as solving the mass transport equations to find the concentrations. This approach is very efficient. Taking the derivative of Equation [4.33] with respect to k_{dis} and λ results in the sensitivity equations for these parameters:

$$\frac{\partial(S_\alpha)}{\partial t}+\frac{\partial}{\partial x_i}(v_i S_\alpha)-\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial S_\alpha}{\partial x_j}\right)=\frac{\partial R_{s/n}}{\partial \alpha} \qquad [5.24]$$

where, α represents either k_{dis} or λ, S_α represents the sensitivity coefficients with respect to either k_{dis} or λ and $R_{s/n}$ represents the sink/source term and the reaction term. The term on the right-hand side of Equation [5.24] can be expanded by:

$$\frac{\partial R_{s/n}}{\partial \alpha}=\begin{cases}\left(C_s^{eq}-C_s\right)-\left(k_{dis}+\lambda\right)S_\alpha & ,\text{if } \alpha=k_{dis} \text{ and } C_s^{eq}>C_s\\ -\lambda S_\alpha & ,\text{if } \alpha=k_{dis} \text{ and } C_s^{eq}\le C_s\\ -C_s-\left(k_{dis}+\lambda\right)S_\alpha & ,\text{if } \alpha=\lambda \text{ and } C_s^{eq}>C_s\\ -\lambda S_\alpha-C_s & ,\text{if } \alpha=\lambda \text{ and } C_s^{eq}\le C_s\end{cases} \qquad [5.25]$$

where, C_s is the concentration of the contaminant species *s* (e.g. benzene), and C_s^{eq} is the equilibrium concentration of the contaminant species *s*, calculated by Equation [4.34]. Similar to Equations [4.46] through [4.49], Equation [5.24] can be discretized by finite difference method with an explicit time weighting:

$$S_{i,j}^{n+1}\cong S_{i,j}^{n}-\Delta t\left[\left(\frac{(D_{xx})_{i+\frac{1}{2},j}}{\Delta x^2}+\frac{(D_{xx})_{i-\frac{1}{2},j}}{\Delta x^2}+\frac{(D_{yy})_{i,j+\frac{1}{2}}}{\Delta y^2}+\frac{(D_{yy})_{i,j-\frac{1}{2}}}{\Delta y^2}\right)\right.$$
$$\left.+\left(\frac{(v_x)_{i+\frac{1}{2},j}-(v_x)_{i-\frac{1}{2},j}}{2\Delta x}+\frac{(v_y)_{i,j+\frac{1}{2}}-(v_y)_{i,j-\frac{1}{2}}}{2\Delta y}\right)\right]S_{i,j}^{n}$$
$$+\Delta t\left[\left(\frac{(D_{xx})_{i+\frac{1}{2},j}}{\Delta x^2}+\frac{(D_{yx})_{i,j+\frac{1}{2}}-(D_{yx})_{i,j-\frac{1}{2}}}{4\Delta x\Delta y}-\frac{(v_x)_{i+\frac{1}{2},j}}{2\Delta x}\right)S_{i+1,j}^{n}\right.$$

$$
\begin{aligned}
&+\Delta t\left[\left(\frac{\left(D_{xx}\right)_{i-\frac{1}{2},j}}{\Delta x^2}+\frac{\left(D_{yx}\right)_{i,j-\frac{1}{2}}-\left(D_{yx}\right)_{i,j+\frac{1}{2}}}{4\Delta x\Delta y}+\frac{\left(v_x\right)_{i-\frac{1}{2},j}}{2\Delta x}\right)S^n_{i-1,j}\right.\\
&+\Delta t\left[\left(\frac{\left(D_{yy}\right)_{i,j+\frac{1}{2}}}{\Delta y^2}+\frac{\left(D_{xy}\right)_{i+\frac{1}{2},j}-\left(D_{xy}\right)_{i-\frac{1}{2},j}}{4\Delta x\Delta y}-\frac{\left(v_y\right)_{i,j+\frac{1}{2}}}{2\Delta y}\right)S^n_{i,j+1}\right.\\
&+\Delta t\left[\left(\frac{\left(D_{yy}\right)_{i,j-\frac{1}{2}}}{\Delta y^2}+\frac{\left(D_{xy}\right)_{i-\frac{1}{2},j}-\left(D_{xy}\right)_{i+\frac{1}{2},j}}{4\Delta x\Delta y}+\frac{\left(v_y\right)_{i,j-\frac{1}{2}}}{2\Delta y}\right)S^n_{i,j-1}\right.\\
&+\Delta t\left[\frac{\left(D_{xy}\right)_{i+\frac{1}{2},j}+\left(D_{yx}\right)_{i,j+\frac{1}{2}}}{4\Delta x\Delta y}\right]S^n_{i+1,j+1}-\Delta t\left[\frac{\left(D_{xy}\right)_{i+\frac{1}{2},j}+\left(D_{yx}\right)_{i,j-\frac{1}{2}}}{4\Delta x\Delta y}\right]S^n_{i+1,j-1}\\
&-\Delta t\left[\frac{\left(D_{xy}\right)_{i-\frac{1}{2},j}+\left(D_{yx}\right)_{i,j+\frac{1}{2}}}{4\Delta x\Delta y}\right]S^n_{i-1,j+1}+\Delta t\left[\frac{\left(D_{xy}\right)_{i-\frac{1}{2},j}+\left(D_{yx}\right)_{i,j-\frac{1}{2}}}{4\Delta x\Delta y}\right]S^n_{i-1,j-1}\\
&+\Delta t\begin{cases}\left(C^{eq}_s-C^n_{i,j}\right)-\left(k_{dis}+\lambda\right)S^n_{i,j} & ,\text{if } \alpha=k_{dis} \text{ and } C^{eq}_s>C_s\\ -\lambda S^n_{i,j} & ,\text{if } \alpha=k_{dis} \text{ and } C^{eq}_s\le C_s\\ -C^n_{i,j}-\left(k_{dis}+\lambda\right)S^n_{i,j} & ,\text{if } \alpha=\lambda \text{ and } C^{eq}_s>C_s\\ -\lambda S^n_{i,j}-C^n_{i,j} & ,\text{if } \alpha=\lambda \text{ and } C^{eq}_s\le C_s\end{cases}
\end{aligned}
\qquad [5.26]
$$

where, n and n+1 represent the current and next time steps, the terms D_{xx}, D_{yy} , D_{xy} and D_{yx} are the components of the dispersion tensor and they may be approximated by Equations [4.47] and [4.48], and the terms v_x and v_y are the velocity components at the cell interfaces which can be calculated by Equations [4.27] and [4.28].

As discussed in Chapter 4, simulation of the advective term, together with all other terms in an Eulerian framework may result in advective fronts (plume edges) that look too smooth and unrealistic. To investigate the impact of this smoothing effect on calculation of sensitivity coefficients, one may compare the calculated sensitivities by sensitivity equations to the sensitivities calculated by perturbation approach. For this purpose, a synthetic study site has been created with known flow and transport properties and boundary conditions. Figure 5-6 shows the synthetic study site that has a heterogeneous hydraulic conductivity

with a log-normal Gaussian distribution with a mean of -10.5 ($\log_e$ m/s), standard deviation of 0.8 ($\log_e$ m/s), and a spherical variogram with a nugget equal to 0.05 and a range equal to 30 m. No-flow boundary conditions are assigned at the west and the east of the site and constant head boundary conditions equal to 3.6 m and 2.0 m are assigned at the north and the south of the site, respectively. In terms of mass transport properties and boundary conditions, a rectangular NAPL source zone was considered at the north of the site with a uniform substrate (e.g. benzene) soil concentration equal to 100 mg/Kg, assuming that the initial mass fraction of the substrate in the NAPL is equal to 0.01. The substrate solubility, substrate and inert molecular weights are set equal to 0.00178 g/cm^3, 78.1 and 101 g/mole, respectively. Dry soil density, total porosity and effective porosity are set equal to 1.6 g/cm^3, 0.35 and 0.3, respectively. The dissolution rate and first-order biodegradation rates are set equal to 0.0011day^{-1}, and 0.0044day^{-1}, respectively. Zero dispersive flux boundary conditions are assigned at all boundaries. The longitudinal and transverse dispersivities are set equal to 1.0m and 0.2m, respectively. Figure 5-7 shows a few snap-shots of the development of the plume, which has been simulated for 10 years.

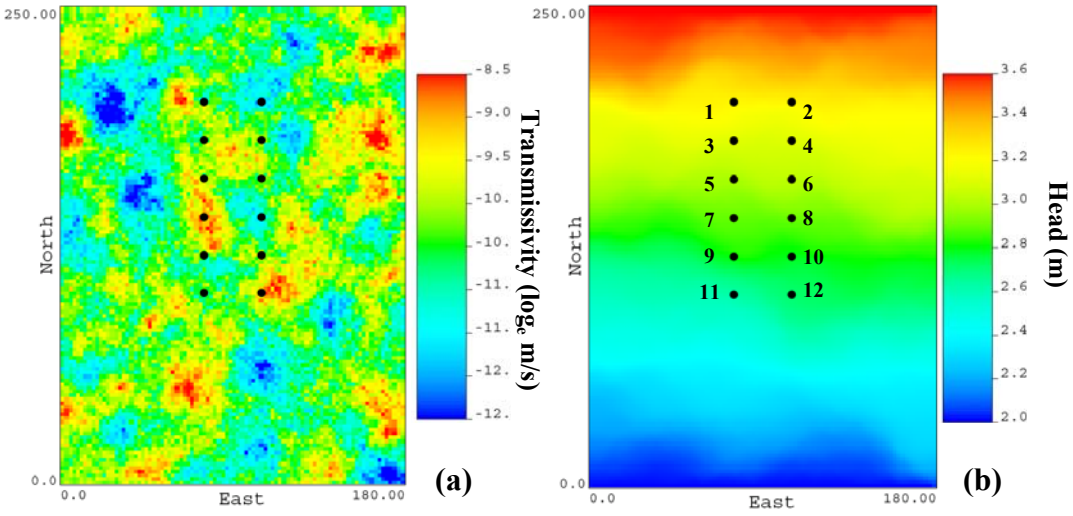


Figure 5-6: (a) Reference hydraulic conductivity field, and (b) the associated piezometric head response. The monitoring locations numbered 1 to 12 are used to record the simulated concentrations and plot the objective function surface in Section 5.2.3.

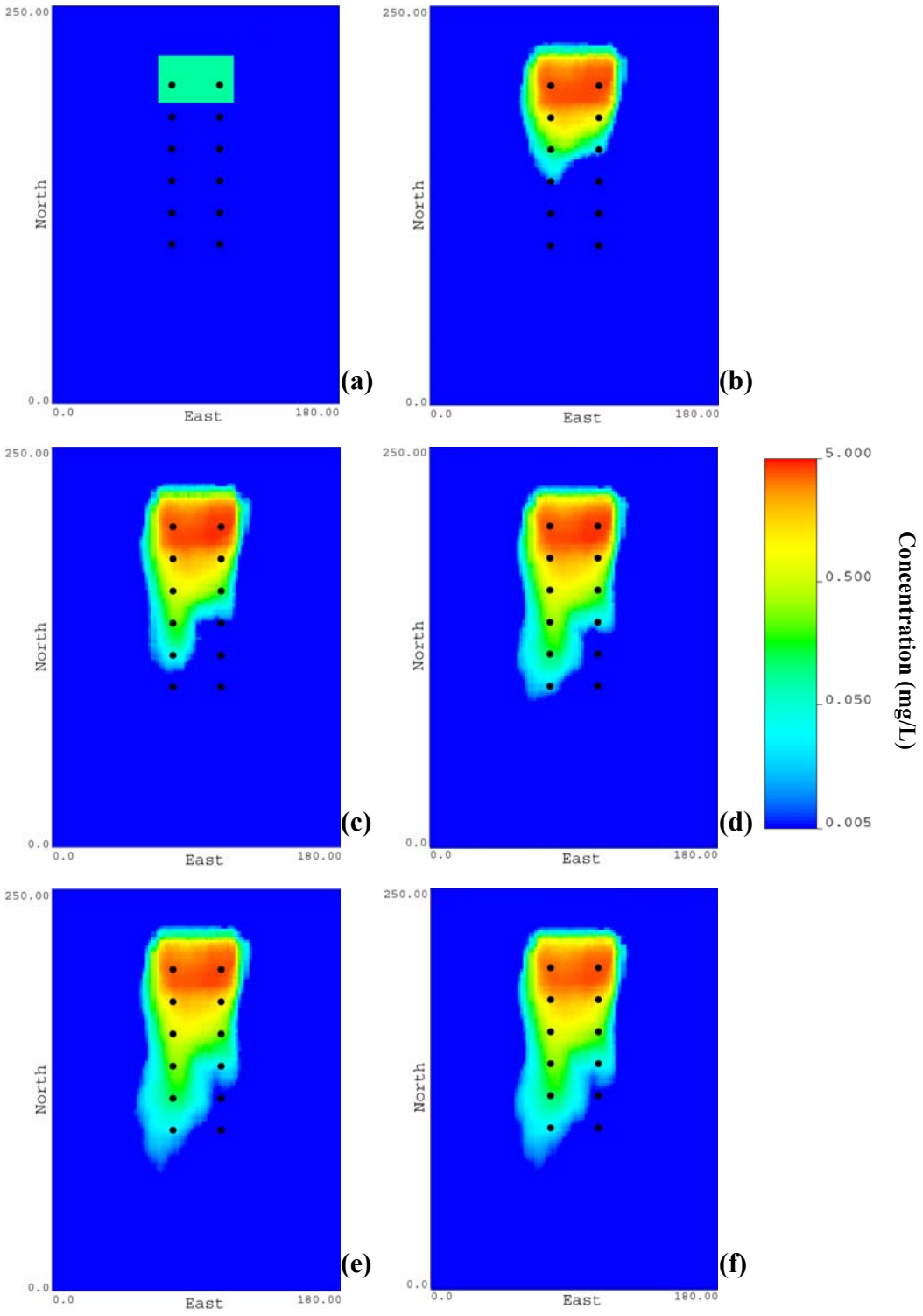


Figure 5-7: Six snap shots from the development of the dissolved contaminant plume after (a) 1 day, (b) 1 year, (c) 2 years, (d) 3 years (e) 7.5 years and (f) 10 years.

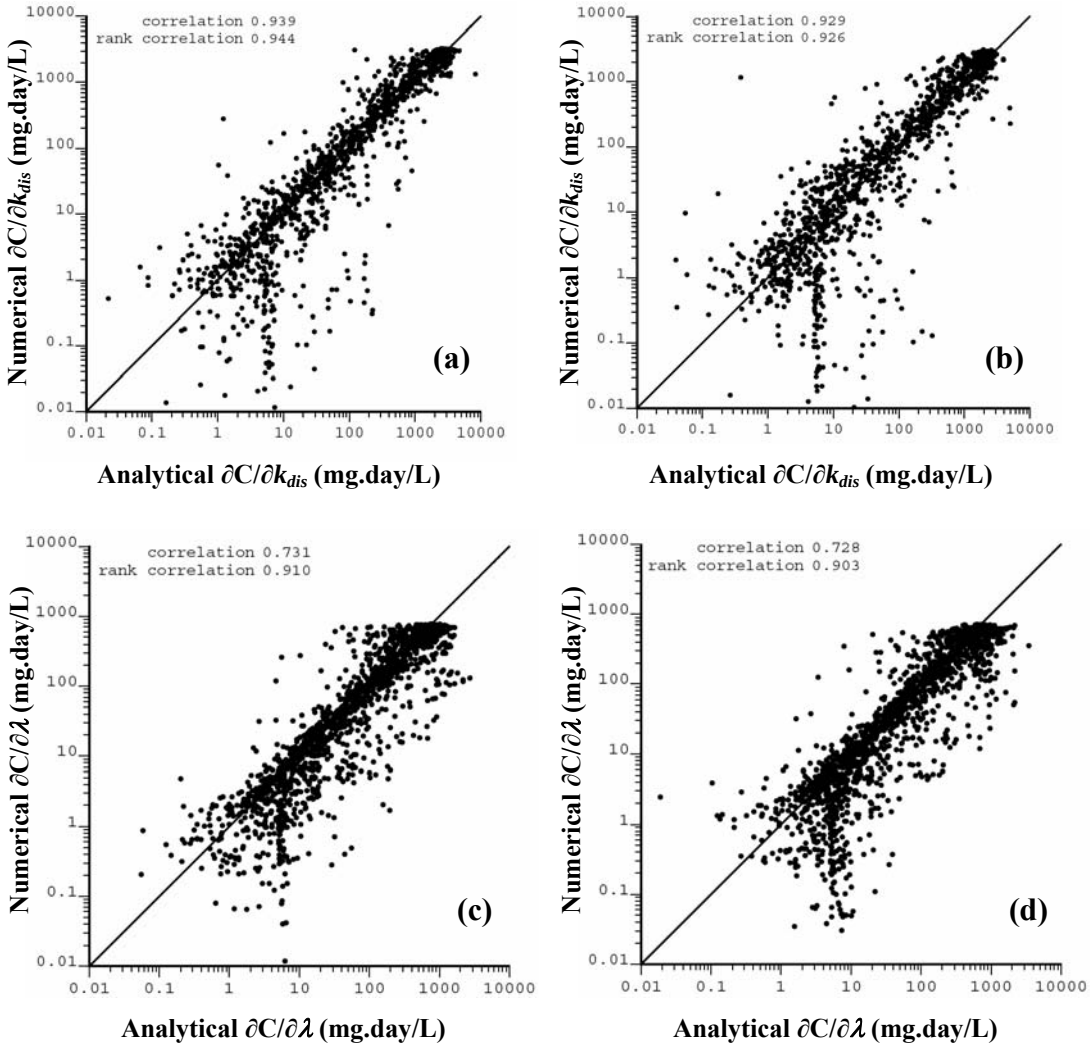


Figure 5-8: Comparison between the numerically (using sensitivity equations) and analytically (using perturbation approach) calculated sensitivity coefficients with respect to (a) k_{dis} after 5 years, (b) k_{dis} after 10 years, (c) λ after 5 years, and (d) λ after 10 years.

Figure 5-8 shows cross-plots of the sensitivity coefficients calculated over the modeling domain using sensitivity equations versus the coefficients calculated by perturbation approach with respect to k_{dis} and λ at two different time steps.

One can observe that there exists a good overall agreement between the calculated sensitivities, which preserves the relative sensitivity of the observations with

respect to the estimated parameters (high rank correlation between the attributes in all cases). There is a better agreement between the calculated sensitivities at the locations with higher concentrations (the areas within the plume). At the edges of the plume, however, the values of the calculated sensitivities may differ up to one order of magnitude. This can be due to and affected by particle insertion/deletion mechanism that is used in the method of characteristics and the size of perturbations used in the study.

5.2.2. Sensitivity analysis

In this work, a decoupled inverse problem has been parameterized in terms of distribution of hydraulic conductivity, geometry of the source zone, dissolution rate constant and biodegradation rate constant. In order to attain a better understanding of the relationships between the model parameters and the response variables, a simple sensitivity analysis has been performed. For this purpose, a 160 m × 300 m synthetic study site has been created (Figure 5-9) with a Gaussian hydraulic conductivity distribution with a mean (in natural logarithmic unit) equal to -11.12 $\log_e$-m/s and two levels of heterogeneity with standard deviations of 1.0 $\log_e$- m/s and 2.1 $\log_e$- m/s. A spherical variogram with a nugget effect equal to 0.1 and a range of 25.0 m has been used to generate the reference hydraulic conductivity fields. The flow boundary conditions include fixed head boundaries at the north and the south of the site equal to 3.5 m and 2.0 m, respectively; and no-flow boundary conditions at the west and the east of the site. The uncertainty in the areal extent of the NAPL source zone was characterized by the distance-function approach (Chapter 3). The uncertainty in dissolution and first order biodegradation rates are modeled by log-normal distributions, as shown in Figure 5-9. The mean and log-normal standard deviation of k_{dis} and λ are assumed to be 0.035 day^{-1}, 1.76 $\log_e day^{-1}$, 0.0031 day^{-1} and 0.37 $\log_e day^{-1}$, respectively. These values are selected to represent the variability in these parameters based on the values commonly found in the literature for benzene (Aronson and Howard 1997,

Waddill and Widowson 1998 and Christ et al. 2006). Zero-dispersive flux is assigned to all boundaries as the transport boundary conditions. All other transport parameters are set equal to those associated with Figure 5-6.

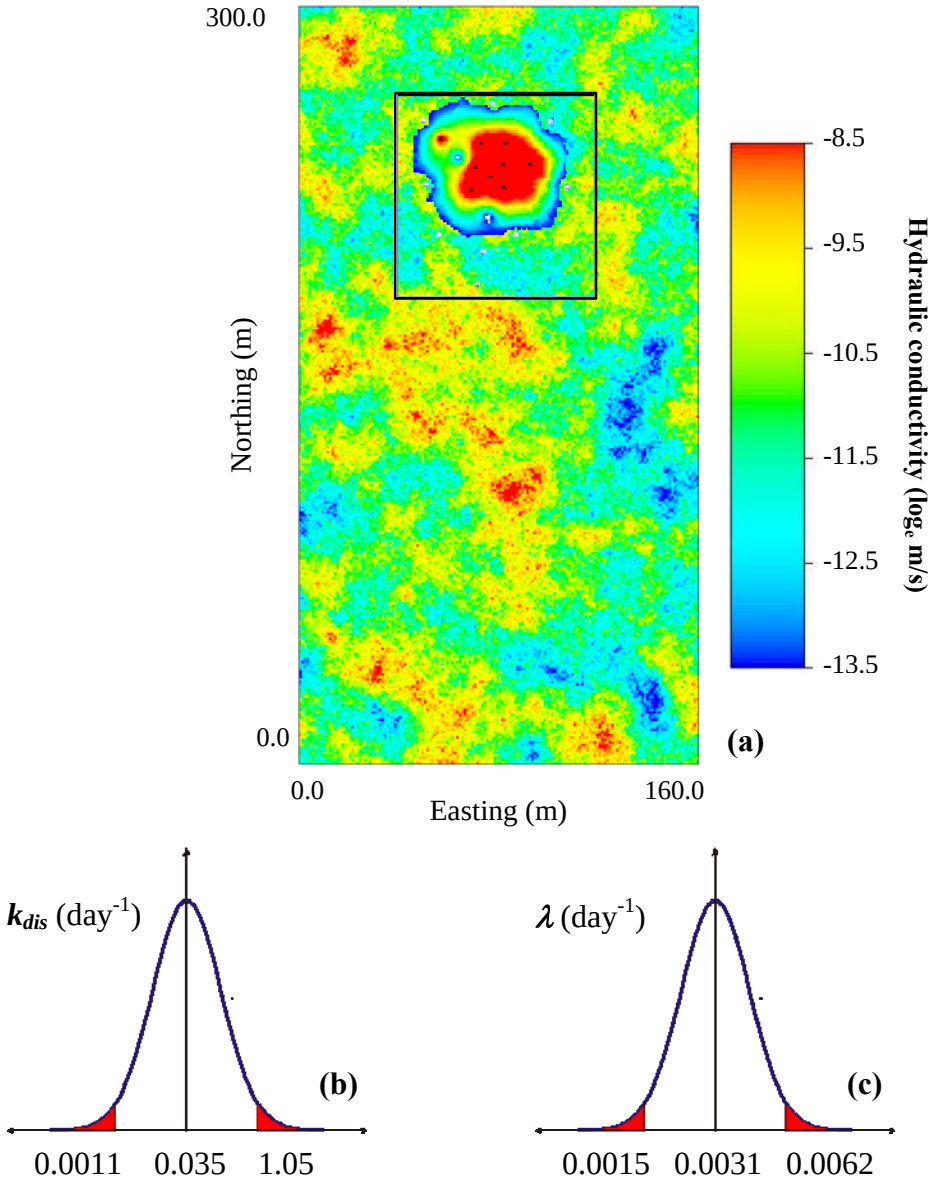


Figure 5-9: (a) The reference hydraulic conductivity field and the uncertainty band for the areal extent of the source zone, (b) log-normal distribution for k_{dis} with given p_{05}, p_{50} and p_{95} quantiles and (b) log-normal distribution for λ with given p_{05}, p_{50} and p_{95} quantiles.

The simple sensitivity analysis involves running the mass transport simulator for 30 years on the hydraulic conductivity fields with the two levels of heterogeneity (σ_T = 1.0 $\log_e$-m/s and σ_T = 2.1 $\log_e$-m/s) for each of p_{05}, p_{50}, and p_{95} quantiles for source geometry, dissolution rate and first-order biodegradation rate, while all other parameters are set to their base case values (p_{50} value). Figures 5-10 and 5-11 show the variations of two important TOR state variables, namely the total mass loaded into a mildly heterogeneous aquifer (σ_T = 1.0 $\log_e$-m/s) and the plume length over a period of 30 years due to changes in the values of dissolution rate constant, biodegradation rate constant and the areal extent of the source zone. Figures 5-12 and 5-13 show the similar plots for a highly heterogeneous aquifer (σ_T = 2.1 $\log_e$-m/s). Looking at the Figures 5-10 through 5-13, one may observe that: (1) there is highly non-linear relationship between the dissolution rate and the state variables, that is, the state variables have larger sensitivities with respect to changes in dissolution rate when the value of the dissolution rate constant is relatively small (in the orders Christ et al. 2006 suggested for field applications); (2) the length of the plume is much more affected by changes in the biodegradation rate constant, compared to the impact of changes in the dissolution rate constant and source geometry (for the given levels of heterogeneity) have smaller effect on the length of the plume; however, this depends on the value of dissolution rate too; (3) the field-scale heterogeneity substantially affects the length of the plume and although it has smaller impact on the amount of mass loaded into the aquifer, it affects the sensitivity of the mass loaded into the aquifer with respect to the changes in the dissolution rate; (4) the sensitivity of the amount of mass loaded into the aquifer with respect to changes in the dissolution rate constant becomes smaller with time. It should be noted that the two standard deviation values for the hydraulic conductivity field used in the analysis correspond to the observed standard deviations reported by Binsariti (1980) for an alluvial basin aquifer (σ_T = 1.0 $\log_e$-m/s) and reported by Rehfeld et al. (1989) for a fluvial sand and gravel aquifer (σ_T = 2.1 $\log_e$-m/s).

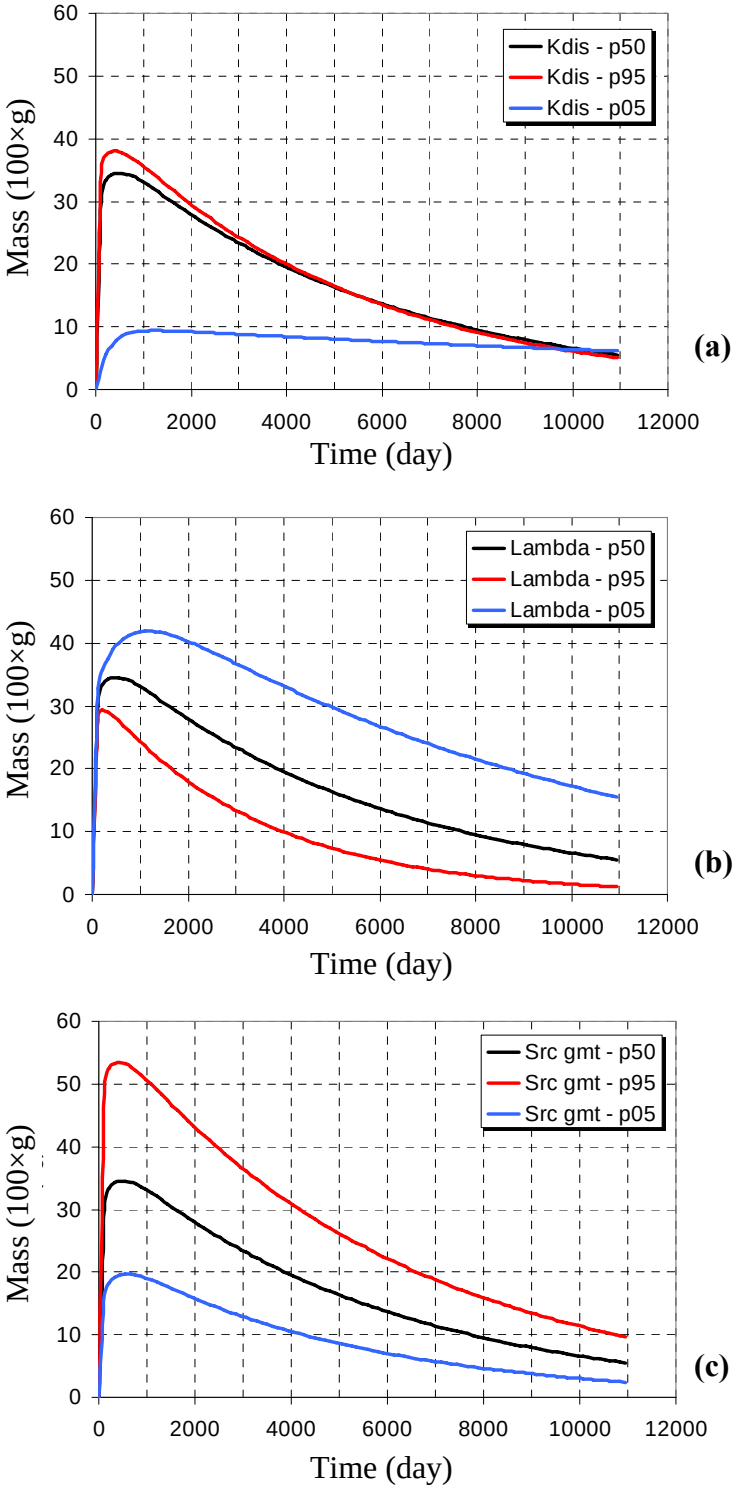


Figure 5-10: Changes in the variations of total mass loaded into a mildly heterogeneous aquifer over time due to changes in (a) dissolution rate, (b) first-order biodegradation rate and (c) the areal extent of the source zone.

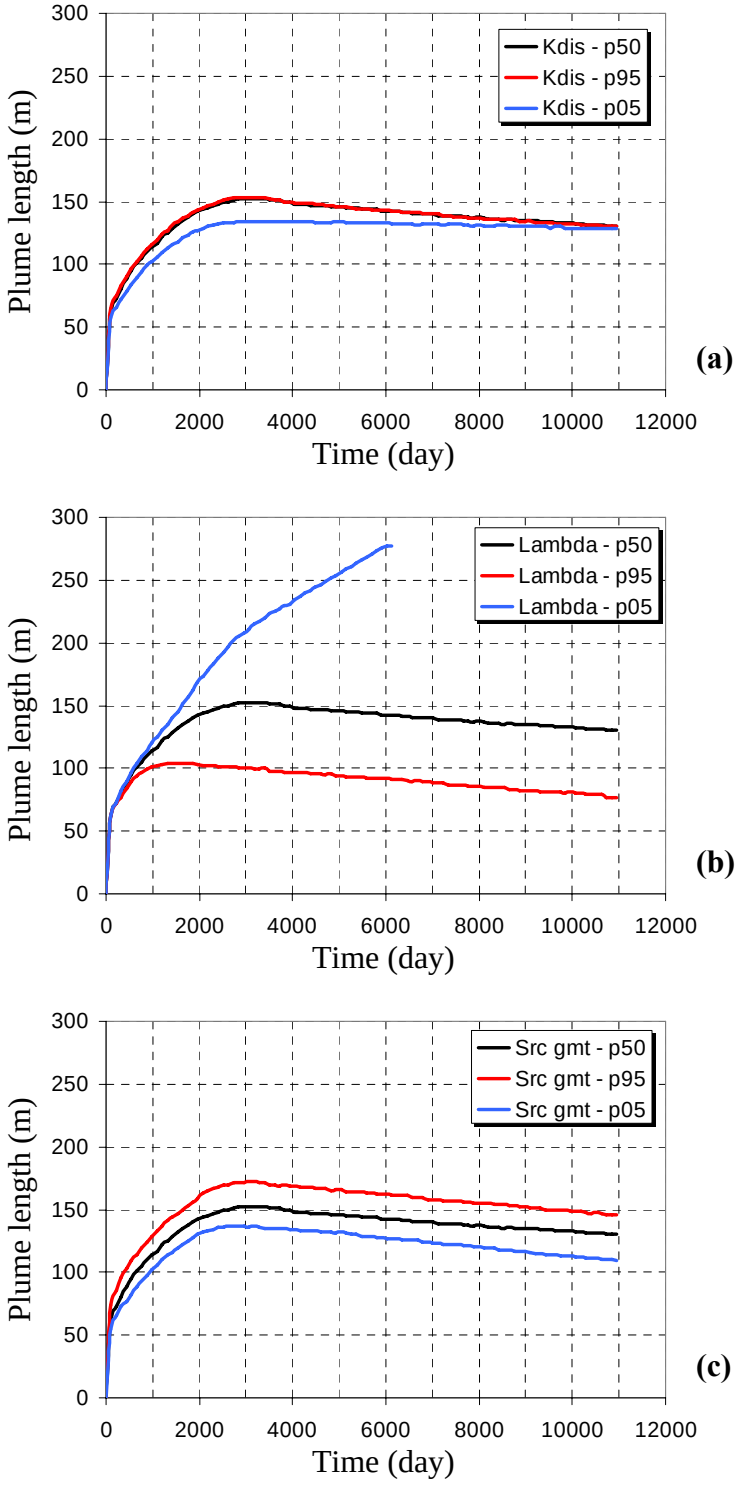


Figure 5-11: The changes in the variations of the plume length over time in a mildly heterogeneous aquifer due to changes in the values of (a) dissolution rate constant, (b) first-order decay rate and (c) the areal extent of the source zone.

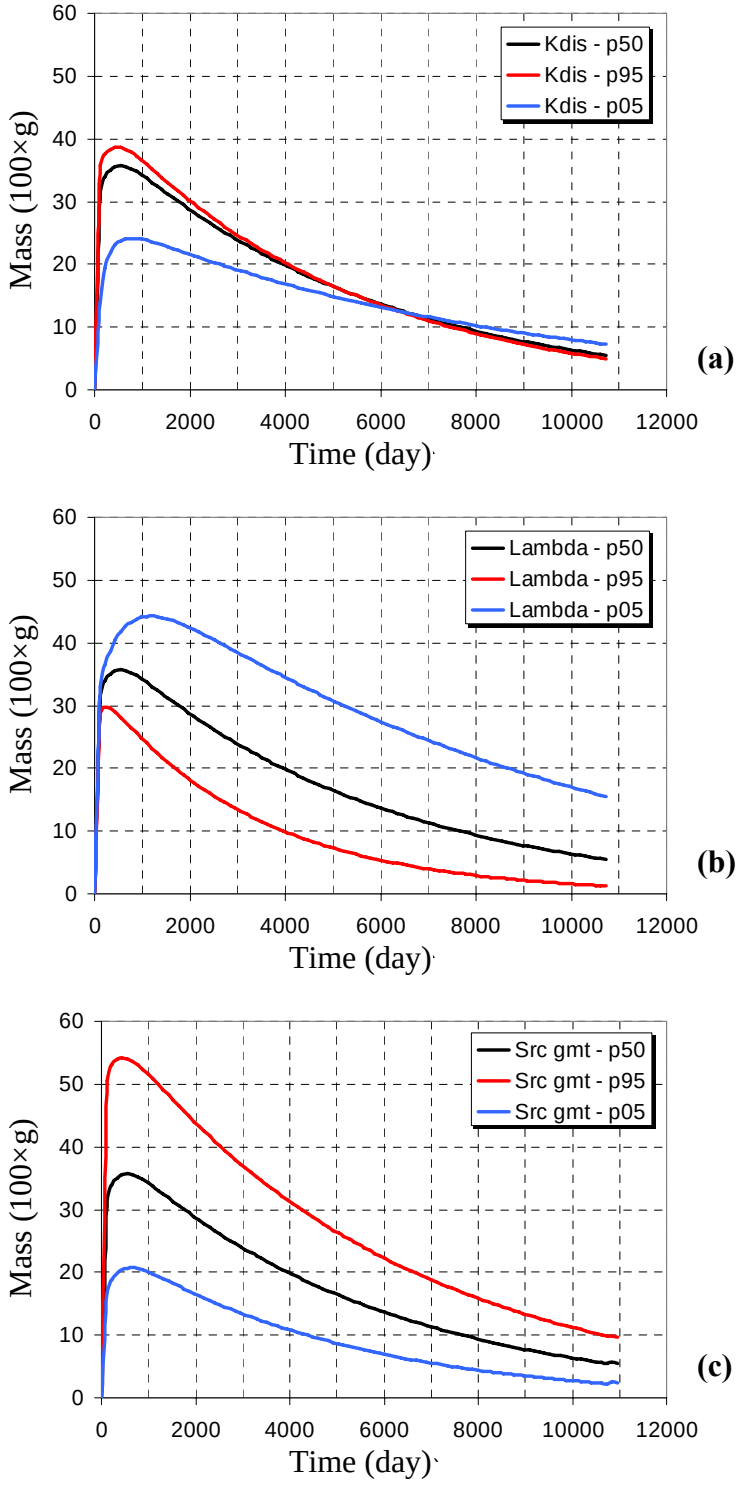


Figure 5-12: Changes in the variations of total mass loaded into a highly heterogeneous aquifer over time due to changes in (a) dissolution rate, (b) first-order biodegradation rate and (c) the areal extent of the source zone.

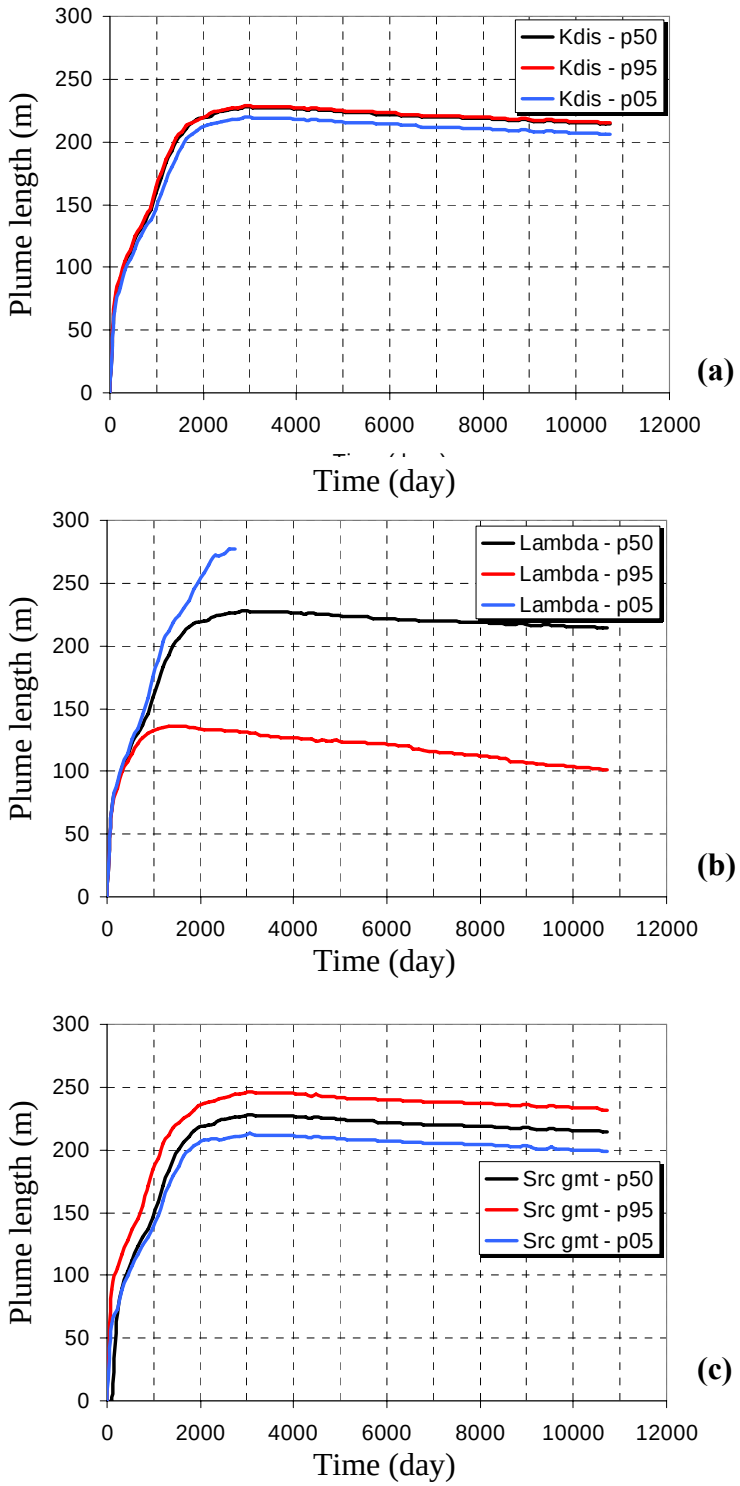


Figure 5-13: The changes in the variations of the plume length over time in a highly heterogeneous aquifer due to changes in the values of (a) dissolution rate constant, (b) first-order decay rate and (c) the areal extent of the source zone.

5.3 Stability of the inverse problem

The problem of simultaneous estimation of source dissolution rate and first-order biodegradation rate is a potentially unstable (prone to parameter non-uniqueness and local minima) inverse problem (Essaid et al. 2003). As pointed out originally by Carrera and Neuman (1986) and reviewed by Carrera et al. (2005) and Friedel (2005), the four potential reasons for parameter non-uniqueness are lack of numerical precision, numerical dispersion, global and local minima in the parameter space and correlation among model parameters. In the context of this thesis, the first two causes of parameter non-uniqueness are avoided by development and application of a Lagrangian mass transport simulator with appropriate precision (Chapter 4). The existence of local minima and high correlation among model parameters are potential problems associated with simultaneous estimation of dissolution rate and first-order biodegradation rate that are investigated in the following 1D and 2D examples.

5.3.1. Correlation between the parameters: 1D case

To investigate the correlations between the parameters, first, a 1D example is presented where the sensitivity of concentrations with respect to the changes in the dissolution rate and first-order biodegradation rate ($S_{k_{dis}}$ and S_{λ}) are calculated and their relationship is studied. For a one-dimensional case, Equation [4.33] can be re-written by:

$$\frac{\partial(C_x)}{\partial t} = \frac{\partial}{\partial x}\left(D_{xx}\frac{\partial C_x}{\partial x}\right) - \frac{\partial}{\partial x}(v_x C_x) + \max\left[0, k_{dis}\left(C^{eq} - C_x\right)\right] - \lambda C_x \qquad [5.27]$$

where, only the velocity and dispersion coefficient in the *x* direction have been retained and the subscript *s* that represents solute species has been replaced by *x* to show that concentration is a function of space. Also, in Equation [5.27], the dispersion coefficient D_{xx} can be replaced by $\alpha_L|v_x|$, where α_L is longitudinal

dispersivity and v_x is the local velocity. Differentiation of Equation [5.27] with respect to k_{dis} and λ results in:

$$\frac{\partial S_\lambda}{\partial t} = \alpha_L v_x \frac{\partial^2 S_\lambda}{\partial x^2} + \left(\alpha_L \frac{\partial v_x}{\partial x} - v_x\right)\frac{\partial S_\lambda}{\partial x} - \left(k_{dis} + \lambda + \frac{\partial v_x}{\partial x}\right)S_\lambda - C_x \quad [5.28]$$

$$\frac{\partial S_{k_{dis}}}{\partial t} = \alpha_L v_x \frac{\partial^2 S_{k_{dis}}}{\partial x^2} + \left(\alpha_L \frac{\partial v_x}{\partial x} - v_x\right)\frac{\partial S_{k_{dis}}}{\partial x} - \left(k_{dis} + \lambda + \frac{\partial v_x}{\partial x}\right)S_{k_{dis}} + \max\left[0, k_{dis}\left(C^{eq} - C_x\right)\right] \quad [5.29]$$

where, all terms have been defined previously. Also, one can differentiate the Equation [5.27] with respect to the local velocity v_x. After rearranging, one can find the partial differential equation for the sensitivity of the local concentration with respect to the local velocity field (S_{v_x}):

$$\frac{\partial S_{v_x}}{\partial t} = \alpha_L v_x \frac{\partial^2 S_{v_x}}{\partial x^2} + \left(\alpha_L \frac{\partial v_x}{\partial x} - v_x\right)\frac{\partial S_{v_x}}{\partial x} - \left(k_{dis} + \lambda + \frac{\partial v_x}{\partial x}\right)S_{v_x} + \alpha_L \frac{\partial^2 C_x}{\partial x^2} - \frac{\partial C_x}{\partial x} \quad [5.30]$$

where, all terms have been defined previously.

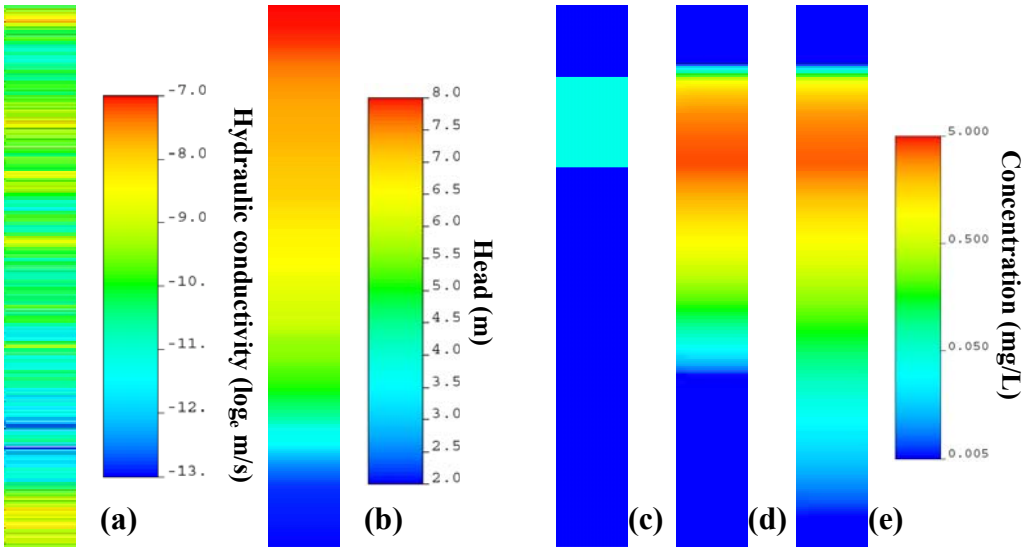


Figure 5-14: (a) The reference (a) transmissivity, (b) steady-state head and concentration (c) after 2 days, (d) after 732 days, and (e) after 2562 days.

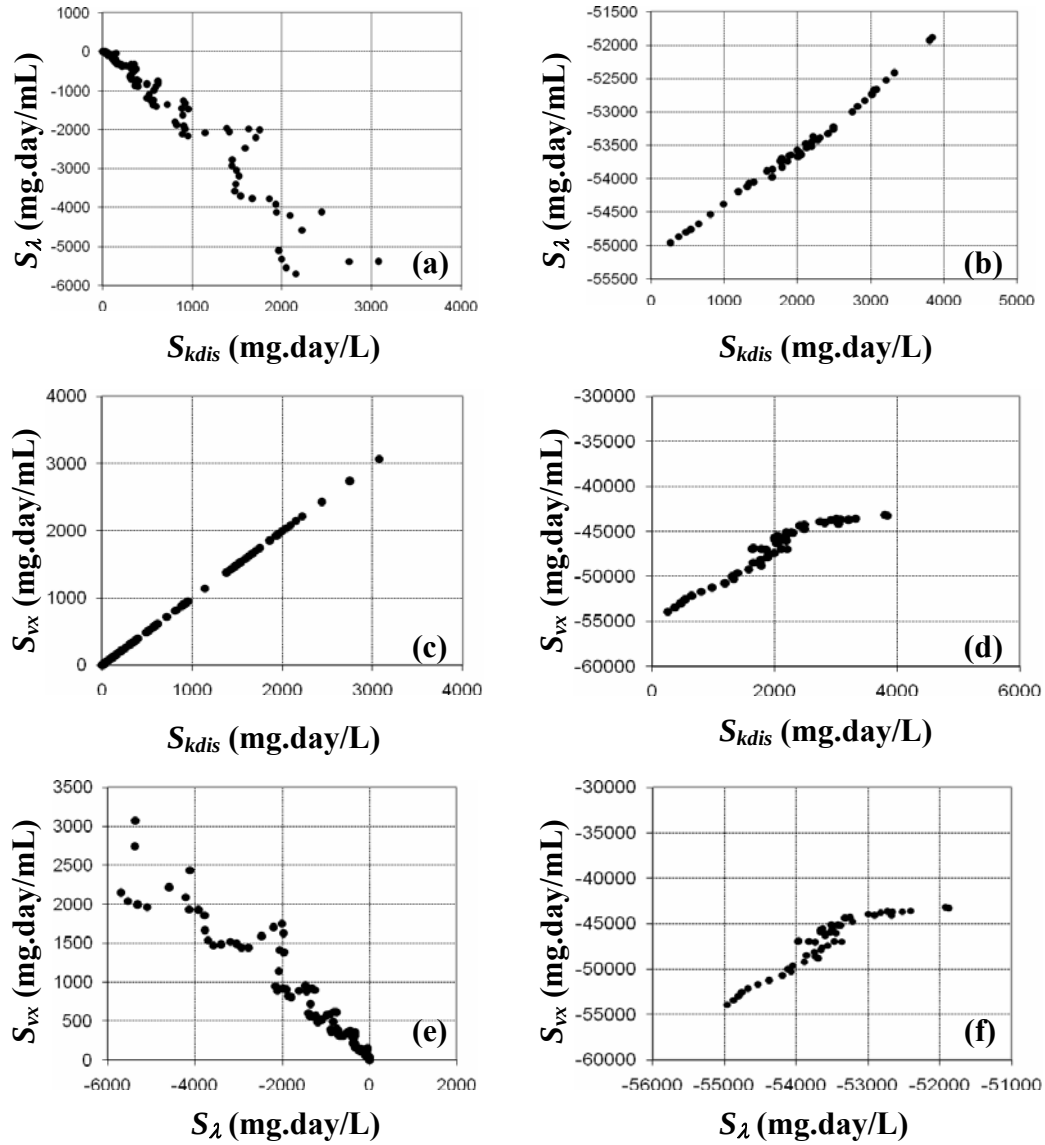


Figure 5-15: The cross-plots between (a) S_λ and $S_{k_{dis}}$ away from the source, (b) S_λ and $S_{k_{dis}}$ in the source zone, (c) S_{v_x} and $S_{k_{dis}}$ away from the source, (d) S_{v_x} and $S_{k_{dis}}$ in the source zone, (e) S_{v_x} and S_λ away from the source, and (f) S_{v_x} and S_λ in the source zone. The cross-plots show the correlation between the sensitivities at time step 2562 days for k_{dis} and λ values equal to 0.0015 day^{-1} and 0.0044 day^{-1}, respectively.

Equations [5.28], [5.29] and [5.30] can be discretized by finite difference method and solved numerically. Figure 5-14 shows the reference 1D hydraulic conductivity and head distribution and the development of the 1D plume until

reaching to a semi-steady-state condition. The length of the 1D simulation domain is 300m, and the reference k_{dis} and λ are 0.0015 day^{-1} and 0.0044 day^{-1}. The flow and transport boundary conditions include fixed head boundaries of 8m and 2m, and zero dispersive flux boundaries at both ends. All other parameters are the same as the ones used in the synthetic example in Figure 5.7.

Figures 5-15-a, b show (1) there exists a high correlation (for the given values of the parameters) between the sensitivities of the concentrations with respect to perturbations of dissolution rate and first-order biodegradation rate inside and outside of the source zone; and (2) the correlation between these parameters differs inside and outside of the source both in terms of magnitude of the sensitivities and in terms of the trends in the correlations (positive or negative). As shown in the subsequent 2D example, the observed changes in the correlations outside and inside of the source can be used to avoid instabilities in the inverse problem. In Figures 5-15-c, d, e, f, the cross-plots between the sensitivities of concentrations with respect to perturbations of local velocity field and the two transport parameters are presented. According to these figures, a very high correlation between these parameters is observed. Nevertheless, making any conclusive argument about requirement of decoupling the flow and transport inverse problems from this observation relies on the assumption of independency of local velocities at different locations, which is an invalid assumption. Thus, although one can not make a conclusive argument about decoupling of the inverse problems (subject to further research), the observations in Figures 5-15-c, d, e, and f show that when the observation locations are sparse and the range of spatial correlation of hydraulic conductivity is small, there may potentially be a high correlation between these parameters that may have an adverse effect on the stability of the coupled problem. This is subject for future research.

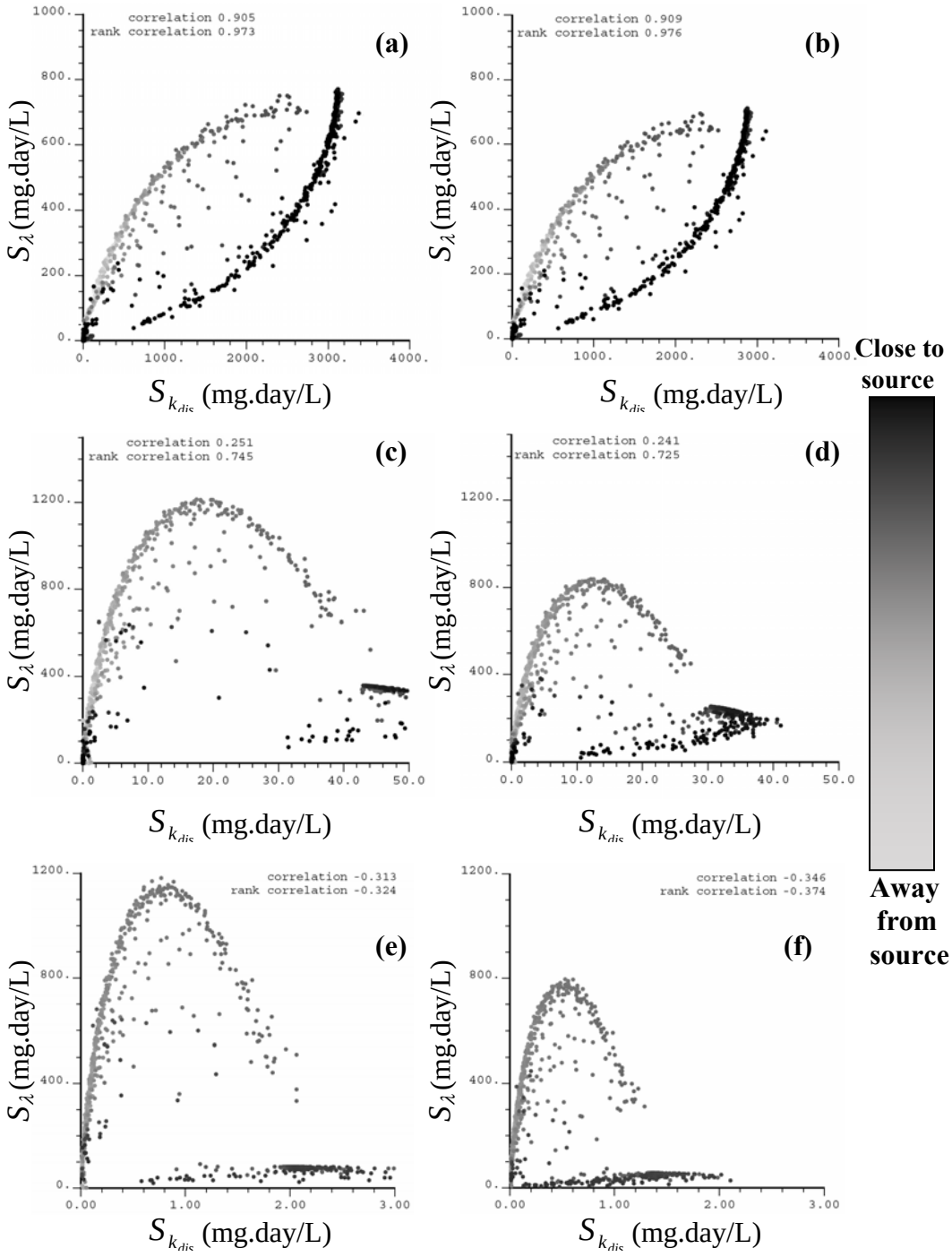


Figure 5-16: The cross-plots between the values of $S_{k_{dis}}$ and the absolute values of S_λ for (a) k_{dis}=0.0011day^{-1} and T=5yrs, (b) k_{dis}=0.0011day^{-1} and T=10yrs, (c) k_{dis}=0.011day^{-1} and T=5yrs, (d) k_{dis}= 0.011day^{-1} and T=10yrs, (e) k_{dis} = 0.175day^{-1} and T=5yrs, (f) k_{dis} = 0.175day^{-1} and T=10yrs. The gray-scale color bar shows the location of the calculated sensitivity coefficients.

5.3.2. Correlation between the parameters and existence of local minima: 2D case

The correlations between the parameters and existence of local minima can also be investigated using the example presented in Figures 5-7 and 5-8. Figure 5-16 shows the cross-plots between the values of the $S_{k_{dis}}$ (always positive) and the absolute values of S_{λ} (always negative) in the 2D example (Figure 5-6) for different values of dissolution rate. In the cross-plots shown in Figure 5-16, the location of the calculated sensitivity coefficients is highlighted using a gray-scale color bar, where the black end represents the points closer to the source zone, and light-grey end represents the points more distant from the source zone. Similar to what is observed in Section 5.2.2, Figure 5-16 shows the value of $S_{k_{dis}}$ strongly depends on the value of the dissolution rate, and as the value of dissolution rate becomes smaller, larger correlation is observed between the sensitivity coefficients inside and outside of the source. Both $S_{k_{dis}}$ and S_{λ} sensitivities are also variable in time. The most important observations and conclusions related to Figures 5-16 are: (1) for large values of the dissolution rate, the correlation between the two parameters becomes smaller and parameter non-uniqueness becomes less of an issue. For this case, however, parameter insensitivity may become the problem; (2) for smaller values of dissolution rate (which are more feasible for real field applications according to Essaid et al. (2003) and Christ et al. 2006), a very high correlation exists between the two sensitivities. It is also observed that there are two different correlations between the two parameters depending on the location of the calculated sensitivities (similar to the 1D case). The fact that there exist two different correlations can be used together with definition of weights inverse proportional to the simulated concentration to improve the stability of the inverse problem (Equation [5.23]).

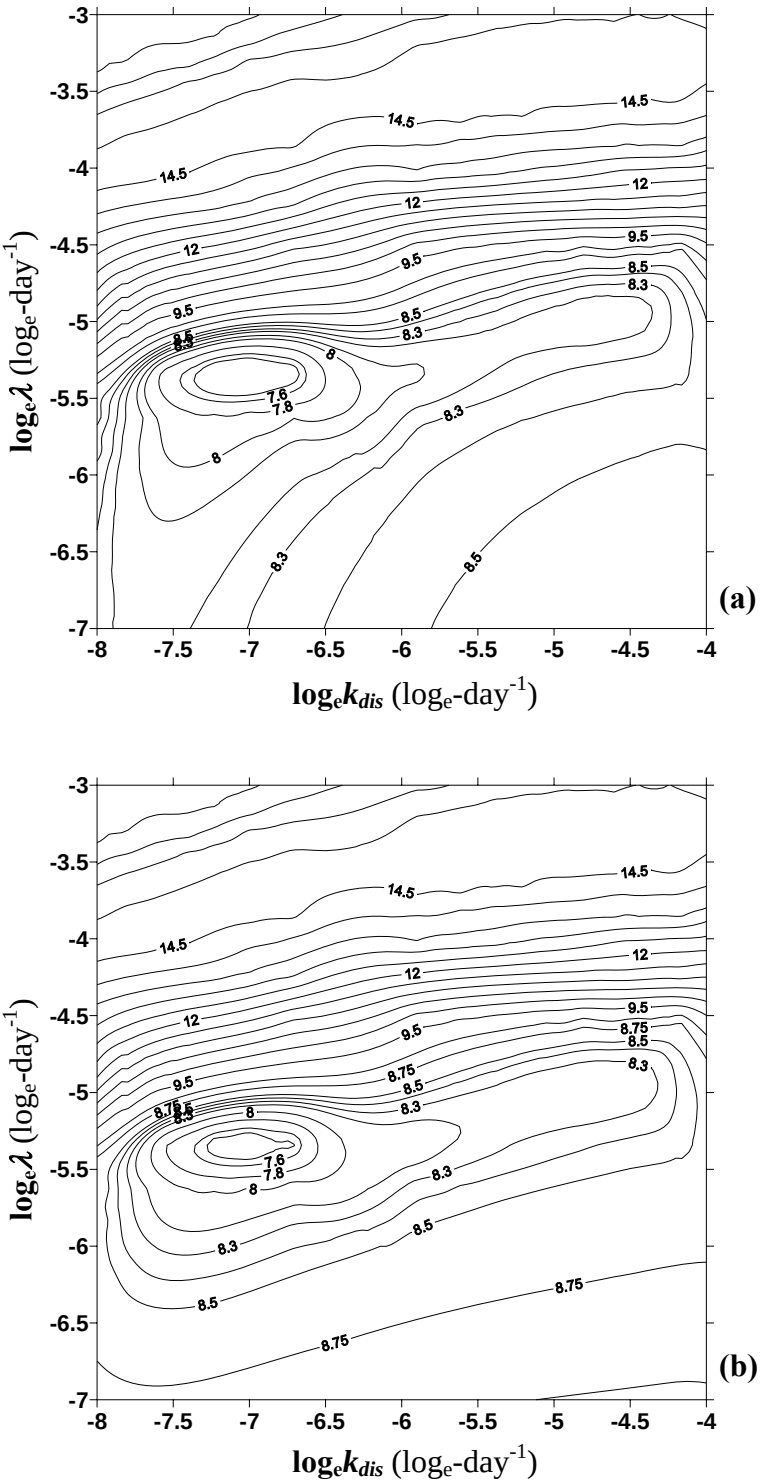


Figure 5-17: The objective function surfaces plotted based on observations in wells 1 to12, (a) without any prior information and (b) with prior information.

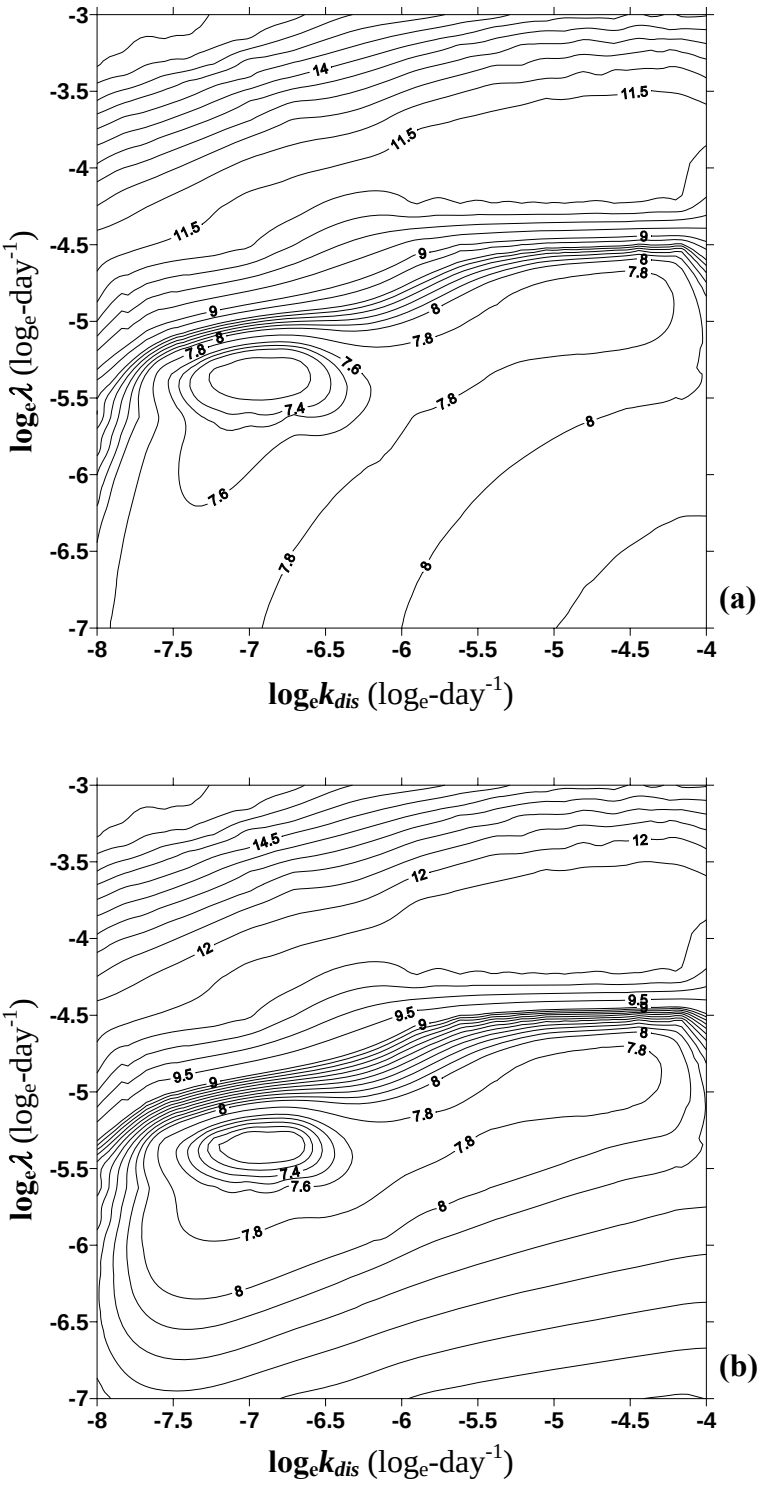


Figure 5-18: The objective function surfaces plotted based on observations in wells 1 to 4, and 9 to 12 (a) without any prior information and (b) with prior information.

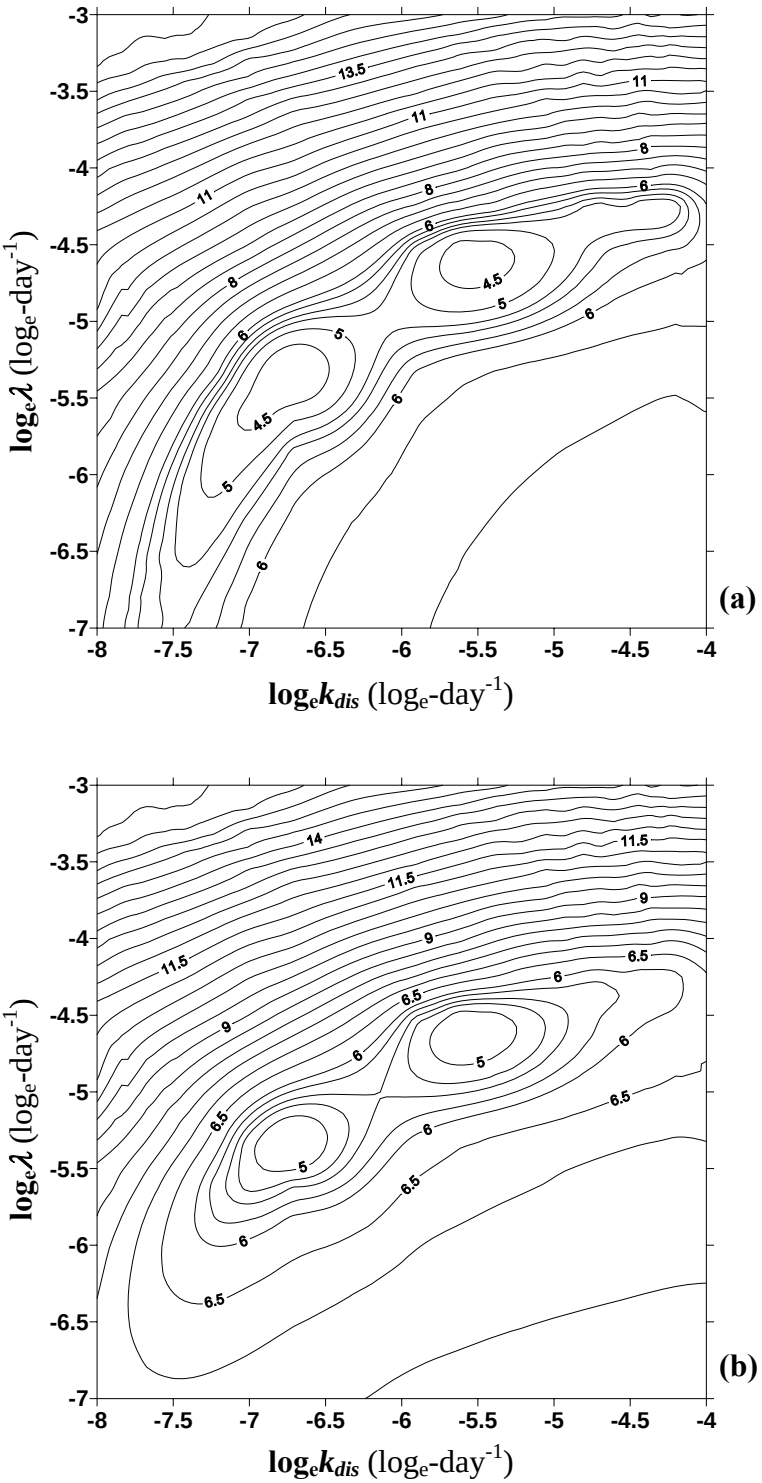


Figure 5-19: The objective function surfaces plotted based on observations in wells 1 to 4 (a) without any prior information and (b) with prior information.

This can be achieved through designing the monitoring network in such a way that observation locations sample the locations close to the source zone as well as the locations well-away from the source, while the weighting scheme in the objective function has to be defined in such a way (Equation [5.15]) that the information contained in concentrations distant from the source zone should be preserved adequately. The fact that designing an appropriate layout for the monitoring network improves the stability of the inverse problem for small values of dissolution rate can be investigated through plotting the objective function surfaces (Figures 5-17, 5-18 and 5-19). The synthetic aquifer shown in Figure 5-6 with 12 monitoring locations is used for this purpose. The monitoring data sampled from the synthetic plume are recorded when the plume is under a steady-state condition at the time steps of 8 years, 8.5 years, 9 years, 9.5 years and 10 years. As shown in Figures 5-18 to 5-20, the objective function surfaces have been plotted with and without prior information, using the equations [5.16] and [5.15], respectively. The true parameter values used to create the reference case include k_{dis} = 0.0011 day^{-1} and λ = 0.0044 day^{-1}. The prior information are included in calculation of the objective function using Equations [5.16] and [5.17]; and they are considered to have log-normal distributions with a mean (k_{dis}^{PR}) of 0.001 day^{-1} and standard deviation ($\sigma_{\ln k}$) of 1.0 $\log_e$day^{-1} for the dissolution rate and a mean (λ^{PR}) of 0.0045 day^{-1} and standard deviation ($\sigma_{\ln \lambda}$) of 0.3 $\log_e$day^{-1} for the first-order biodegradation rate. In Figure 5-17, monitoring data includes samples from all 12 wells in Figure 5-6. In Figure 5-18 monitoring data includes samples from wells 1 to 4 at the upstream of the site and wells 9 to 12 at the downstream of the site; and in Figure 5-17 monitoring data includes samples from wells 1 to 4 at the upstream of the site (in the source and close to the source zone). It is evident from the Figures 5-17 to 5-19 that if the layout for the monitoring network is designed in such a way that the upstream (near the source zone) as well as downstream (near the edge of the plume) are sampled, the stability of the problem will significantly be improved. Comparing Figures 5-18

and 5-19, one may also observe that inclusion of the middle wells in calculation of the objective function has little effect on the improvement of the stability of the problem. It can also be observed that inclusion of prior information has a secondary effect on the stability of the inverse problem. In other words, although a good set of priori information are used to create the objective function surface in Figure 5-19-b, it has not avoided the problem of non-uniqueness that manifests itself as local minima in this case. Comparing Figure 5-17-b to Figure 5-17-a and Figure 5-18-b to Figure 5-18-a, one can observe that inclusion of good prior information only slightly improves the stability of the inverse problem.

In summary, in this chapter, details of a Monte Carlo type decoupled inverse modeling was presented that can be used to characterize the uncertainty in the field-scale dissolution rate and first-order biodegradation rate. The inverse modeling presented uses sequential self calibration technique (Gomez-Hernandez et al. 1997) to condition the hydraulic conductivity to head observation data. Assuming full statistical independence, joint realizations of hydraulic conductivity and source geometry were constructed and a gradient-based optimization approach was implemented to simultaneously estimate the dissolution rate and first-order biodegradation rate. Simultaneous estimation of these parameters is subject to numerical instabilities, associated with parameter non-uniqueness. The potential instability problems were investigated and it was observed that creating a wide-spread monitoring network (with appropriate weighting scheme in the objective function based on Equation 5-23) can effectively help to mitigate the non-uniqueness problems associated with high correlation between the parameters. In Chapter 6, a synthetic example is presented to show the effects of tailoring the estimation of first-order biodegradation rate and dissolution rate in reducing the uncertainty in the prediction of the state variables.

CHAPTER 6

PARAMETER ESTIMATION EXAMPLE

A synthetic example is presented in this chapter to investigate the performance of the Monte Carlo type decoupled inverse modeling in characterization of uncertainty in the dissolution rate and first-order biodegradation rate and to study the effects of error in observed data on the modeling outcomes. In Section 6.1, three reference study sites with three different source zone sizes are created and the groundwater flow and mass transport are simulated. A synthetic hydraulic conductivity dataset, two different head observation datasets with two different levels of measurement error, and four concentration datasets are sampled from the reference study sites. Applying the SSC approach, the sampled hydraulic conductivity and head data are used to create multiple realizations of hydraulic conductivity field conditioned to both hydraulic conductivity and head measurements. The distance function approach, developed in Chapter 3, is used to create multiple realizations of areal extent of the source zone. In Section 6.2, inverse modeling is implemented to estimate the values of dissolution rate and first-order biodegradation rate constants for the sets of joint realizations of source geometry and hydraulic conductivity fields. The performance of the methodology is investigated through studying the distributions of the estimated parameters and source zone sizes and comparing the variations of the state variables (e.g. plume length and mass loaded into the aquifer) through time with those of the reference study sites. The simulated state variables are also compared to the results of a set of Monte Carlo simulations performed using k_{dis} and λ distributions that represent the range of variability that may be observed under realistic field conditions. The effects of head and concentration measurement errors on the estimation of k_{dis} and λ and the predictions of the state variables are also investigated.

6.1 The reference case and Monte Carlo simulations

Figure 6-1 shows the reference study site with the sampling locations, the suspected source zone area, the reference hydraulic conductivity field and the associated head response. The reference hydraulic conductivity field shown in Figure 6-1-b has a Gaussian distribution in natural logarithmic units with a mean of -10.1 $\log_e$m/s, standard deviation of 1.2$\log_e$m/s, and a spatial correlation defined by a spherical variogram with a nugget effect equal to 0.1 and a range of 32.0 m. The modeling domain is 250 m by 160 m, which is descritized by 2.0 m × 2.0 m squared shape grid cells. The flow boundary conditions involve fixed head boundary conditions at the north and the south of the site equal to 4.0 m and 2.0 m, respectively. At the east and west of the site, no-flow boundary conditions are assigned. As shown in Figure 6-1-a, there are a total of 40 observation wells where piezometric heads (steady-state) and concentrations are sampled. There are 11 of these wells (shown by blue circles), with hydraulic conductivity measurements. The solid black wells indicate the observation wells where residual NAPL has been observed. Figure 6-2-a shows the calibrated uncertainty band for the given well arrangement and suspected source area. The associated optimal values of scaling factor α and separation factor β are 3.56 and 15.86, respectively. To investigate the performance of the methodology when the actual source size deviates from the average source size that is characterized by the distance function approach, three source sizes corresponding to lower quartile, median and upper quartile of the calibrated uncertainty band are considered as reference cases for source geometry. Figure 6-2-b shows the CDF of the source sizes associated with the calibrated uncertainty band in Figure 6-2-a and the selected quartiles. According to Figure 6-2-b, the reference source sizes (p_{25}, p_{50} and p_{75} quartiles) have areas equal to 643 m^2, 938 m^2, and 1395 m^2. Figures 6-3 to 6-5 show the simulated plumes for the three reference source sizes.

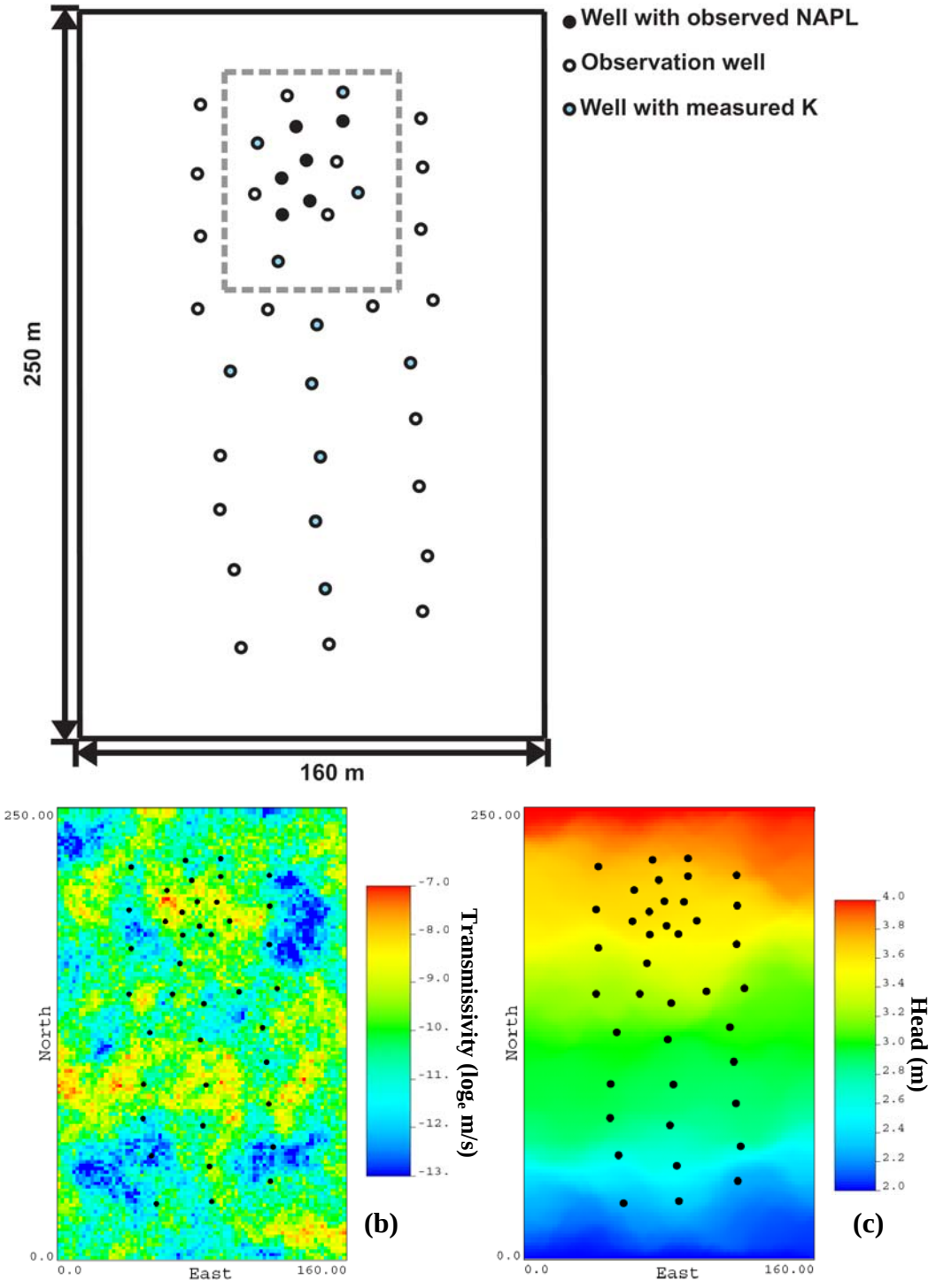


Figure 6-1: (a) The reference study site with monitoring locations and suspected source zone area (dashed box), (b) the reference hydraulic conductivity field, and (d) the reference hydraulic head distribution.

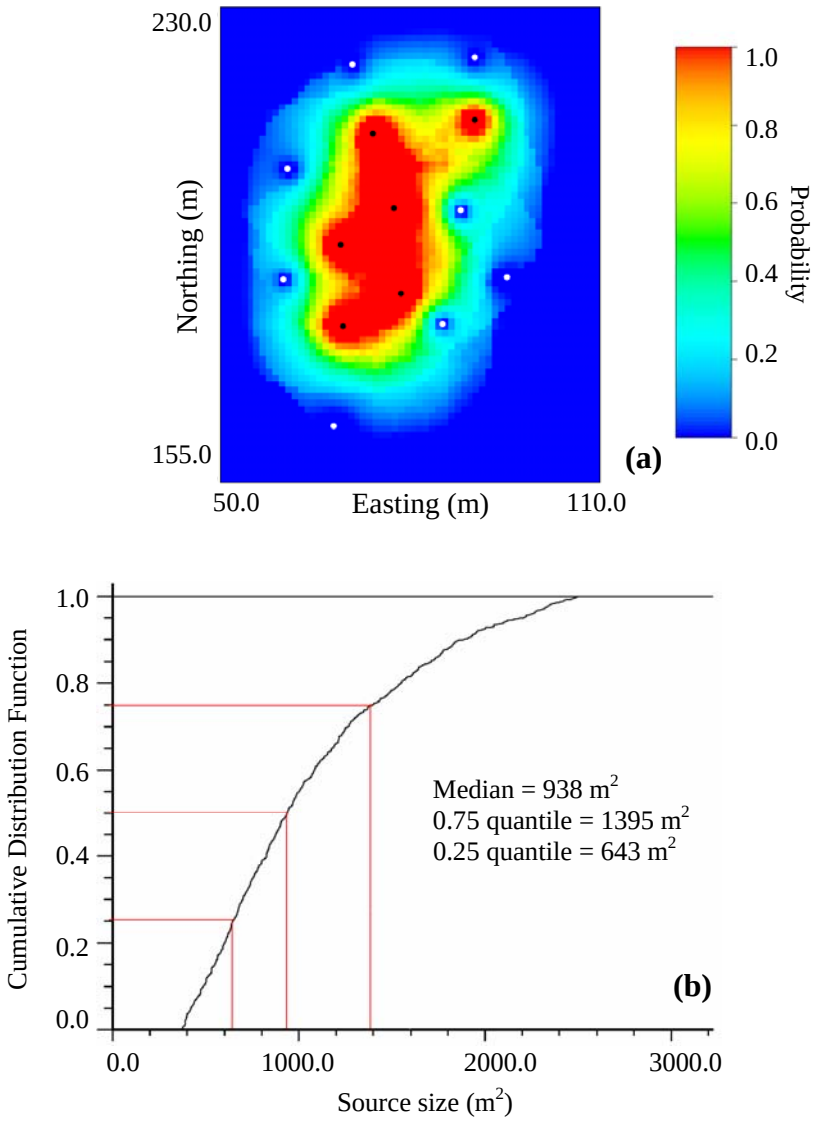


Figure 6-2: (a) The calibrated band of uncertainty for the contaminant source zone, and (b) the CDF of the source size.

For simplicity, it has been assumed that the distribution of residual NAPL (soil concentration) within the areal limits of the source zone is uniform. Following the work presented in Chapter 3, variability within areal limits can easily be incorporated. The uniform soil concentration of NAPL is set to 10g/Kg. The initial mass fraction of the substrate (e.g. Benzene) in NAPL is equal to 0.01.

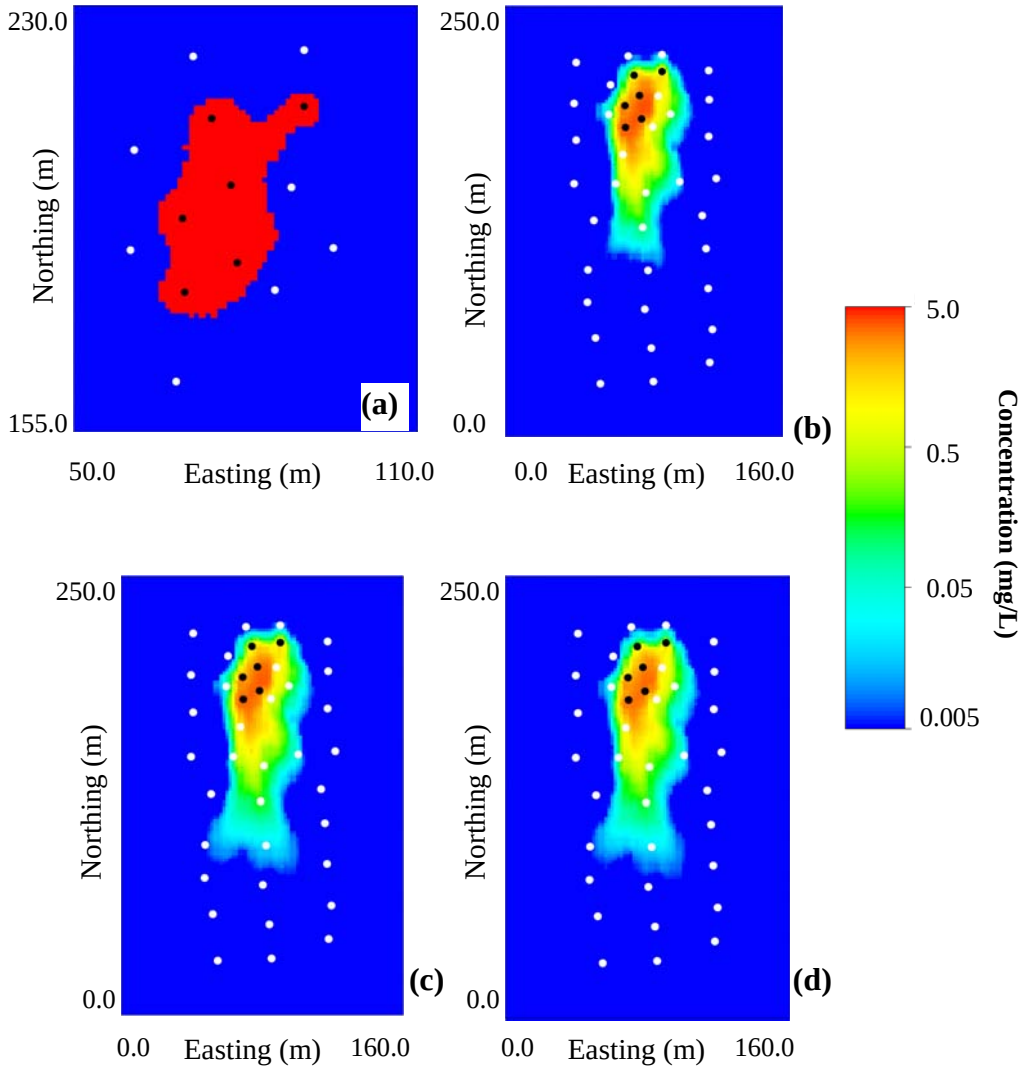


Figure 6-3: (a) The smaller source zone size corresponding to p_{25} of the calibrated uncertainty band, (b) the simulated plume after 550 days, (c) the simulated plume after 1281 days, and (d) simulated plume after 2562 days.

Similar to the example in Figure 5-7, the substrate solubility, substrate and inert molecular weights are set equal to 0.00178 g/cm^3, 78.1 and 101 g/mole, respectively. Dry soil density, total porosity and effective porosity are set equal to 1.6 g/cm^3, 0.35 and 0.3, respectively. The longitudinal and transverse dispersivities, dissolution rate and first-order biodegradation rates are equal to 1.5

m, 0.3 m, 0.0015 day^{-1} and 0.006 day^{-1}, respectively. Zero dispersive flux boundary conditions are assigned at all boundaries. Two synthetic observed datasets for piezometric heads are created by sampling from the reference piezometric head distribution and subsequent addition of Gaussian noise. The first set of head observations is considered to be error-free (with standard deviation of head measurement error σ_{nH} equal to 0.0). The second head observation dataset is considered to be noisy by addition of Gaussian noise with a mean of zero and standard deviations of σ_{nH} =0.20 m.

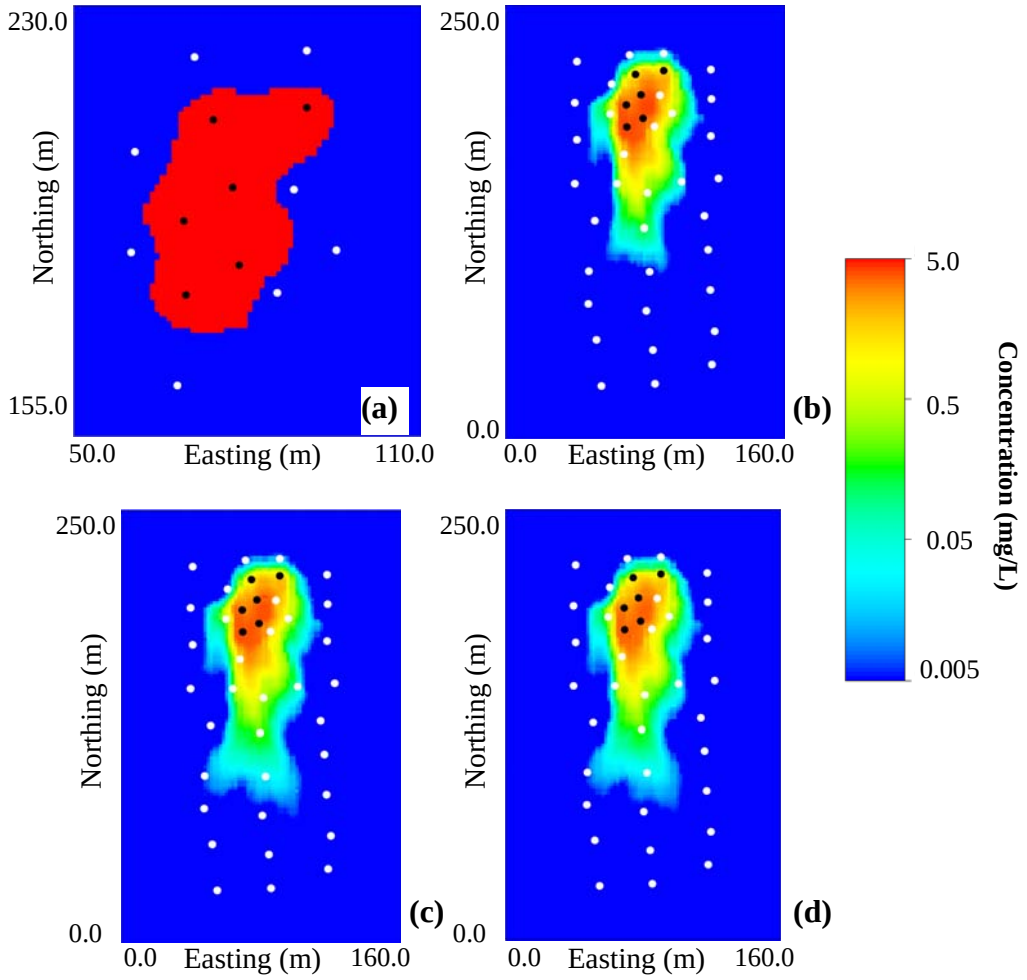


Figure 6-4: (a) The medium source zone size corresponding to p_{50} of the calibrated uncertainty band, (b) the simulated plume after 550 days, (c) the simulated plume after 1281 days, and (d) simulated plume after 2562 days.

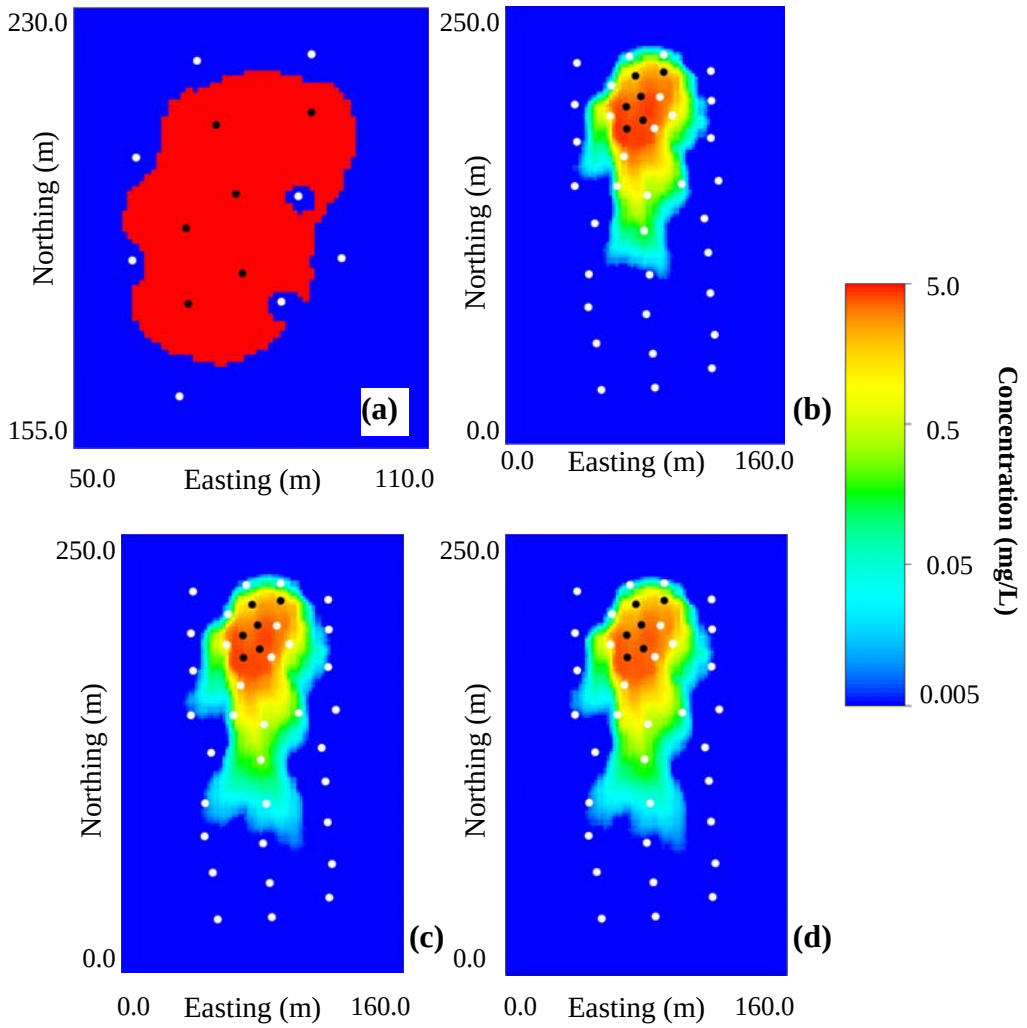


Figure 6-5: (a) The larger source zone size corresponding to p_{75} of the calibrated uncertainty band, (b) the simulated plume after 550 days, (c) the simulated plume after 1281 days, and (d) simulated plume after 2562 days.

Applying the SSC technique, two sets of 300 realizations of hydraulic conductivity field conditioned to both hydraulic conductivity and head measurements are constructed form two levels of head measurement errors (σ_{nH} =0.0 m and σ_{nH} =0.20 m), and combined with realizations of source geometry to create two sets of 300 joint realizations that are used in subsequent estimation of k_{dis} and λ.

In terms of the synthetic concentration datasets, three error-free concentration datasets are sampled from the simulated plumes (a total of 520 samples at 40 observation wells over a period of two years – from 1825 to 2555 days from the start of the simulations) for the reference cases shown in Figures 6-3 to 6-5. To investigate the effects of errors in measured concentrations on the modeling outcomes, Gaussian noise is added to the synthetic concentration dataset sampled from the first reference case with smaller source size. The added Gaussian noise has a coefficient of variation equal to cv_{nc} =0.35.

To study the importance of tailoring the estimation of first-order biodegradation rate constant and dissolution rate to realizations of source geometry and hydraulic conductivity, the results of the decoupled inverse modeling including the simulated state variables should be compared to the available field-scale parameter estimation techniques. Due to the fact that the proposed methodology is aimed to be an advanced screening tool for characterization of uncertainty in the field-scale parameters, its outcomes should be compared to the outcomes of similar screening tools commonly applied to the field. For this purpose, a set of Monte Carlo simulations (MCS) are performed with (1) realizations of hydraulic conductivity conditioned to conductivity and head data by SSC, (2) realizations of source extend characterized by the DF algorithm in Figure 6-2-a, (3) the values of first-order biodegradation rate drawn from a distribution reported by Bauer et al. (2006), and (4) the values drawn from a distribution of dissolution rate constant representing the uncertainty in a realistic field condition.

Bauer et al. (2006) showed that the field-scale method of normalization to a recalcitrant co-contaminant (Wiedemeier et al. 1996) that corrects for the effects of uncertainty in the value of longitudinal dispersivity (reviewed in Chapter 7) gives the closest estimate to the true value of the first-order biodegradation rate constant. For an aquifer with a log-normal hydraulic conductivity distribution

with a mean of -9.54 $\log_e$m/s and a standard deviation of 1.3 $\log_e$m/s, Bauer et al. (2006) showed that the method of normalization to a recalcitrant co-contaminant overestimates the true first-order rate (on average) by a factor of two, while the standard deviation of the normalized rates is equal to 2. Similar results were found by Bauer et al. (2007) for the improved method of Stenback et al. (2003) with off-centerline measurements. Figure 6-6-b shows a distribution of first-order biodegradation rate similar to the distribution observed by Bauer et al. (2006) based on normalization to a recalcitrant co-contaminant.

As mentioned before, the field-scale distribution of dissolution rate constant is quite uncertain with no documented range in the literature that can be directly applied as prior distribution. Based on the work of Powers et al (1994) and a pore network model, the work of Dillard et al. (2001) gives a series of correlations for estimation of dissolution rate for field applications and gives an estimate of the dissolution rate constant using the modified Peclet number calculated by:

$$Pe' = \frac{v_{avg} d_{50}}{D_m} \qquad [6.1]$$

where, v_{avg} is the average groundwater velocity, d_{50} is the median grain size and D_m is the diffusion coefficient of the NAPL species. The average groundwater velocity for the reference site can be calculated by:

$$v_{avg} = K_G.i \qquad [6.2]$$

where, K_G and i are the geometric average of hydraulic conductivity measurements and the hydraulic gradient, respectively. Equation [6.2] results in the value of average groundwater velocity to be equal to 3.09×10^{-7} m/s. Assuming the median grain size to be 2.5×10^{-3} m and the diffusion coefficient (for benzene) to be equal to 1.1×10^{-9} m^2/s, the Peclet number is calculated to be equal to 0.7. According to Dillard et al. (2001), for a Peclet number equal to 0.7 and assuming a low NAPL saturation, the range of dissolution rate constant obtained by the pore network model will be approximately between 0.8 and 8.

This range of values for k_{dis} is unrealistic in field-scale conditions (Parker and Park 2004, Christ et al. 2006) and also results in insensitivity of the state variables to the changes in the dissolution rate constant (Section 5.3.2). Thus, to investigate the effect of variability in the dissolution rate, a uniform distribution with an order of magnitude variability (which seems to be a lower bound to variability in this parameter based on the observations in Dillard et al. 2001 and Christ et al. 2006) and a mean equal to 0.0066 day^{-1} (computed by Essaid et al. (2003) for Bemidji site) is considered (Figure 6-6-a). Figure 6-7 shows the variations of the mass loaded into the aquifer and the plume length in time for the reference case (Figure 5-24) as well as the mean and quartiles of the state variables based on the results of the Monte Carlo simulations with 100 joint realizations of hydraulic conductivity field and source geometry and the values of k_{dis} and λ drawn from the distributions in Figures 6-6-a and 6-6-b. Figure 6-8 shows the probability map for the concentrations to exceed a threshold value of 0.005 mg/L (water quality standard for benzene). Figures 6-7 and 6-8 show that the MCS result in large uncertainties in the dimensions of the simulated plume as well as the mass loaded into the aquifer. The distribution of the parameters shown in Figure 6-6 and the MCS results shown in Figures 6-7 and 6-8 will be compared to the results of the inverse modeling presented in the next section.

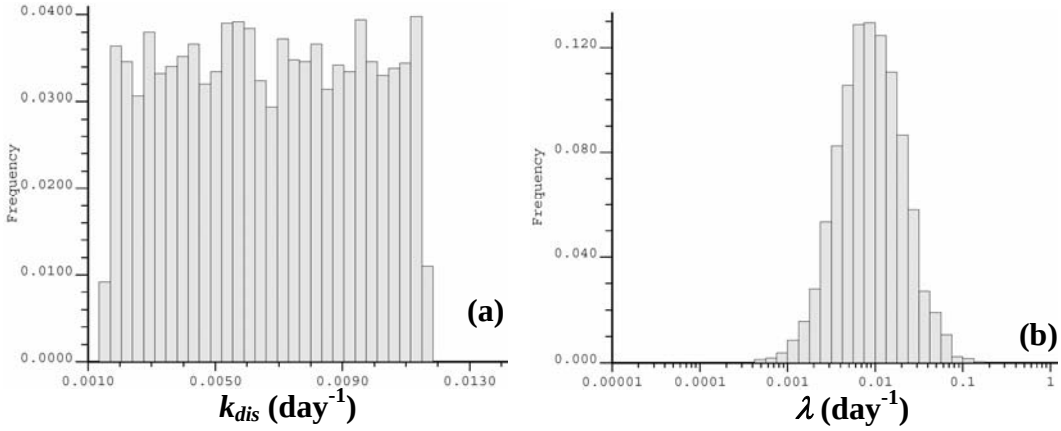


Figure 6-6: The distribution of uncertainty in (a) the dissolution rate constant and (b) the first-order biodegradation rate constant, used in the subsequent MCS.

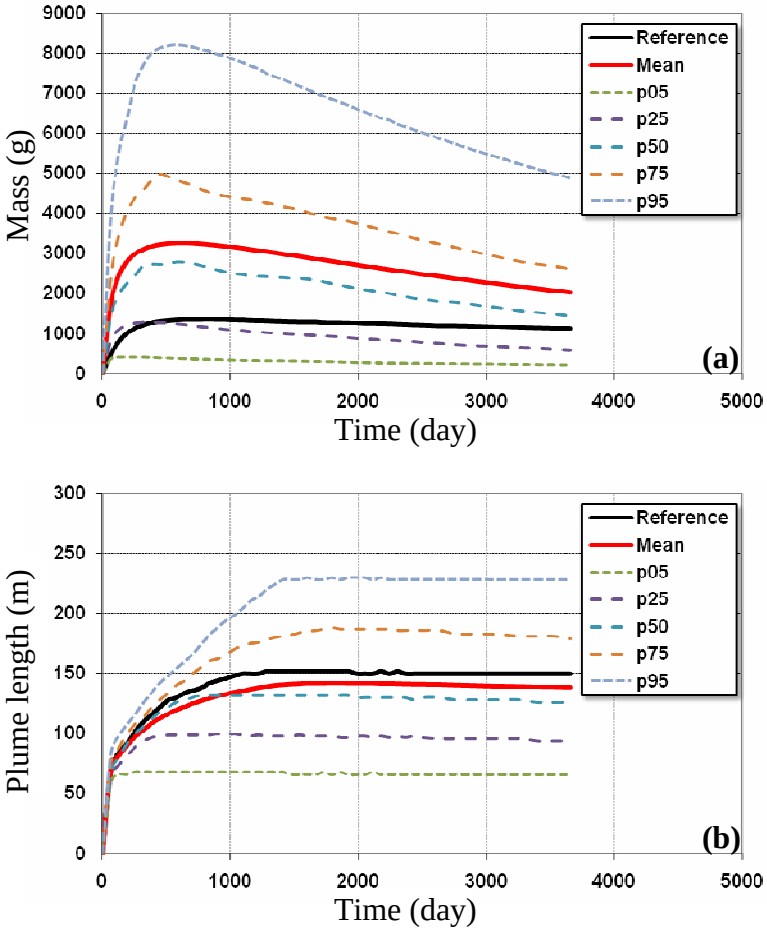


Figure 6-7: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (p_{50} source size) (a) total mass loaded into the aquifer and (b) plume length based on the results of the MCS. The reference curve is associated with the median source size (Figure 5-24)

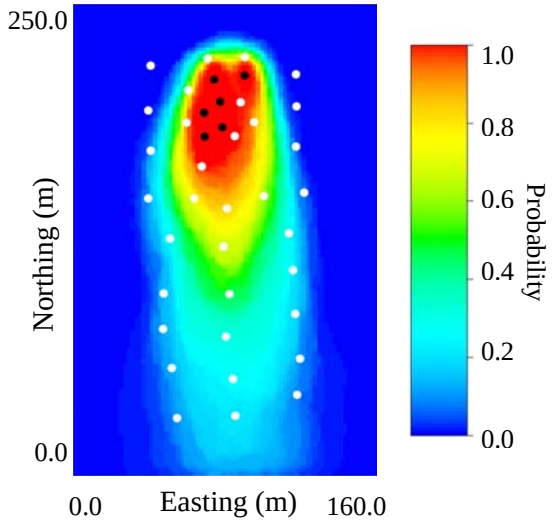


Figure 6-8: The probability of concentrations exceeding 0.005 mg/L based on the results of the Monte Carlo simulations.

6.2 Inverse modeling results

6.2.1. Error-free observations, without ranking

Figures 6-9, 6-11, and 6-13 show the histograms of the estimated k_{dis} and λ for three sets of 100 joint realizations that are calibrated to concentration measurements from the three reference cases. It has been assumed that the head and concentration measurements are error-free. Table 6-1 shows the summary statistics including non-linear confidence intervals. Figures 6-10, 6-12 and 6-14 show the variations of the mass loaded into the aquifer and the plume length through time for three sets of 100 realizations corresponding to the three reference cases. Figure 6-15 shows the probability maps for concentrations to exceed a threshold value of 0.005 mg/L for the three reference cases. According to Figures 6-9-a, 6-11-a and 6-13-a and Table 6-1, as expected, the dissolution rate constant is slightly under-estimated for the case with the smaller reference source zone and slightly over-estimated for the case with the larger reference source zone. In other words, there exists a negative correlation between the source size and the dissolution rate (Figure 6.17). For the case with the medium source zone size, the estimated k_{dis} is very close to the reference value (0.0015 day^{-1}). In all three cases, the approach on average slightly underestimates the value of λ. According to Figures 6-10, 6-12, and 6-14, the proposed approach significantly reduces the uncertainty in the state variables (comparing to the results of the Monte Carlo simulations in Figure 6-7). Although the ensemble of realizations on average over/under-estimates the reference values, in all three cases and for both state variables, the reference curve falls within the 90% non-linear confidence interval. Comparing Figure 6-15 to Figure 6-8, it is also evident that the estimation of dissolution rate and first-order biodegradation rate for joint realizations of hydraulic conductivity and source geometry using concentration data can significantly reduce the uncertainty in the dimensions of the plume.

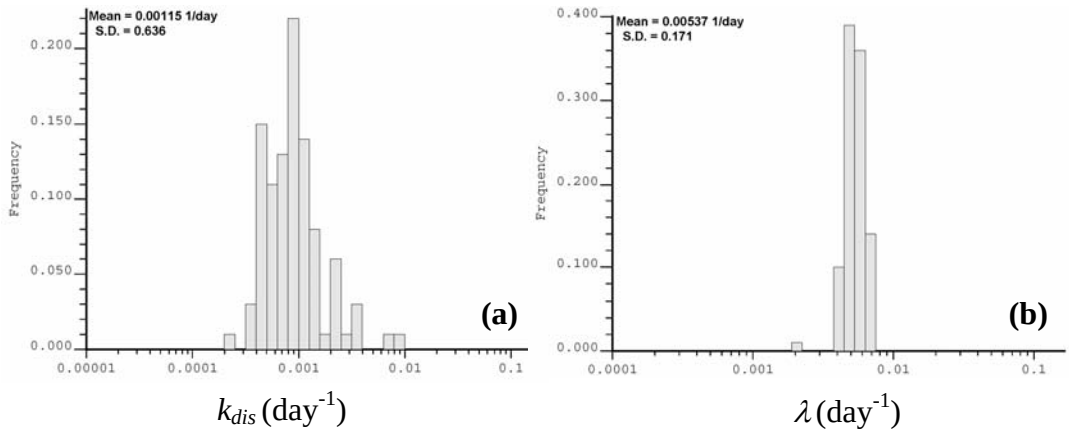


Figure 6-9: The histograms of (a) k_{dis} and (b) λ for the case with σ_{nH} = 0.0 m and cv_n = 0.0 and the smaller reference source size.

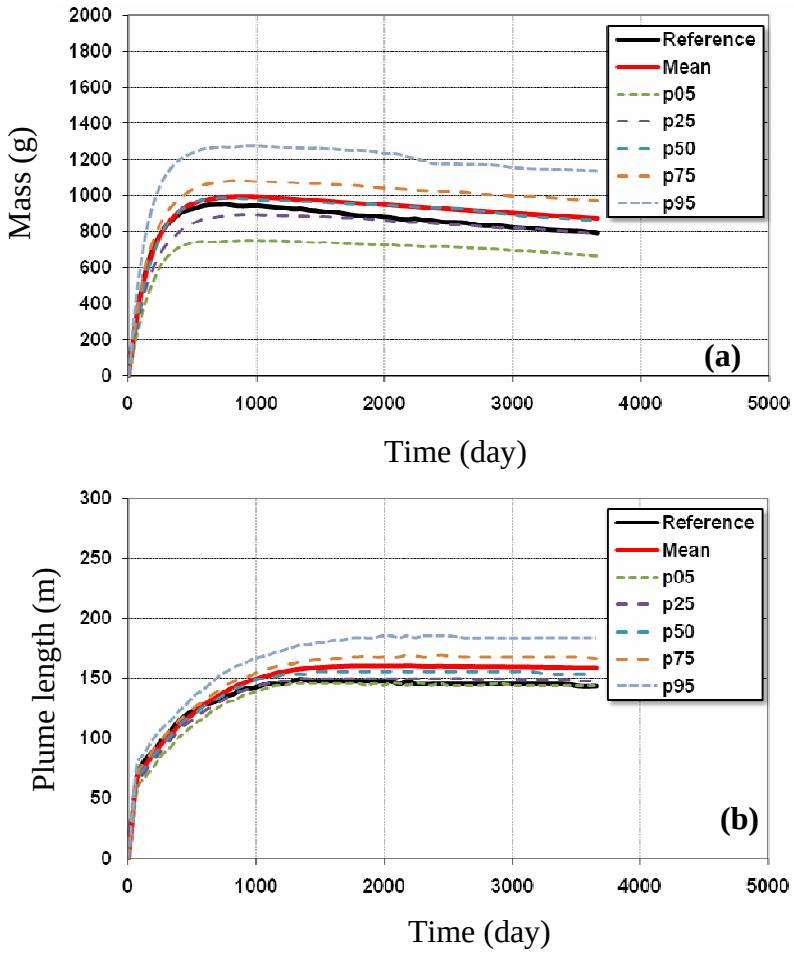


Figure 6-10: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the smaller reference source size.

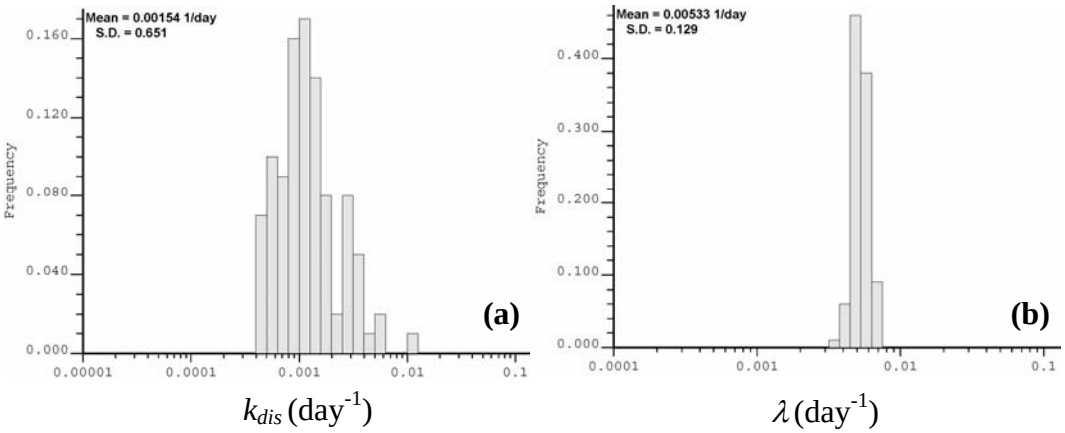


Figure 6-11: The histograms of (a) k_{dis} and (b) λ for the case with σ_{nH} = 0.0 m and cv_n = 0.0 and the medium reference source size.

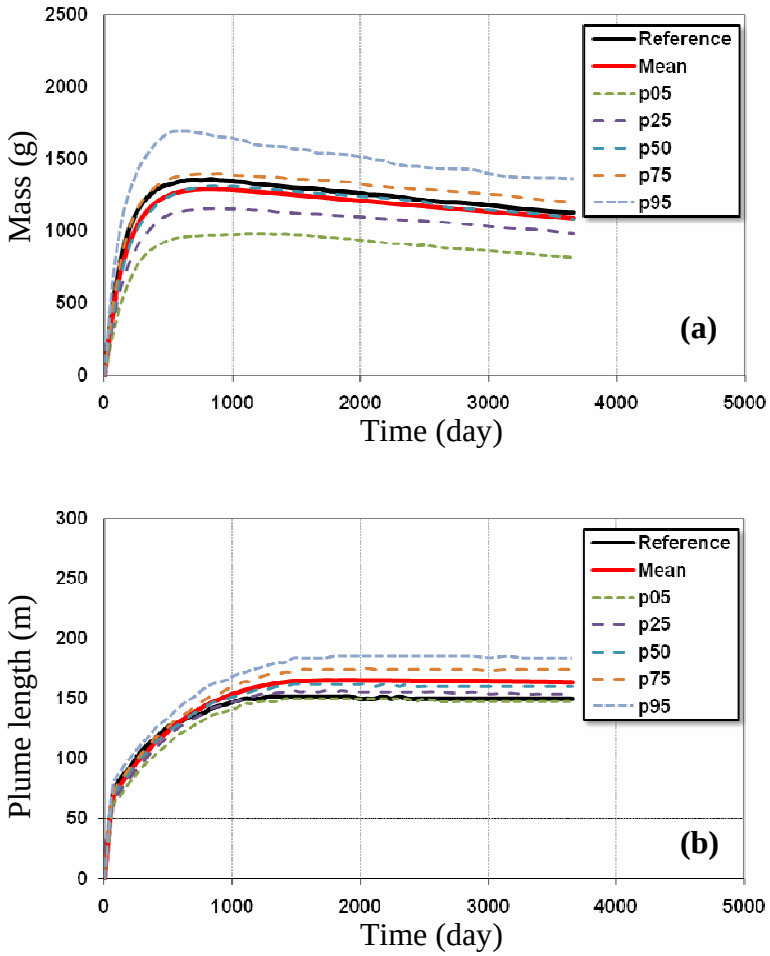


Figure 6-12: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the medium reference source size.

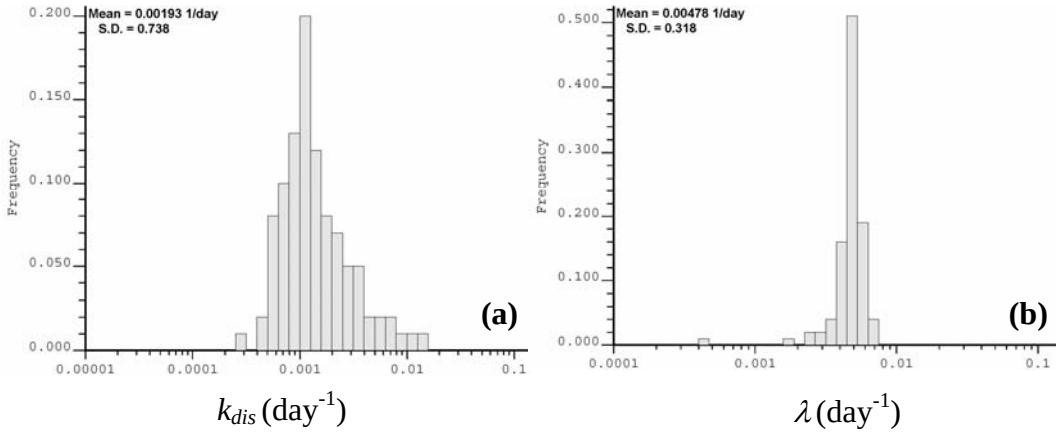


Figure 6-13: The histograms of (a) k_{dis} and (b) λ for the case with σ_{nH} = 0.0 m and cv_n = 0.0 and the larger reference source size.

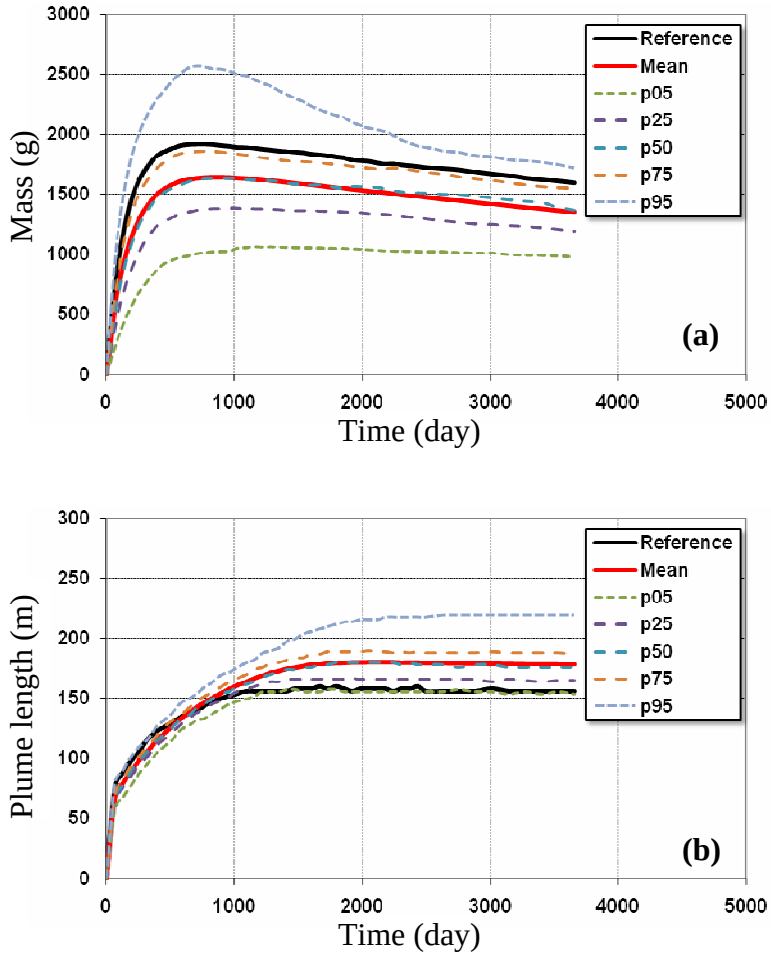


Figure 6-14: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the larger reference source size.

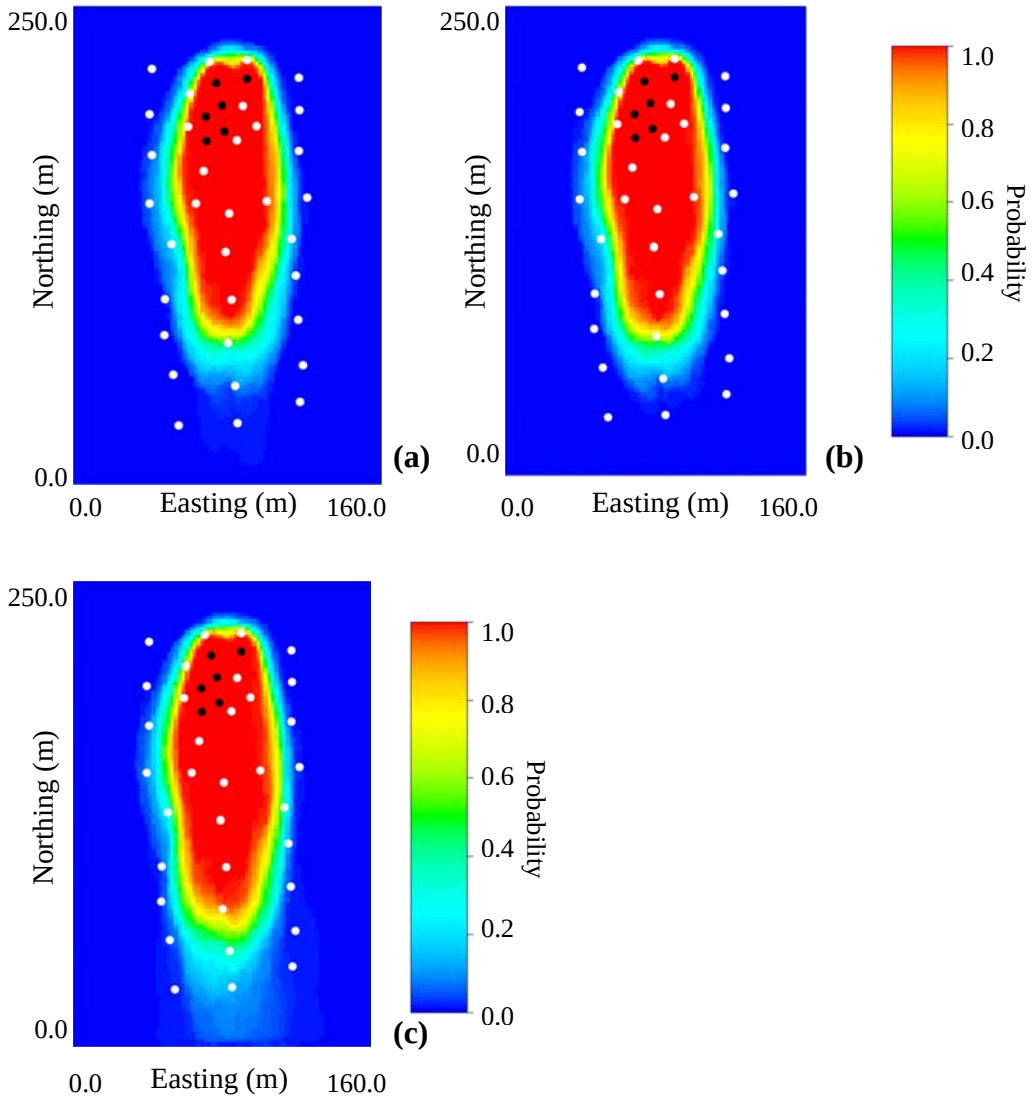


Figure 6-15: The probability of concentrations exceeding 0.005 mg/L after conditioning to concentrations for (a) smaller source zone, (b) medium source zone, and (c) larger source zone.

For both the estimated parameters and the state variables, larger uncertainties are observed for the reference case with the larger source zone. Nevertheless, for all three reference cases the estimated parameters and simulated state variables show a better match to reference values than the available field techniques (Beyer et al. 2006, Beyer et al. 2007).

	Reference Source Size		
	Smaller	Medium	Larger
Mean k_{dis} (day^{-1})	0.00115	0.00154	0.00193
Standard deviation k_{dis}	0.636	0.651	0.738
Median k_{dis} (day^{-1})	0.00086	0.00108	0.00119
p_{05} - k_{dis} (day^{-1})	0.00040	0.00048	0.00055
p_{95} - k_{dis} (day^{-1})	0.00277	0.00372	0.00579
Mean λ (day^{-1})	0.00537	0.00533	0.00478
Standard deviation λ	0.171	0.129	0.318
Median λ (day^{-1})	0.00530	0.00528	0.00485
p_{05} - λ (day^{-1})	0.00424	0.00443	0.00313
p_{95} - λ (day^{-1})	0.00662	0.00651	0.00615

Table 6-1: The summary statistics including mean, standard deviation, median and p05 and p95 quantiles (representing 90% confidence interval) for the estimated transport parameters before ranking the realizations

	Reference Source Size		
	Smaller	Medium	Larger
Mean k_{dis} (day^{-1})	0.00186	0.00156	0.00133
Standard deviation k_{dis}	0.586	0.701	0.575
Median k_{dis} (day^{-1})	0.00134	0.00112	0.00091
p_{05} - k_{dis} (day^{-1})	0.00071	0.00048	0.00055
p_{95} - k_{dis} (day^{-1})	0.00496	0.00616	0.00221
Mean λ (day^{-1})	0.00551	0.00519	0.00509
Standard deviation λ	0.121	0.115	0.128
Median λ (day^{-1})	0.00539	0.00506	0.00498
p_{05} - λ (day^{-1})	0.00452	0.00445	0.00405
p_{95} - λ (day^{-1})	0.00664	0.00628	0.00620

Table 6-2: The summary statistics including mean, standard deviation, median and p05 and p95 quantiles (representing 90% confidence interval) for the estimated transport parameters after ranking and screening the realizations

6.2.2. Error-free observations, with ranking

The observed over-estimation and under-estimation of the state variables is partially due to unresolved uncertainties in the source size which can not be fully handled by adjusting the values of k_{dis} and λ by the model. Thus, a ranking-based screening can be applied (Similar to the work of Poeter and McKenna 1995) to choose from the set of realizations based on the values of the modified objective function, and to decrease the uncertainty in the source zone sizes previously characterized by the distance-function approach (Figure 6-2-a). As discussed in Section 5.2.1, the objective functions defined by Equations [5.15] and [5.16] are not appropriate for ranking the realizations conditioned to concentration data. It is observed in this example that the modified objective function, presented in Equation [5.18], gives a more appropriate measure for ranking the realizations that can be used to decrease the uncertainty in source dimensions and result in better match to the reference parameter values and state variables. Due to the fact that the modified objective function is a dimensionless number, it can be considered a robust measure for ranking that is largely independent of overall level of concentrations in the field. To investigate the effectiveness of ranking on the reduction of uncertainty and to have enough realizations to explore the space of uncertainty, 300 joint realizations of hydraulic conductivity (conditioned to head data with σ_{nH} =0.0 m) and source geometry are constructed and the concentrations sampled from the three reference cases (with cv_{nc} =0.0) are used to estimate the values of dissolution rate constant and first-order biodegradation rate for each realization. Figures 6-16-a, b and c show the CDF of the source sizes for 100 realizations (out of 300 realizations) that have the smallest values of modified objective function defined by Equation [5.18] for the three reference cases.

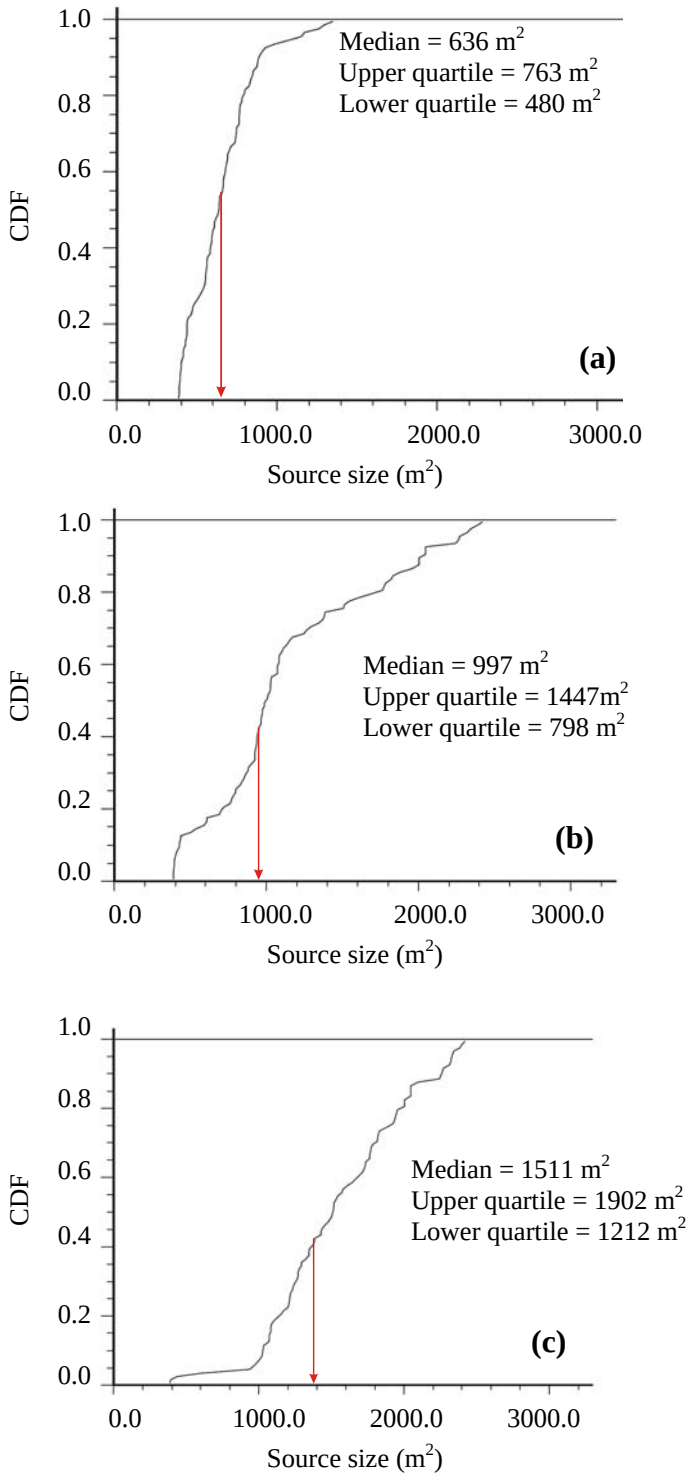


Figure 6-16: The CDF of the source sizes of the 100 accepted realizations after ranking based on the modified objective function value for (a) smaller, (b) medium, (c) larger source. The arrows show the reference size for each case.

Comparing Figure 6-16 to Figure 6-2-b, one can observe that ranking and screening the realizations can effectively reduce the uncertainty in the source zone sizes for each reference case. To further investigate the effect of ranking, one may also look at Figure 6-17 where the cross-plots of dissolution rates and source size quantiles for the two reference cases with smaller and larger source sizes are shown. The color-scale represents the rank of each realization based on the value of the modified objective function (black shows lower values of the modified objective function, higher ranks and therefore accepted realizations). For this example, Figure 6-17-a, b show that (1) there is a negative correlation between the size of the source the estimated dissolution rate; and (2) ranking of the realizations can effectively identify the joint realizations that have not properly converged in optimization (solid circles) and the joint realizations that have source sizes that deviate from the reference source size (dashed circles). In Figures 6-17-a, b the reference source sizes correspond to p_{25} and p_{75} quantiles (0.25 and 0.75) on the horizontal axis, respectively. Figures 6-18 and 6-19 show the cross-plots between the dissolution rate and first-order biodegradation rate constant and the cross-plots between the source size quantile and first-order biodegradation rate (after ranking) for the smaller and larger reference source sizes. According to Figure 6-18, there exists a positive correlation between first-order biodegradation rate and dissolution rate. Looking at Figure 6-19, one observes that there is little correlation between the values of source quantile and first-order biodegradation rate. The observations in Figures 6-17, 6-18 and 6-19 justify the importance of simultaneous characterization of uncertainty in source areal extent, source dissolution rate and first-order biodegradation rate, that is, the dissolution rate is adjusted with respected to the source size to control the concentrations in the upstream portion of the plume, while biodegradation rate is adjusted with respect to dissolution rate to control the concentrations at the upstream as well as downstream of the site.

Figures 6-20 to 6-26 and Table 6-2 show the estimated parameters, the associated state variables and the probability maps after ranking and screening based on the values of the modified objective function. In these figures, one can observe that ranking and screening the realizations can effectively reduce the uncertainty and mitigate the bias in the estimated parameters and the predicted state variables. An important related issue is to investigate the effect of the number of accepted realizations on the distribution of the source zone sizes after screening. For this purpose, a sensitivity analysis is performed, where the best 50, 100 and 150 realizations are selected (based on the value of modified objective function) and the CDF of the source sizes are compared in Figures 6-27, 6-28 and 6-29, where one can see ranking and screening any number of realizations reduces the uncertainty.

According to Figures 6-16 to 6-29, one can conclude that (1) reduction in the uncertainty of the source zone sizes appears to be achievable by ranking and screening the realizations based on the values of the modified objective function (Equation [5.18]); (2) for this purpose, an appropriate number of realizations should be selected; and (3) by reducing the uncertainty in the source zone sizes, there will be reductions in the associated uncertainty in the estimated parameters values and the state variables. Although the choice of the number of realizations seems to be problem-dependent, ranking the realizations based on the values of the modified objective function is considered to be effective in providing a general idea about the size of the source zone. As discussed above, although the Equation [5.18] represents a dimensionless number and seems to be a robust measure for ranking that is largely independent of overall level of concentrations at the site, future research in this area is needed.

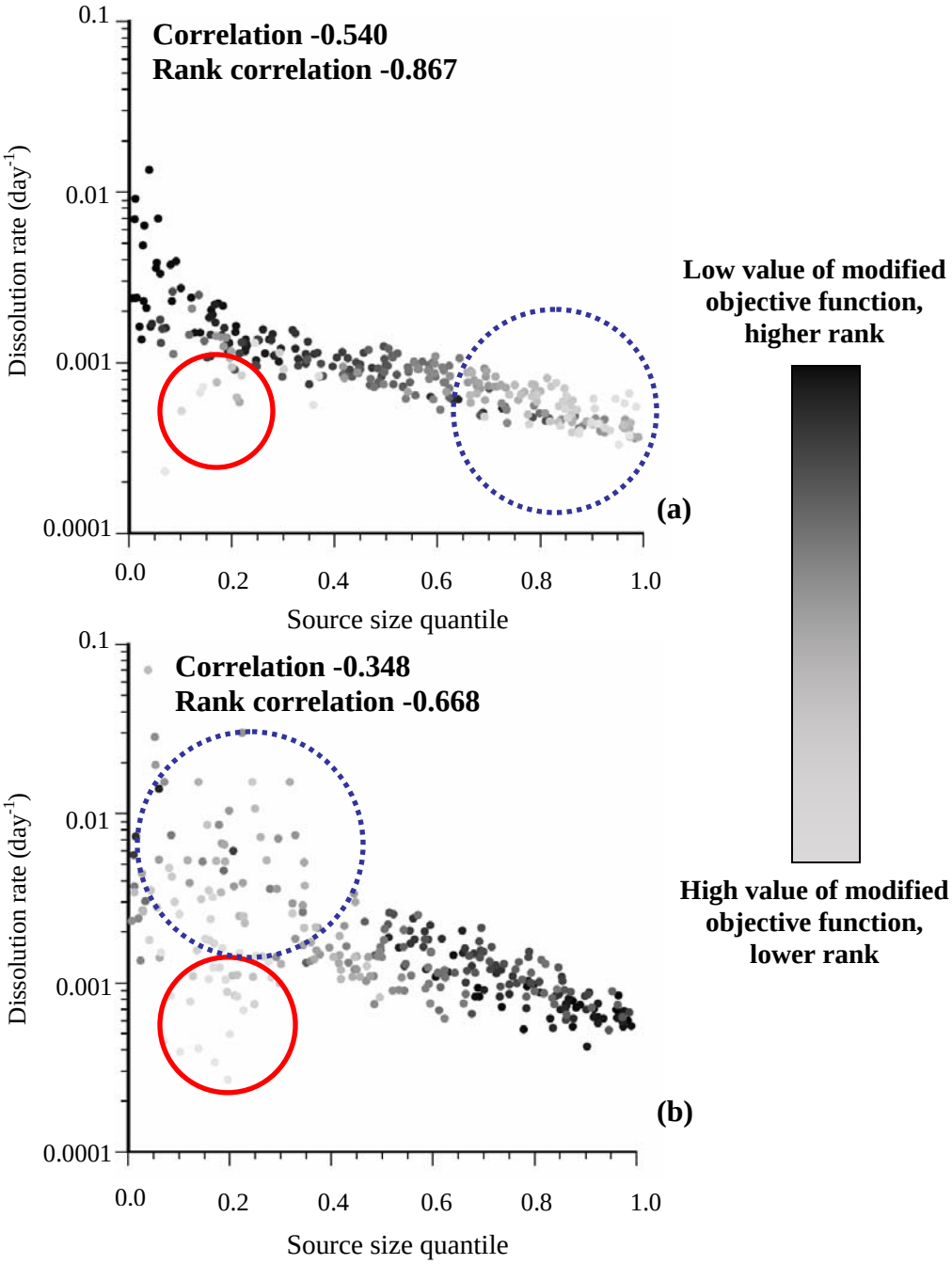


Figure 6-17: The cross-plots between the source size quantiles and the estimated dissolution rate constant for the reference case with (a) smaller source size (p_{25}), and (b) larger source size (p_{75}). The solid circles show the realizations that are likely not converged and the dashed circles show the realizations that their source sizes significantly deviate from the reference source size. The color scale shows the rank of realizations based on their modified objective function value.

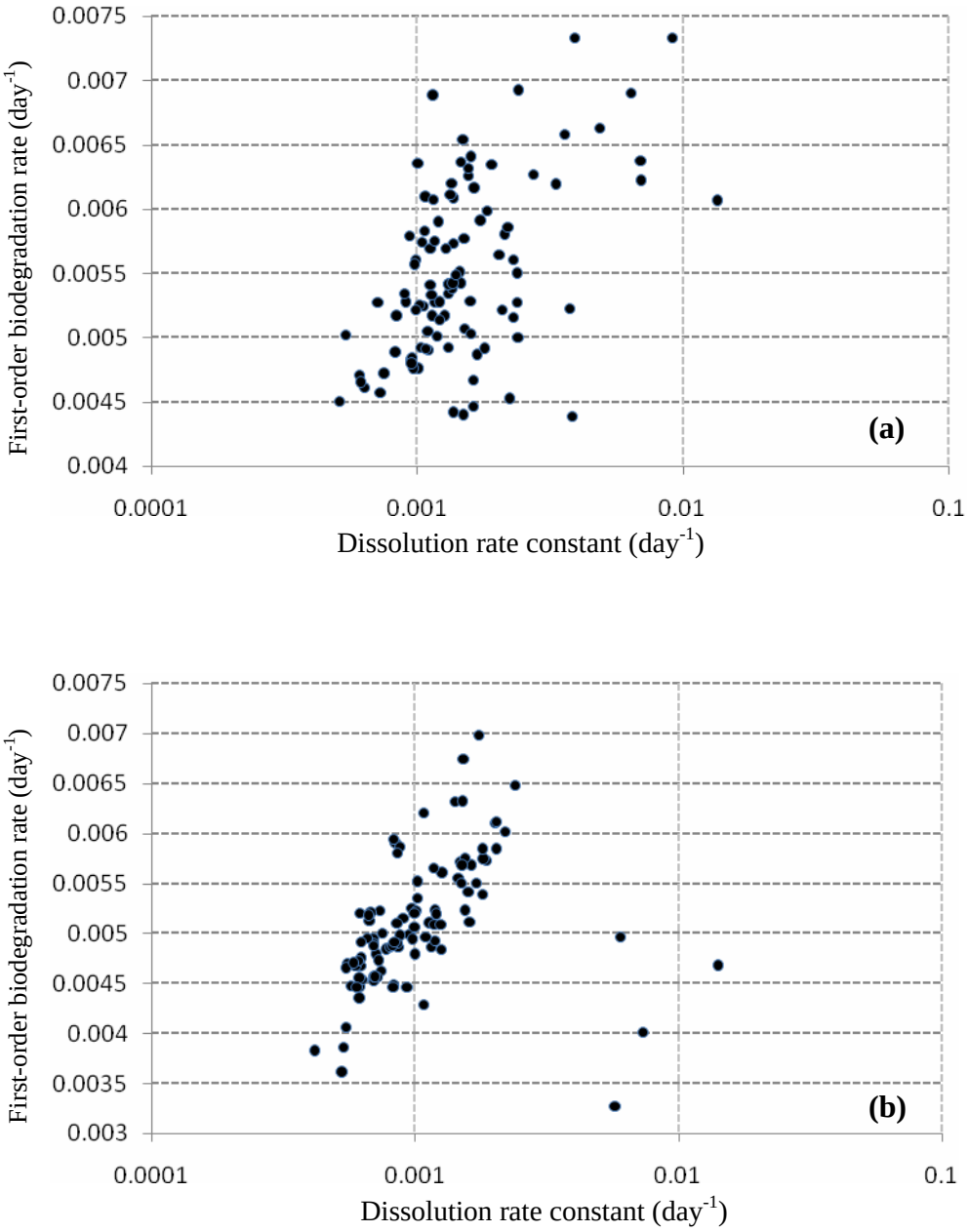


Figure 6-18: The cross-plots between the dissolution rate constant and firs-order biodegradation rate constant for the reference case with (a) smaller source size (p_{25}) with a correlation coefficient equal to 0.438; and (b) larger source size (p_{75}) with a correlation coefficient equal to 0.744 (after removing the outliers)

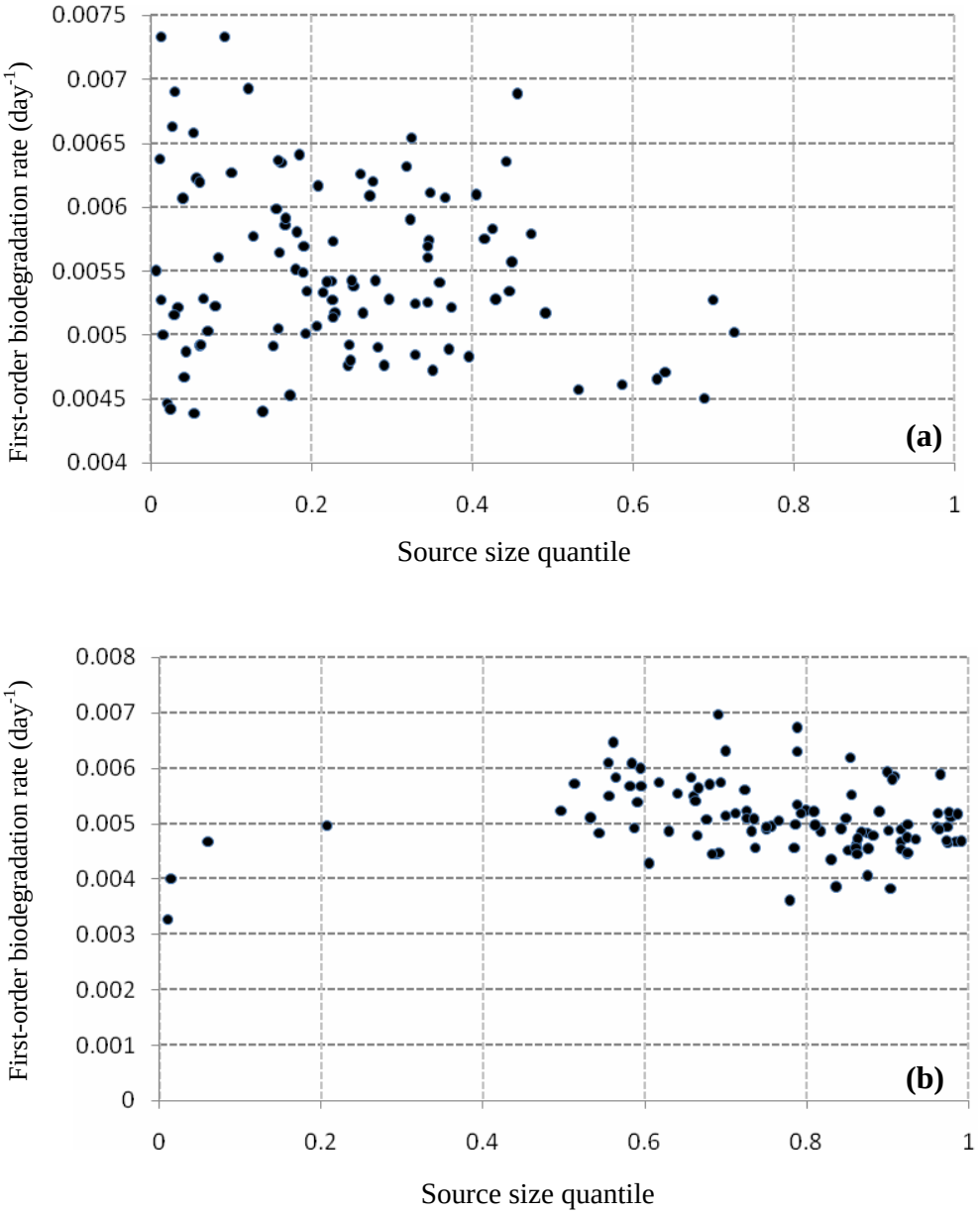


Figure 6-19: The cross-plots between the source size quantile and first-order biodegradation rate constant for the reference case with (a) smaller source size (p_{25}), and (b) larger source size (p_{75})

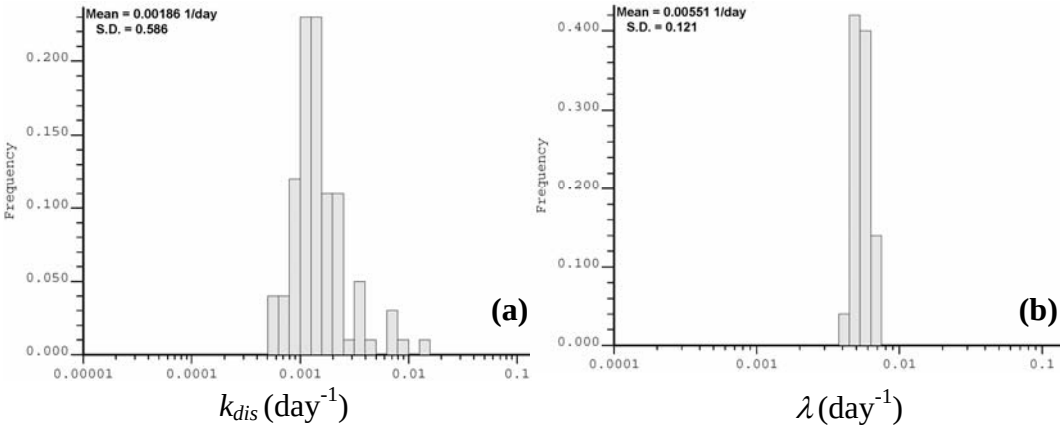


Figure 6-20: The histograms of (a) k_{dis} and (b) λ for the accepted realizations after ranking, based on the reference case with the smaller source size.

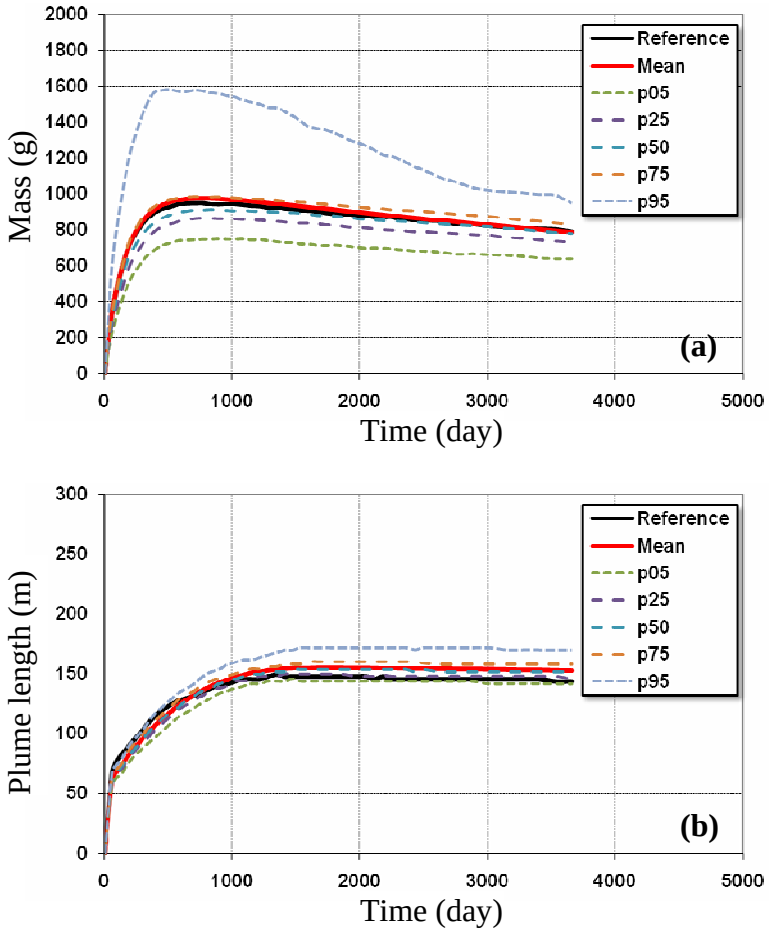


Figure 6-21: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the smaller source size after ranking.

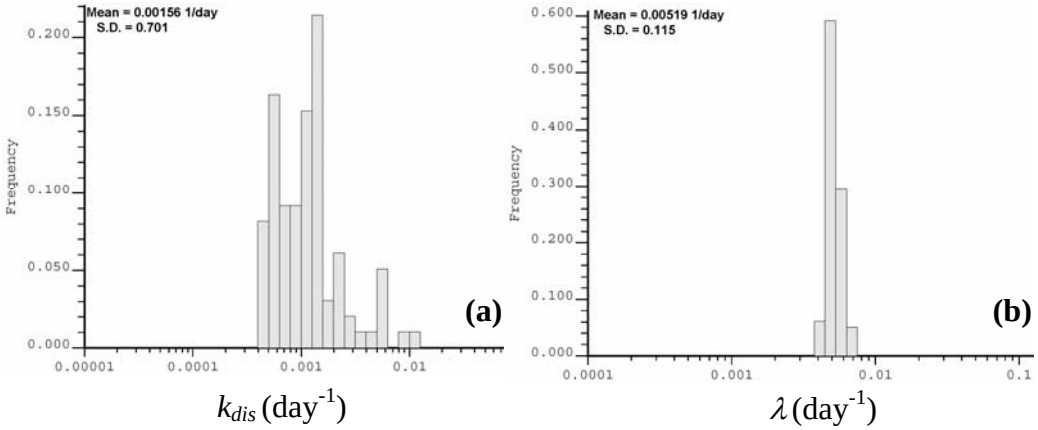


Figure 6-22: The histograms of (a) k_{dis} and (b) λ for the accepted realizations after ranking, based on the reference case with the medium source size.

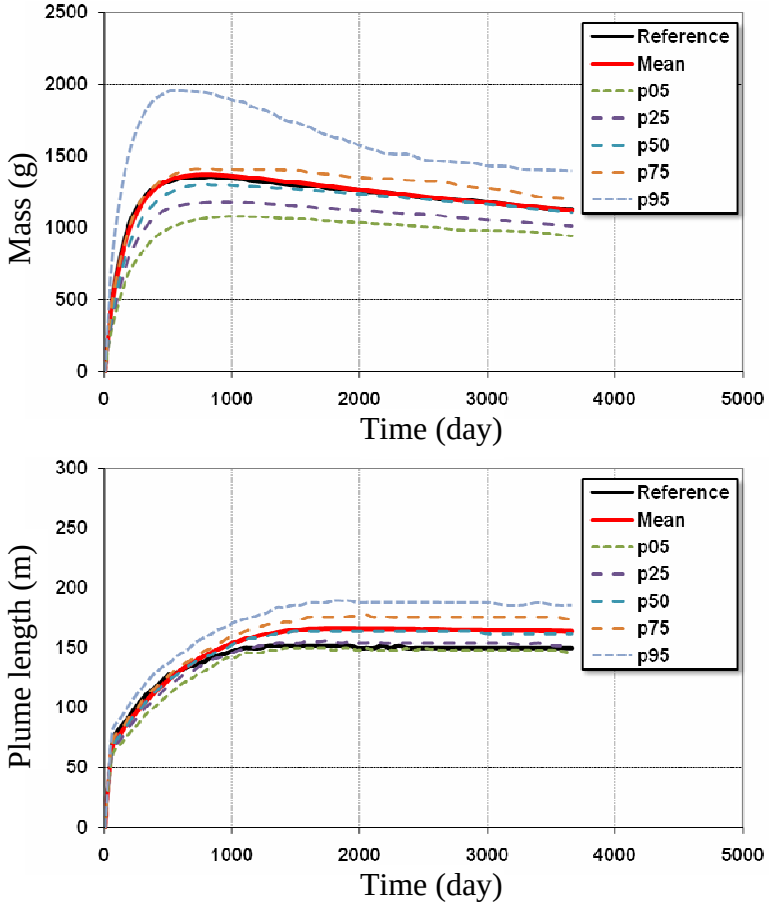


Figure 6-23: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the medium source size after ranking.

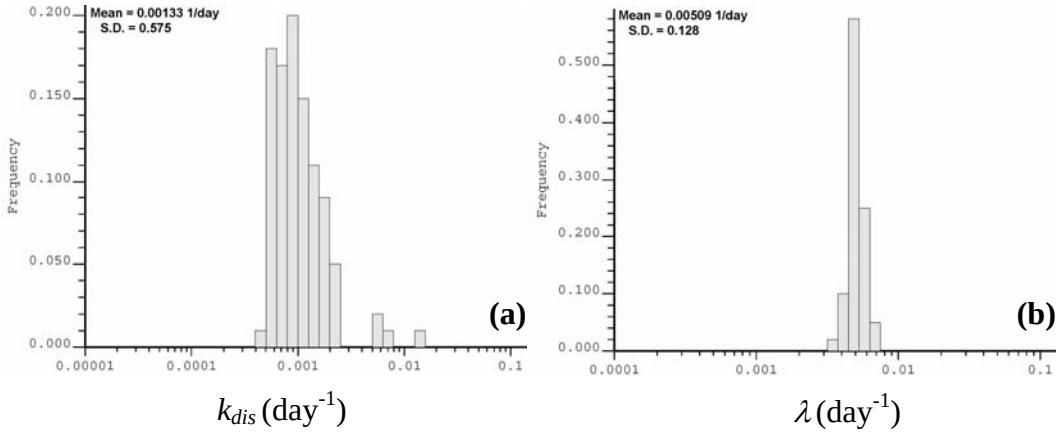


Figure 6-24: The histograms of (a) k_{dis} and (b) λ for the accepted realizations after ranking, based on the reference case with the larger source size.

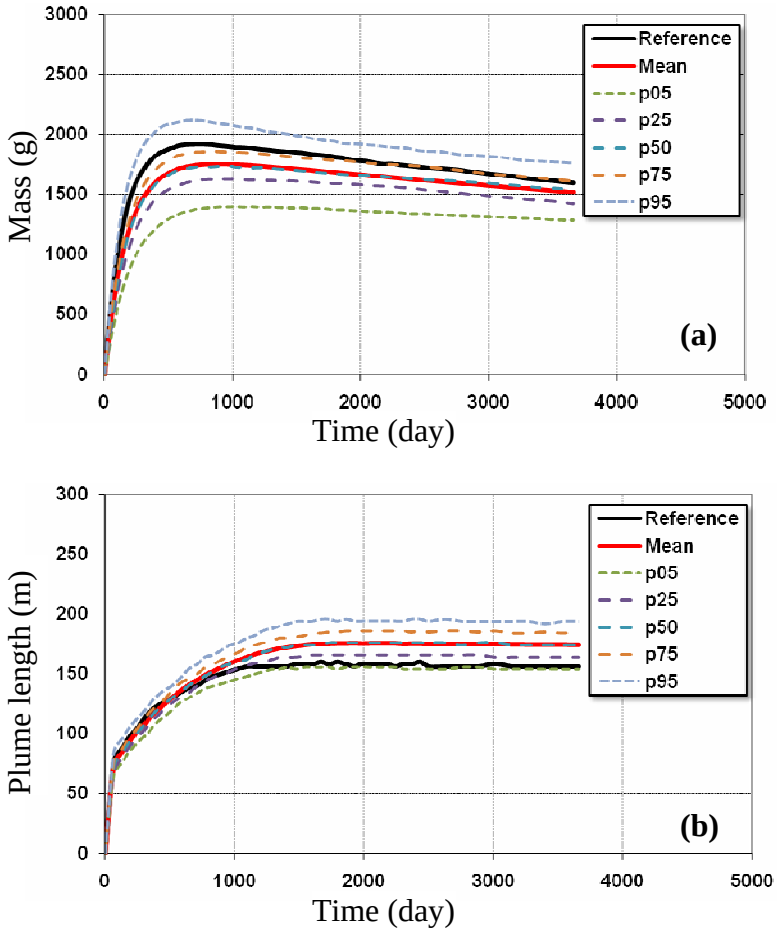


Figure 6-25: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the larger source size after ranking.

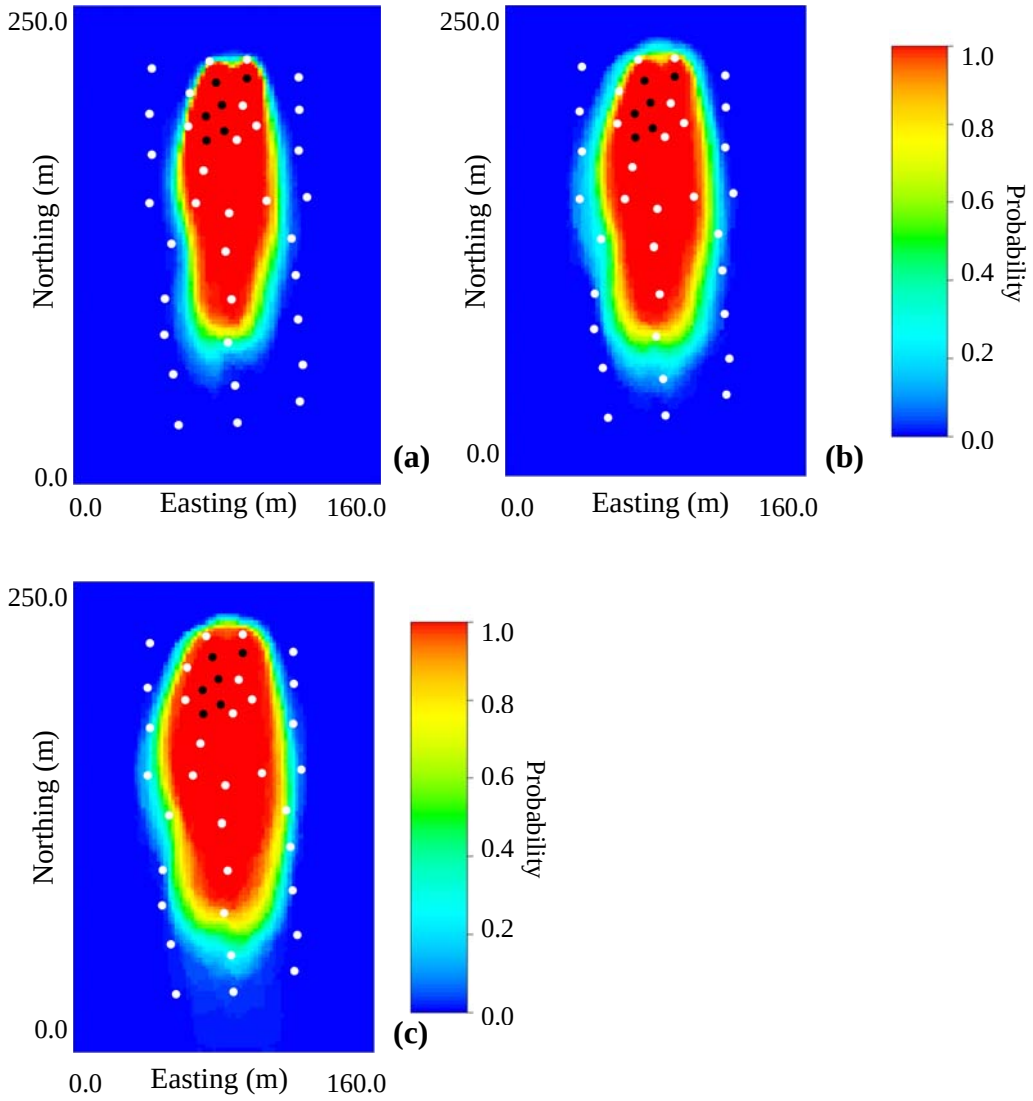


Figure 6-26: The probability of concentrations exceeding 0.005 mg/L after ranking for (a) smaller source zone, (b) medium source zone and (c) larger source zone

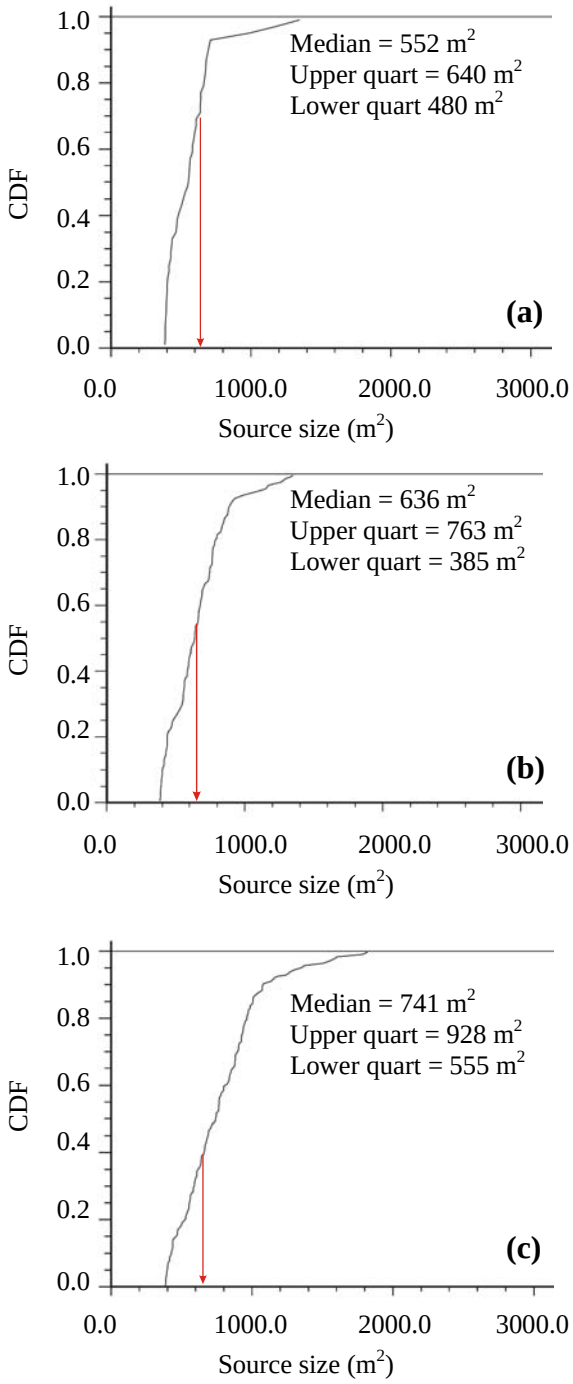


Figure 6-27: The CDF of the source sizes associated with the first (a) 50, (b) 100, and (c) 150 realizations after ranking based on the modified objective function value for the reference case with the smaller source size. The red arrows show the reference source size.

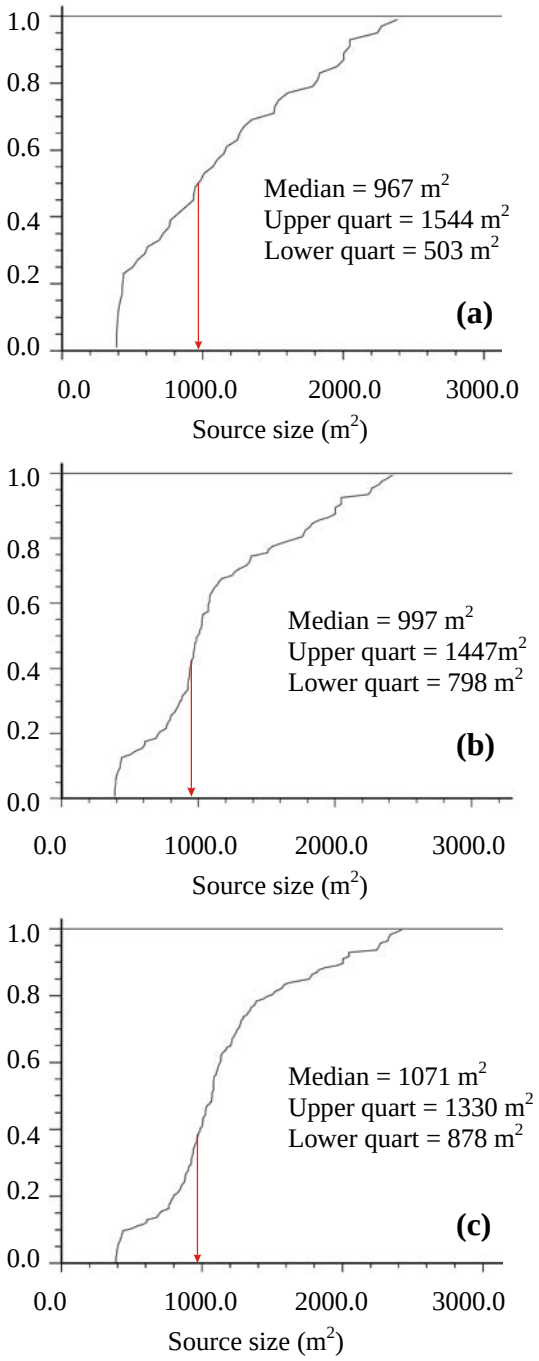


Figure 6-28: The CDF of the source sizes associated with the first (a) 50, (b) 100, and (c) 150 realizations after ranking based on the modified objective function value for the reference case with the medium source size. The red arrows show the reference source size.

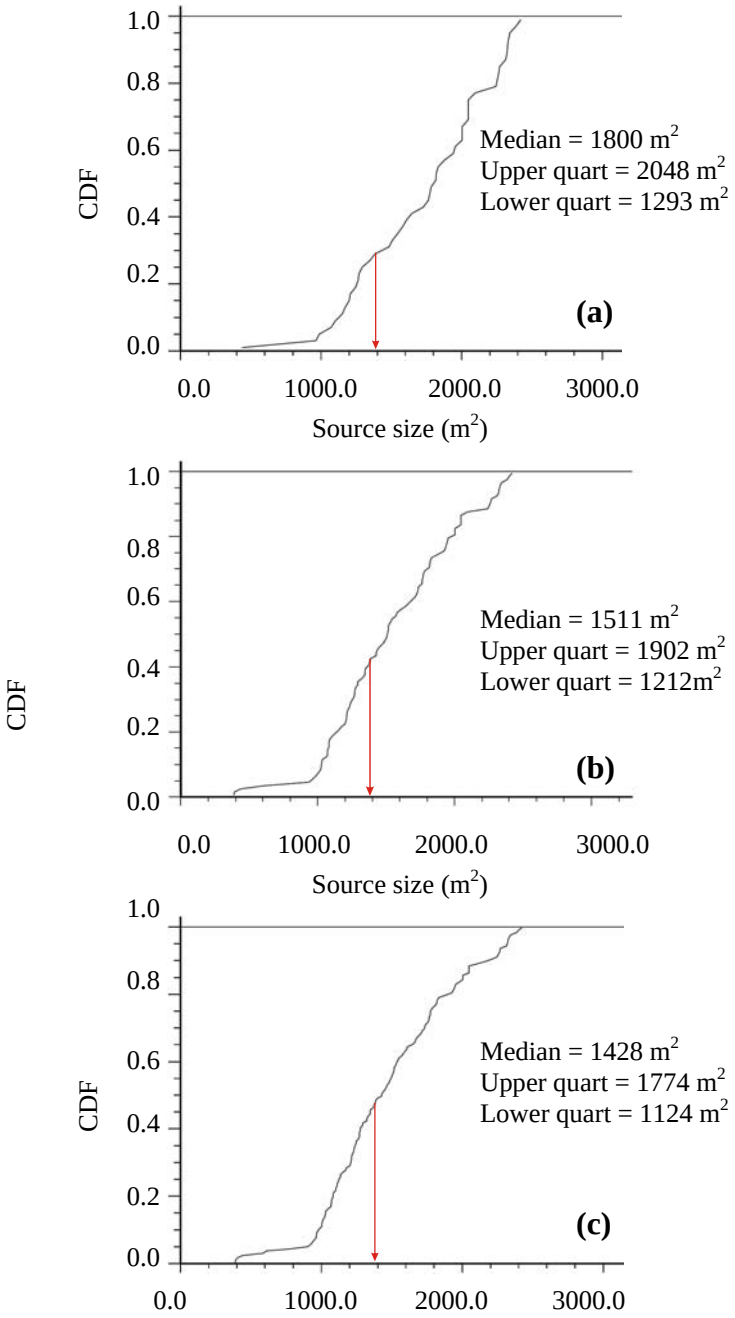


Figure 6-29: The CDF of the source sizes associated with the first (a) 50, (b) 100, and (c) 150 realizations after ranking based on the modified objective function value for the reference case with the larger source size. The red arrows show the reference source size.

6.2.3. The effect of error in observations

In practice, it is quite rare to consider the observation data as error-free. Due to the fact that the proposed methodology is a decoupled approach, one may generate hydraulic conductivity realizations honoring a particular level of error in head observations (measure of fit *s* close to one); and then estimate the rate constants for the joint realizations, while calibrating to concentrations with a particular level of error in the data values. In the subsequent analysis, Gaussian noise (with a relatively large standard deviation/coefficient of variation) has been added to the head observations and concentration measurements associated with the smaller reference case. It is assumed that (1) a good knowledge of magnitude of error exists in the observations, (2) the error in observations is Gaussian noise with a mean equal to zero (no systematic bias is introduced), and (3) errors at different locations and for heads and concentrations are independent of each other. As it can be observed in the following example, uncertainty in the estimated parameters and the predicted state variables increases with an increase in the measurement errors.

Comparing Figures 6-30, 6-31 and 6-32 to Figures 6-9, 6-10 and 6-15-a, one can observe that introducing measurement errors to heads and concentrations results in (1) considerable increase in the standard deviation of the estimated parameters, (2) larger deviation of the average biodegradation rate constant from the reference value, (3) introducing more uncertainty and bias in the estimation of the length (and width) of the plume and (4) increasing the uncertainty in the estimation of mass loaded into the aquifer. Although introducing measurement errors (with relatively large standard deviations) can considerably increase the uncertainty in the response variables, the observed uncertainties are still smaller than the uncertainties observed in the Monte Carlo simulation results (Figures 6-7 and 6-8).

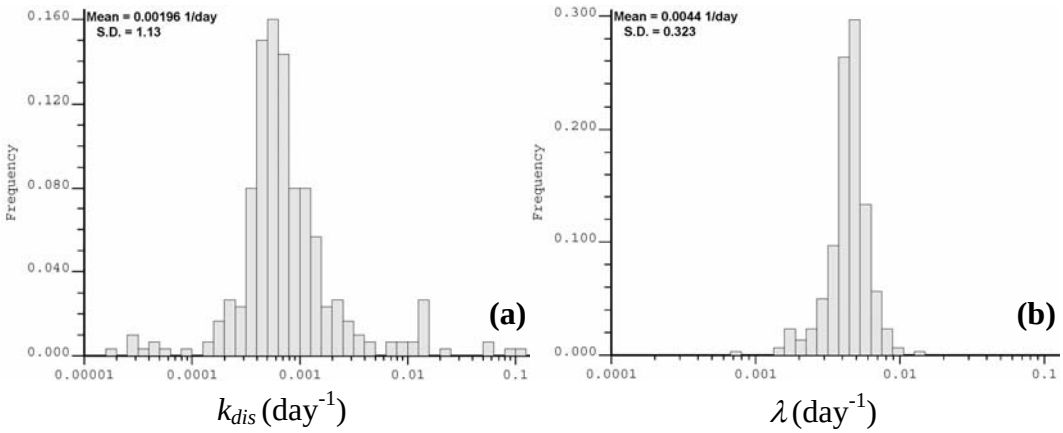


Figure 6-30: The histograms of (a) k_{dis} and (b) λ for the case with σ_{nH} = 0.2 m and cv_n = 0.3 and the smaller reference source size.

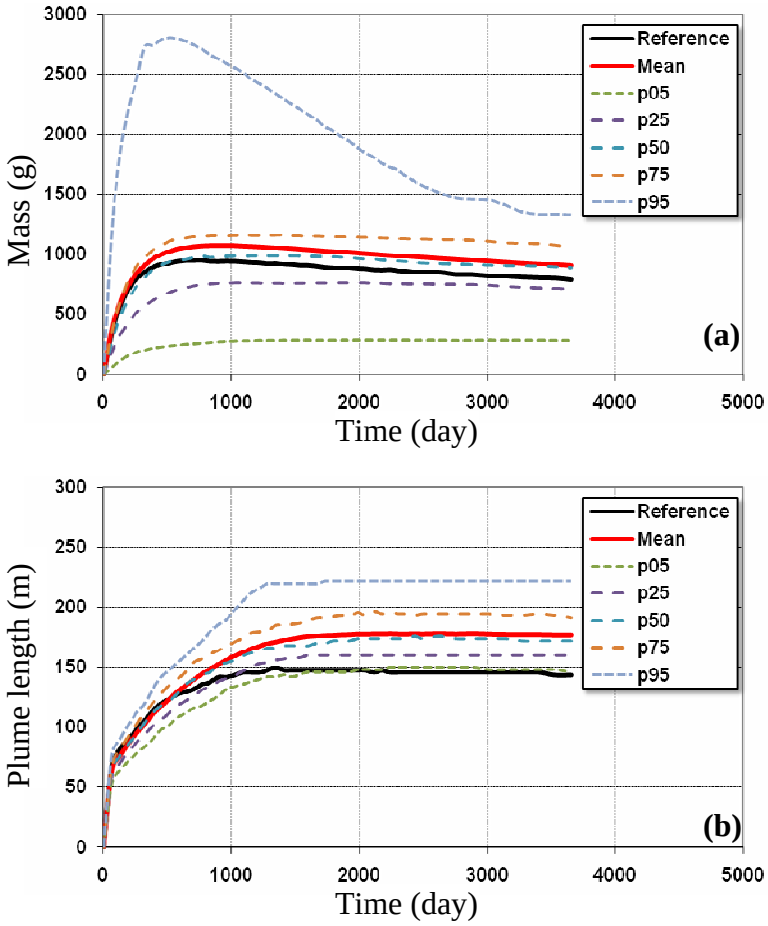


Figure 6-31: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the case with σ_{nH} = 0.2 m and cv_n = 0.3 and the smaller reference source size.

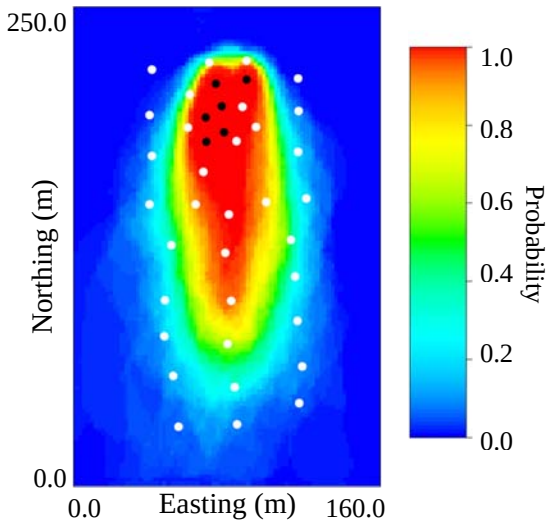


Figure 6-32: The probability of concentrations exceeding 0.005 mg/L for the case with σ_{nH} = 0.2 m and cv_n = 0.3 and the smaller reference source size.

6.2.4. The effect of error in the time of release

Up to this point, it has been assumed that a good knowledge of time of release (or the time that NAPL comes into contact with groundwater) exists in all the simulations. In the following example, it is shown that wrong estimates for the start of simulations has little impact on the modeling outcomes (prediction of current and future state of plume length and mass), as long as the concentration measurements are sampled when the plume is in steady-state condition. The existing concentration measurements (form the reference case with the smaller source) are sampled between 5 to 7 years from the start of simulations. In this example, it is assumed that the practitioner makes a wrong estimate of the release date to be 3 years earlier than the actual release date and uses the decoupled approach to characterize the uncertainty in the transport parameters and the state variables. Comparing Figures 6-33, 6-34 and 6-35 to Figures 6-9, 6-10 and 6-15-a, one can observe that the misinterpretation starting time has a minor impact on

the prediction of uncertainty in the state variables and the probability map for the plume.

In summary, in Chapter 6 a synthetic example was presented to investigate the performance of the Monte Carlo type decoupled inverse modeling presented in Chapter 5 in characterizing the uncertainty in the dissolution rate constant and first-order biodegradation rate. Three reference cases with different source sizes were considered and head and concentration data were sampled. The observed data were used to estimate the values of the parameters and simulate the state variables. The modeling outcomes were compared to the results of a set of Monte Carlo simulations, and observed that calibrating to concentration data can effectively reduce the uncertainties in the state variables. In all cases, the reference state variables fell within 90% confidence interval defined by the ensemble of realizations. The bias observed in the prediction of the state variables was deemed to be partially due to large variability in the source zone size. To reduce the uncertainty in the source size and mitigate the bias in the predictions, ranking and screening of the realizations based on the value of objective function was implemented. Finally, the effects of error in observations and starting time of simulations were also investigated.

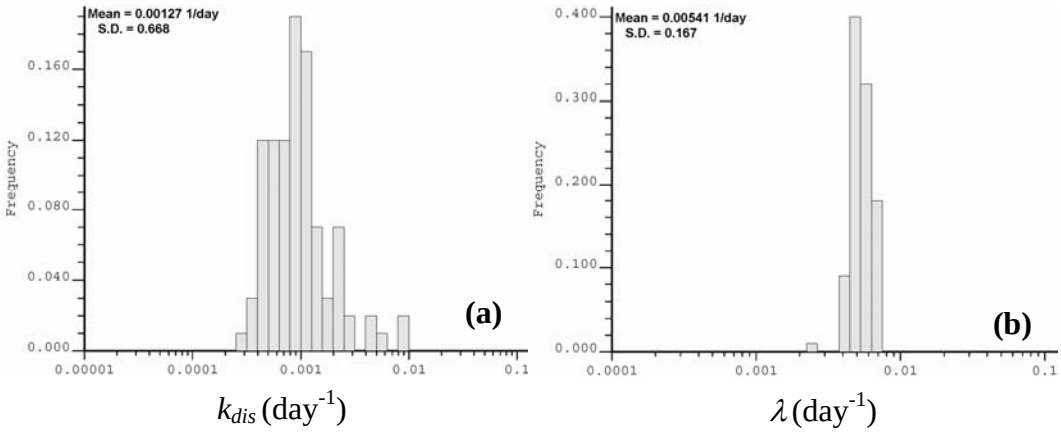


Figure 6-33: The histograms of (a) k_{dis} and (b) λ for the case with σ_{nH} = 0.0 m and cv_n = 0.0 and the smaller reference source size, when the start of simulations is misestimated for 3 years.

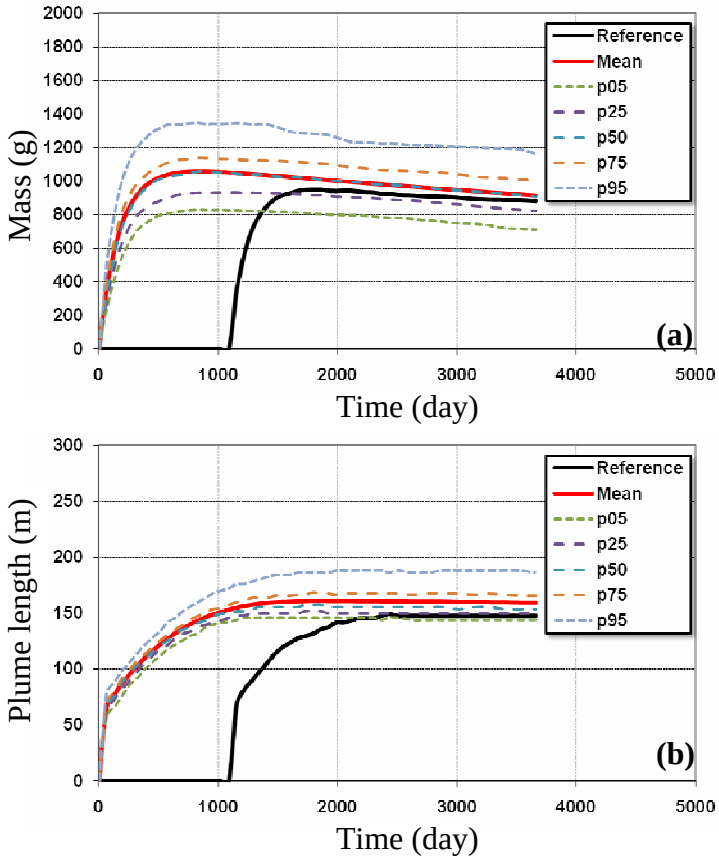


Figure 6-34: The variations of simulated (p_{05}, p_{25}, p_{50}, p_{75}, and p_{95} quantiles of ensemble of realizations) and reference (a) total mass loaded into the aquifer and (b) plume length for the smaller reference source size, when the start of simulations is misestimated for 3 years.

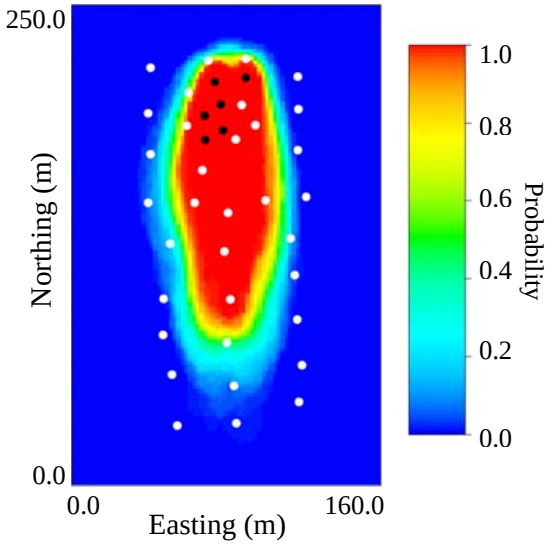


Figure 6-35: The probability of concentrations exceeding 0.005 mg/L for the smaller reference source size, when the start of simulations is misestimated for 3 years.

CHAPTER 7

UNCERTAINTY IN FIELD-MEASURED BIODEGRADATION RATE CONSTANT

The biodegradation rate constant of organic contaminants is particularly important for decision making and management of contaminated sites. Field approaches to estimate the rate constants by using concentration-distance relations along the plume centerline are commonly used; however, rate constants estimated in this manner are affected by heterogeneity in the hydraulic conductivity field, observation errors in hydraulic head measurements, uncertainty in field-scale dispersivities, uncertainty in source properties, as well as other environmental drivers including precipitation and the associated infiltration, temperature and variations in groundwater elevation.

As detailed in Chapter 2, a number of previous works dealt with uncertainty in field-measured biodegradation rate constants. The most important works in this area include the works of Odencrantz et al. (2003), Stenback et al. (2003), Bauer et al. (2006), Beyer et al. (2007) and Ukankus and Unlu (2008). These researchers showed that the field estimated first-order biodegradation rate constant is prone to uncertainty due to (1) uncertainty in field dispersivities, (2) uncertainty in field-scale groundwater velocity, and (3) heterogeneity and uncertainty in the transmissivity field that results in missing the plume centerline. They also showed that the level of uncertainty in the estimated first-order biodegradation rate can be up to a few orders of magnitude (overestimation) and depends on the degree of heterogeneity, the correlation range of transmissivity field, the width of the source zone and the level of uncertainty in average flow velocity and dispersivity values.

Stenback et al. (2003) and Beyer et al. (2007) showed that utilizing concentrations sampled off the centerline of the plume and calibrating a 2D analytical mass transport model against the sampled concentrations may or may not reduce the degree of over-estimation in the rate constants. This improvement, however, depends on the width of the plume.

The work that is presented in this chapter follows a similar approach as that presented by Bauer et al. (2006), Beyer et al. (2007) and Ukankus and Unlu (2008), where a number of transmissivity realizations were generated to represent the heterogeneity and uncertainty in the transmissivity field. They then estimated the biodegradation rate constant (by one of the field approaches) for each realization and compared it to the reference value. In their works, they also assumed that the source zone maintains a constant concentration (fixed boundary condition) throughout the process.

Almost all of the field estimation techniques (reviewed in the next paragraphs) for first-order biodegradation rate constant assume a constant source concentration. In reality, soil and aqueous concentration in the source zone area continuously change due to dissolution of NAPL into groundwater and transport of contaminants downstream of the source. The concentrations in the source zone area (and the downstream concentrations) depend on factors such as the size of the source zone, groundwater flow velocity within the source and dissolution rate of NAPL into groundwater. Thus, a more realistic approach for characterization of uncertainty in the field measured first-order rate constant must account for uncertainties in the dissolution rate constant as well as groundwater flow velocity within the source zone.

Another issue associated with most of the previous works is the fact that the transmissivity realizations used in the studies are not conditioned to head

observations and therefore may not reproduce the observed head measurements at observation locations. In fact, the work of Beyer et al. (2007) is the only related work that avoids the need for explicit conditioning to head data by simulating the heads and concentrations for a limited number of realizations (20 realizations) and designing a different layout for monitoring networks for each of the generated realizations based on their associated head and concentration fields. Although, their work was capable of showing the effect of the "human factor" on the level of uncertainty in the estimated first-order biodegradation rates, the number of realizations that was used in their work was small. At a real site, the centerline of the plume is determined by professionals by reviewing head and concentration values at the observation locations and the wells along the plume centerline are laid out mostly based on the contours of groundwater surface and professional judgment. Thus, the extremely large uncertainties in the values of estimated first-order biodegradation rate that are observed in the work of Bauer et al. (2006) is partially due to using realizations that are conditioned to point transmissivity measurements and unconditional to head data.

Section 7.1 reviews three most commonly used centerline techniques for field estimation of first-order biodegradation rate constant. Section 7.2 investigates the uncertainties associated with the estimated rates in 1D and 2D synthetic examples. The synthetic examples investigate the uncertainty associated with concentration of dissolved contaminants in the source zone as well as the uncertainties associated with transmissivity and flow fields and their impacts on the estimated first-order biodegradation rate.

7.1 Field estimation of biodegradation rate constant

There are several field approaches to determine the site-specific biodegradation rate coefficients. These methods include mass balances, the concentration-distance approaches such as the technique of Buscheck and Alcantar (1995), normalization of contaminant concentrations to that of a recalcitrant contaminant present in the initial release and the use of in-situ microcosms.

The method of Buscheck and Alcantar is based on an analytical solution to one-dimensional, steady-state contaminant transport involving advection, longitudinal dispersion, sorption, and first-order biodegradation:

$$C(x) = C_0 \exp\left\{\left(\frac{x}{2\alpha_L}\right)\left[1-\left(1+\frac{4\lambda\alpha_L}{v_c}\right)^{-1/2}\right]\right\} \qquad [7.1]$$

where, $C(x)$ is concentration at a distance x downstream of the source, C_0 is the point source concentration, λ is first-order biodegradation rate, α_L is longitudinal dispersivity and v_c is the contaminant velocity. Buscheck and Alcantar (1995) identified a trend according to which the contaminant concentration decreases. This trend is recognized to change exponentially along the plume centerline as a function of the distance from the source and it can be described by:

$$C(x) = C_0 \exp\left(-k\frac{x}{v}\right) \qquad [7.2]$$

where, k is the lumped decay coefficient incorporating dilution, sorption, biodegradation, etc; and v is the groundwater velocity. Combining Equations [7.1] and [7.2], Buscheck and Alcantar (1995) derived an equation for estimation of first-order biodegradation rate constant:

$$\lambda = -\left(\frac{v_c}{4\alpha_L}\right)\left\{\left[1 - 2\alpha_L\left(\frac{\ln\left(\frac{C(x)}{C_0}\right)}{\Delta x}\right)\right]^2 - 1\right\} \qquad [7.3]$$

where, the term $\ln(C(x)/C_0)/\Delta x$ is the slope of a line obtained from a log-linear plot of the concentration versus distance along the centerline of the plume, which is calculated for any number of concentration measurements by linear regression.

Equation [7.3] shows that any estimate of λ by this method is directly affected by uncertainties in the values of longitudinal dispersivity and the average flow velocity. Also, Equation [7.3] has been derived under the assumption of existence of a point source. Thus, a good estimate of the location of the point source is another important factor in the accuracy of the approach. As will be discussed in Section 7.2, when there is uncertainty in the distribution of hydraulic conductivity, the estimated rate constants may also be affected by uncertainty in the value of dissolution rate constant. In this case, the dependency of the estimated rate constants to rate-limited dissolution of NAPL is indirect and manifests itself in slight variations in the slope of $\ln(C(x)/C_0)/\Delta x$ line. If sorption is negligible and the biodegradation occurs solely in the aqueous phase, v_c in Equations [7.1] and [7.3] can be replaced by groundwater seepage velocity, which can be determined based on Darcy's law.

Zhang and Heathcoat (2003) proposed a modified version of Buscheck and Alcantar approach which is based on the solution to Domenico's analytical solution to the two-dimensional transport equation including first-order decay (Domenico 1987) and accounts for finite source width, longitudinal and transverse dispersion. In the modified approach, the biodegradation rate constant is calculated by:

$$\lambda = -\left(\frac{v_c}{4\alpha_L}\right)\left\{\left[1 - 2\alpha_L\left(\frac{\ln\left(\frac{C(x)}{C_0\beta}\right)}{\Delta x}\right)\right]^2 - 1\right\} \quad \text{with} \quad \beta = erf\left(\frac{W_S}{4\sqrt{\alpha_T \Delta x}}\right) \qquad [7.4]$$

where, W_S is the source width perpendicular to the average flow direction, Δx is the average distance between the observation wells and α_T is transverse dispersivity.

Another popular technique is the method of normalization to a recalcitrant co-contaminant proposed by Wiedemeier et al. (1996) and is calculated based on one-dimensional transport equation with a modification. In 1D, the steady state solution for the concentration profile (based on advection and first-order degradation only) can be rearranged to yield the first-order degradation rate constant as:

$$\lambda = -\left(\frac{v_c}{\Delta x}\right)\ln\left(\frac{C(x)}{C_0}\right) \qquad [7.5]$$

As pointed out by Newell et al. (2002), Equation [7.5] yields the attenuation rate rather than degradation rate constant as it does not account for dispersion, explicitly. Wiedemeier et al. (1996) proposed a modified approach based on Equation [7.5] with a correction factor based on the ratio of upgradient concentration to the downgradient concentration of a tracer at the same wells. This modification accounts for dispersion of the plume and ensures the calculated biodegradation is not affected by off-centerline measurements. The modified form of Equation [7.5] is calculated by:

$$\lambda = -\left(\frac{v_c}{\Delta x}\right)\ln\left(\left(\frac{C(x)}{C_0}\right)/\left(\frac{\hat{C}(x)}{\hat{C}_0}\right)\right) \qquad [7.6]$$

where, $\hat{C}_0$ and $\hat{C}(x)$ are the source and downgradient concentration of the tracer.

The method of Buscheck and Alcantar based on its original 1D presentation (Buscheck and Alcantar 1995) is widely used in practice and therefore used in this chapter. In Section 7.2, the effect of uncertainty in the dissolution rate of NAPL and different magnitudes of error in observed head on the estimated first-order biodegradation rate constant is investigated. In the 1D example, the method of Buscheck and Alcantar (1995) is used to estimate the first-order rate constant under homogenous and heterogeneous conditions with a constant source concentration and variable rate-limited dissolution. In the 2D example, the same method is used with three different levels of heterogeneity in the aquifer transmissivity, two different magnitude of error in hydraulic head measurements, and with and without uncertainty in the dissolution rate constant.

7.2 Investigation scenario

7.2.1. Uncertainty in the source concentration: 1D case

As pointed out in the previous section, the method of Buscheck and Alcantar is developed based on the analytical solution to a 1D transport problem. In order to investigate the effects of uncertainties in the concentration of the source on the estimated rate constants, the hydraulic conductivity field and the dissolution rate constant on the estimated first-order biodegradation rate, two 1D contaminant transport problems with homogenous and heterogeneous hydraulic conductivity fields are created. Figure 7-1 shows the distributions of hydraulic conductivity and the associated concentration profiles under semi-steady-state conditions. The length of the 1D field is 300 m, and a hydraulic gradient equal to 0.01 is applied to the system through fixed-head boundary conditions to simulate the hydraulic head field.

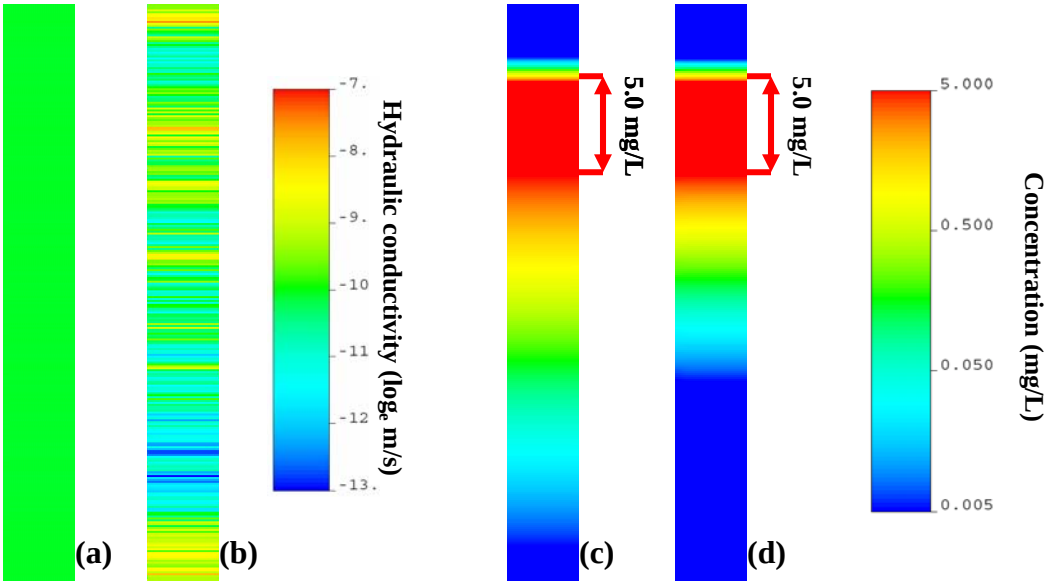


Figure 7-1: The hydraulic conductivity distribution for the 1D (a) homogenous and (b) heterogeneous cases and the associated concentration profiles under semi-steady-state condition for (c) homogenous and (d) heterogeneous cases.

In the 1D example, the homogenous field has a uniform hydraulic conductivity of 4.1×10^{-5} m/s and the heterogeneous case has a log-normal hydraulic conductivity distribution with the mean of 4.1×10^{-5} m/s, standard deviation of 1.2 in natural logarithmic units and a spatial correlation range of 32 m. To simulate the contaminant concentration profile, a 50 m long line source with constant concentration equal to 5 mg/L is considered for both homogenous and heterogeneous cases (Figure 7-1). The development of the 1D plume is simulated for 7 years, when it becomes under semi-steady-state condition. The reference values for longitudinal dispersivity and biodegradation rate constant for both the homogenous and heterogeneous cases are 1.5 m and 0.0044 day^{-1}, respectively. The reference cases (black curves) shown in Figures 7-2, 7-3 and 7-4 are the reference concentration profiles for the homogenous and heterogeneous cases under semi-steady-state condition. The other concentration profiles (realizations 1 to 4) shown in Figure 7-2 are the first four realizations out of a total of 100 generated realizations that are associated with different values of dissolution rate constant. Realizations 1 to 4 shown in Figure 7-3 are the first four realizations corresponding to different hydraulic conductivity distributions. The reference case shown in Figure 7-4 is the same as the reference case shown in Figure 7-3, and the realizations in Figure 7-4 are calculated for uncertain dissolution rate constant and uncertain hydraulic conductivity field. The uncertainty in the value of dissolution rate constant is simulated by assuming a log-normal distribution with a mean of 0.001 day^{-1} and standard deviation of 1.0 in natural logarithmic units.

As stated above, good estimates of the values of seepage velocity, longitudinal dispersivity and the location of the point source are necessary for application of the method of Buscheck and Alcantar. In the 1D example presented in this section, the true values of seepage velocity and dispersivity are assumed to be known for the homogenous case. For the heterogeneous case however, these values must be estimated. An estimate for seepage velocity is obtained by

computing the effective hydraulic conductivity and known hydraulic gradient (equal to 0.01 for all cases). The effective hydraulic conductivity is calculated as the harmonic average of the reference hydraulic conductivity distribution. The reference hydraulic conductivity distribution is unknown in practice. However, it is considered known in this work to find an effective hydraulic conductivity value that is reasonable for all unconditional realizations. As a result, the estimated values for the effective hydraulic conductivity and the average seepage velocity for the heterogeneous case are 2.1×10^{-5} m/s and 6.05×10^{-2} m/day. The location of source is another important factor, when estimating the first-order rate by any center-line approach. It is however, prone to uncertainty. In this work, two different locations for the source are considered: the middle of the line source in Figure 7-1 and the downstream edge of the line source in Figure 7-1.

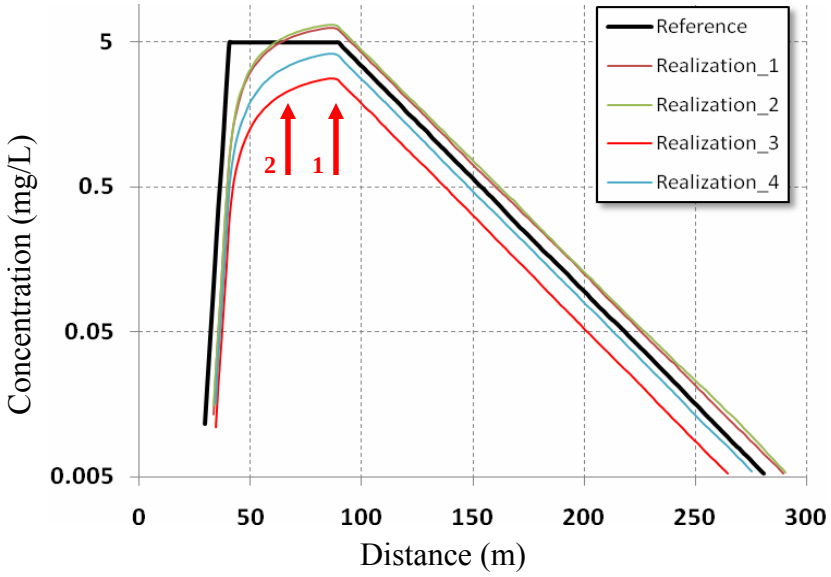


Figure 7-2: The concentration profile for the reference homogenous case, as well as concentration profiles for four realizations with homogenous hydraulic conductivity and variable dissolution rate constants. The arrows show the locations of the two point sources used in the subsequent estimation of first-order biodegradation rate constants.

In Figure 7-2, it can be observed that changes in the dissolution rate constant has minor impact on the slope of $\ln(C(x)/C_0)/\Delta x$ line in the homogenous case, and therefore does not affect the estimation of first-order rate constant by any of the centerline approaches. This observation is also depicted in Table 7-1, where the estimated (mean) values of first-order rate constant, the standard deviation and minimum and maximum values of the estimated rate constants for the ensemble of 100 realizations are presented for all different scenarios. Table 7-1 shows that the method of Buscheck and Alcantar can obtain an excellent estimate of the first-order rate constant for the homogenous case with a known and/or uncertain value of the dissolution rate constant, and with the point source considered to be at the edge of the line source. As expected, considering the point source to be at the forward edge of the source may result in slight underestimation of the true rate constant.

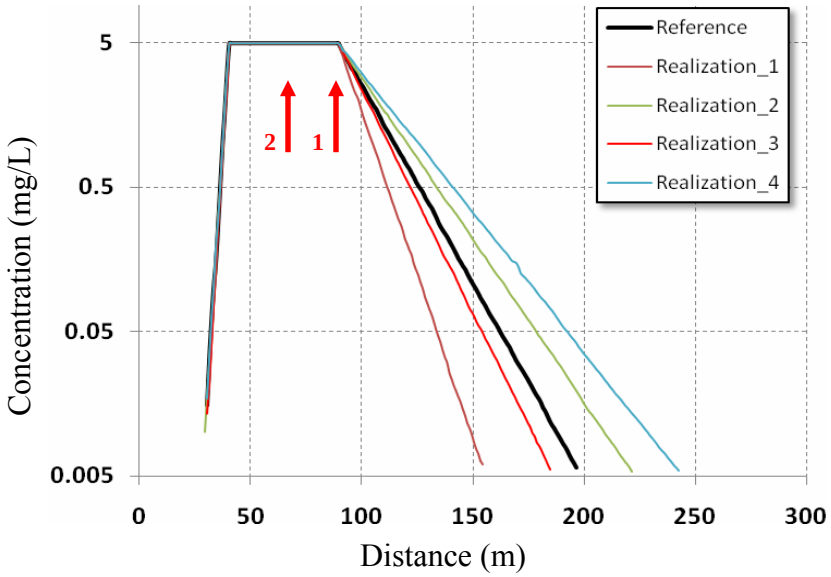


Figure 7-3: The concentration profile for the reference heterogeneous case, as well as concentration profiles for four realizations with heterogeneous hydraulic conductivity and constant source concentration. The arrows show the locations of the two point sources used in the subsequent estimation of first-order biodegradation rate constants.

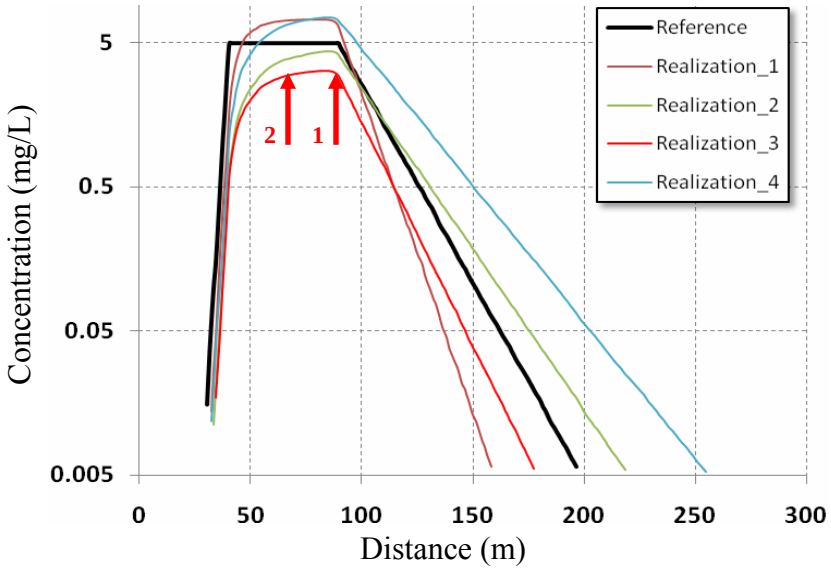


Figure 7-4: The concentration profile for the reference heterogeneous case, as well as concentration profiles for four realizations with heterogeneous hydraulic conductivity and variable dissolution rate constants. The arrows show the locations of the two point sources used in the subsequent estimation of first-order biodegradation rate constants.

Figure 7-3 shows that the uncertainty in the hydraulic conductivity field may result in a large amount of variability in the slope of the $\ln(C(x)/C_0)/\Delta x$ line that translates into large uncertainties in the value of the estimated first-order rate (Table 7-1). This is an extreme case, as no conditioning data are used to generate the hydraulic conductivity fields. As the number of conditioning data increases, there will be a reduction in the uncertainty of the estimated rate constants. Nevertheless, this observation quantifies how heterogeneity in the hydraulic conductivity field translates into uncertainty in and over/underestimation of the estimated rate constants. This is due to inherent uncertainty in the estimation of effective hydraulic conductivity for the heterogeneous aquifers; and it is other than the notion of 'missing the centerline of the plume' that has been illustrated in the previous works in the literature (Wilson et al. 2004).

In Figure 7-4, the concentration profiles of the realizations with uncertain hydraulic conductivity and uncertain dissolution rate constant are illustrated. Unlike the homogenous case, uncertainty in the dissolution rate increases the uncertainty in the estimated first-order rate constants. According to Table 7-1, by introducing variability in the dissolution rate, the standard deviation of the ensemble of the estimated first-order rates increases from 0.242 to 0.271 which can be explained by variability in the pattern of source dissolution in heterogeneous hydraulic conductivity fields. Heterogeneity in the hydraulic conductivity, on average, results in overestimation of the true rate constant.

Hydraulic Conductivity	Source Location	Source Concentration	Estimated first-order degradation rate			
			Mean	S.D.	Min	Max
Homogenous	Edge	Known-fixed	4.46×10^{-3}	-	-	-
Homogenous	Center	Known-fixed	3.93×10^{-3}	-	-	-
Homogenous	Edge	Uncertain	4.41×10^{-3}	0.013	4.13×10^{-3}	4.48×10^{-3}
Homogenous	Center	Uncertain	3.76×10^{-3}	0.014	3.50×10^{-3}	3.80×10^{-3}
Heterogeneous	Edge	Known-fixed	6.02×10^{-3}	0.242	1.80×10^{-3}	6.85×10^{-3}
Heterogeneous	Center	Known-fixed	5.13×10^{-3}	0.245	1.55×10^{-3}	5.89×10^{-3}
Heterogeneous	Edge	Uncertain	5.55×10^{-3}	0.271	1.82×10^{-3}	7.91×10^{-3}
Heterogeneous	Center	Uncertain	4.74×10^{-3}	0.285	1.42×10^{-3}	6.81×10^{-3}

Table 7-1: The estimated (mean) values of the first-order biodegradation rate constant as well as the standard deviation (in natural logarithmic units) and minimum and maximum of the rate constants estimated for different scenarios.

7.2.2. Uncertainty in the source concentration: 2D case

The method of Buscheck and Alcantar (1995) has been developed based the solution to 1D mass transport problem. In practice, however, it is applied to real 2D and 3D transport problems. As reviewed in Chapter 2 and at the beginning of this chapter, a few works reported in the literature deal with uncertainties associated with application of the method of Buscheck and Alcantar to 2D transport problems. A synthetic reactive transport problem is created in this work to investigate the impact of head measurement errors and uncertainty in the dissolution rate constant on the estimation of the first-order rate by the method of Buscheck and Alcantar for the heterogeneous aquifers.

The synthetic aquifer developed for this work consists of a 2D heterogeneous aquifer with two different levels of heterogeneity and a NAPL source zone with known geometry located at the North of the site (Figure 7-5). The flow and transport boundary conditions are all assumed to be known and include no-flow boundary conditions at the East and West of the site, fixed head boundaries at the north and south of the site and fixed dispersive flux boundary conditions at the north and south of the site. As a result of the boundary conditions used, a hydraulic gradient equal to 0.005 is applied to the site. The distribution of hydraulic conductivity follows a log-normal distribution with a mean of 4.1×10^{-5} m/s and standard deviations of 1.0, 1.5 and 2.0 (in natural logarithmic units) that represent three different levels of heterogeneity in the synthetic aquifer. The range of spatial correlation structure is defined by a spherical variogram with a nugget effect equal to 0.1 and a range of 25 m. As shown in Figure 6-5, a 30 m by 30 m NAPL source zone is considered to be located at the north of the site. The reference value for the dissolution rate constant is equal to 1.0×10^{-3} day^{-1}. In order to investigate the effect of uncertainty in the dissolution rate constant on the estimated first order rates, two different scenarios are considered: (1) a constant concentration boundary condition equal to 5 mg/L over the source zone, and (2) a

source zone with soil NAPL concentration equal to 10g/Kg, the initial mass fraction of the substrate (e.g. BTEX) equal to 0.01, and a variable dissolution rate constant that follows a log-normal distribution with a mean of 0.001 day^{-1} and standard deviation of 1.0 in natural logarithmic units. Figure 7-6 shows the distribution of hydraulic head for the three different levels of heterogeneity in the hydraulic conductivity fields as well as the layout of the centerline and off-centerline wells for each case. In practice, a representative design for the layout of the centerline wells is often obtained based on the location of the source zone and the contour lines of the hydraulic head distribution that is determined using piezometric heads observed at a number of wells in the appraisal stage. Therefore, it is expected that error in the head observations affects the locations of the centerline wells and may result in missing the centerline of the plume.

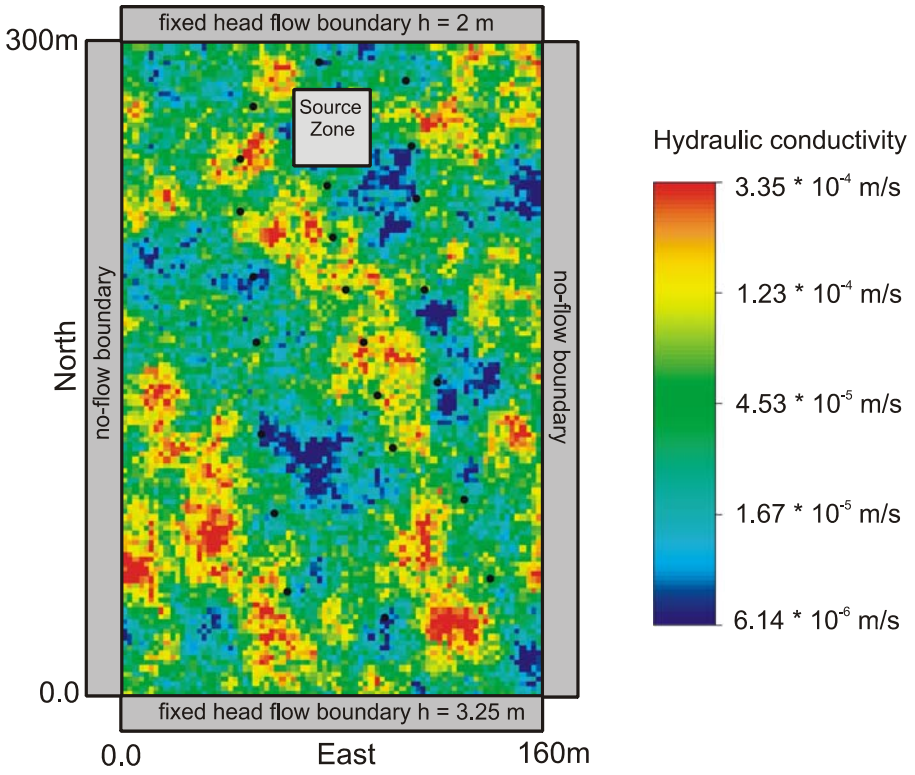


Figure 7-5: The reference hydraulic conductivity field (the case with σ_{lnK} = 1.0), the location of the source zone and the types and values of the flow boundary conditions in the synthetic reactive contaminant transport problem.

To investigate the potential uncertainty in the estimated rate constants, 100 realizations of the hydraulic conductivity field conditioned to both hydraulic conductivity and head measurements are generated with different levels of heterogeneity in hydraulic conductivity (σ_{lnK} =1.0, 1.5 and 2.0) and two different levels of error in head measurements (σ_H = 0.10 m and 0.15 m). Given the fact that there are two different scenarios for the source zone concentration (constant aqueous concentration and constant soil NAPL concentration with variable rate limited dissolution), a total of 12 scenarios are created (3 different levels of heterogeneity, 2 different levels of errors in the head measurements and 2 different scenarios for source aqueous concentration). Figure 7-7 shows the plumes of the dissolved contaminants for the cases of σ_{lnK} = 1.0, σ_{lnK} = 1.5 and σ_{lnK} = 1.5. The plumes are simulated for 7 years and all of are under semi-steady-state condition. It should be noted that the layout of the centerline wells has been designed separately for each specific case.

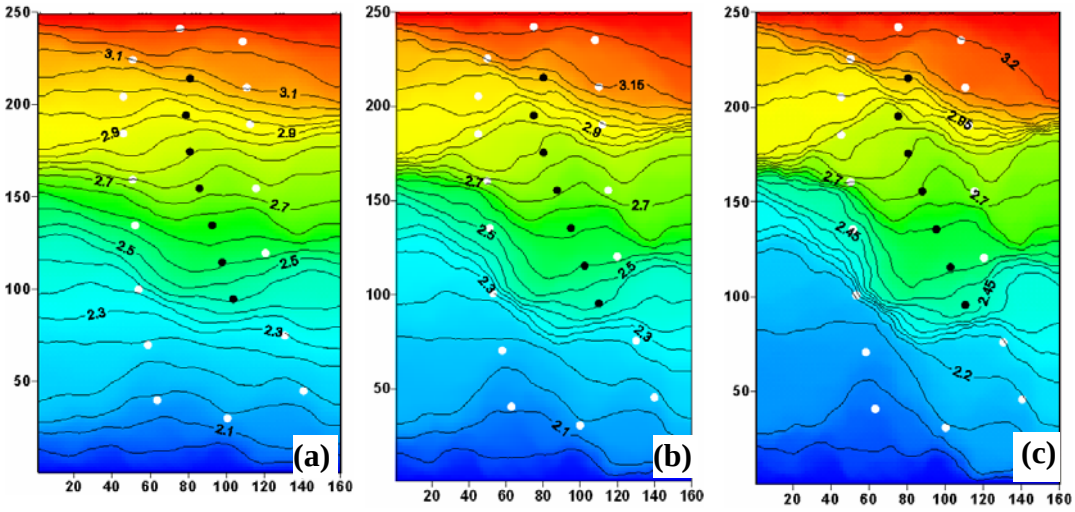


Figure 7-6: The hydraulic head contours for the reference synthetic aquifers with log-normal hydraulic conductivity distributions with a mean equal to 4.1 × 10^{-5} m/s and standard deviations equal to (a) σ_{lnK} =1.0, (b) σ_{lnK} =1.5 and (c) σ_{lnK} =2.0 in natural logarithmic units. The centerline wells are depicted by black circles.

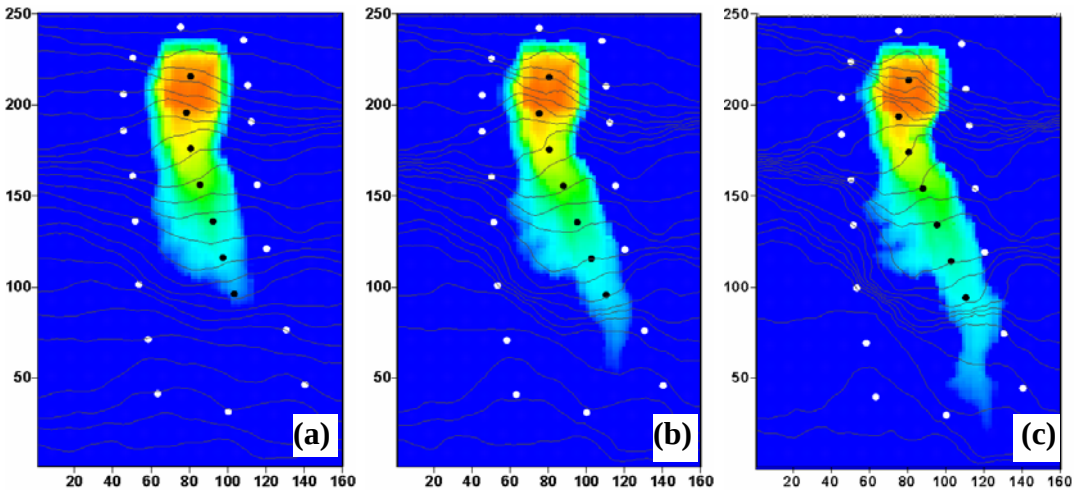


Figure 7-7: The development of the dissolved plumes after 7 years, when reaching to a semi-steady-state condition for the reference synthetic aquifers with log-normal hydraulic conductivity distributions with a mean equal to 4.1×10^{-5} m/s and standard deviations equal to (a) σ_{lnK} =1.0, (b) σ_{lnK} =1.5 and (c) σ_{lnK} =2.0 in natural logarithmic units.

In practice, the contours of hydraulic heads are obtained by professionals either through hand contouring or by calibration of a numerical groundwater model to site-specific conditions. The calibration of numerical models to site-specific conditions is implemented by trial and error calibration or by automated approaches. The quality of calibration to the available head observations is always affected by the level of details that is incorporated in the model through parameterization as well as the calibration techniques used. In any case, as the true distribution of hydraulic conductivity is inaccessible, the level of precision and details in the head contour lines as shown in Figures 7-6 and 7-7 can never be achieved in practice. The layout of the centerline wells are, in turn, affected by errors in the head contours which translates into uncertainty in the first-order rate constants estimated by any centerline approach.

Generally speaking, the magnitude of errors in the hydraulic head surface obtained by hand-contouring or simulated by a calibrated groundwater surface is

affected by a number of factors: (1) errors in piezometric head observations, (2) the number of piezometric head observations and (3) the quality of calibration to piezometric head observations, which is a function of the technique used and the level of parameterization details considered. In this work, only the effects of error in piezometric head observations are investigated; and it is assumed that the number of piezometric head data is adequate and calibration to the set of available piezometric head observations is consistent with the level error in observations. The erroneous head observation datasets that are used in this work are created by sampling hydraulic heads from the reference cases (Figure 7-7) and adding Gaussian noise. Table 7-2 shows the scenarios that are simulated in this work to evaluate the effects of error in head observations and source concentration. For each scenario, 100 realizations with measure of fit (*s*) close to one are created and used in the subsequent Monte Carlo simulations. For each realization, the development of the plume is numerically simulated and the first-order biodegradation rate constant is estimated through the method of Buscheck and Alcantar and by using the concentration measurements at the centerline wells.

Scenario	σ_{lnK}	σ_H	Source concentration	Scenario	σ_{lnK}	σ_H	Source concentration
1	1.0	0.10	C_0=5 mg/L	7	1.5	0.15	C_0=5 mg/L
2	1.0	0.10	$\sigma_{k_{dis}}$ = 1.0	8	1.5	0.15	$\sigma_{k_{dis}}$ = 1.0
3	1.0	0.15	C_0=5 mg/L	9	2.0	0.10	C_0=5 mg/L
4	1.0	0.15	$\sigma_{k_{dis}}$ = 1.0	10	2.0	0.10	$\sigma_{k_{dis}}$ = 1.0
5	1.5	0.10	C_0=5 mg/L	11	2.0	0.15	C_0=5 mg/L
6	1.5	0.10	$\sigma_{k_{dis}}$ = 1.0	12	2.0	0.15	$\sigma_{k_{dis}}$ = 1.0

Table 7-2: The description of different scenarios that are studied to investigate the effects of head observation and source concentration errors on the estimated values of first-order biodegradation rate

As stated above, to apply the method of Buscheck and Alcantar (1995), one must estimate the values of the seepage velocity and field-scale longitudinal dispersivity. Following a common practice in contaminant hydrogeology, the

value of longitudinal dispersivity is set to be 0.1 times the observed plume length that is calculated for each level of heterogeneity (because the plume lengths are different for different values of σ_{lnK}). The value of the average seepage velocity is calculated as:

$$v_c = \frac{K_G}{\theta}\frac{\Delta h}{\Delta l} \qquad [7.7]$$

where, K_G is the geometric average of hydraulic conductivity values sampled at the centerline wells, θ is effective porosity and $\Delta h/\Delta l$ is hydraulic gradient. In this work, it is assumed that no uncertainty exists in the values of porosity and the value of hydraulic gradient.

Table 7.3 shows the geometric average, standard deviation, p25, p50, p75 quartiles and the width of 50% confidence interval for the normalized estimated first-order biodegradation rate constants (normalized to the reference value used in the simulations), based on the results of Monte Carlo simulations.

	Normalized first-order rate constant					
Scenario	**Average**	**Standard deviation**	**p_{25}**	**p_{50}**	**p_{75}**	**width of 50% C.I.**
1	2.30	0.471	1.73	2.09	2.45	0.72
2	2.61	0.560	1.73	2.14	4.57	2.84
3	2.46	0.499	1.82	2.17	2.6	0.78
4	3.12	0.551	1.81	2.28	4.79	2.98
5	3.53	0.687	2.44	2.95	3.71	1.27
6	4.91	0.712	2.41	3.05	7.29	4.88
7	3.81	0.630	2.51	3.4	5.68	3.17
8	4.3	0.687	2.55	3.58	7.91	5.36
9	4.42	0.686	2.68	3.92	5.57	2.89
10	4.72	0.719	2.64	4.01	8.42	5.78
11	4.64	0.693	2.76	4.16	6	3.24
12	5.11	0.735	2.84	4.35	10.28	7.44

Table 7-3: Average, standard deviation, p_{25}, p_{50}, p_{75} quartiles and the width of 50% confidence interval of the distribution of normalized estimated first-order biodegradation rate constants for different scenarios. The values of the first-order rates are normalized to the reference value (0.0048 day^{-1}).

In Table 7-3, first, it is observed that application of the method of Buscheck and Alcantar (1995) for heterogeneous aquifers often results in significant over-estimation of the true rate constant, which is a result of missing the centerline of the plume and uncertainty in the values of average seepage velocity and longitudinal dispersivity. It is also observed that the average over-estimation of the true rate constant as well as the uncertainty about the true value increase as the level of heterogeneity increases in the aquifer.

Similar to the 1D case, incorporating variability in the dissolution rate constant results in more uncertainty (larger standard deviation and considerably larger width of the 50% confidence interval) in the estimated first-order rate constants. This is due to variability in the dissolution pattern of the NAPL source zone under uncertain and heterogeneous distribution of hydraulic conductivity. Although the effects of variability in the dissolution rate constant is smaller for moderately heterogeneous aquifers, they become more significant for highly heterogeneous aquifers. This is important to note that the average overestimation also becomes larger when incorporating the variability in the source concentration. The range of correlation structure may also have an impact on the uncertainty of the estimated rate constant due to its impact on the dissolution pattern of the NAPL source. This is not, however, investigated in this thesis.

Another important observation is related to the effects of uncertainty and error in head observations. It is evident that the error in head observations has an impact on the degree of over-estimation and the level of uncertainty in the estimated rates. In all cases, one can observe that (50% confidence intervals in Table 7-3) introducing Gaussian noise to the head observation data and/or variability in the source concentration may have a similar impact on the estimated first-order rates to that of additional heterogeneity in the aquifer material. This shows the importance of calibrating the hydraulic conductivity realizations to good-quality

head data that can reduce the uncertainty in the estimated rate constants. Also, comparing the results in Table 7-3 to the results obtained by Bauer et al. (2006) shows that larger uncertainties (on average) reported are due to application of hydraulic conductivity realizations that are not conditioned to head data.

CHAPTER 8

CONCLUSIONS AND FUTURE RESEARCH

From a practical point of view, risk management for natural attenuation of petroleum hydrocarbon contaminants requires a sound understanding of the variability in the factors that control the influx of contaminants into the aquifer and the factors that control the destructive removal of contaminants from the aquifer. Estimation of these parameters using concentration measurements at sparse sampling locations involves uncertainties that are overlooked by the available field estimation techniques. Linking the estimation of these parameters to distributions of uncertainty in the source geometry and hydraulic conductivity can effectively characterize the uncertainty in these parameters and reduce the uncertainties in the state variables, being plume length and dissolved mass.

The central idea in this thesis has been to tailor the estimation of dissolution rate and first-order biodegradation rate to joint realizations of source size and hydraulic conductivity and to develop a simple modeling approach that can be used as a screening tool for characterization of uncertainty (non-linear confidence intervals) in these parameters and the associated state variables.

8.1 Summary and Conclusions

In Chapter 2 of this thesis, natural attenuation mechanisms, principles and modeling approaches were reviewed. Then, a thorough literature review on stochastic modeling of natural attenuation, inverse modeling and parameter estimation was presented. In the last section of Chapter 2, the parameterization of natural attenuation of PHCs, as used in this thesis, was presented.

In the first part of Chapter 3, a case study was presented where CPT data and soil behavior type charts were used to define categorical and continuous variables to construct a probabilistic model for hydrogeological properties of a real aquifer. A hydrogeological conceptual model was then developed based on this probabilistic model. The definition of hydro-facies based on CPT data may potentially be used in Monte Carlo simulations and/or calibration of groundwater flow models. The definition of continuous variables based on CPT data was subsequently used in calibration and prediction of localized distribution of residual NAPL contamination given soil texture as a secondary data source.

In the second part of Chapter 3, the concepts of distance-function and uncertainty band were introduced. Then, the uncertainty band was calibrated against a large number of synthetic plumes using a downhill simplex optimization. A close match between the calibrated probabilities and the target probabilities showed the convergence of the technique. The proposed distance function approach is deemed to be useful to model the areal limits of non-stationary continuous or categorical random variables and to provide a prior probability map for source zone sizes that can be used in Monte Carlo simulations of contaminant transport or subsequent inverse modeling studies.

In the third part of Chapter 3, data integration approaches were employed to create a probabilistic model for localized distribution of residual NAPL within the areal limits of a source zone for a real aquifer. Secondary data sources such as the distance to water table and soil texture were defined using calibration tables and combined based on the assumptions of full data independence and conditional independence. Two different data integration techniques, namely Bayesian updating and sequential indicator simulation with locally varying means were employed to integrate primary hard data and the secondary data sources. A cross-

validation study was then implemented to compare the performance of different techniques used. According to the results of cross-validation, inclusion of secondary data sources improved the prediction of both contaminated and uncontaminated locations. Improvements due to incorporation of distance to water elevation data seemed to be more significant compared to soil texture data. The results obtained based on the assumptions of full data independence and conditional independence appeared to provide almost identical results confirming that the two data sources are fully independent. After incorporation of hard data, it seemed that Bayesian updating approach over-estimated the contaminated locations (with smaller global proportions) and under-estimated the uncontaminated locations. The degree of overestimation of contaminated locations was smaller using SIS-LVM approach, as it improved the prediction of uncontaminated locations over global proportions as well. As a general conclusion, it should be noted that although the integration of secondary data sources showed an overall improvement in predictive ability, the distance to groundwater can be considered a more robust secondary data source as it is less dependent on the number of wells and their spatial arrangement. In fact, the improvements achieved by the inclusion of soil texture data in the analysis represent an upper bound to what can be achieved by the incorporation of soil data, since all the collected CPT data are used in construction of 3D map for conditional probabilities. In any case, inclusion of soil texture data can still be considered useful due to the fact that more reliable estimates of correlation structure (variogram) of soil texture data are often available comparing to presence/absence of contamination.

Chapter 4 of this thesis presented the development and verification of simple and flexible 2D numerical flow and mass transport simulators that were subsequently used in the parameter estimation step. The 2D groundwater flow simulator was based on the finite volume approach, the particle tracking code was developed

based on 2D semi-analytical particle tracking, and the mass transport simulator was developed based on the method of characteristics. Rate-limited NAPL dissolution, advection, dispersion and first-order biodegradation were coded in the mass transport simulator. All the flow and mass transport codes were verified against commercial software under homogeneous and heterogeneous conditions.

Chapter 5 presented simple Monte Carlo type decoupled inverse modeling for characterization of uncertainty (non-linear confidence interval) in dissolution rate and first-order biodegradation rate constants. The justification for the decoupled approach was due to avoiding the computational cost of calculation of sensitivity coefficients in a multi-state system (while the significance of incremental value of such coupling was unknown) and avoiding potential instability problems associated with simultaneous perturbations of hydraulic conductivity field, dissolution rate and first-order biodegradation rate. As the first step of the decoupled approach, the sequential-self calibration technique was reviewed and a case study was presented to compare the conditional realizations in terms of mismatch to head observations with measurement errors. As the second step of the decoupled approach, a gradient-based inverse modeling was implemented to apply concentration measurements for simultaneous estimation of dissolution rate and first-order biodegradation rate. Calculation of sensitivity coefficients and optimization of the objective function were based on sensitivity equations and the modified Gauss-Newton approach, respectively. In a sensitivity analysis, it was observed that the length of the plume is primarily affected by the value of first-order biodegradation rate constant and the level of heterogeneity in the aquifer material. Size of the source and the dissolution rate of NAPL had smaller impacts on the length of the plume; however, this depends on the value of dissolution rate. Similarly it was observed that the sensitivity of the mass loaded into the aquifer to dissolution rate depends on the value of dissolution rate as well as heterogeneity of the aquifer material. The biodegradation rate constant and source geometry

showed profound impact on the loaded mass but this sensitivity was not affected by the heterogeneity of the aquifer material. The stability of the inverse problem was also studied in Chapter 5 through 1D and 2D examples. According to these examples, for large values of dissolution rate, the correlation between the two parameters (dissolution rate and first-order biodegradation rate) becomes smaller and parameter non-uniqueness becomes less of an issue. For this case, however, parameter insensitivity may become the problem. For smaller values of dissolution rate that are more feasible for real field applications (Essaid et al. 2003 and Christ et al. 2006), a very high correlation exists between the two sensitivities. However, it was observed that correlation between the two parameters depends on the location of the calculated sensitivities. This can be used to improve the stability of the inverse problem through designing the monitoring network in such a way that observation network samples the locations close to the source zone as well as the locations well away from the source, while the weighting scheme in the objective function has to be defined inverse proportional to simulated concentrations to preserve the importance of the concentrations at the downstream as well as upstream edge of the plume. This observation was also validated through plotting the objective function surfaces.

Chapter 6 presented a synthetic example to investigate the performance of the decoupled inverse modeling approach in characterizing the uncertainty in the dissolution rate and first-order biodegradation rate and reducing the uncertainty in the associated state variables. Three reference cases with three source sizes were considered. First, a set of Monte Carlo simulations were implemented whose results were subsequently compared to the results of the inverse modeling methodology. The Monte Carlo simulations were performed using realizations of hydraulic conductivity conditioned to head data, realizations of source geometry drawn from the calibrated uncertainty band, first-order biodegradation rate values drawn from a log-normal distribution similar to what was observed by Bauer et al.

(2006) using the approach of normalization to a recalcitrant co-contaminant, and dissolution rate constant values drawn from a uniform distribution with one-order of magnitude variability that seems to be a lower bound for the variability in this parameter based on the observations of Dillard et al. (2001) and Christ et al. (2006) for field applications of pore network models and upscaled models for NAPL dissolution. Comparing the results of the Monte Carlo simulations to the results of inverse modeling for each reference case, it was observed that tailoring the estimation of first-order biodegradation rate and dissolution rate to distributions of uncertainty in the source geometry and hydraulic conductivity field results in characterization of uncertainty in these parameters and significant reduction of uncertainty in the state variables, being mass loaded into the aquifer and the dimensions of the plume. Although the reference values always fell within 90% confidence interval, a bias was observed in the prediction of the reference state variables by the ensemble of simulated realizations. This bias was deemed to be partially due to large variabilities in the size of the source zone which cannot be fully resolved by adjusting the values of dissolution rate and first-order biodegradation rate. In this work, it was observed that ranking and screening the conditional realizations based on the value of modified objective function (Chapter 5) can effectively reduce the uncertainties in the size of the source zone and uncertainty and bias in the prediction of the state variables. The value of the modified objective function is deemed to be independent of overall level of concentrations in the modeling domain due to the fact that it is a dimension-less number and a normalization that takes place by defining the weights inverse proportional to the value of simulated concentrations. On the other hand, the number of realizations that are kept after ranking and screening may be considered ‘problem dependent’ for a large part. Despite this apparent subjectivity, it was observed through a sensitivity analysis that ranking the realizations and keeping any number of realizations can still be useful as it can give a general idea about the size of the source, while reducing its uncertainty.

The importance of simultaneous characterization of uncertainty in the parameters was investigated through cross-plots of the parameters, where it was observed that a positive correlation exists between the values of dissolution rate and first-order biodegradation rate, a negative correlation exists between the values of dissolution rate and source size quantile and little correlation exists between first-order biodegradation rate and source size quantile. To investigate the effects of errors in observation data on the modeling outcomes, relatively large levels of observation errors (Gaussian noise with a mean of zero and pre-specified standard deviation/coefficient of variation) were added to head and concentration observations and the transport parameters were estimated. According to the results, existence of Gaussian noise in the data resulted in an increase in the uncertainty and bias of the estimated parameters and the predicted state variables. Comparing these results to the results of Monte Carlo simulations indicated that even if the observed data are subject to a relatively large level of Gaussian noise, the uncertainty in the predicted state variables are still smaller than the results of Monte Carlo simulations. Finally, to investigate the effect of misestimation of starting time of simulations on the modeling outcomes, a three year lag was considered in the starting time of the simulations, while it was assumed that the data have been collected under steady-state condition. As it was expected, misestimation of the time of release had negligible impact on the modeling outcomes.

In Chapter 7, a series of simple 1D and 2D Monte Carlo simulations were performed to investigate the effects of uncertainty in the source concentration and measurement error in head data on the estimates of first-order biodegradation rate constant obtained by a widely used concentration-distance relationship. The method of Buscheck and Alcantar (1995) has been developed based on an analytical solution to a one-dimensional transport equation. But, even in a 1D problem, the results show that this method may on average over-estimate the true

rate constant by 36% (with a standard deviation equal to 0.242) when the true source location and source concentration are known but the 1D domain is heterogeneous. This is mainly due to uncertainty in the value of groundwater velocity. Uncertainty in the value of source concentration can further increase the standard deviation (uncertainty) to 0.271. Similar to the 1D example, the objective of the 2D example was to show the effect of variability in source concentration on the uncertainty in the estimated first-order biodegradation rate constants. The effect of head observation error was also studied. For the purpose of this study, three levels of heterogeneity, two levels of head observation error and the effects of existence of uncertainty in source concentration were considered (a total of 12 scenarios). The results showed that (1) depending on the level of heterogeneity, on average, the method of Buscheck and Alcantar may overestimate the true rate constant between a factor of 2 to 5 with standard deviations ranging from 0.47 to 0.74 in natural logarithmic units; (2) introducing variability in the source concentration has a profound impact on the over-estimation of and the uncertainty in the estimated rate constants; (3) although error in head observations increase the uncertainty in the estimated rates, their impact seem to be smaller as compared to variability in the source concentration; and (4) comparing the results to the results obtained by Bauer et al. (2006) shows that larger uncertainties (on average) reported in their work are due to application of hydraulic conductivity realizations that are not conditioned to head data. The general conclusion for Chapter 7 is: application of concentration-distance relationships to heterogeneous aquifers may result in significant overestimation of the rate constants while this uncertainty is adversely affected by uncertainty in the value of source concentration; and this uncertainty/overestimation can be reduced by calibrating groundwater models to good quality head observation data in the appraisal stage.

8.1 Recommendations for Future Research

In the development and implementation of the decoupled inverse problem, it has been assumed that the biodegradation of petroleum hydrocarbons in the reference cases is subject to first-order decay kinetics. The rate of such first-order model and its uncertainty were then estimated using the decoupled approach. It is well-known that the use of first-order kinetics may be problematic in some situations as it can be a poor representation of processes occurring in contaminated aquifers (Bekins et al. 1998, Schafer et al. 2004). Thus, future research is needed to investigate (1) the performance of the simple model (based on first-order decay) presented in this thesis in prediction of the future behavior of the reference cases that are subject to more sophisticated biodegradation kinetics such as Michaelis-Menten (Simkins and Alexander, 1984) and Monod kinetics, and (2) the performance of the presented approach in estimation of the parameters associated with such more sophisticated models (Michaelis-Menten).

In this work, constant dissolution rate and constant first-order biodegradation rate were adopted due to the fact that the model was aimed to be a screening-level model for field-scale applications with small/average domain. These assumptions are valid for the sites with relatively smaller size. For larger sites (with large source areas), however, it is more appropriate to consider spatial (and/or temporal) variability in these parameters. As discussed in Chapter 5, joint estimation of these parameters is prone to instability problems. These instability problems are a major factor to be considered in parameterization of spatial variability. The other important factor is related to avoiding the over-parameterization of the model. In fact, the number of perturbation locations and correlation length for variability in these parameters (especially first-order biodegradation rate) should be carefully selected to avoid any over-parameterization. It should be noted that consideration of spatial variability in the first-order biodegradation rate constant through zonation can be also useful, as the

zonation can be designed based on approximate locations of zones associated with different electron acceptors for plumes under steady-state condition. One approach to select the number of zones (regardless of electron-acceptor distribution) is using residual analysis or other approaches such as Fisher's information matrix as used by Medina and Carrera (1996).

In Chapter 5 of this thesis inverse modeling was implemented to characterize the uncertainty in the dissolution rate and first-order biodegradation rate by combining sequential self-calibration approach, distance function approach and a gradient-based optimization technique. The motivation for decoupling the inverse modeling process was to (1) avoid the cumbersome task of calculating sensitivities in a multi-state system while the potential benefits of such coupling in parameter estimation are unknown; and (2) to avoid the adverse effects that this coupling may have on potentially unstable problem of simultaneous estimation of dissolution and first-order biodegradation rates, as observed by Essaid et al. (2003). The investigation of the potential value of coupling the estimation of flow and transport parameters in parameter uncertainty and prediction of the state variables is an area for future research. A closely related subject would be studying the potential instability problems (if any) of coupling the estimation of flow and transport parameters (dissolution rate and first-order biodegradation rate). The author expects that the stability of such coupled inverse problem is a function of (1) arrangement of monitoring locations, and (2) the correlation range of hydraulic conductivity field relative to the number of monitoring locations. It is expected that larger correlation range and larger number of monitoring locations improves the stability of the problem.

The distance function approach presented in Chapter 3 is a simple and flexible approach in characterization of uncertainty in the areal limits of the stationary

variables. Extension of this approach to 3D applications and capability of integration of secondary data sources are other areas for future research.

Finally, as a possible future research, the presented approach should be applied to a real aquifer to investigate the performance of the proposed model in real field applications.

BIBLIOGRAPHY

Ahlstrom, S., H. Foote, R. Arnett, C. Cole, and R. Serne (1977), *Multicomponent Mass Transport Model: Theory and Numerical Implementation*, Tech. Rep. BNWL 2127, Battelle Pacific Northwest Laboratories, Richmond, WA.

Alcolea, A., J. Carrera, A. Medina (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, Advances in Water Resources, 29(11), 1678-1689.

Alvarez, P.J.J., and W.A. Illman (2006), *Bioremediation and Natural Attenuation: Process Fundamentals and Mathematical Models*, John Wiley & Sons, Hoboken, NJ.

American Petroleum Institute (2004), *API Interactive LNAPL Guide Version 2*, Environmental Systems and Technologies, Blacksburg, VA.

Anderson, M.P., and W.W. Woessner (1992), *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, Academic Press, San Diego, 381 p.

Anderman E.R., and M.C. Hill (1999), A new multistage groundwater transport inverse method: Presentation, evaluation, and Implications, *Water Resources Research*, 35(4), 1053-1063.

Aral, M.M., J. Guan, M.L., Masila (2001), Identification of contaminant source location and release history in aquifers, *Journal of Hydrologic Engineering*, 6, 225-234.

Armstrong, J. E., C. V. Deutsch, and K. W. Biggar (2003), Geostatistical Assessment of Cone Penetrometer Test Data for Contaminant Assessment Based on Ultraviolet Induced Fluorescence. *Proceeding of the 56th Canadian Geotechnical Conference and the 4th Joint CGS/IAH-CNC Groundwater Specialty Conference*, Winnipeg, MB, Canada, September 29-October 1. 17 pp.

Atmadja, J., and A.C. Bagtzoglou (2001), State of the Art Report on Mathematical Methods for Groundwater Pollution Source Identification, *Environmental Forensics*, 2, 205-214.

Axelsson, O., and V. Barker (1984), *Finite Element Solution of Boundary Value Problems*, Academic Press, Orlando, Florida, pp. 1-63.

Aziz, C.E., C.J. Newell, J.R. Gonzales, P.E. Haas, T.P. Clement, and Y. Sun (1999), *BIOCHLOR Natural Attenuation Decision Support System User's Manual*, Prepared for the U.S. Air Force Center for Environmental Excellence, Brooks AFB, San Antonio, TX.

Baptista, A.E. de M.(1987), *Solution of Advection-Dominated Transport by Eulerian-Lagrangian Methods Using the Backwards Method of Characteristics*, PhD Thesis, MIT, Cambridge, MA.

Bauer, S., M. Bayer-Raich, T. Holder, C. Kolesar, B. Muller and T. Ptak (2004), Quantification of Groundwater contamination in an urban area using integral pumping tests, *Journal of Contaminant Hydrology*, 75(3-4), 183-213.

Bauer, S., C. Beyer, and O. Kolditz (2006), Assessing measurement uncertainty of first order degradation rates in heterogeneous aquifers, *Water Resources Research*, 42, W01420, doi: 10.1029/2004WR003878.

Bauer, S., C. Beyer, and O. Kolditz (2007), Einfluss von Heterogenität und Messfehler auf die Bestimmung von Abbauraten erster Ordnung-eine Virtueller-Aquifer-Szenarioanalyse, *Grundwasser*, 12, 3-14.

Bekins, B.A., B.E. Rittmann, and J. A. MacDonald (2001), Natural Attenuation Strategy for Groundwater Cleanup Focuses on Demonstrating Cause and Effect, *USGS Eos Transactions*, AGU, 82, 5, pp. 57-58.

Bennett, G. D., and C. Zheng (1995), *Applied Contaminant Transport: Theory and Practice*, Van Nostrand Reinhold, New York, NY.

Beyer, C., S. Bauer, and O. Kolditz (2006), Uncertainty assessment of contaminant plume length estimates in heterogeneous aquifers, *J. of Contaminant Hydrology*, 87, 73-95.

Beyer C., C. Chen, J. Gronewold, O. Kolditz, and S. Bauer (2007), Determination of first-order degradation rate constants from monitoring networks, *Ground Water*, 45(6): 774-785.

Binsariti, A.A. (1980), Statistical analysis and stochastic modeling of the Cartaro aquifer in southern Arizona, *PhD. Thesis*, Tucson: Department of Hydrology and Water Resources, University of Arizona.

Blandford, T.N., and P.S. Huyakorn (1991), *A Modular Semi-Analytical Model for the Delineation of Wellhead Protection Areas, Version 2.0*, U.S. Environmental Protection Agency, Washington, DC.

Borden, R.C., P.B., Bedient (1986), Transport of dissolved hydrocarbons influenced by oxygen-limited biodegradation: 1.Theoretical development, *Water Resources Research*, 22(13), 1973-1982.

Borden, R.C., P.B., Bedient, M.D., Lee, C.H., Ward, J.T. Wilson (1986), Transport of dissolved hydrocarbons influenced by oxygen-limited biodegradation: 2. Field application, *Water Resources Research*, 22(13), 1983-1990.

Borden, R.C., R.A. Daniel, L.E. LeBrun, and C. Davis (1997), Intrinsic bioremediation of MTBE and BTEX in a gasoline contaminated aquifer, *Water Resources Research*, 33(5), 1105-1115.

Bouner, E. J., B. E. Rittmann, and P. L. McCarty (1981), Anaerobic degradation of halogenated 1- and 2-carbon organic compounds, *Environmental Science and Technology*, 15, 596-599.

Bradford, L.A., L.M. Abriola, and K.M. Rathfelder (1998), Flow and entrapment of dense nonaqueous phase liquids in physically and chemically heterogeneous aquifer formations, *Advances in Water Resources*, 22(2), 117-132.

Bridge, J.S., and M.R. Leeder (1979), A simulation model of alluvial stratigraphy, *Sedimentology*, 26(5): 617-644.

Buscheck, T.E., and C. M. Alcantar (1995), Regression techniques and analytical solutions to demonstrate intrinsic bioremediation, in Intrinsic Bioremediation, edited by R.E. Hinchee, T.J. Wilson, and D. Downey pp. 109-116, Battelle Press, Columbus, Ohio.

Caers, J. (2008), Distance-Based Random Field Models: Theory and Applications, *8th International Geostatistics Congress*, Santiago, Chile, December 1-5.

Caers, J., T. Hoffman (2006), The probability perturbation approach: a new look at Bayesian inverse modeling, Mathematical Geology, 38(1), 81-100.

Capilla J.E., J. Rodrigo, and J.J. Gomez-Hernandez (1999), Simulation of non-Gaussian transmissivity fields honoring piezometric data and integrating soft and secondary information, *Mathematical Geology*, 31(7), 907-927.

Carey, G.R., P.J. van Geel, J.R. Murphy, E. A. McBean, and F. A. Rover (1998), Full-scale field application of a coupled biodegradation-redox model (BIOREDOX), In *Natural attenuation of chlorinated solvents*, G. B. Wickramanayake, and R. H. Hinchee (Eds.), Batelle Press, Columbus, OH, 213-218.

Carrera, J. and S.P. Neuman (1986) Estimation of aquifer parameters under transient and steady-state condition. 1. Maximum likelihood method incorporating prior information, *Water Resources Research*, 22(2), 199-210.

Carrera, J. and S.P. Neuman (1986) Estimation of aquifer parameters under transient and steady-state condition. 2. Uniqueness, stability and solution algorithms, *Water Resources Research*, 22(2), 211-227.

Carrera, J., F. Navarrina, L. Vives, J. Heredia, and A. Medina (1990), Computational aspects of the inverse problem, *In Proc. Of VIII international conference on computational methods in water resources*, CMP, 513-523.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L.J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeology Journal*, 13(1), 206-222.

Chapelle, F.H., P.M. Bradley, D.R. Lovely, and D.A. Vroblesky (1996), Measuring rates of biodegradation in a contaminated aquifer using field and laboratory methods, *Ground Water*, 34(4), 691-698.

Chapelle, F.H., M.A. Widdowson, J.S. Brauner, E. Mendez, and C. Casey (2003), *Methodology for Estimating Times of Remediation Associated with Monitored Natural Attenuation*, U.S. Geological Survey, Water Resources Investigations Report 03-4057.

Chen, M.J., A.A. Keller, D.X. Zhang, Z.M. Lu, and G.A. Zyvoloski (2006), *Water Resources Research*, 42(3), W03425.

Cheng, R.T., V. Casulli, and S.N. Milford (1984), Eulerian-Lagrangian solution of the convection-dispersion equation in natural coordinates, *Water Resources Research*, 20(7), 944-952.

Chiang, C.Y., J.P. Salanitro, E.Y. Chai, J.D. Colthart, and C.L. Klein (1989), Aerobic biodegradation of benzene, toluene, and xylene in a sandy aquifer – data analysis and computer modeling, *Groundwater*, 27, 823 – 834.

Christensen, O.F., G. Cassiani, P.J. Diggle, P. Riberio, and G. Anreotti (2004), Statistical estimation of the relative efficiency of natural attenuation mechanisms in contaminated aquifers, *Stoch. Envir. Res. and Risk Ass.*, 18: 339-350.

Christ, J.A., C.A. Ramsburg, K.D. Pennell, and L.M. Abriola (2006) Estimating mass transfer from dense nonaqueous phase liquid source zones using upscaled mass transfer coefficients: An evaluation using multiphase numerical simulations, *Water Resources Research*, 42(11), W11420.

Chu, M., P. Kitanidis, P.L. McCarty (2007) Dependence of lumped mass transfer coefficient on scale and reactions kinetics for biologically enhanced NAPL dissolution, *Advances in Water Resources*, 30(6-7), 1618-1629.

Cirpka, O.A., P.K. Kitanidis (2000), Sensitivity of temporal moments calculated by the adjoint-state method and joint inversing of head and tracer data, *Advances in Water Resources*, 24(1), 89-103.

Clement, T.P. (1997), *RT3D: A Modular Computer Code for Simulating Reactive Multi-species Transport in 3-Dimensional Groundwater Systems*, Prepared for the office of Environmental Management U.S. Department of Energy, Richland, WA.

Cooley R.L., and R.L. Naff (1990), *Regression modeling of ground-water flow*, Techniques of Water-Resources Investigations of the US Geological Survey, 232 pp.

Cowan, E., R. Beatson, H. Ross, W. Fright et al (2003), Practical Implicit Geological Modeling, *Proceedings of 5th International Mining Geology Conference*, 17-19 November, Bendigo, Victoria, Australia.

D'Affonseca, F.M., P. Blum, M. Finkel, R. Melzer, P. Grathwohl (2008), Field-scale characterization and modeling of contaminant release from a coal tar source zone, *Journal of Contaminant Hydrology*, 302: 120-139.

Dagan, G. (1985), Stochastic modeling of groundwater flow by unconditional and conditional probabilities: The inverse problem, *Water Resources Research*, 21(1), 65–72.

Dagan, G., and Y. Rubin (1988), Stochastic identification of recharge, transmissivity and storativity in aquifer transient flow: A quasi-steady approach, *Water Resources Research*, 24(10), 1698–1710.

Dagdelen, K., and K., Turner (1996), Importance of Stationarity of Geostatistical Assessment of Environmental Contamination, *Geostatistics for environmental and geotechnical applications,* Volume 1283, ASTM, 280 pp.

Datta-Gupta, A., S. Yoon, I. Barman, and D.W. Vasco (1998), Streamline-based production data integration into high-resolution reservoir models, *Journal of Petroleum Technology*, 50(12), 72-76.

Deutsch, C.V., and A.G. Journel (1998), *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, NY.

Deutsch, C.V., (2002), *Geostatistical Reservoir Modeling.* Oxford University Press, 376 pp.

Deutsch, C.V. (2006), A sequential indicator simulation program for categorical variables with point and block data, *Computers and Geosciences* 32(10): 1669-1681.

Dillard, L.A., H.I. Essaid, W.N. Herkelrath (1997), Multiphase flow modeling of a crude-oil spill site with a bimodal permeability distribution, *Journal of Contaminant Hydrology*, 48 (1–2), 89–119.

Dillard, L.A., M.J. Blunt (2000), Development of pore network simulation model to study nonaqueous phase liquid dissolution, *Water Resources Research*, 36(2), 439-454.

Dillard, L.A. H.I. Essaid, M.J. Blunt (2001), A functional relation for field-scale nonaqueous phase liquid dissolution developed using a pore network model, *Journal of Contaminant Hydrology*, 48(1-2), 89-119.

Domenico, P.A. (1987), An analytical model for multidimensional transport of a decaying contaminant species, *J. of Hydrology*, 91, 49-58.

Douglas, J.B. and R.S. Olsen (1981), Soil Classification using Electric Cone Penetrometer, *Symposium on Cone Penetration Testing and Experience*, Geotechnical Engineering Division, ASCE, St. Louis, pp. 209-227.

Downey, D. C., Hinchee, R. E., and R. N. Miller (1999), *Cost-effective remediation and closure of petroleum-contaminated sites*, Battelle Press, Columbus, OH.

DuPont, T., R. P. Kendall, and H. H. Rachford (1968), An approximate factorization procedure for solving self-adjoint elliptic difference equations, *SIAM J. Numerical Analysis* 5, 559-573.

Ellis, B., and K. Gorder (1997), Intrinsic bioremediation: An economic option for cleaning up contaminated land, *Chemistry and Industry*, 3, 95-99.

Elkateb, T., R. Chalaturnyk, P.K. Robertson (2003), Simplified geostatistical analysis of earthquake-induced ground response at the Wildlife site, California, *Canadian Geotechnical Journal*, 40, 16-35.

Essaid, H.I., W.N. Herkelrath, and K.M. Hess (1993), Simulation of fluid distributions observed at a crude-oil spill site incorporating hysteresis, oil entrapment, and spatial variability of hydraulic properties, *Water Resources Research*, 29(6) 1753-1770.

Essaid, H.I., and K.M. Hess (1993), Monte Carlo simulations of multiphase flow incorporating spatial variability of hydraulic properties, *Ground Water*, 31 (1), 123–134.

Essaid, H.I., I.M. Cozzarelli, R.P. Eganhouse, W.N., Herkelrath, BA. Bekins, G.N. Delin (2003), Inverse modeling of BTEX dissolution and biodegradation at the Bemidji, MN crude-oil spill site, *Journal of Contaminant Hydrology*, 67(1-4), 269-299.

Federal Remediation Technologies Roundtable (1999), *Natural Attenuation*, http://www.frtr.gov//matrix2/section4/4_35.html, accessed 18 February 2007.

Fetter, C.W. (1993), *Contaminant hydrogeology*, Macmillan Publishers, New York, NY.

Friedel, M.J. (2005), Coupled inverse modeling of vadose zone water, heat, and solute transport: calibration constraints, parameter nonuniqueness, and predictive uncertainty, *Journal of Hydrology*, 312, 148-175.

Fu, J., J.J. Gomez-Hernandez (2008), A Blocking Markov Chain Monte Carlo Method for Inverse Stochastic Hydrogeological Modeling, *Mathematical Geosciences*, 41(2): 105-128.

Garder, A.O., D.W. Peaceman, and A.L. Pozzi (1964) , Numerical calculation of multi-dimensional miscible displacement by the method of characteristics, *Society of Petroleum Engineers Journal*, 6(2), 175-182.

Gomez-Hernandez, J.J., A. Sahuquillo, J.E. Capilla (1997). Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data – 1.Theory, *Journal of Hydrology*, 203(1-4): 162-174.

Gomez-Hernandez, J.J., Hendricks-Franssen, HJ, A. Sahuquillo (2003), Stochastic conditional inverse modeling of subsurface mass transport: A brief review and the self-calibrating method, *Stochastic Environmental Research and Risk Assessment*, 17(5): 319-328.

Goovaerts, P. (1997), *Geostatistics for natural resources evaluation*, Oxford University Press, New York, NY, 483 pp.

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

Graham, W., D. McLaughlin (1989a), Stochastic analysis of non-stationary subsurface solute transport: 1. unconditional moments, *Water Resources Research*, 25(2), 215-232.

Graham, W., D. McLaughlin (1989b), Stochastic analysis of non-stationary subsurface solute transport: 2. conditional moments, *Water Resources Research*, 25(2), 2331-2355.

Grindrod, P., and M. D. Impey (1991), Fractal field simulations of tracer migration within the WIPP Culebra Dolomite, *Intera Inf. Technol.*, Henley-upon-Thames, UK.

Gustafsson, I. (1978), A class of first order factorization methods, *BIT*, 18, 142.

Gutjahr, A. L., and J. R. Wilson (1989), Co-kriging for stochastic flow models, *Transport in Porous Media*, 4(6), 585–598.

Gutjahr, A., B. Bullard, S. Hatch, and L. Hughson (1994), Joint conditional simulations and the spectral method approach for flow modeling, *Stochastic Hydrol. Hydraul.*, 8(1), 79–108.

Harvey, C.F., and S.M. Gorelick (1995), Mapping hydraulic conductivity – sequential conditioning with measurements of solute arrival time, hydraulic-head and local conductivity, *Water Resources Research*, 31(7): 1615-1626.

Hendricks Franssen, H.J., J.J. Gomez-Hernandez, J.E. Capilla, A. Sahuquillo (1999a), Joint simulation of transmissivity and storativity fields conditional to steady-state and transient hydraulic head data, *Advances in Water Resources*, 23(1), 1-13.

Hendricks Franssen, H.J., E.F. Cassiraga, J.J. Gomez-Hernandez, A. Sahuquillo, J.E. Capilla (1999b) Inverse modeling of groundwater flow in a 3D fractured medium. In: *GeonENV II – Geostatistics for Environmental Applications* (ed. by J.J. Gomez-Hernandez, A. Soares and R. Froidevaux), 283-294, Kluwer, Dorrecht, The Netherlands.

Hendricks Franssen, H.J., J.J. Gomez-Hernandez, A. Sahuquillo (2003), Coupled inverse modeling of groundwater flow and mass transport and the worth of concentration data, *Journal of Hydrology*, 281(4), 281-295.

Hendricks Franssen, H.J., J.J. Gomez-Hernandez (2003), Impact of measurement errors in stochastic inverse conditional modeling by the self-calibrating approach, *Advances in Water Resources*, 26(5), 501-511.

Hendricks Franssen, H.J., P. Brunner, P. Makobo, and W. Kinzelbach (2008), Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images, *Water Resources Research*, 44(1), W01419.

Higgins, I. J., and P. D. Gilbert (1978), The biodegradation of hydrocarbons, *The Oil Industry and Microbial Ecosystems*, Heyden and sons, London, pp. 80-115.

Hill, M.C., and C.R. Tiedeman (2007), *Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty*, John Wiley & Sons, Hoboken, NJ.

Hoeksema, R. J., and P. K. Kitanidis (1984), An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resources Research*, 20(7), 1003–1020.

Houlding, S.W. (1994), *3D Geoscience Modeling: Computer Techniques for Geological Characterization*, Springer, Berlin Heidelberg New York.

Hu, L.Y. (2000), Gradual deformation and iterative calibration of Gaussian related stochastic models, Mathematical Geology, 32(1): 87-108.

Huang, H., B.X. Lu, X.H. Wen, C, Shirley (2004), Stochastic inverse mapping of hydraulic conductivity and sorption partitioning coefficient fields conditioning on non-reactive and reactive tracer test data, Water Resources Research, 40(1), W01506.

Huntley, D., and G.D. Beckett (2002), Persistence of LNAPL sources: relationship between risk reduction and LNAPL recovery, *Journal of Contaminant Hydrology*, 59, 3–26.

Imhoff, P.Y., P.J. Jaffe, and G.F. Pinder (1994), An experimental study of complete dissolution of a non-aqueous phase liquid in saturated porous media, *Water Resources Res.*, 30(2), 307-320.

Jarsjo, J. and M. Bayer-Raich (2008), Estimating plume degradation rates in aquifers: Effect of propagating measurement and methodological errors, *Water Resources Research*, 44, W02501, doi:10.1029/2006WR005568.

Journel, A.G. (2002), Combining knowledge from diverse sources: an alternative to traditional data independence hypotheses, *Mathematical Geology* 34(5): 573-596.

Journel, A.G., and J.J. Gomez-Hernandez (1993) Stochastic imaging of the Wilmington clastic sequence, *SPE Formation Evaluation*, 8(1): 33-40.

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

King, M., W.G. Barker, J.F. Devlin, B.J. Butler (1999), Migration and natural fate of coal tare creosote plume: 2. Mass balance and biodegradation indicators, *Journal of Contaminant Hydrology*, 39, 281-307.

Kitanidis, P.K. and E.G. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resources Research*, 19(3), 677-690.

Kitanidis, P. K., and R. W. Lane (1985), Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, *Journal of Hydrology*, 79(1–2), 53–71.

Konikow, L.F., and J.D. Bredehoef (1978), *Computer Model of Two-Dimensional Solute Transport and Dispersion in Ground Water*, U.S. Geological Survey Water-Resources Investigations Book 7, Chapter C2.

Kram, M.L., A.A. Keller, S.M. Massick, L.E. Laverman (2004), *Soil & Sediment Contamination*, 13(2): 103-118.

Kubert M., M. Finkel (2006) Contaminant mass discharge estimation in groundwater based on multi-level point measurements: A numerical evaluation of expected errors, *Journal of Contaminant Hydrology*, 84, 55-80.

Kueper B.H., E.O. Frind (1991), 2-phase flow in heterogeneous porous-media: 1. Model development, *Water Resources Research*, 27(6): 1049-1057.

Kueper B.H., J.I. Gerhard (1995), Variability of point source infiltration rates for two-phase flow in heterogeneous porous media, *Water Resources Research*, 31(12), 2971-2980.

Landmeyer, J. E., F. H. Chapelle, P.M. Bradly, J.F. Pankow, C.D. Church, and P. G. Tratnyk (1998), Fate of MTBE relative to benzene in a gasoline-contaminated aquifer, *Groundwater Monitor Rev.*, 18(4), 93-102.

Langmuir, D. (1997), *Aqueous Environmental Geochemistry*, Prentice Hall, Upper Saddle River, NJ.

LaVenue, M., G. deMarsily (2001), Three-dimensional interference test interpretation in a fractured aquifer using the pilot point inverse method, *Water Resources Research*, 37(11), 2659-2675.

Letniowski, F.W. (1989), An overview of Preconditioned Iterative Methods for Sparse Matrix Equations, *Research Report CS-89-26*, Department of Applied Mathematics, University of Waterloo, 35 p.

Li, B.L., and T.C.J. Yeh (1999) Cokriging estimation of the conductivity field under variably saturated flow conditions, Water Resources Research, 35(12), 3663-3674.

Ling, M., and H.S. Rifai (2007), Modeling natural attenuation with source control at a chlorinated solvent dry cleaner site, *Ground Water Monitoring and Remediation*, 27(1): 108-121.

Lu, G., T.P. Clement, C. Zheng, and T. H. Wiedemeier (1999), *Natural Attenuation of BTEX compounds: model development and field-scale application*, Ground Water, 37, 707-717.

Mahar, P.S., B. Datta (2001), Optimal identification of groundwater pollution sources and parameter estimation, *Journal of Water Resources Planning and Management* 131(1): 45-57.

Mayer, A.S., and C.T. Miller (1996), The influence of mass transfer characteristics and porous media heterogeneity on nonaqueous phase dissolution, *Water Resources Res.*, 32(6), 1551-1567.

Mayer, A.S. and C.L. Huang (1999) Development and application of a coupled-process parameter inversion model based on the maximum likelihood estimation method, *Advances in Water Resources*, 22(8), 841-853.

McDonald, M.G., and A.W. Harbaugh (1988), *A Modular Three-Dimensional Finite-Difference Ground-Water Flow Model*, U.S. Geological Survey Techniques of Water Resources Investigations, Book 6.

McKenna, S.A., J. Doherty, D.B. Hart (2003), Non-uniqueness of inverse transmissivity field calibration and predictive transport modeling, *Journal of Hydrology*, 281, 265-280.

McLaughlin, D., and L. R. Townley (1996), A reassessment of the groundwater inverse problem, *Water Resources Research*, 32(5), 1131–1161.

McLennan, J.A. (2004) Using the Variogram to establish the stratigraphic correlation structure, *Sixth Annual Report of the Centre for Computational Geostatistics*, University of Alberta, Canada.

McLennan, J.A. (2007), *The decision of stationarity*, PhD Thesis, University of Alberta, Edmonton, Alberta, Canada.

McNab, W.W., and B.P. Dooher (1998), A critique of a steady-state analytical method for estimating contaminant degradation rates, *Groundwater*, 36(6), 983-987.

Medina, A., J. Carrera, and G. Galarza (1990), Inverse modeling of coupled flow and solute transport problems, *ModelCARE 90: Calibration and reliability in groundwater modeling*, IAHS Publishers, 195, 185-194.

Medina, A., and J. Carrera (1996), Coupled estimation of flow and solute transport parameters, *Water Resources Research*, 32(10): 3063-3076.

Mendez, E., M. Widdowson, S. Brauner, S., Chapelle, F., and C. Casey (2004), Natural Attenuation Software (NAS): A computer program for estimating remediation times of contaminated groundwater, *Proceedings of Tenth International Conference on Development and Application of Computer Techniques to Environmental Studies, ENVIROSOFT X, Jun 2-4*, WIT Press, Southampton, United Kingdom, 185-194.

Miles, B., R. Maji, E.A. Sudicky, G. Teutsch, A. Peter (2008), A pragmatic approach for estimation of source-zone emissions at LNAPL contaminated sites, Journal of contaminant hydrology, 96(1-4): 83-96.

Mishra, S., J.C. Parker, N. Singhal (1989), Estimation of soil hydraulic properties and their uncertainty from particle size distribution data, *Journal of Hydrology*, 108, 1-18.

Monod, J. (1949), The growth of bacterial cultures, *Ann Rev Microbiology*, 3, 371-394.

Moore, C., and J. Doherty (2005), Role of the calibration process in reducing model predictive error, *Water Resources Research*, 41(5), W05020.

National Research Council (2000), *Natural Attenuation for Groundwater Remediation*, National Academy Press, Washington, DC.

Nambi I.M., S.E. Powers (2003), Mass transfer correlations for nonaqueous phase liquid dissolution from regions with high initial saturations, *Water Resources Research*, 39(2), AN1030.

Nelder, J.A., and R. Mead (1965), A simplex method for function minimization, *Computer Journal*, 7, 308-313.

Neuman, S.P. (1981), An Eulerian-Lagrangian numerical scheme for the dispersion-conversion equation using conjugate space-time grids, *J. of Computational Physics*, 41, 270-294.

Neuman, S.P. (1984), Adaptive Eulerian-Lagrangian finite element method for advection-dispersion, *Int. J. Numerical Methods in Engineering*, 20, 321-337.

Neupauer, R.M., and J.L. Wilson (2001), Adjoint-derived location and travel time probabilities for multi-dimensional groundwater system, *Water Resources Research* 37(6): 1657-1668.

Neupauer, R.M., and R. Lin (2006), Identifying sources of a conservative groundwater contaminant using backward probabilities conditioned on measured concentrations, *Water Resources Research* 42(3): W03424.

Newell, C. J., R. K. McLeod, and J. Gonzales (1996), *BIOSCREEN Natural Attenuation Decision Support System*, Report EPA/600/R-96/087, US EPA-Office of Research and Development, Washington, DC.

Newell, C. J., H.S. Rifai, J.T. Wilson, J.A. Connor, J.A. Aziz, and M.P. Suarez (2002), Calculation and use of first-order rate constants for monitored natural attenuation studies, *U.S. EPA Ground Water Issue*, *EPA/540/S-02/500*, US Environment Protection Agency, Washington, DC.

Newman, M.A., K. Hatfield, J. Hayworth, P.S.C. Rao, and T. Stauffer (2006) Inverse characterization of NAPL source zones, *Environmental Science and Technology*, 40(19): 6044-6050.

Odencrantz, J.E., R.A. Vogl, and M.D. Varljen (2003) Natural Attenuation rate clarifications: The true picture is in the details, Soil and Sediment Contamination, 12(5), 663-672.

Parker, J.C., A.K. Katyal, J.J. Kaluarachchi, R.J. Lenhard, T.J. Johnson, K. Jayaraman, K. Unlu, and J.L. Zhu (1991), *Modeling multiphase organic chemical transport in soils and groundwater*, Report, EPA/600/2-91/042 U.S. Environment Protection Agency, Washington, DC.

Parker, J.C., M. Islam (2000), Inverse modeling to estimate LNAPL plume release timing, Journal of Contaminant Hydrology, 45(3-4), 303-327.

Parker J.C. and E. Park (2004), Modeling field-scale dense nonaqueous phase liquid dissolution kinetics in heterogeneous aquifers, *Water Resources Research*, 40(5), W05109.

Peter, A., A. Steinbach, R. Liedl, T. Ptak, W. Michaelis, G. Teutsch (2004) Assessing microbial degradation of o-xylene at field-scale from reduction in mass flow rate combined with compound-specific isotope analysis, *Journal of Contaminant Hydrology*, 71, 127-154.

Pepper, J.W., A.O. Wright, J.E. Kenny (2002), In situ measurements of subsurface contaminants with a multi-channel laser induced fluorescence system, *Spectrochimica acta. Part A, Molecular and biomolecular spectroscopy*, 58(2): 317-331.

Pfannkuch, H.O. (1984), Determination of the contaminant source strength from mass exchange processes at the petroleum-ground-water interface in shallow aquifer systems, *Proc. NWWA/API Conf. on Petroleum Hydrocarbons and Organic Chemicals in Groundwater – Prevention, Detection, and Restoration*, NWWA, Worthington, Ohio, 5-7 Nov.

Pitter, P., and J. Chudoba (1990), *Biodegradability of organic substances in the Aquatic Environment*, CRC Press, Boca Raton, FL.

Poeter, E.P., S. McKenna (1995), Reducing uncertainty associated with ground water flow and transport predictions, *Ground Water*, 33(6), 899-904.

Pollock, D.W. (1988) Semi-analytical computation of pathlines for finite difference models, *Ground Water*, 26(6), 743-750.

Powers, S.E., L.M. Abriola, and W.J. Weber (1994), An experimental investigation of nonaqueous phase liquid dissolution in saturated subsurface systems: Transient mass transfer rates, *Water Resources Res.*, 30(2), 321-332.

Prickett, T.A., T.G. Naymik, and C.G. Lonnquist (1981), *A Random Walk Solute Transport Model for Selected Groundwater Quality Evaluations*, Bulletin 65, Illinois State Water Survey, Champaign, IL.

Pyrcz, M., O. Catuneanu, and C.V. Deutsch (2005), Stochastic surface-based modeling of turbidite lobes, *AAPG Bulletin*, 89(2) 177-191.

RamaRao, B. S., A. M. LaVenue, G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields, 1, Theory and computational experiments, *Water Resources Research*, 31(3), 475–493.

Rehfeldt, K.R., L.W. Gelhar, J.B. Southhard and A.M. Dasinger (1989), *Estimates of macro-dispersivity based on the analysis of hydraulic conductivity variability at the MADE site*, EPRI EN-6405, Project 2485-5, Palo Alto, CA.

Rifai, H. S., R.C. Borden, J.T. Wilson, and C.H. Ward (1995), Intrinsic bioattenuation for subsurface restoration, In *Intrinsic Bioremediation*, R.E. Hinchee, F.J. Brockman, and C.M. Vogel (Eds.), Battelle Press, Columbus, OH, 3(1), 1-29, 1995.

Rifai, H.S., C. J. Newell, J. R. Gonzales, S. Dendrou, L. Kennedy, and J. T. Wilson (1997), *BIOPLUME III Natural Attenuation Decision Support System Version 1.0 User Manual*, Prepared for the U.S. Air Force Center for Environmental Excellence, Brooks Air force Base, San Antonio, TX.

Rifai, H. S., C. J. Newell, J. R. Gonzales, and J. T. Wilson (2000), Modeling Natural Attenuation of Fuels with Bioplume III, *Journal of Environmental Engineering*, 126(5), 428-438.

Rifai, H. S., and T. Rittaler (2005), Modeling natural attenuation of benzene with analytical and numerical models, *Biodegradation*, 16, 291-304.

Robertson, P.K. (1986), Soil Profiling and Classification, In *CPT Application Guide*, Gregg In-Situ Inc.

Robertson, P.K., R.G. Campanella, D. Gillespie, and J. Greig (1986), Use of Piezometer Cone Data, *Proceedings of the ASCE Specialty Conference on In Situ'86: Use of In Situ Tests in Geotechnical Engineering*, Blacksburg, Virginia, pp. 1263-1280.

Robin, M. J. L., A. L. Gutjahr, E. A. Sudicky, and J. L. Wilson (1993), Cross-correlated random field generation with the direct Fourier transform method, *Water Resources Research*, 29(7), 2385–2397.

Rojas-Avellaneda, D., and J.L. Silvan-Cardenas (2006), Performance of geostatistical interpolation methods for modeling sample data with non-stationary mean, *Stochastic Environmental Research and Risk Assessment*, 20(6): 455-467.

Rubin, Y. (1991a), Transport in heterogeneous porous media-prediction and uncertainty, *Water Resources Research*, 27(7): 1723-1738.

Rubin, Y. (1991b), The spatial and temporal moments of tracer concentration in disordered porous media, *Water Resources Research*, 27(11): 2845-2854.

Rubin, Y., and G. Dagan (1987), Stochastic identification of transmissivity and effective recharge in steady groundwater flow, 1, Theory, *Water Resources Research*, 23(7), 1185–1192.

Russell, T.F., and M.F. Wheeler (1983), Finite Element and finite difference methods for continuous flows in porous media, In *SIAM The Mathematics of Reservoir Simulation*. R.E. Ewing (Ed.), 35-106.

Sahuquillo A., H.J. Hendricks Franssen, J.J. Gomez-Hernandez, J.E. Capilla (1999), Computational Aspects of the Coupled Inversion of Groundwater Flow and Mass Transport, In: Stauffer et al. (eds.) *ModelCARE 99*, IAHS Press.

Sciortino, A., T.C. Harmon and W.G. Yeh (2000), Inverse modeling for locating dense nonaqueous pools in groundwater under steady flow conditions, *Water Resources Research*, 36(7): 1723-1735.

Scheidt C., K. Park, J. Caers (2008) Defining a random function from a given set of model realizations, *8th International Geostatistics Congress*, Santiago, Chile, December 1-5.

Shewchuk, J.R. (1994), An introduction to the Conjugate Gradient Method without the agonizing pain, School of Computer Science, Carnegie Mellon University, 64 p. (Available by anonymous FTP to WARP.CS.CMU.EDU).

Sidauruk, P.A., H.D. Cheng and D. Ouzar (1998), Groundwater contaminant source and transport parameter identification by correlation coefficient optimization, *Ground Water*, 26(2), 208-214.

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

Sonnenborg, T.O., P. Engesgaard, D. Rosbjerg (1996), Contaminant transport at a waste residue deposit .1. Inverse flow and non-reactive transport modeling, *Water Resources Research*, 32(4): 925-938.

Stenback, G.A., S.K. Ong, S.W. Rogers, and B.H. Kjartonson (2004), Impact of transverse and longitudinal dispersion on first-order degradation rate constant estimation, *Journal of Contaminant Hydrology*, 73, 3-14, doi: 10.1016/j.jconhyd.2003.11.004.

Strauss, P. (1998), *Natural Attenuation of Organic Compound*, http://cpeo.org/pubs/fsna2o.html, accessed 21 February 2007.

Strecker, E.W., W.S. Chu (1986), Parameter identification of groundwater contaminant transport model, *Ground Water*, 24(1) 56-62.

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

Sun, N.Z. (1994), *Inverse Problems in Groundwater Modeling*, Springer Publishers, New York, NY.

Sun, N.Z., W.W.G. Yeh (1990a), Coupled inverse problems in groundwater modeling: 1.sensitivity analysis and parameter-identification, *Water Resources Research*, 26(10): 2507-2525.

Sun, N.Z., W.W.G. Yeh (1990b), Coupled inverse problems in groundwater modeling: 2.identifiability and experimental design, *Water Resources Research*, 26(10): 2527-2540.

Surampalli, R., and S. Banerji (2002), Long-term performance monitoring at natural attenuation site, *Pract. Periodical Hazardous, Toxic, Radioactive Waste Management,* 6(3), 173-176.

Suthersan, S. S. (2002), *Natural and Enhanced Remediation Systems*, Lewis Publishers, Boca Raton, FL.

Thornton, S.F., D.N. Lerner, and S.A. Banwart (2001), Assessing the natural attenuation of organic contaminants in aquifers using plume-scale electron and carbon balances: model development with analysis of uncertainty and parameter sensitivity, *J. of Contaminant Hydrology*, 53, 199-232.

Tonkin, M., J. Doherty, C. Moore (2007), Efficient nonlinear predictive error variance for highly parameterized models, *Water Resources Research*, 43(7), W07429.

Ucankus, T., K. Unlu (2008), The effect of aquifer heterogeneity on natural attenuation rate of BTEX, *Environmental Geology*, 54: 759-776.

Umari, A., R. Willis, P.L.F. Liu (1979), Identification of aquifer dispersivities in 2-dimesnsional transient groundwater contaminant transport-optimization approach, *Water Resources Research*, 15(4): 815-831.

US EPA (1999a), *Use of Monitored Natural Attenuation at Superfund, RCRA Corrective Action, and Underground Storage Tank Sites*, Report 9200.4-17P, US EPA - Office of Solid Waste and Emergency Response, Washington, DC.

US EPA (1999b), *Monitored Natural Attenuation of Petroleum Hydrocarbons*, Report EPA/600/F-98/021, US EPA - Office of Research and Development, Washington, DC.

Van Rooy, D., A. Keidser, D. Rosbjerg (1989), Inverse modeling of flow and transport, in *Groundwater Contamination*, IAHS Publishers, 185: 11-23.

Van Cauwenberghe, L., and D.S. Roote (1998), *In-situ Bioremediation*, Report TO-98-01, Groundwater Remediation Technologies Analysis Center, Pittsburgh, PA.

Van Genuchten, M. (1980), A closed form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Science Society of America Journal*, 44, 892–898.

Vecchia, A.V., and R.L. Cooley (1987), Simultaneous confidence and prediction intervals for nonlinear regression models with application to a groundwater flow model, *Water Resources Research*, 23(7), 1237-1250.

Waddill, D. W., and M. A. Widdowson (1998), *SEAM3D: A Numerical Model for Three-Dimensional Solute Transport and Sequential Electron Acceptor-Based Bioremediation in Groundwater*, Final Report, Virginia State University, Department of Civil and Environmental Engineering, Blacksburg, VA.

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

Wagner, B.J., S.M. Gorelick (1986), A statistic methodology for estimating transport parameters: Theory and applications to one-dimensional advective-dispersive systems, Water Resources Research, 22(8), 1303-1313.

Wen, X.H., C.V. Deutsch, and A.S. Cullick (2002), Construction of geostatistical aquifer models integrating dynamic flow and tracer data using inverse technique, *Journal of Hydrology*, 255(1-4), 151-168.

Wen, X.H., C.V. Deutsch, A.S. Cullick (2003), Inversion of dynamic production data for permeability: fast streamline-based computation of sensitivity coefficients of fractional flow rate, *Journal of Hydrology*, 281(4), 296-312.

Wiedemeier, T. H., J.T. Wilson, D. H. Kampbell, R. N. Miller, and J. E. Hansen (1995), *Technical Protocol For Implementing Intrinsic Remediation with Long-Term Monitoring for Natural Attenuation of Fuel Contamination Dissolved in Groundwater (Volume I)*, AFCEE, Technology Transfer Division, Brooks Air Force Base, San Antonio, TX.

Wiedemeier, T. H., M.A. Swanson, J.T. Wilson, D. H. Kampbell, R. N. Miller, and J. E. Hansen (1996), Approximation of biodegradation rate constants for monoaromatic hydrocarbons (BTEX) in ground water, *Ground Water Monitoring and Remediation*, 16(3), 186-194.

Wiedemeier, T.H., H.S. Rifai, C.J. Newell, and J.T. Wilson (1999), *Natural Attenuation of Fuels and Chlorinated Solvents in the Subsurface*, John Wiley & Sons, New York.

Wilson, J.L., and L. Liu (1994), Backward Tracking to Find the Source of Pollution, *Waste Management from Risk to Remediation*, 1: 181-199.

Wilson, R.D., S.F. Thornton, and D.M. Mackay (2004), Challenges in monitoring the natural attenuation of spatially variable plumes, *Biodegradation Journal*, 15(6): 459-469.

Yeh, G.T. (1990), A Lagrangian-Eulerian method with zoomable hidden fine-mesh approach to solving advection-dispersion equations. *Water Resources Research*, 26(6): 1133-1144.

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

Yong, R. N., and C. N. Mulligan (2004), *Natural attenuation of contaminants in soil*, CRC Press, Boca Raton, FL.

Yoon, G.L., M.W. O'Neill (1999), Estimation of driven pile resistance using geostatistical methods, *Probabilistic Engineering Mechanics*, 14(1999), 205-211.

Zhang, Z., and M.T. Tumay (2003), Non-Traditional Approaches in Soil Classification Derived from the Cone Penetration Test. In *VanMarcke E, Fenton GA (eds) Probabilistic Site Characterization at the National Geotechnical Experimentation Sites*. ASCE, Reston VA, pp 101-149.

Zhang, Y.K. and R.C. Heathcote (2003), An improved method for estimation of biodegradation rate with field data, *Ground Water Monitoring and Remediation*, 23(3), 112-116.

Zheng, C. (1989), *PATH3D, A Groundwater Path and Travel Time Simulator, Version 3.0 User's Manual*, S.S. Papadopoulos & Associates, Inc., Bethesda, MD.

Zheng, C. (1990), *MT3D: A Modular Three-Dimensional Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems*. Report to the U.S. Environment Protection Agency, Ada, OK.

Zheng, C., and G.D. Bennett (1995), *Applied Contaminant Transport Modeling: Theory and Practice*, Van Nostrand Reinhold Pub., New York, NY.

Zheng, C., and P.P. Wang (1999), *MT3DMS: A Modular three-dimensional multi-species transport model for simulation of advection, dispersion and chemical reactions of contaminants in groundwater systems, Documentation and User's Guide*, Prepared for Army Corps of Engineers, Washington, DC.

Zhu, J. (2001), Transport and fate of nonaqueous phase liquid (NAPL) in variably saturated porous media with evolving scales of heterogeneity, *Stochastic Environmental Research and Risk Assessment*, 15(6), 447-461.

Zhu, J., and J.F. Sykes (2004), Simple screening models of NAPL dissolution in the subsurface, *Journal of Contaminant Hydrology*, 72, 245–258.

Zimmerman, D.A., G. de Marsily, C.A. Gotway, et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resources Research*, 34(6): 1373-1413.

Appendix A

IMPLEMENTATION OF DISTANCE BASED SIMULATION

An alternative way of generating multiple Gaussian realizations with their measure of fit *s* (as defined in Chapter 5) close to one would be to use 'distance-based' simulation (Scheidt et al. 2008 and Caers 2008) that builds on the concepts of multidimensional scaling (MDS), Karhunen-Loeve (KL) expansion, kernel principal component analysis (KPCA), and modeling and simulation in the metric and feature spaces. According to Scheidt et al. (2008) and Caers (2008), using distance-based simulation, one may choose a limited number of realizations that have a desired response and expand the set of realizations with similar responses while being conditioned to static data and reproduce the histogram and variogram. The idea in this appendix is to examine the performance of this approach in expanding the set of acceptable realizations conditioned to head data by SSC.

MDS is applied to reduce the dimensionality of the geostatistical realizations from N dimensional space to L dimensional space, where N and L are the number of grid nodes and number of realizations. KL expansion, which relies on eigen-value decomposition of the covariance matrix, can be used to generate new realizations in the metric space that is defined by pair-wise differences in the values of *s* (Chapter 5). When a non-Euclidean distance (e.g. the measure of fit *s*) is used as the measure of pair-wise difference between the realizations, KL expansion must be implemented in the feature space, where Gaussian/linear type modeling becomes more appropriate. KPCA is used to transform the set of realizations to the feature space. A back-transformation (pre-image problem) is then required to find the corresponding set of realizations in the Cartesian space.

In this work, the methodology proposed by Scheidt et al. (2008) and Caers (2008) is implemented in the context of using measure fit as the measure of distance (difference) between the realizations. The distance based simulation starts by generating multiple geostatistical realizations conditioned to all transmissivity and head observations with the values of measure of fit close to one. This can be achieved by conditioning a limited number of realizations to head data using the SSC algorithm, post-processing them, and select L realizations with desired response. Then, the pair-wise distance between the selected realizations is defined as the difference in the values of measure of fit for all realizations. Assuming $\mathbf{D} = [d_{ij}]$ to be the pair-wise distance matrix defined based on the measure of fit, the centered dot-product matrix $\mathbf{B}$ can be calculated by:

$$\mathbf{B} = \mathbf{HAH} \qquad [A.1]$$

where, $\mathbf{H}$ is the centering matrix and $\mathbf{A}$ is the equivalent dot-product matrix whose values are calculated by:

$$a_{ij} = -\frac{1}{2} d_{ij}^2 \qquad [A.2]$$

The centering matrix $\mathbf{H}$ in Equation [A.1] is defined by:

$$\mathbf{H} = \mathbf{I} - \frac{1}{L}\mathbf{1}\mathbf{1}^T \quad \text{with} \quad \mathbf{1} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \end{bmatrix} \qquad [A.3]$$

In the next step, the realizations are mapped from a high dimensional space (N dimensions) into a lower dimensional space (up to L dimensions). For this purpose, eigenvalue decomposition of the dot product matrix is performed:

$$\mathbf{B} = (\mathbf{HX})(\mathbf{HX})^T = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T \qquad [A.4]$$

where, $\mathbf{X}$ is an $L \times N$ matrix containing the original L realizations as its rows, $\mathbf{V}$ is the matrix containing the eigenvectors and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. After eigen-value decomposition of the dot-product matrix, the matrix of mapped realizations into a lower dimensional space ($\hat{\mathbf{X}}_d$) is calculated by Equation [A.5]. The overall process is termed multi-dimensional scaling.

$$\hat{\mathbf{X}}_d = \mathbf{V}_d \boldsymbol{\Lambda}_d^{1/2} \qquad [A.5]$$

where, $\boldsymbol{\Lambda}_d$ is the matrix containing the *d* largest eigenvalues and $\mathbf{V}_d$ is the matrix of the corresponding eigenvectors. The subscript *d* varies between 1 and the number of realizations *L*. According to Caers (2008), if the Euclidean distance measure is used between realizations with Gaussian distributions, the duality between the dot-product matrix and covariance matrix can be used to generate new realizations as linear combinations of existing realizations through KL expansion:

$$\mathbf{x} = \mathbf{m} + \left(\mathbf{HX}\right)^T \mathbf{V}_d \frac{1}{\sqrt{L}} \mathbf{y} \qquad [A.6]$$

where, **y**, **m** and **x** represent the vectors of standard Gaussian deviates, the mean of the Gaussian field and the vector of the new Gaussian field which has been calculated as a linear combination of the existing realizations, respectively. When non-Euclidean distance measures are used such as the measure of fit, it is observed that the problem becomes non-Gaussian and the new realizations generated by KL expansion do not reproduce the desired response being the measure of fit close to one. To overcome this problem, Scheidt et al. (2008) proposed (1) transforming the realizations into feature space using a radial basis function (RBF), (2) performing KL expansion in the feature space and (3) back transforming the realizations to Cartesian space by solving the pre-image problem.

Transforming the realizations to the feature space ($\hat{\mathbf{x}}_{d,i} \mapsto \boldsymbol{\varphi}\left(\hat{\mathbf{x}}_{d,i}\right)$) is a complex problem, and according to Scheidt et al. (2008), involves a high dimensional multivariate function $\boldsymbol{\varphi}$ whose determination is not easy. However, as shown in Equations [A.4], [A.5] and [A.6], to perform KL expansion, we only need eigenvalue decomposition of the matrix of dot-product in the feature space, which can be calculated by kernel RBF as:

$$K_{ij} = k\left(\hat{\mathbf{x}}_{d,i}, \hat{\mathbf{x}}_{d,j}\right) = \left\langle \boldsymbol{\varphi}\left(\hat{\mathbf{x}}_{d,i}\right), \boldsymbol{\varphi}\left(\hat{\mathbf{x}}_{d,j}\right) \right\rangle = \exp\left(- \frac{\left(\hat{\mathbf{x}}_{d,i} - \hat{\mathbf{x}}_{d,j}\right)^T \left(\hat{\mathbf{x}}_{d,i}, \hat{\mathbf{x}}_{d,j}\right)}{\sigma^2} \right) \qquad [A.7]$$

where, σ is the tuning parameter. The K_{ij}'s are collectively form the components of a matrix termed the Gram matrix **K**. In the feature space, new realizations can be generated based on the existing transformed realizations through KL expansion:

$$\mathbf{C}_\phi = \frac{1}{L}\boldsymbol{\Phi}\boldsymbol{\Phi}^T = \mathbf{V}_{C,\phi}\boldsymbol{\Lambda}_{C,\phi}\mathbf{V}_{C,\phi}^T \;\Rightarrow\; \boldsymbol{\varphi}_{new}\left(\mathbf{x}\right) = \mathbf{V}_{C,\phi}\boldsymbol{\Lambda}_{C,\phi}^{1/2}\mathbf{y} \qquad [A.8]$$

where, $\boldsymbol{\Phi}$ is the matrix of unknown existing transformed realizations into the feature space, $\mathbf{C}_\phi$ represents the covariance matrix of the transformed realizations, $\mathbf{V}_{C,\phi}$ is the matrix that involves the eigenvectors for covariance matrix, and $\boldsymbol{\Lambda}_{C,\phi}$ is the corresponding diagonal matrix of the eigenvalues. Similar to Equations [A.4], [A.5] and [A.6], the duality between the dot product matrix and the covariance matrix in the feature space can be used to derive the relationship between their eigenvalues and eigenvectors:

$$\boldsymbol{\Lambda}_{C,\phi} = \frac{1}{L}\boldsymbol{\Lambda}_\phi \quad , \quad \mathbf{V}_{C,\phi}^T = \boldsymbol{\Phi}^T\mathbf{V}_\phi\boldsymbol{\Lambda}_\phi^{1/2} \qquad [A.9]$$

where, $\mathbf{V}_\phi$ is the matrix involving the eigenvectors for the dot-product matrix, and $\boldsymbol{\Lambda}_\phi$ is the corresponding matrix of eigenvalues. Combining Equations [A.8] and [A.9], new realizations in the feature space can be expressed as linear combination of existing realizations also in the feature space:

$$\boldsymbol{\varphi}^{new}\left(\mathbf{x}\right) = \boldsymbol{\Phi}^T\boldsymbol{\alpha} \qquad [A.10]$$

where, a represents the vector of coefficients whose components can be calculated by:

$$\alpha_i = y_i\lambda_{\phi,i}^{1/2}\sum_{j=1}^{d} v_{\phi,i,j} \quad \text{with} \quad 2 \le d \le L \quad \text{and} \quad i = 1, ..., L$$

where, $v_{\phi,i,j}$ represents a component of the matrix of the eigenvectors for the dot-product matrix, y_i is a component of the vector of Gaussian deviates, and $\lambda_{\phi,i}$ is a component of diagonal matrix of eigenvalues for the dot-product matrix. The new set of realizations generated in the feature space must be back-transformed to Cartesian space. As the inverse of $\boldsymbol{\varphi}$ is not explicitly known, this back-transformation translates into an ill-posed inverse problem (also known as pre-image problem) that is formulated as an optimization problem:

$$\hat{\mathbf{x}}_{d,new} = \arg\min_{\hat{\mathbf{x}}_{d,new}} \left\| \boldsymbol{\varphi}\left(\hat{\mathbf{x}}_{d,new}\right) - \boldsymbol{\Phi}^T \boldsymbol{\alpha} \right\| \qquad [A.11]$$

Following the recommendations by Schoelkopf and Smola (2002), the fixed-point method can be used to solve the pre-image problem. Implementation of the fixed point method results in the expression for back-transformation of the newly generated realizations from the feature space to the MDS space (up to L dimensions):

$$\hat{\mathbf{x}}_{d,new} = \sum_{i=1}^{L} \beta_i\left(\hat{\mathbf{x}}_{d,i}\right)\hat{\mathbf{x}}_{d,i} \quad \text{with} \quad \beta_i\left(\hat{\mathbf{x}}_{d,i}\right) = \frac{\alpha_i k'\left(\hat{\mathbf{x}}_{d,new}, \hat{\mathbf{x}}_{d,i}\right)}{\sum_{j=1}^{L} \alpha_j k'\left(\hat{\mathbf{x}}_{d,new}, \hat{\mathbf{x}}_{d,j}\right)} \qquad [A.12]$$

where, k' is the derivative of the kernel RBF in Equation [A.7]. According to Scheidt et al. (2008), one can perform unconstrained optimization to find the newly generated realizations into the N dimensional Cartesian space. The unconstrained optimization is appropriate for Gaussian fields and is performed through using the same weights of Equation [A.12] to the realizations in the Cartesian space:

$$\mathbf{x}_{new} = \sum_{i=1}^{L} \beta_i\left(\hat{\mathbf{x}}_{d,i}\right)\mathbf{x}_i \quad \text{with} \quad \beta_i\left(\hat{\mathbf{x}}_{d,i}\right) = \frac{\alpha_i k'\left(\hat{\mathbf{x}}_{d,new}, \hat{\mathbf{x}}_{d,i}\right)}{\sum_{j=1}^{L} \alpha_j k'\left(\hat{\mathbf{x}}_{d,new}, \hat{\mathbf{x}}_{d,j}\right)} \qquad [A.13]$$

In Equations [A.12] and [A.13], the weights sum to one, which ensures the reproduction of conditioning static data at the data locations. Also, it is observed that the calculation of the weights depends on the new realizations (in the MDS

space) themselves. Thus, an iterative procedure is employed in this work to quantify the weights. To find the weights to generate a new realization, one can start from one of the existing realizations (mapped into the MDS space ($\hat{\mathbf{x}}_{d,j}$)) each time as an initial guess for $\hat{\mathbf{x}}_{d,new}$, and calculate a new vector of weights $\boldsymbol{\beta}$. The new vector of weights $\boldsymbol{\beta}$ is then used to find an updated $\hat{\mathbf{x}}_{d,new}$ using Equation [A.12]. This process is repeated until the difference in $\hat{\mathbf{x}}_{d,new}$ in consecutive iterations becomes small and the convergence is achieved. Then, the same weights are used in Equation [A.13] to find the desired number of output realizations, using a given number of input realizations in the Cartesian space. The methodology does not always result in governance, when calculating the weights. Also, some of the generated realizations may closely resemble the input realizations. So, a post-processing is almost always required. Despite these shortcomings, given the fact that the algorithm is significantly faster than conditioning new realizations to head data by inverse modeling, it can be considered a useful methodology.

Figures A-1-a, b, c show the projection of the 1000 Gaussian realizations into a 3D MDS space based on a Euclidean distance measure, measure of fit as the distance measure, and after transforming the non-Euclidean projection of the realizations into the feature space, respectively. It can be observed that the disorganized cloud of points in Figure A-1-b becomes close to linear, which is more appropriate for modeling (KL expansion), after transforming to the feature space.

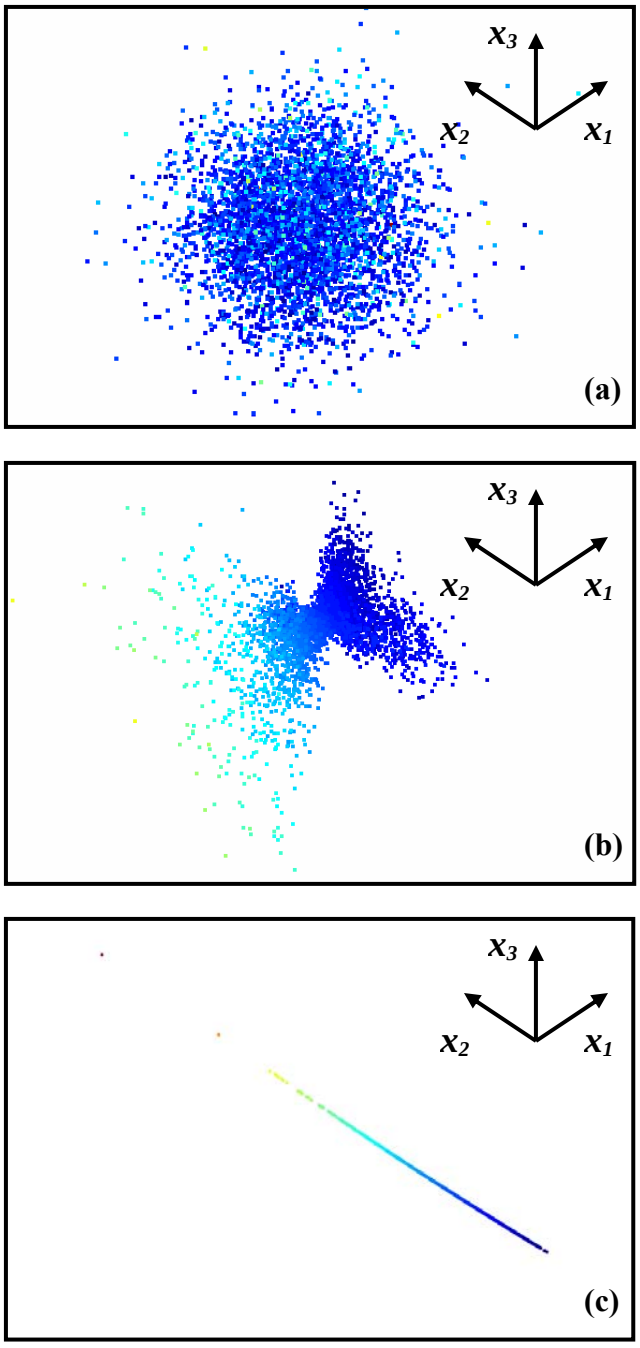


Figure A-1: The 3D mapped realizations into the MDS space (a) based on Euclidean distance between the realizations, (b) based on non-Euclidean distance (difference in measure of fit), and (c) after transforming to the feature space by kernel RBF.

As pointed out previously, when the level of measurement error is small, there may be only a small number of realizations that are accepted by screening. However, a large number of realizations are often needed for subsequent Monte Carlo type analysis. Given the low acceptance rate of the realizations, obtaining a large number of realizations with measure of fit close to one through running the SSC algorithm can be extremely demanding in terms of the CPU time. A faster alternative would be the distance-based simulation (explained above) to expand the set of acceptable realizations. As an example, two sets of 50 realizations that their values of measure of fit are closest to one are selected among the realizations that are conditioned to head data (by SSC) with error standard deviations of 0.10 m and 0.15 m. It should be noted that all the realizations conditioned to head data by SSC have Gaussian distributions. The sets of realizations are used as input for the distance based simulations, and the sets of acceptable realizations are expanded to 500 realizations. The difference in the values of measure of fit is applied as the non-Euclidean measure of distance. Figure A-2 shows the histograms of the values of the measure of fit for the expanded sets of realizations. Figure A-3 shows two example transmissivity realizations generated by distance based simulations, Figure A-4 shows the variogram reproduction check for 20 realizations generated by the SSC algorithm and distance based simulation in North-South direction. Comparing the realizations in Figure A-3 to Figure 5-3-a (Chapter 5), one can observe that the SSC algorithm and subsequent distance based simulation reproduce the overall structure of the reference transmissivity field.

Although distance based simulation may be considered useful in expanding the set of acceptable realizations, there are some serious issues that must be taken into account. First, solving the pre-image problem by fixed point method is only acceptable when the realizations are Gaussian; and even Gaussian realizations do not ensure the convergence of the approach.

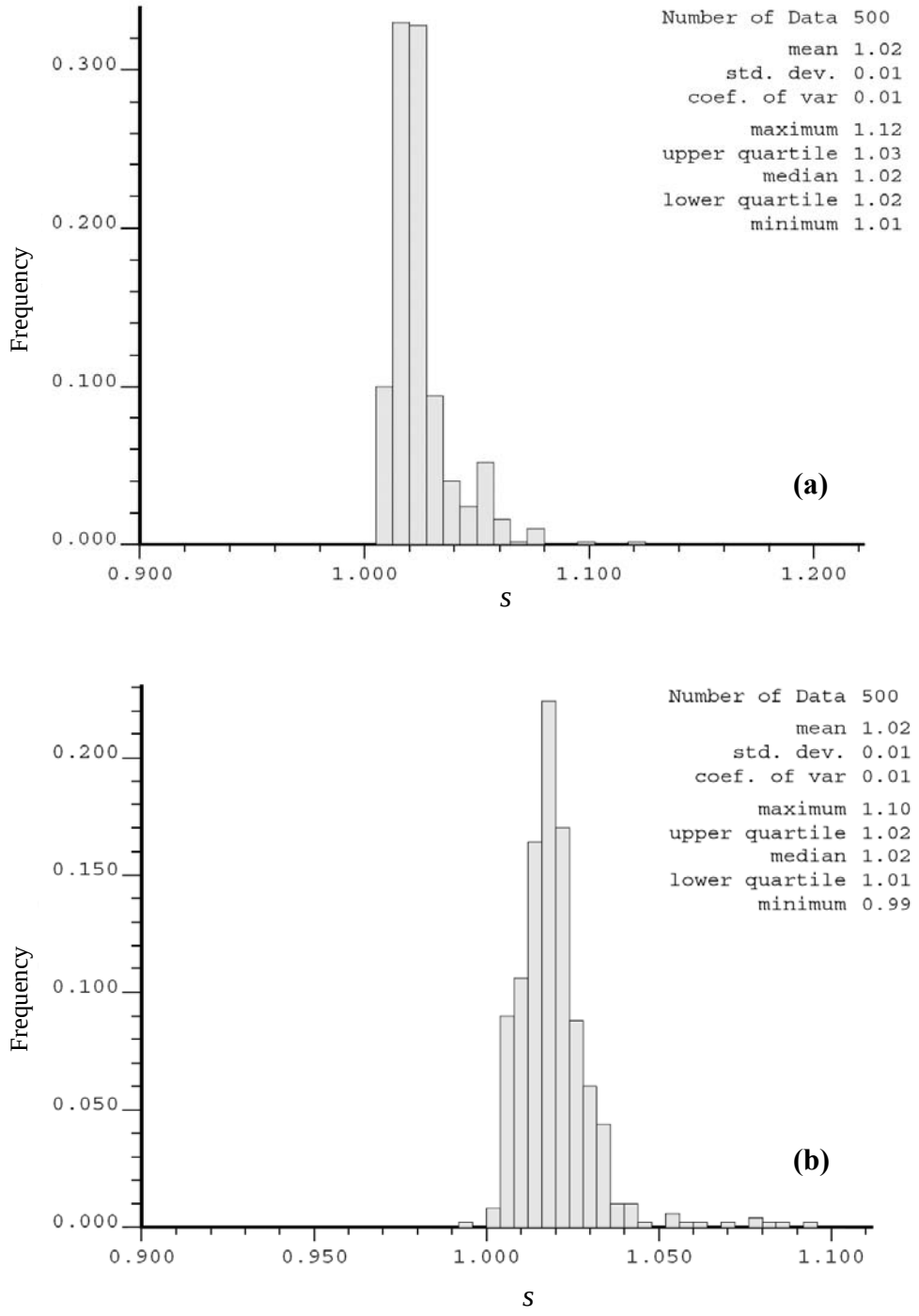


Figure A-2: The histograms of the values of measure of fit for the sets of 500 realizations generated by distance based simulation for (a) standard deviation of error σ equal to 0.10m, and (b) standard deviation of error σ equal to 0.15m.

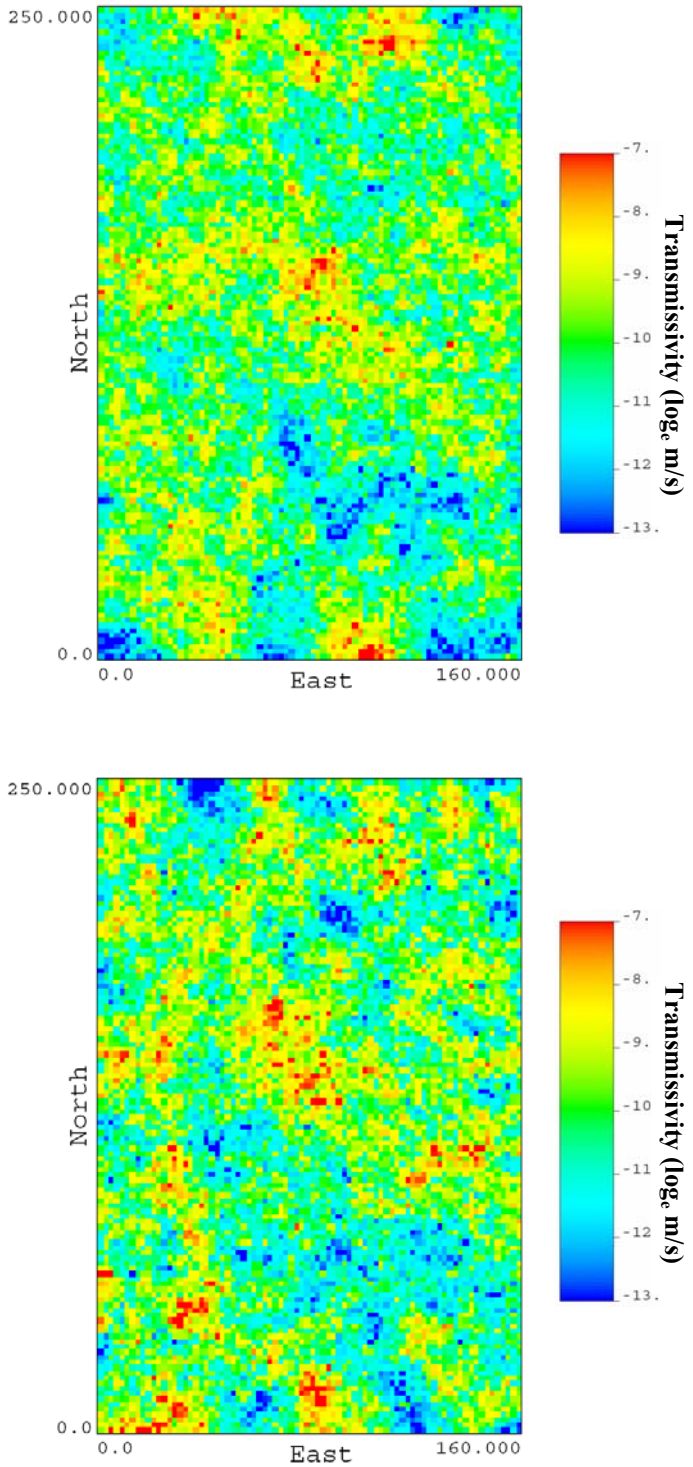


Figure A-3: Two example realizations generated by distance-based simulations for a standard deviation of error σ equal to 0.15m.

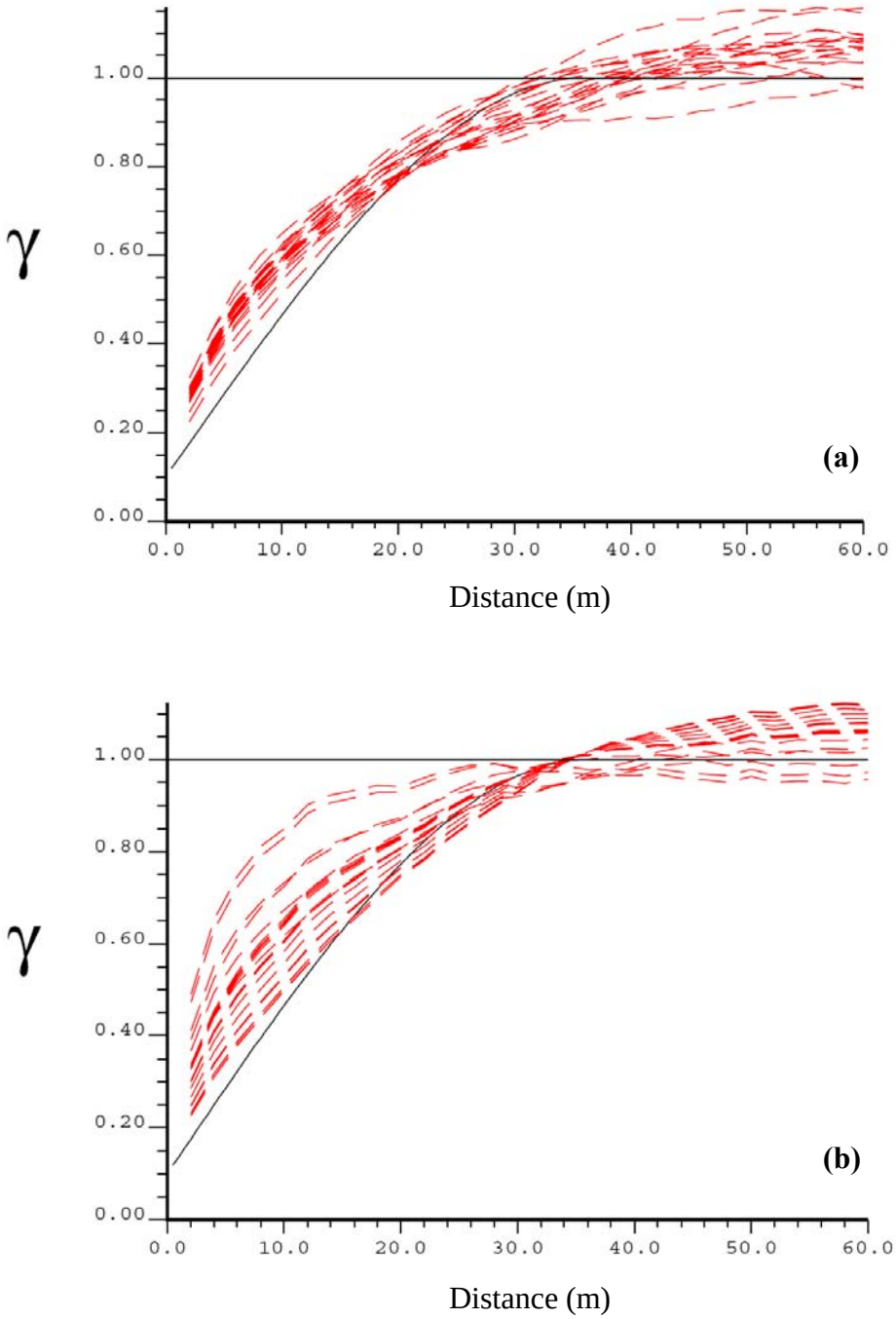


Figure A-4: The variogram reproduction check for (a) the first 20 realizations generated by the SSC algorithm and (b) the first 20 realizations generated by distance-based simulation.

Also, in terms of the reproduction of basic statistics, the generated realizations show a reasonable reproduction of variogram in most cases. The histogram reproduction, however, is not guaranteed. Although the Gaussian shape of the histogram and its mean are almost always reproduced, the standard deviation of output realizations may considerably over- or under-estimate the true standard deviation. It is also observed that many of the generated realizations may closely resemble a few of the input realizations. Due to these issues, and based on Author's experience, the distance-based simulation (in the current level of development) should be used with care and post-processing (acceptance/rejection) of the generated realizations is needed.

Appendix B

PARAMETER FILES FOR BOUNDARY SIMULATION AND PARAMETER ESTIMATION CODES

This appendix presents the parameter files for two codes developed for boundary modeling under uncertainty namely **mlimit2d** and **ubcalib2d** as well as the parameter file for the parameter estimation code **optkna**.

Figure B-1 shows the parameter file for **milimit2d** code that can be used to generate a large number of synthetic realizations to be applied in the subsequent calibration of uncertainty band as discussed in Chapter 3. Line 1 involves the name of the data file containing the location and status of the wells deemed inside and outside of the source zone. Line 2 involves the column numbers. Line 3 involves the name of the data file including the location and status of the control points. Line 4 involves the associated column numbers. Lines 5 and 6 involve the discretization parameters for the 2D domain. The number of realizations to be generated is entered in line 7. Line 8 involves the random number seed. Line 9 involves the starting angle as well as angle resolution for the directional search. The starting angle and in particular the angle resolution control the short-scale variations in the shape of the generated synthetic plumes. Line 10 involves the nugget and exponent for inverse distance interpolation. Line 11 gives the upper and lower bounds for the search radius. The optimum search radius is found within this range. Line 12 involves the number of increments for search radius. Lines 13 and 14 involve the file names for the debugging output and the output file.

```
                  Parameters for MLIMIT2D
                  ***********************

START OF PARAMETERS:
1    ../data/wells.dat                -file with location and status of existing wells
2    1   2   3                        -   columns for X and Y coordinates and status
3    ../data/cpw.dat                  -file with location and status of control points
4    1   2   3                        -   columns for X and Y coordinates and status
5    30     0.5     1.0               -nx,xmn,xsiz
6    30     0.5     1.0               -ny,ymn,ysiz
7    100                              -number of realizations to generate
8    69069                            -random number seed
9    0.0    24                        -start and resolution of directional search angle
10   0.2    1.5                       -nugget and exponent for ID interpolation
11   20.    100.                      -   minimum and maximum search radius
12   160                              -   number of increments for search radius
13   mlimit2d.dbg                     -file for debugging output
14   mlimit2d.out                     -file for output particle paths
```

Figure B-1: The parameter file for **mlimit2d** code

Figure B-2 shows the parameter file for **ubcalib2d** code that uses multiple realizations of synthetic limits and calibrates the uncertainty band for the given well arrangement and the desired quintiles by a simplex optimization approach. Line 1 involves the name of the data file containing the location and status of the wells deemed inside and outside of the source zone. Line 2 involves the associated column numbers. Line 3 involves the name of the file with control points. The column numbers are entered in line 4. The name of the file with multiple synthetic limits (generated by **mlimit2d** code) is entered in line 5. Line 6 involves the associated column number. The number of realizations to be used in optimization is entered in line 7. Lines 8 and 9 involve the discretization parameters. Line 10 involves the number of control points to be assigned at the boundaries. Line 11 involves the number of quantiles to be used in optimization. Line 12 involves a list of quantiles. The initial guesses for scaling and separation factors are given in line 13. Line 4 involves the characteristic length the maximum change in the separation factor in the first optimization iteration. Line 15 involves the random number seed. The optimization tolerance and the minimum relative improvements (in four consecutive iterations) are given in line 16. Line 17 involves the maximum number of iterations for optimization. Line 18 involves the nugget and exponent for inverse distance interpolation. Lines 19 and 20 involve the maximum and minimum search radius and the number of increments for

search radius to be used in calibration of search radius. Lines 21 and 22 involve the name of the files for debugging file and the output file.

```
                  Parameters for UBCALIB2D
                  ************************

START OF PARAMETERS:
1  ../data/wells.dat                -file with existing wells
2  1   2   3                        -   columns for X and Y and status
3  ../data/cpw.dat                  -file with control points
4  1   2   3                        -   columns for X and Y and status
5  ../data/mlimit2d.out             -file with multiple synthetic limits
6  3                                -   column for IDF
7  100                              -   number of realizations to use
8  30     0.5     1.0               -nx,xmn,xsiz
9  30     0.5     1.0               -ny,ymn,ysiz
10 100    100                      -number of control points on boundaries
11 5                                -number of quantiles to use
12 0.1  0.3  0.5  0.7  0.9          -   quantiles to use in calibration
13 0.5     0.0                      -initial guess scaling, seperation factors
14 5.0                              -characteristic length
15 69099                            -random number seed
16 0.003   0.01                     -optimization target,min rel improvement
17 50                               -maximum number of iterations
18 0.2    1.5                       -nugget and exponent for ID interpolation
19 25.    100.                      -   minimum and maximum search radius
20 160                              -   number of increments for search radius
21 ubcalib2d.dbg                    -file for debugging output
22 ubcalib2d.out                    -file for output
```

Figure B-2: The parameter file for **ubcalib2d** code

Figure B-3 shows the parameter file for the parameter estimation code **optkna** that is used to jointly estimate the values of dissolution rate and first-order biodegradation rate. Line 1 involves the name of the file with observed concentrations. Line 2 involves the column numbers for coordinates, time step, solute concentration (in mg/L) and the associated coefficient of variation. Lines 3 and 4 involve the file name for initial conditions and the associated solute concentration. Lines 5 and 6 involve the file name for source zone distribution and the soil concentration. The trimming limits are given in line 7. Line 8 involves the values of dry soil density and total porosity. Lines 10 and 11 involve the substrate solubility and mass fraction as well as the molecular weights for the solute and the tracer. These parameters are used to control the dissolution of the substrate from the NAPL source. The initial guess as well as the prior value for the dissolution rate constant is given lines 11 and 12. The standard deviation

associated with the dissolution rate is given in line 13. Lines 14 and 15 involve the filename for the transmissivity data (in natural logarithmic units) and the associated column number. The value of effective porosity is given in line 16. Lines 17 and 18 involve the file with calculated heads and the associated column number. Lines 19 to 22 involve the names and column numbers for flow and transport boundary condition. Line 23 and 24 involve the discretization parameters. Line 25 involves the lower and higher number of random particles per cell (Chapter 4). Line 26 involves the random number seed. Lines 28 and 29 involve the relative cell concentration gradient and the absolute minimum and maximum number of particles per cell, as well as the value of grid Courant number. Line 30 involves the longitudinal and transverse dispersivities. Lines 31, 32 and 33 involve the initial guess for biodegradation rate constant, prior value for biodegradation rate constant and the associated standard deviation in natural logarithmic units. Line 34 involves the number of time steps and total simulation time (in days). Line 35 involves the flag to report the sensitivities. Line 36 involves the output mode (in terms of being in arithmetic or logarithmic units). The output file name for sensitivities, the number of time steps and the time steps to report the sensitivities are given in lines 37, 38 and 39. The minimum value for concentration is given line 40. In every time step, all cells/particle with concentrations below this value are set to zero. Line 41 and 42 involves the file names for debugging output and the debugging level. Line 43 involves the filename for optimization results. The optimization parameters are entered in the lines 44 to 48. The maximum number of iterations is entered in line 44. The absolute minimum and maximum values of dissolution rate and first-order biodegradation rates are entered in lines 45 and 46. The maximum allowed change for each parameter is given in line 47 and the overall relative tolerance (relative to initial value of the objective function), minimum relative tolerance in consecutive iterations and minimum number of times to consecutive iterations are given in line 48.

```
                  Parameters for OPTKNA
                  *********************

START OF PARAMETERS:
1    ../data/observ.dat              -file with observed concentrations
2    1   2   3   4   5               -columns for X,Y,time step,conc(mg/L),cov
3    ../data/initial.dat             -file with initial condition
4    3                               -   column for solute conc(gr/cm3)
5    ../data/source.dat              -file with source zone distribution
6    3                               -   column for NAPL conc(gr/gr)
7    -1.0e21   1.0e21                -   trimming limits
8    1.6       0.35                  -dry soil density(gr/cm3),total porosity
9    0.00178   0.01                  -substrate solubility(gr/cm3),mass fraction
10   78.1      101.56                -substrate,tracer molecular weight(gr/mole)
11   0.035                           -initial guess for dissolution rate constant
12   0.035                           -prior value for dissolution rate constant
13   1.00                            -dissolution rate standard deviation(log_e)
14   ../data/Kxy.dat                 -file with transmissivity data
15   3                               -   column for Txy
16   0.3                             -effective porosity
17   ../data/heads.out               -file with calculated heads
18   1                               -   column for calculated head
19   ../data/FBC.dat                 -file with flow boundary condition
20   1   2                           -   columns for ID and head/flux
21   ../data/TBC.dat                 -file with flow boundary condition
22   1   2                           -   columns for ID and TBC
23   30     0.5     1.0              -nx,xmn,xsiz
24   30     0.5     1.0              -ny,ymn,ysiz
25   0      12                       -lower and higher no of random cell particles
26   67097                           -   random number seed
27   -1.0e5                          -relative cell concentration gradient
28   4      24                       -minimum and maximum no of particles
29   1.0                             -grid courant number
30   3.0    0.6                      -longitudinal and transverse dispersivity(m)
31   0.0025                          -initial guess for biodegradation rate constant
32   0.0025                          -prior value for biodegradation rate constant
33   0.50                            -biodegradation rate standard deviation(log e)
34   365    3650                     -number of time steps, simulation time(days)
35   1                               -report the sensitivities (0=no,1=yes)
36   0                               -   output mode (0=arithm,1=logarithm)
37   sensit.out                      -   output file for sensitivities
38   3                               -   number of times to report sensitivities
39   100   200   300                 -   time steps to report sensitivities
40   0.001                           -minimum concentration
41   optkna.dbg                      -file for debugging output
42   1                               -   debugging level
43   optkna.out                      -file for optimization results
44   25                              -   maximum number of iterations
45   1.0e-5    1.0                   -   absolute min,max dissolution rate(1/day)
```

Figure B-3: The parameter file for **optkna** code