Uncertainty Analysis and the Identification of the Contaminant Transport and Source Parameters for a Computationally Intensive Groundwater Simulation

by

Yong Yin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada, 2009

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Transport parameter estimation and contaminant source identification are critical steps in the development of a physically based groundwater contaminant transport model. Due to the irreversibility of the dispersion process, the calibration of a transport model of interest is inherently ill-posed, and very sensitive to the simplification employed in the development of the lumped models. In this research, a methodology for the calibration of physically based computationally intensive transport models was developed and applied to a case study, the Reich Farm Superfund site in Toms River, New Jersey.

Using HydroGeoSphere, a physically based transient three-dimensional computationally intensive groundwater flow model with spatially and temporally varying recharge was developed. Due to the convergence issue of implementing saturation versus permeability curve (van Genuchten equation) for the large scale models with coarse discretization, a novel flux-based method was innovated to determined solutions for the unsaturated zone for soil-water-retention models. The parameters for the flow system were determined separately from the parameters for the contaminant transport model. The contaminant transport and source parameters were estimated using both approximately 15 years of TCE concentration data from continuous well records and data over a period of approximately 30 years from traditional monitoring wells, and compared using optimization with two heuristic search algorithms (DDS and MicroGA) and a gradient based multi-start PEST.

The contaminant transport model calibration results indicate that overall, multi-start PEST performs best in terms of the final best objective function values with equal number of function evaluations. Multi-start PEST also was employed to identify contaminant transport and source parameters under different scenarios including spatially and temporally varying recharge and averaged recharge. For the detailed, transient flow model with spatially and temporally varying recharge, the estimated transverse dispersivity coefficients were estimated to be significantly less than that reported in the literature for the more traditional approach that uses steady-state flow with averaged, less physically based recharge values. In the end, based on the Latin Hypercube sampling, a methodology for comprehensive uncertainty analysis, which accounts for multiple parameter sets and the associated correlations, was developed and applied to the case study.

Acknowledgements

I would like to express my sincere thanks and appreciation to my supervisor, Dr. Jon Sykes, for guidance, and for providing me with excellent facilities to pursue my work, and for ensuring financial support throughout my studies.

I would also like to express my special and great thanks to P.Eng. Stefano, Normani for his valuable advice and friendship throughout my study. I appreciate his important suggestions and countless patient hours spent on my thesis work. His unrelenting support and outstanding work ethic also made the work much enjoyable. I also thank Dr. Mikko Jyrkama for helping me set up the model. I am also grateful to Dr. Bryan Tolson for providing me with the DDS method and many useful comments.

My sincere thanks are extended to the members of my committee for their constant interest in my work, and for many insightful comments during this research. They are: Dr. Neil Thomson, Dr. Bill Annable, Dr. Edward Sudicky, and Dr. Allan Woodbury.

I would also like to acknowledge financial support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and Nuclear Waste Management Organization (NWMO). This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca).

Special thanks go to the graduate students in the Water Resources Group for their help, the laughs, and the memories.

Finally, I would like to express my deepest gratitude to my parents and my wife, Hua Guan for their constant love and support.

Contents

List of Tables

List of Figures

List of Abbreviations

ANN	Artificial Neural Networks
BTC	BreakThrough Curve
CDF	Cumulative Distribution Function
CERCLA	Comprehensive Environmental Response, Compensation, and Liability Act
CI	Confidence Interval
COV	Coefficient Of Variation
DAFB	Dover Air Force Base
DDS	Dynamically Dimensioned Search
DEM	Digital Elevation Model
DNAPL	Dense NonAqueous Phase Liquid
DTML	Dover Township Municipal Landfill
FEFLOW	Finite Element subsurface FLOW system
FORM	First-Order Reliability Method
FOSM	First-Order Second Moment method
GIS	Geographic Information System
GL	Gloucester Landfill
GLUE	Generalized Likelihood Uncertainty Estimation
HELP3	Hydrologic Evaluation of Landfill Performance Model Version 3
LHC	Latin HyperCube
MCMC	Markov chain Monte Carlo
MIBBE	Marching-Jury Backward Beam Equation
MicroGA	Micro Genetic Algorithm
MMR	Massachusetts Military Reservation
MPI	Malcolm Pirnie Inc.
MRE	Minimum Relative Entropy
NAPL	NonAqueous Phase Liquid
NJDEP	New Jersey Department of Environmental Protection
NJGS	New Jersey Geological Survey

NLS	Non-linear Least-Squares method
PCE	Tetrachloroethylene
PEST	Parameter ESTimation method
PGA	Progressive Genetic Algorithm
PGL	Projected Augmented Lagrangian
SAN	Styrene-AcryloNitrile
QR	Quasi-Reversibility
TR	Tikhonov Regularization
SCE	Shuffled Complex Evaluation
SCS	Soil Conservation Service
Sharcnet	Shared Hierarchical Academic Research Computing NETwork
SORM	Second-Order Reliability Method
SSE	Sum of Square Errors
SWIFT3	Sandia Waste-Isolation Flow and Transport Model
TCE	Trichloroethylene
U.S.EPA	United States Environmental Protection Agency
UCC	Union Carbide Corporation
USGS	United States Geological Survey
UWTR	United Water Toms River
VOC	Volatile Organic Compound

List of Symbols

α	air entry pressure head [L^{-1}]
α_l	longitudinal dispersivity [L]
α_t	transverse dispersivity [L]
β	pore size distribution index [dimensionless]
β_p	compressibility of the bulk aquifer [M L T^2]
β_w	compressibility of water [M L T^2]
δ_{ij}	Kronecker delta [dimensionless]
Γ_{ex}	volumetric fluid exchange rate [L^3 L^{-3} T^{-1}]
γ	specific weight of water [M L^{-2} T^{-2}]
μ	viscosity of the groundwater [M $L^{-1}T^{-1}$]
ω_m	volumetric fraction of the total porosity [dimensionless]
$\overline{C}_{obs,i}$	average of the observed concentrations [M L^{-3}]
ψ	pressure head [L]
ρ	density of the groundwater [M L^{-3}]
ρ_b	bulk density [M L^{-3}]
σ_e^2	error variance between the observed and simulated concentrations
σ_{obs}^2	variance of the observed concentrations
τ	tortuosity [dimensionless]
θ_n	volumetric fraction of NAPL in the system [dimensionless]
θ_n^0	initial volumetric fraction of NAPL in the system [dimensionless]
θ_r	residual volumetric water content [dimensionless]
θ_s	saturated volumetric water content [dimensionless]
$\widehat{k}$	mass transfer rate coefficient [T^{-1}]
$\|q\|$	magnitude of the Darcy flux [L T^{-1}]
a_0	interfacial area per unit volume of porous medium [L^{-1}]
C	contaminant concentration [M L^{-3}]
$C_0(t)$	residual NAPL zone solute concentration at time t [M L^{-3}]
$C_{obs,i}$	observed concentrations [M L^{-3}]

C_s	NAPL solubility [M L^{-3}]
$C_{sim,i}$	simulated concentrations [M L^{-3}]
C_{solid}	adsorbed concentration in the solid phase [dimensionless]
$C_{TCE,initial}$	initial source TCE concentration at the source [M L^{-3}]
d_{50}	median grain diameter of porous medium [L]
D_{ij}	hydrodynamic dispersion tensor [L^2 T^{-1}]
D_L	free liquid diffusivity of the organic species in water [L^2 T^{-1}]
D_w	free solution diffusion coefficient [L^2 T^{-1}]
g	gravitational acceleration [L T^{-2}]
J_0	interphase mass transfer flux [M L^{-2} T^{-1}]
K'	equilibrium distribution coefficient [L^3 M^{-1}]
k_f	mass transfer coefficient [L T^{-1}]
$K_{ij,eff}$	effective unsaturated hydraulic conductivity [L T^{-1}]
K_{ij}	hydraulic conductivity tensor [L T^{-1}]
k_{ij}	permeability tensor of the porous medium [L^{-2}]
K'_{source}	equivalent K' at the source [L^3 M^{-1}]
k_r	relative permeability [dimensionless]
$M(t)$	NAPL mass remaining at time t [M]
M_0	initial NAPL residual mass [M]
n	porosity of the porous media [dimensionless]
N	shaping factor
N_{obs}	number of the observations
Q	volumetric fluid flux per unit volume [L^3 L^{-3} T^{-1}]
q_i	Darcy flux [L T^{-1}]
R	retardation factor [dimensionless]
Re	Reynolds number using Darcy velocity [dimensionless]
Re'	Reynolds number using interstitial velocity [dimensionless]
S_n	NAPL saturation [dimensionless]
S_w	degree of water saturation [dimensionless]
S_{wr}	residual saturation [dimensionless]
S_e	effective saturation [dimensionless]

S_s	specific storage coefficient of the porous medium [L^{-1}]
Sh	Sherwood number [dimensionless]
U_i	uniformity index [dimensionless]
v_i	linear velocity [L T^{-1}]
w_i'	equivalent weighting factor [dimensionless]
w_i	weighting factor [dimensionless]
x	strength of deemphasizing factor [dimensionless]
x/d_p	distance into the residual saturation zone [dimensionless]
z	elevation head [L]

Chapter 1

Introduction

Groundwater is a vital national resource, which in 1995 supplied drinking water for 46% of the overall population in the United States and 99% of the population in the rural areas [*U.S. EPA*, 2002]. In some parts of the world, groundwater serves as the only reliable source of drinking and irrigation water. However, this vital resource is vulnerable to contamination from specific sources such as leaking underground storage tanks, septic systems, spills, landfills, and industrial facilities. With the development of technology, more and more human activities are polluting the groundwater system.

In the 1970s, the events that happened in Love Canal, New York and Valley of the Drums, Kentucky attracted people's attention and showed that past contaminant disposal practices might endanger public health and the environment. To clean up the sites contaminated by hazardous wastes, the U.S. Congress passed the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA or Superfund) in 1980. The law authorizes remedial response action at the sites on the National Priorities List (NPL) with highest hazard rank. To date, there are a large portion of Superfund sites on the NPL that are associated with groundwater contamination.

Due to the limited funds available for the cost of cleanup, the U.S.EPA (United States Environmental Protection Agency) must investigate and find the responsible private par-

ties liable for the contamination, and partition the remediation costs among them in a way consistent with their degree of culpability. Unfortunately, because of the slow movement of groundwater through aquifers, groundwater contamination is frequently discovered long after it has occurred. It is not easy to obtain sufficient evidence in identification of responsible private parties through the commonly used forensic techniques, such as chemical profiling, aerial photography interpretation, federal agency records, private parties' records [*Morrison*, 2000a,b; *Atmadja and Bagtzoglou*, 2001b], compositional analysis, tracer use, and contaminant distribution [*Michalak*, 2001]. This could be confirmed by the fact that "in 1994, the National Academy of Sciences estimated that over a trillion dollars, or approximately $4,000 per person in the U.S., would be spent in the next 30 years on clean-up of contaminated soil and groundwater" [*U.S. EPA*, 1999]. There is a need for additional forensic evidence to aid in the identification process, such as the source release history obtained through inverse modelling methods.

As the direct input of contaminants into the groundwater system, the source release information is essential to gain further insight about the current contaminant distribution and predict the future spread of a contaminant plume. These could provide important information for almost all of the groundwater problems concerning contaminant transport, including determination of responsible parties' liability for damage caused to downstream victims, reliability analysis of contamination issues and the design of remediation systems [*Michalak*, 2001; *Michalak and Kitanidis*, 2004b]. One example is to optimize the sampling well deployment. Generally, the existing sampling wells do not have the most efficient locations for identifying the contaminant sources in terms of extent, location, and duration. To reduce the relatively high cost of installing monitoring wells, the new sampling wells can be optimally placed with the aid of preliminary contaminant source data. The additional measurements can, in turn, be utilized to more efficiently and accurately update the source information activity [*Mahar and Datta*, 1997].

Acknowledging the significance of contaminant source information, considerable work has been devoted to it during the past 20 years (refer to Section 2.2 and Table 2.1of this thesis for the details). Many inverse modelling methods, ranging from deterministic to stochastic approaches, have been developed to reconstruct the source release history.

However, many of the attempts to characterize a contaminant source are restricted to lumped groundwater flow and transport models with simplified representation of hydrologic and transport processes. For instance, *Skaggs and Kabala* [1994] employed Tikhonov regularization to reconstruct the nonnegative release history of a plume for a 1-D contaminant transport problem in a saturated homogeneous aquifer. *Woodbury and Ulrych* [1996]; *Woodbury et al.* [1998] applied the minimum relative entropy (MRE) to a practical case study with a 3-D constant velocity and constant dispersivity system at the Gloucester Landfill in Ontario, Canada. *Alapati and Kabala* [2000] utilized a non-linear least-squares method (NLS) without regularization to determine the parameters in the source release function for a 1-D homogeneous system. A more detailed discussion of inverse modelling attempts will be presented in Chapter 2. Due to the irreversibility of the dispersion process, the calibration of transport model of interest is inherently ill-posed, and very sensitive to the simplification employed in the development of the lumped models. The case study, that is investigated in this thesis, require the analysis of a regional-scale three-dimensional variably saturated transient flow system. The assumption of steady-state flow would yield non-representative results. Typical of field studies, measured concentrations contain uncertainty and are available for only the later stage of contamination.

Numerical modelling of physically based, field-scale groundwater flow and transport processes requires high computing demand. Hundreds of realizations of the forward modelling are needed for most of the optimization algorithms. Thus, computation in the calibration of such a model is exceptionally intensive. To date, few reported studies have compared the algorithm performance for complex, field-scale applications in the groundwater field.

For many field-scale groundwater contaminant problems, the expense of data acquisition combined with the inherent difficulty of obtaining representative values can result in a sparse or incomplete data base. For many sites, data are restricted to the property owned by the contaminators. Observed values can be uncertain. This is particularly true at the margins of a plume where concentrations at instrument detection limit are obtained from wells with a large screened interval. This sparsity of information can lead to high uncertainty in the estimation of model parameters. Even a perfectly calibrated model

cannot entirely remove uncertainty. Neglecting this uncertainty tends to cast serious doubt on the credibility of the results. For instance, in the forward modelling case of Woodrow Stering et al. versus Velsicol Chemical Corporation, two groups of reputable modelers hired by plaintiffs and defendants presented different results only caused by different loading rate factors [*Michalak*, 2001]. Owing to the fact that inverse modelling is even more sensitive to the parameter perturbation, performing a comprehensive uncertainty analysis is essential to facilitate the partitioning of the responsibility of clean-up of contaminated sites and the prediction of the ongoing contaminant plume. Additionally, ill-posedness of the inverse modelling results in the non-uniqueness of the estimated parameters with correlation. However, no research has been carried out to quantify the uncertainty, not only incorporating multiple parameter sets but also honouring the parameter correlations in the groundwater field.

1.1 Research Objectives

The primary objective of this thesis is the development of an effective methodology for the practical calibration of a physically based, field-scale, computationally intensive groundwater flow and contaminant transport model. The specific objectives of this research were to:

1. Develop a computationally efficient approach to calibrate a three-dimensional, physically based, contaminant transport model;
2. Compare the performance of different optimization algorithms in the analysis of a field-scale computationally intensive groundwater problem through the application to the case study: the Reich Farm Superfund site;
3. Investigate the algorithms for an ill-posed problem and a noisy objective function response surface;
4. Evaluate the impact of the simplified representation of the hydrologic processes on the estimated parameter values;

5. Evaluate the impact of different contaminant source characterization on the calibration results;

6. Quantify the influence of recovered uncertain contaminant source on the forward modelling of contaminant transport.

1.2 Outline of the Thesis

This thesis consists of five additional chapters, as follows:

Chapter 2 provides the underlying theory employed by HydroGeoSphere to describe variably-saturated groundwater flow and contaminant transport, a review of the parameter estimation approaches, and a detailed discussion of the selected optimization algorithms.

Chapter 3 presents the development of a three-dimensional, variably saturated, transient groundwater flow model for the Reich Farm Superfund site near Toms River, New Jersey (Figure 3.1). The flow model, which incorporates a spatially and temporally varying recharge [*Jyrkama et al.*, 2002] derived by the hydrologic model HELP3 in conjunction with GIS, was manually calibrated through trial and error by *Sykes and Normani* [2002]. A particle tracking was conducted to estimate the travel time from the Reich Farm site to the Toms River Municipal Parkway well field. This thesis builds on the work of *Jyrkama et al.* [2002] and *Sykes and Normani* [2002] with an emphasis on the contaminant system. The revision of the spatially and temporally varying recharge model and the flow domain calibration procedure is beyond the scope of this thesis.

Chapter 4 presents the formulation of the optimization problem, including objective function definition and the decision variables with box constraints. Multi-start Parameter ESTimation method (PEST), Dynamically Dimensioned Search (DDS) and Micro Genetic Algorithm (MicroGA) parameters estimation algorithms were employed to calibrate the parameters of the transport model for the transient flow field given by the flow model [*Sykes and Normani*, 2002] described in Chapter 3. A comparison among those algorithms was conducted in terms of the performance measure and the computation time. The transport

models under four different scenarios with spatially and temporally varying and averaging recharge were calibrated to evaluate the impact of the averaging scheme on the estimated parameter values. Additionally, the impact of different types of contaminant source characterization on the estimated transport parameters and the calibration performance was evaluated. The investigation of the behaviour of the three parameter estimation algorithms for a computationally simple, generic groundwater system is beyond the scope of this thesis. The scope is restricted to solely the investigation of a computationally intensive problem.

Chapter 5 presents a Latin hypercube simulation to assess the uncertainty of the TCE breakthrough curves at selected pumping wells and the influent to the air stripper for the TCE contaminated wells at the Parkway well field. The uncertainty of several transport parameters was characterized by probability density functions obtained from the calibration results in Chapter 4. The Latin hypercube sampling method honours the parameter correlations as well as accounts for multiple parameter sets with the likelihood estimated on a basis of its associated performance measure.

Chapter 6 presents the conclusions of this thesis, and recommendations for future work.

Chapter 2

The HydroGeoSphere Model and Parameter Estimation Algorithms

2.1 HydroGeoSphere Model Description

In this research, the code-HydroGeoSphere [*Therrien et al.*, 2004b], developed from FRAC3DVS [*Therrien et al.*, 2004a], has been chosen, because the source code is accessible, the model is readily available at no cost, and it is provided with technical support by the developers. HydroGeoSphere is a distributed-parameter and fully-integrated model designed to solve three-dimensional variably-saturated subsurface and surface flow and solute transport equations in granular or discretely-fractured media. Although the model is capable of handling dual porosity simulations, for the purpose of this computationally intensive regional study, an equivalent porous media approximation was assumed. To solve the density-dependent non-linear flow equation, a Picard iterative solver is utilized; however, only density-independent flow is investigated in this work.

HydroGeoSphere uses the control volume finite element technique which not only conserves mass, but also is capable of precisely delineating the details of boundaries. Upstream weighting of the relative permeabilities and fully-implicit temporal weighting scheme em-

ployed in HydroGeoSphere ensure a monotone solution for saturated flow models. There is no stability criterion for the time step or grid spacing for the implicit solution. Due to the first-order accuracy of upstream weighting, an adaptive time stepping procedure is used to enhance the efficiency of the solution process. Given these beneficial features, a few modifications were still necessary for transport model calibration, and are discussed further in the Sections 3.4.5 and 4.3.

The assumptions on which HydroGeoShpere is based in order to solve the governing flow equation include: porous media grains are non-deformable; the system being described in the model is under isothermal conditions; and the air phase, where present, is infinitely mobile. HydroGeoShpere is formulated in terms of Richards' Equation. In order to describe the three-dimensional variably-saturated flow, the following form of Richards' Equation was used:

$$- \bigtriangledown \cdot (\omega_m q_i) + \sum \Gamma_{ex} \pm Q = \omega_m \frac{\partial}{\partial t}(\theta_s S_w) \tag{2.1}$$

where ω_m [dimensionless] is the volumetric fraction of the total porosity of the porous medium, q_i [L T^{-1}] is the fluid flux, Γ_{ex} represents the volumetric fluid exchange rate [L^3 L^{-3} T^{-1}] between the subsurface domain and any applicable model supported domain types, Q [L^3 L^{-3} T^{-1}] represents the volumetric fluid flux per unit volume, S_w [dimensionless] represents the degree of water saturation, and θ_s is the saturated water content [dimensionless]. The volumetric fraction always will be equal to 1.0 for single porosity calculations. The term Q is used to represent a source or a sink. In Richards' Equation, the fluid flux q_i [L T^{-1}] is given by:

$$q_i = -K_{ij} k_r \bigtriangledown (\psi + z) \tag{2.2}$$

where $k_r = k_r(S_w)$ represents the relative permeability [dimensionless] of the porous medium with respect to the degree of water saturation (S_w), ψ is the pressure head [L], and z is the elevation head [L]. The saturated water content is assumed to be equal to the porosity. The hydraulic conductivity tensor K_{ij}[L T^{-1}] is given by

$$K_{ij} = \frac{\rho g}{\mu} k_{ij} \tag{2.3}$$

where g is the gravitational acceleration [L T^{-2}], μ is the viscosity of the groundwater [M L^{-1} T^{-1}], k_{ij} is the permeability tensor of the porous medium [L^{-2}] and ρ is the density of the groundwater [M L^{-3}]. While not invoked in this study, the density of the groundwater can be dependent on a concentration C [M L^{-3}] of a given solute such that $\rho = \rho(C)$.

Water saturation is related to the water content θ [dimensionless] by the relationship

$$S_w = \frac{\theta}{\theta_s} \tag{2.4}$$

In Equation (2.1), the parameter Γ_{ex} is expressed as a unit volume of the other domain types. The possible domain types may be surface wells, tile drains, discrete fractures or dual continuum.

In Equation (2.1), the primary solution variable is the pressure head. To solve for this variable, a constitutive relationship is required to relate the pressure head to other secondary variables such as the saturation and permeability terms. The saturation can be related to the pressure using the [*Brooks and Corey*, 1964] relationship:

$$\begin{aligned} S_w = S_{wr} + (1 - S_{wr})|\alpha\psi|^{-\beta} \text{ for } \psi < -1/\alpha \\ S_w = 1 \text{ for } \psi \geq -1/\alpha \end{aligned} \tag{2.5}$$

and the relative permeability is described by:

$$k_r = S_e^{2/\beta+3} \tag{2.6}$$

where α [L^{-1}] is the inverse of the air entry pressure head, β [dimensionless] is a number which characterizes the pore size distribution, and S_e is the effective saturation. The effective saturation is determined by $S_e = (S_w + S_{wr})/(1 - S_{wr})$, with S_{wr} referring to the residual water saturation [dimensionless].

The following pressure-saturation relationship was described by [*Van Genuchten*, 1980]:

$$\begin{aligned} S_w = S_{wr} + (1 - S_{wr}) \left[1 + |\alpha\psi|^{\beta}\right]^{-\nu} \text{ for } \psi < -1/\alpha \\ S_w = 1 \text{ for } \psi \geq -1/\alpha \end{aligned} \tag{2.7}$$

with the permeability being described by:

$$k_r = S_e^{(l_p)} \left[1 - \left(1 - S_e^{1/\nu}\right)^{\nu}\right]^2 \tag{2.8}$$

where:

$$\left(v = 1 - \frac{1}{\beta}\right), \quad \beta > 1 \tag{2.9}$$

and where α and β are obtained by fitting (2.7) and (2.8) to experimental data.

The description of subsurface flow in the saturated zone is done by expanding the storage term on the right hand side of (2.1) to relate a change in storage to a change in fluid pressure through compressibility terms. This requires the assumption that the bulk compressibility of the porous medium is constant for saturated conditions. For unsaturated conditions, it is assumed that the compressibility effects on the storage of water is negligible when compared with the changes in saturation. Following [*Cooley*, 1971; *Neumann*, 1973], the following expression for the storage term is developed:

$$\frac{\partial}{\partial t}(\theta_s S_w) \approx S_w S_s \frac{\partial \psi}{\partial t} + \theta_s \frac{\partial S_w}{\partial t} \tag{2.10}$$

where S_s is the specific storage coefficient of the porous medium [L^{-1}].

The solute mass conservation equation is written in terms of concentration as:

$$\frac{\partial}{\partial x_i}\left(\theta D_{ij} \frac{\partial C}{\partial x_j}\right) - \frac{\partial}{\partial x_i}\left(q_i C\right) = \theta \frac{\partial C}{\partial t} + \rho_b \frac{\partial C_{solid}}{\partial t} \tag{2.11}$$

where the Darcy flux q_i is computed by solving (2.1), ρ_b is the bulk density, C_{solid} is the adsorbed concentration in the solid phase and D_{ij} is the hydrodynamic dispersion tensor [*Bear*, 1988]:

$$\theta D_{ij} = (\alpha_l - \alpha_t)\frac{q_i q_j}{|q|} + \alpha_t |q| \delta_{ij} + \theta \tau D_w \delta_{ij} \tag{2.12}$$

where α_l and α_t are the longitudinal and transverse dispersivities respectively, $|q|$ is the magnitude of the Darcy flux, τ is the tortuosity, D_w is the free solution diffusion coefficient

or simply the diffusion coefficient and δ_{ij} is the Kronecker delta. The pore water diffusion coefficient is obtained by τD_w. In literature, the pore water diffusion coefficient is also referred to as the diffusion coefficient of the porous medium [*Bear*, 1988]. A linear Freundlich adsorption isotherm is normally assumed between the concentration in the solid phase and the concentration in the aqueous phase:

$$C_{solid} = K'C \tag{2.13}$$

where K' is the equilibrium distribution coefficient [$\mathrm{L^3M^{-1}}$]. Substitution of (2.13) into (2.11) gives:

$$\frac{\partial}{\partial x_i}\left(\theta D_{ij}\frac{\partial C}{\partial x_j}\right) - \frac{\partial}{\partial x_i}\left(q_i C\right) = \theta\left(1 + \frac{\rho_b}{\theta}K'\right)\frac{\partial C}{\partial t} \tag{2.14}$$

More frequently, a dimensionless retardation factor, R given by [*Freeze and Cherry*, 1979], is used as the indicator to describe the linear adsorption isotherm between the aqueous and solid phases:

$$R = 1 + \frac{\rho_b}{\theta}K' \tag{2.15}$$

In Equation (2.14), it is important to note that the dispersion and advection terms are dependent on the Darcy velocity. The water content appears solely as a multiplier on the accumulation term with the multiplier also including the adsorption-isotherm.

2.2 Methods for the Identification of Contaminant Transport and Source Parameters

Groundwater flow and contaminant transport models are widely used to predict contaminant plume migration and attenuation, to assess the adverse impacts at potential receptors, and to design efficient strategies for mitigating groundwater contamination problems. The contaminant source information reconstructed by the calibration of the transport model is increasingly used as forensic evidence to identify responsible parties for the contamination and to assign liability among them, due to the limited funds available for the cleanup of contaminated sites. Because the evolution of the contaminant plume is determined by the

transport parameters, contaminant source information, and the groundwater flow field, parameter identification becomes a critical step in the application of the models. Unfortunately, parameter identification is an inherently difficult process due to the dispersive term in the advective-dispersion equation. Inverse modelling of contaminant transport in groundwater is an irreversible and ill-posed problem, which is unstable and sensitive to errors in data [*Skaggs and Kabala*, 1994; *Woodbury and Ulrych*, 1996; *Snodgrass and Kitanidis*, 1997]. Thus, a considerable effort has been devoted to the estimation of contaminant source release history and transport parameters. The first attempt to formulate and solve this inverse problem was initiated by *Gorelick et al.* [1983] using a linear optimization model. Following this work, a variety of inverse modelling techniques have been proposed, including the estimation of the release history of a known source, identification of the location of sources, and recovery of the historical contaminant distribution. An extensive literature review of commonly used inverse methods has been presented by *Atmadja and Bagtzoglou* [2001a].

In general, inverse modelling techniques can be divided into three broad categories based on their mathematical formulations and applications [*Liu and Ball*, 1999]: full estimation methods, parameter estimation methods, and backward tracking. For the cases with no knowledge of source history, except for the lower and upper concentration bounds, full estimation methods, which provide a function estimate of the history of contamination, are more appropriate and therefore received a large amount of attention in the past decade. If prior knowledge of the source model for a contaminant is known, applying the parameter estimation methods is more efficient and accurate. In practical situations, the cases with instantaneous point source that can be solved by the method of backward tracking are not very common, with this inhibiting the method's usefulness. However, for certain cases, the method of backward tracking is still an efficient way to obtain information about the prior position of contamination or travel time of contamination from an upgradient location. Table 2.1 provides a brief summary of inverse modelling methods for reconstructing source release history. A detailed review of the inverse modelling methods for finding the history of contamination will be presented in the following paragraphs.

2.2.1 Full estimation methods

Tikhonov regularization

As a regularized least squares method, Tikhonov regularization (TR) enables the ill-posed inverse problem to be a well-posed optimization problem by adding a regularization term to the objective function. Then the function to be minimized is

$$\min \left(\|y - Gs\|^2 + \alpha^2 \|Ls\|^2 \right) \tag{2.16}$$

where y is the vector of measured concentrations; s is the vector to be solved, representing the source concentration for each time step in this context; G is a matrix of weighted kernel function, which is the solution of the advective-dispersion equation; L is the regularization operator; and α is the regularization weight, which determines the relative weight between these two terms when solving this minimum objective function. Therefore, the Tikhonov method actually improves the stability of the inverse problem at the expense of accuracy [*Neupauer et al.*, 2000]. Normally the regularization operator is expressed as a measure of smoothness of s

$$\|Ls\|^2 = \int_{t_a}^{t_b} \left(\frac{d^n s}{dt^n} \right)^2 dt \tag{2.17}$$

where t is the time; t_a and t_b denote the starting and ending time for the contaminant source respectively; d^n/dt^n is the n^{th} derivative of s. A value of $n = 2$ is most commonly used to obtain a smooth solution [*Provencher*, 1982]. Because the value of α plays a key role to the accuracy of the regularized solution, several methods have been suggested for finding an optimal value for α. Most methods require prior knowledge of measurement error statistics, which is rarely available. One applicable method of estimating α employed by *Skaggs and Kabala* [1994]; *Liu and Ball* [1999] was developed by *Provencher* [1982], which seeks an optimal balance between reducing the variance of the solution and biasing the solution.

Table 2.1: Brief Summary of Methods for Identifying the History of Contaminant Source (extracted from *Atmadja and Bagtzoglou* [2001a]; *Michalak and Kitanidis* [2004b])

Method	Reference	Source type	Application domain	Site	Limitation
Tikhonov regularization	*Skaggs and Kabala* [1994]	point	1-D homogeneous	hypothetical	Biased by regularization term; restrict to continuous source; deterministic approach
	Skaggs and Kabala [1998]	point	1-D homogeneous	hypothetical	
	Liu and Ball [1999]	interface	1-D nonuniform	DAFB[a]	
	Neupauer et al. [2000]	point	1-D homogeneous	hypothetical	
Quasi-reversibility	*Skaggs and Kabala* [1995]	point	1-D homogeneous	hypothetical	Biased by stabilization term; complete information of contaminant distribution at some time; deterministic approach
	Bagtzoglou and Atmadja [2003]	point	1-D heterogeneous	hypothetical	
Minimum relative entropy	*Woodbury and Ulrych* [1996]	point	1-D homogeneous	hypothetical	Sensitive to specified error level
	Woodbury et al. [1998]	patch or point	3-D homogeneous	GL[b] and hypothetical	
	Neupauer et al. [2000]	point	1-D homogeneous	hypothetical	
Marching-jury backward beam equation	*Atmadja and Bagtzoglou* [2001a]	point	1-D nonuniform	hypothetical	Requirement of initial and final conditions and error bounds
	Bagtzoglou and Atmadja [2003]	point	1-D heterogeneous	hypothetical	
Geostatistically based methods	*Snodgrass and Kitanidis* [1997]	point	1-D homogeneous	hypothetical	Potential source must be known *a prior*
	Michalak and Kitanidis [2002]	interface	1-D homogeneous	DAFB[a]	
	Michalak and Kitanidis [2003]	interface	1-D nonuniform	DAFB[a]	
	Michalak and Kitanidis [2004a]	interface	1-D homogeneous	DAFB[a]	
	Michalak and Kitanidis [2004b]	patch	3-D heterogeneous	hypothetical	
Progressive genetic algorithm	*Aral et al.* [2001]	point	3-D heterogeneous	hypothetical	Requirement of one observation data for each time step, and more observation data than unknown variables

Continued on next page

Table 2.1: (continued)

Method	Reference	Source type	Application domain	Site	Limitation
Backward tracking methods	*Wilson and Liu* [1994]	point	1-D homogeneous	hypothetical	Only applicable to instantaneous point contaminant source
	Neupauer and Wilson [2001]	point	2-D homogeneous	hypothetical	
	Neupauer and Wilson [2003]	point	1-D homogeneous	hypothetical	
	Neupauer and Wilson [2004]	point	3-D heterogeneous	MMR[c]	
Parameter estimation methods	*Gorelick et al.* [1983]	point	2-D heterogeneous	hypothetical	Restrict to the evaluation of a few parameters; cannot provide a function estimate of contaminant sources or distribution
	Wagner [1992]	point	2-D nonuniform	hypothetical	
	Sonnenborg et al. [1996]	patch	2-D heterogeneous	Vestskoven site[d]	
	Alapati and Kabala [2000]	point	1-D homogeneous	hypothetical	
	Mahar and Datta [1997, 2000, 2001]	multi-point	2-D homogeneous	hypothetical	
	Sidauruk et al. [1998]	point	2-D homogeneous	hypothetical	
	Sciortino et al. [2000]	patch	3-D homogeneous	experimental	

[a] Dover Air Force Base, Delaware, USA

[b] Gloucester Landfill, Ottawa, Ontario, Canada

[c] Massachusetts Military Reservation, Massachusetts, USA

[d] Vestskoven site, Denmark

Skaggs and Kabala [1994] applied Tikhonov regularization to reconstruct the contaminant source release history. They studied a 1-D contaminant transport problem in a saturated homogeneous aquifer and attempted to recover the nonnegative release history of a plume with prior knowledge of point source location. The result shows that "the regularized solutions are insensitive to round off errors but their accuracy is heavily dependant on the plume measurement errors and the extent to which the plume has dissipated" [*Skaggs and Kabala*, 1994]. It would be impossible to reconstruct the history of a plume that has dispersed to background concentration levels. Finally, they pointed out that the accuracy of plume measurements has more weight on the recovered plume history than that of the parameter estimates. Thereafter, a case study was conducted by *Liu and Ball* [1999] using this technique to identify the contaminant source history from measured contaminant concentrations within a low permeability aquitard at Dover Air Force Base. The forward modelling of contaminant transport based on the recovered contaminant source history gave similar results as the measured data for both PCE and TCE, except for the deeper TCE concentrations. Although the TR technique performs well in the hypothetical example and field case applications described above, it needs to be remembered that the introduction of the regularization term to the objective function will bias the analysis [*Provencher*, 1982; *Skaggs and Kabala*, 1994]. Further, because the commonly used second-order TR attempts to fit a smooth function, TR is not applicable to the case of step source history [*Neupauer et al.*, 2000]. Finally, TR is a deterministic approach, and the arbitrariness in selecting the derivative operator and regularization weight suggests regularized solution is only one solution out of an infinite number possible, not definitely the best [*Woodbury and Ulrych*, 1996].

Minimum relative entropy

The principle of minimum relative entropy (MRE) originated from probability theory and was first introduced by *Kullback* [1959]. As the general form of maximum entropy approach, MRE states that given a prior estimate of the probability density function (pdf), subject to constraints imposed by the observed data, a posterior pdf is chosen to minimize the

relative entropy $H(q,p)$ in probability space.

$$H(q,p) = \int q(x) \ln \left[\frac{q(x)}{p(x)} \right] dx \tag{2.18}$$

where, $q(x)$ and $p(x)$ are prior and posterior pdfs respectively. In another words, MRE attempts to minimize the "distance", in an information sense, between the prior and the new constraint [*Woodbury and Ulrych*, 1993]. This principle has been applied to a wide variety of fields. *Woodbury and Ulrych* [1993] originally applied the principle of MRE to the forward probabilistic modelling of contaminant transport problem in groundwater hydrology. In their work, MRE was used to estimate the prior pdf of hydrologic parameters subject to the unimodular constraint and the expected value constraints with the base level knowledge of a joint boxcar pdf (uniform distribution between upper and lower bounds). The prior estimate of a parameters' pdf, which minimizes the relative entropy to the boxcar pdf, is a multivariate truncated exponential.

Woodbury and Ulrych [1996] extended the MRE's application to recover the release history of a groundwater contaminant. By additionally imposing a discrete set of observed data constraints to the prior estimated pdf derived in *Woodbury and Ulrych* [1993] and applying the same derivation procedure, the posterior pdf is a multivariate truncated exponential as well. They applied this inverse method to the same hypothetical problem as that of *Skaggs and Kabala* [1994]. For noise free data, MRE tends to reconstruct the source release history indistinguishable from the true history and performs better than TR. For data with noise, if only the noise level is known exactly, MRE is able to recover the salient features of the source history. Otherwise, MRE is sensitive to the specified noise level[*Neupauer et al.*, 2000]. If the data contains unknown noise, *Ulrych and Woodbury* [2003] shows that the noise level in MRE can be estimated from the data themselves, analysis of data residuals, and a rigourous approach using the real cepstrum and the Alaike Information Criterion, and MRE performs well on problems with delta-like sources. The advantage of MRE is that due to the assumption of the joint boxcar pdf, it is very easy to enforce parameter lower (nonnegativity) and upper bounds (solubility limit), which are often hard to handle for other inverse methods. In addition, another advantage of using the MRE approach is that once the source release history is determined, future behavior of the plume can be predicted on the basis of probabilistic analysis. *Woodbury et al.* [1998]

applied MRE to the practical case study problem with a 3-D constant velocity and constant dispersivity system at the Gloucester Landfill in Ontario, Canada. "The relative entropy measure is shown to be very useful in indicating reduction in uncertainty between the posterior and prior as a result of the new information provided by the physical constraints and data" [*Woodbury et al.*, 1998].

Geostatistically based methods

Geostatistical approach is a methodology for the analysis of spatially correlated data, including various techniques such as kriging, which assigns weights to the samples in a manner to minimize the estimation variance [*Kitanidis*, 1997]. It is mainly used for spatial interpolation [*Kitanidis and Shen*, 1996]. *Kitanidis* [1995] developed a quasi-linear geostatistical theory for inverse analysis. *Snodgrass and Kitanidis* [1997] employed this geostatistical approach combined with Bayesian analysis to the problem of source release history reconstruction. The source function to be estimated is temporally discretised into components, which are assigned a known correlation structure with unknown parameters.

Geostatistical approach is very general and actually includes Tikhonov regularization and many common interpolation schemes as special cases. In addition, it gains the advantage over some other methods because it makes no assumptions about the nature and structure of the unknown source function. It is considered to be robust enough to tackle complex cases with multiple potential sources, spatially varying velocity field, and dispersion coefficients [*Snodgrass and Kitanidis*, 1997]. The solution also incorporates uncertainty in contaminant concentration. Therefore, the confidence interval can be specified. The limitation of this approach is that the location of the potential source must be known *a priori*.

Snodgrass and Kitanidis [1997] modelled the example problem after the commonly used hypothetical example [*Skaggs and Kabala*, 1994]. For the unconstrained case, the solutions compare well with those obtained by *Skaggs and Kabala* [1995] using the method of quasi-reversibility. Enforcing the nonnegativity constraint would improve the performance of the method at the expense of computational efficiency, because more iteration steps are needed

to converge. The constrained solutions compare favourably with those achieved by *Skaggs and Kabala* [1995] using Tikhonov regularization. It has to be noted that the geostatistical approach for both cases has the advantage of providing a measure of estimation uncertainty or confidence interval.

Applications of geostatistical approach to contaminant source identification for the Gloucester Landfill, Ottawa, Ontario, Canada and for the Dover Air Force Base, Delaware were conducted by *Michalak and Kitanidis* [2002, 2003, 2004a], in which a Markov chain Monte Carlo (MCMC) method coupled with Gibbs sampling algorithm and MCMC combined with the application of Lagrange multipliers are implemented to enforce parameter nonnegativity respectively. The last two applications are related to the estimation of the source release history at the interface of two homogenous one-dimensional layers with distinctly different characteristics. The results demonstrate the applicability to field data, as well as its robustness when applied to a non-uniform domain. Furthermore, the incorporation of methods to enforce nonnegative parameter effectively reduces estimation uncertainty. In this manner, an upper constraint (such as the solubility limit in the case of concentration) could also be taken into account, which would narrow the confidence interval further.

Michalak and Kitanidis [2004b] originally extended the geostatistical approach to the recovery of a historical contaminant distribution at a single time. In the work, they made use of the adjoint state method to improve the efficiency of calculating the sensitivity matrix. Another advantage of implementing the adjoint state method is that it allows using existing groundwater transport codes to carry out the adjoint state simulations [*Neupauer and Wilson*, 2001].

Quasi-reversibility

The method of quasi-reversibility (QR) was first developed by *Lattes and Lions* [1969]. Because the advection-dispersion equation for contaminant transport is not reversible and ill-posed with a negative time step, the QR method would replace it by an equation which is close to the original one but stable with reversed time. They substituted the QR op-

erator $(\partial/\partial t) - \Delta - \epsilon\Delta^2$ for the dispersion operator $(\partial/\partial t) - \Delta$, where Δ stands for the Laplacian operator and ϵ is a positive stabilization parameter. The source release function or historical contaminant distribution could be obtained by simulating contaminant transport backward in time. Obviously, the QR method has the same problem as the TR approach. The quasi-solution is biased by the introduction of a stabilization term which makes the equation reversible. When ϵ is relatively small, the quasi-solution will be close to the regular solution. QR requires complete information of contaminant distribution at some time as well, which is actually rarely available but can be obtained through interpolation and extrapolation methods. *Skaggs and Kabala* [1994]; *Bagtzoglou and Atmadja* [2003] performed the comparisons of QR to TR and MJBBE (refer to the following paragraph) respectively. The QR method is consistently less accurate than TR, but it is computationally less expensive and more straightforward [*Skaggs and Kabala*, 1994]. The authors claimed that it is very easy to incorporate heterogeneous parameters in the QR method. However, up to now, no case study solved by the QR or TR approach has ever incorporated the heterogeneous parameters. Finally, *Bagtzoglou and Atmadja* [2003] showed QR to be inferior to MJBBE for most cases, except for the case with homogeneous parameters and with initial data which are characterized by uncertainty.

Marching-jury backward beam equation method

Atmadja and Bagtzoglou [2001a]; *Bagtzoglou and Atmadja* [2003] developed a method named marching-jury backward beam equation (MJBBE) to recover plume spatial distributions and source release history. The results demonstrated that MJBBE is computationally efficient and robust enough to handle heterogeneity. More importantly, it is able to keep the shape and salient features of contaminant plume or release history. However, it cannot be used to recover the plume near the initial time, because the errors increase as the solution is backtracked toward the initial time [*Bagtzoglou and Atmadja*, 2003]. In addition, MJBBE requires too much input information including initial and final conditions and the error bounds between the measured and the exact values, which is almost not available for most field cases.

Progressive genetic algorithm

Aral et al. [2001] proposed a new combinatorial approach, called progressive genetic algorithm (PGA), to solve the nonlinear optimization model that describes the identification of contaminant source location and release history in a 2-D heterogeneous aquifer. They tested this deterministic approach to three different conditions. The computational results indicate its robustness and effectiveness under those conditions. What's more, the PGA method is the only one which has been applied to the identification of the contaminant source location and release history simultaneously [*Aral et al.*, 2001]. But two requirements are necessary to solve the identification problem properly: at least one observation data for each time step; more independent observation data than the number of variables.

2.2.2 Parameter estimation methods

Gorelick et al. [1983] initiated the first attempt to recover sources of groundwater pollution. They assumed the hydrogeologic parameters were known. They made use of least-squares regression and linear programming to determine source locations and magnitudes with the objective function defined by the absolute values of the difference between observed and simulated data. Error analysis were performed to determine the source locations and magnitudes in a probability framework.

Wagner [1992] developed an inverse model for simultaneously estimating the model parameters and characterizing the contaminant source by the method of non-linear maximum likelihood estimation. A distributed source term is included as parameters in a 2-D steady-state groundwater flow and contaminant transport model. He applied the model to investigate a few examples with temporally and spatially varying source release history. The fairly accurate results demonstrate its usefulness for the cases with a simple contaminant release scenario.

A 2-D heterogeneous field case study in Denmark was presented by *Sonnenborg et al.* [1996] for the first time, in which flow and transport parameters are estimated simultaneously with contaminant source strength using nonlinear least squares multiple regression.

In addition, the sensitivity of this optimization approach to steady-state versus transient flow conditions was conducted. The estimated parameters with transient flow differed slightly from those with steady-state flow condition, and resulted in a better fit. However, the convergence rate under transient flow condition was low, and if the initial parameter vector was far away from the optimal value, there existed a high probability of ending with either a local minimum or unrealistic parameter estimates.

Alapati and Kabala [2000] employed a non-linear least-squares method (NLS) without regularization to determine the parameters in source release function for 1-D homogeneous systems. They found this method was very sensitive to noise for gradual release scenarios, but can tolerate a much higher level of measurement noise for catastrophic release scenarios.

Mahar and Datta [1997, 2001] proposed a methodology to determine contaminant sources and the optimal design of a groundwater quality monitoring network for improved identification of contaminant sources. In the methodology, an optimization model was embedded and solved by the projected augmented Lagrangian (PAL) algorithm. They applied this method to a hypothetical 2-D homogeneous, isotropic, and saturated aquifer with conservative contaminant. The effects of parameter uncertainty on the contaminant sources and optimal design were not adequately discussed in their papers. *Mahar and Datta* [2000] also used a nonlinear optimization model for identification of contaminant sources in transient groundwater system.

An analytical solution-based inverse approach was developed by *Sidauruk et al.* [1998] to identify the point source location and transport parameters. After taking the logarithm of an analytical solution, a linear relationship between the plume concentration and the distance from the source could be set up. Therefore, unlike other methods, the objective function of this approach is the correlation coefficients of linear regression. The approach succeeded in estimating source locations and transport parameters for instantaneous and continuous point sources in a 2-D uniform groundwater flow system. Due to the strict assumptions made when deriving the analytical solutions, its application is limited to homogeneous aquifers, uniform flow and simple geometries.

Sciortino et al. [2000] presented a deterministic inverse modelling approach for identi-

fying the location and size of a rectangular DNAPL pool. In this approach, the DNAPL pool at the bottom of a 3-D homogeneous, saturated, and half infinite domain was treated as a Neumann type boundary condition. In other words, the concentration gradient normal to the DNAPL pool was already known. Then the analytical solution describing the solute transport was derived by assuming constant mass transfer coefficient governing the dissolution process across the pool. They formulated the inverse problem as a least squares minimization problem, and then solved it by the Levenberg-Marquardt method coupled with a genetic algorithm to avoid local minima or a non-unique global minimum. Three applications to the controlled bench-scale experiment shows that the inverse solution is sensitive to the location of the observation wells, the type of residuals minimized, and errors in the dispersion coefficients.

In summary, unlike the approaches employed for full-estimation problems, there is no vast difference among the approaches in this category. In fact, almost all of them can be classified as optimization approaches. They usually follow the same procedure: define an objective function often expressed as the squared difference between the simulated and observed data, apply an optimization algorithm to solve the inverse problem subject to constraints (such as nonnegativity, solubility), and sometimes perform a sensitivity analysis in the end. It should be noted that all the approaches discussed above have been applied to estimate the parameters characterizing the contaminant source. Some other stochastic optimization approaches, having successively identified groundwater flow and contaminant transport parameters such as simulated annealing, and tabu search, may be appropriate to solve this kind of problem and need to be examined for their applicability to this field.

2.2.3 Backward tracking for instantaneous point sources

Bagtzoglou et al. [1991, 1992] employed backward location probabilities to identify sources of contamination. They obtained probability distributions using a random walk method by reversing the flow field and leaving the dispersion part unchanged. *Wilson and Liu* [1994] solved the ADE using the stochastic differential equations backwards-in-time. *Wilson and Liu* [1994] kept the dispersion part unchanged and reversed the flow field as well. The

results are interpreted as location and travel time probability distributions of contamination. In addition, they took the example of 1-D model to verify that this approach can be generalized to include first order decay, linear equilibrium adsorption, and non-equilibrium adsorption. *Neupauer and Wilson* [1999, 2001, 2002] developed a mathematical modelling technique to derive the backward probability model in 1-D and multidimensional, heterogeneous aquifers using the adjoint state method. *Neupauer and Wilson* [2003, 2004] extended the model to incorporate reactive transport including first order decay, linear equilibrium sorption, and linear non-equilibrium sorption. However, due to the limitation of an instantaneous point source, this approach cannot find a site's contaminant release history.

Inverse modelling of the field-scale contaminant transport problem is extremely computationally demanding. Although rapid development of computer technology relaxes this type of limitation to some extent, calibration of an inverse modelling problem needs further investigation, especially when it comes to full estimation approaches. Over parameterization frequently associated with full estimation methods tends to cause an automatic calibration method to produce nonsensical results [*Kitanidis*, 1997; *Moore and Doherty*, 2006]. More importantly, due to the lack of data in the early contamination stage and the high noise level in the data of the case study investigated in this thesis (refer to Figure 4.11), full estimation approaches for this specific case study may not be applicable. Backward tracking is not an ideal approach for inverse modelling problems with continuous non-point pollution, reactive contaminants, or a fast changing flow field. Therefore, in this chapter, a selection of regression methods are applied to the calibration of a physically-based three-dimensional variably-saturated transient contaminant transport system with appropriate assumption of the source model. All the computational work was conducted on a Shared Hierarchical Academic Research Computing Network (Sharcnet).

2.3 Selected Parameter Estimation Algorithm

The main focus of this thesis is the performance comparison of the parameter estimation for a computationally intensive transport model with alternative algorithms including both

traditional and newly developed parameter estimation approaches. Parameter estimation (PEST) method, frequently used for the calibration of hydrologic models, is a gradient-based local search algorithm. The micro genetic algorithm (MicroGA) is a heuristic optimization approach extensively applied to a variety of water distribution and groundwater system problems [*Savic and Walters*, 1997; *Reis et al.*, 1997; *Babbar and Minsker*, 2006]. Dynamically dimensioned search (DDS) is another stochastic search algorithm, designed to find multiple good, but not locally optimal solutions [*Tolson and Shoemaker*, 2007]. "Good" means an agreement is achieved between the simulated and observed data, and "multiple" is opposed to one global optimal solution. A description of each method is followed by the reasoning behind its selection and implementation strategy.

2.3.1 Multi-start PEST

Parameter estimation (PEST) method is a model-independent nonlinear parameter estimator. Since its inception in the mid 1990s, PEST has become the industry standard in the calibration of all kinds of environmental problems [*Doherty and Johnston*, 2003; *Moore and Doherty*, 2006]. The use of an innovative calibration algorithm is only justifiable if its performance is superior or comparable with PEST. Thus, PEST is selected and serves as the benchmark optimization algorithm for evaluating alternative approaches.

It is assumed that a natural system can be described by the linear equation:

$$\mathbf{Xb} = \mathbf{c} \tag{2.19}$$

where, $\mathbf{X}$ is a constant matrix, a vector $\mathbf{b}$ holds the system parameters, and the vector $\mathbf{c}$ denotes the system's response or observation in the groundwater context. A least square method is utilized to provide a best linear unbiased estimator of the system parameter vector $\mathbf{b}$. Non-linear models occurring in most cases must be locally linearized by Taylor's theorem with respect to certain values of model parameters $\mathbf{b_0}$ so that the least square theory can apply:

$$\mathbf{c} = \mathbf{c_0} + \mathbf{J}\left(\mathbf{b} - \mathbf{b_0}\right) \tag{2.20}$$

where, $\mathbf{J}$ is the first-order Jacobian matrix denoting the linear relationship between pa-

rameter $\mathbf{b}$ and system response $\mathbf{c}$. If the observation weights are correctly assigned, the sum of square error Φ can be calculated using:

$$\Phi = (\mathbf{c} - \mathbf{c_0} - \mathbf{J}(\mathbf{b} - \mathbf{b_0}))^{\mathbf{t}} \mathbf{Q} (\mathbf{c} - \mathbf{c_0} - \mathbf{J}(\mathbf{b} - \mathbf{b_0})) \tag{2.21}$$

By taking the first-order derivative and setting it equal to 0, the parameter vector $\mathbf{b}$ can be expressed as:

$$\mathbf{b} = \mathbf{b_0} + \mathbf{u}, \quad \mathbf{u} = \left(\mathbf{J^t Q J}\right)^{-1} \mathbf{J^t Q} (\mathbf{c} - \mathbf{c_0}) \tag{2.22}$$

where, $\mathbf{Q}$ is a diagonal matrix whose i^{th} diagonal element is the square of the weight attached to the i^{th} observation, and $\mathbf{u}$ is the parameter upgrade vector. The linearized models are only approximately correct at local point $\mathbf{b_0}$ on the solution space. Then, the validity of its linearity at the solution point $\mathbf{b}$ is not guaranteed, which will question whether the objective function is at its local minimum. An iterative approach must be applied. Figure 2.1, abstracted from the PEST manual, diagrammatically shows the process of iterative convergence towards the objective function minimum.

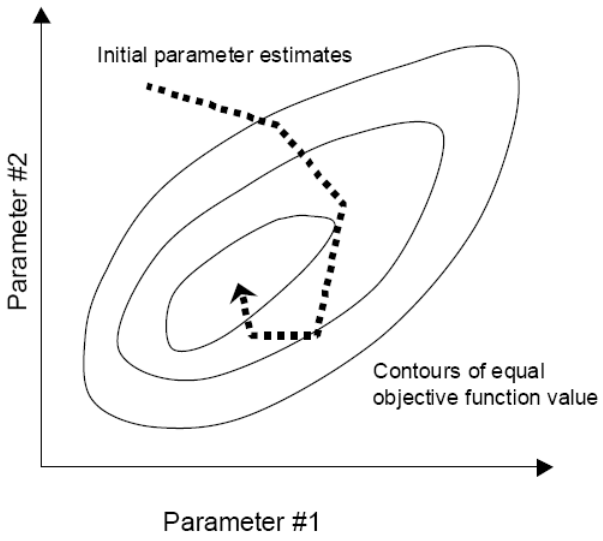

Figure 2.1: Iterative improvement of initial parameter values toward the local optimum

The parameter upgrade vector $\mathbf{u}$ is not aligned with $\mathbf{g}$, the steepest descent of Φ, expressed as:

$$\mathbf{g} = \frac{\partial \mathbf{\Phi}}{\partial \mathbf{b}} = -2\mathbf{J^t Q} (\mathbf{c} - \mathbf{c_0}) \tag{2.23}$$

$\mathbf{g}$ often leads to the phenomenon of "hemstitching" shown in Figure 2.2 (parameter set jumps from side to side of a valley in Φ), especially in cases where parameters are highly correlated.

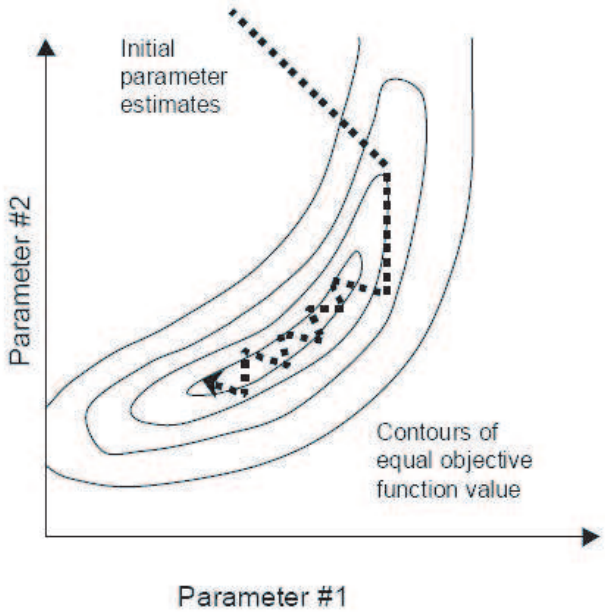

Figure 2.2: The phenomenon of "hemstitching"

The vector $\mathbf{u}$ frequently gives a better parameter upgrade direction than $\mathbf{g}$. Thus a combination of $\mathbf{u}$ and $\mathbf{g}$ tends to provide better performance in the initial stages of the estimation process. The upgrade vector $\mathbf{u}'$ is expressed as:

$$\mathbf{u}' = \left(\mathbf{J^t Q J} + \alpha \mathbf{I}\right)^{-1} \mathbf{J^t Q} \left(\mathbf{c} - \mathbf{c_0}\right) \tag{2.24}$$

where, α is the Marquardt parameter introduced by Levenberg to adjust the upgrade direction, and $\mathbf{I}$ is the identity matrix. A large value of α is assumed for the initial iterations. An upgrade vector closer to the direction of steepest descent results in faster reduction in the objective function. As the estimation process progresses, decreasing α makes it possible to avoid hemstitching. The value of α initially is supplied by the user, and adjusted as the calibration proceeds.

Like all the other gradient based algorithms, the Gauss-Marquardt-Levenberg method presented above is a local search technique. It can only find the local optimum in the

neighbourhood of the solutions around the initial solution. To circumvent the constraint of local search with PEST, a multi-start method was implemented to endow PEST with a stochastic feature. As the name implies, the multi-start method executes multiple times from different initial solutions. In the context of a heuristic search, a multi-start can be generally viewed as a method that iterates between two components: a constructive method to create a new starting solution; and an improvement method to improve this solution by local search. The optimization process operates as follows: a complete solution is obtained through a constructive approach, then switches to implementation of a local search procedure, once a local optimum is achieved within the convergence neighbourhood of the initial solution, compare it with the best solution so far, generate a new starting solution and enter another loop. Thus, a multi-start algorithm switches between these two components while keeping the best solution found throughout the search process. The purpose of diversifying the initial solution and restarting the local search procedure is to drive the search into new regions of the solution space.

The hybrid multi-start PEST employed in this study can be deemed as a combination of a stochastic approach and a conventional gradient-based search algorithm. The PEST approach serves as the solution improvement part. For the constructive component, a selective random generating approach was utilized. The initial solution sheds light on the efficiency of the optimization algorithms. A better guess (initial solution) will generate quicker convergence, especially for the gradient-based algorithms such as PEST. For each constructive run (or pre-PEST module in this case), 10 solutions, which are considered as enough samples to provide a starting point for optimization and to not incur heavy computational cost, were randomly initialized with the parameter constraints enforced, and then the best one with the least objective function value was selected as the starting solution for PEST. The use of 10 solutions, as invoked in this study, is problem specific.

The two bases for the comparisons are: the objective function values because they indicate the quality of the estimates, and the number of function evaluations, which tells the amount of computational resources required.

2.3.2 MicroGA

The genetic algorithm (GA) characterized as adaptive global search heuristics, is based on the Darwinian principle of "survival of the fittest". GA is composed of techniques inspired by evolutionary biology such as inheritance, mutation, selection and crossover. Since its development by John Holland in the 1960s, it has been widely explored in solving practical optimization problems in science and engineering disciplines. Contrary to a necessary large number of individuals in each population for classical GA, MicroGA referring to a small population genetic algorithm with re-initialization was suggested by *Goldberg* [1989]. Once converged for one generation, a new generation is started with the best individual and the remainder of the population is filled with new randomly generated parents. *Goldberg* [1989] suggested that a population size of 3 was sufficient to converge, regardless of the chromosome length (or dimensionality of the optimization problems). MicroGA with re-initialization can avoid premature convergence, overcome stagnation and more rapidly converge to the near optimal region [*Carroll*, 1996; *Krishnakumar*, 1989; *Zhou and Harris*, 2008; *Madadi and Balaji*, 2008]. *Krishnakumar* [1989] compared MicroGA with a population size of 5 against a simple GA with a population size of 50. Faster and better results were obtained on two stationary functions and a real world engineering control problem (a wind shear controller task). *Carroll* [1996] implemented classical GAs and MicroGAs to the optimization problems of chemical oxygen-iodine lasers. The uniform crossover MicroGA with a population size of 5 was identified as the best overall performer in his case study. By implementing MicroGA concepts within a standard GA procedure, the new strategy was demonstrated to have a beneficial effect in overcoming the stagnation encountered in standard GA structure solutions *Zhou and Harris* [2008]. In addition, [*Madadi and Balaji*, 2008] has shown that the method by integrating artificial neural networks (ANN) with MicroGA requires less function evaluations than ANN, and the computational time can be reduced substantially. Thus, the feature of re-initialization enables MicroGA to avoid stagnation and premature convergence, and a relatively small population size makes the algorithm converge faster. For the details on MicroGA algorithm, various types of operators are introduced and discussed in the following [*Goldberg*, 1989]:

1. Individual solutions are randomly generated to form an initial population, based

on the uniform distribution assumption over the entire range of the feasible region. To apply the GA operators in continuous parameter space, a binary-string representation of the solution space is required. The genotype of one parameter is a 6-bit binary digit which contains $2^6 = 64$ schemata. Through a one-to-one mapping of binary integers to real numbers, 64 real values can be represented by their corresponding 64 schemata, respectively.The mapping scheme is generally operated in monotonic linear or logarithmic representation. For example, a parameter with feasible range of [2, 5] is denoted by a 6-bit binary with linear relationship.

Table 2.2: Linear mapping from binary genotypes to decimal real numbers (vice versa)

Index	Binary genotype	Decimal real number
1	000000	2.000
2	000001	2.048
3	000010	2.095
4	000011	2.143
5	000100	2.190
...	...	...
63	111110	4.952
64	111111	5.000

As shown in Table 2.2, 000000 and 111111 represent lower and upper bounds respectively, and the rest are uniformly distributed in the feasible solution space. It might be noted that the length of binary genotypes place a limit on the accuracy with which the solution can be found. The more binary digits it contains, the closer it approaches the real value. Thus, MicroGA is specialized in finding multiple good global solutions as opposed to precise local optima, which makes it especially suitable to ill-posed calibration problems. For multiple parameter problems, all the genotype for each parameter can be united to form a binary $6n$-bit string or individual. The 6 binary digits between $6\,(i-1)+1$ and $6i, i = 1, 2, 3, \ldots, n$ represent the i^{th} parameter.

Each time that the objective function is evaluated, the parameter must be decoded to decimal equivalent. This conversion would tend to increase the computational time for traditional GA, because thousands of individuals are normally suggested to ensure the convergence. In this study, MicroGA requires much fewer individuals, and the forward modelling of contaminant transport tends to take up most of the computing resources. Thus the marginal cost of computational time introduced by converting binary into real form is very trivial.

2. Divide the strings into genotypes for each parameter, convert binary genotypes into decimal real number, evaluate the individuals by running the contaminant transport model, assign objective function values as fitness for each individual, and establish the best individual with the smallest function value. This step is not a typical GA operator, but necessary for all the GAs.

3. Selection is to choose the individuals in the population that will create offspring for the next generation. More emphasis is given to the fitter individuals in hopes that their offspring will have even higher fitness. Strong selection tends to let suboptimal highly fit individuals take over the population, and consequently reduce the diversity. While, weak selection will result in slow evolution as inferior individuals can not be sufficiently screened out. Numerous selection schemes have been proposed in the literature including tournament selection and fitness proportionate selection (or roulette-wheel selection). The MicroGA employs the tournament selection technique for the purpose of computational efficiency. Two candidate individuals are chosen at random from the population. For conventional GAs, the fitter the individual, the more likely it is selected. MicroGA only chooses the better of the two possible parents for mating. Candidate individuals are returned to the population and can be selected again.

4. The crossover operator varies the sequence of a chromosome (or bit string of individual) between two parents to create offspring. It roughly mimics the biological recombination mechanism between two single-chromosome organisms. Among a variety of crossover techniques, a modified single-point crossover method, which generally gets good performance with MicroGA [*Carroll*, 1996], is used to exchange continuous sections of chromosomes. Firstly, a fixed probability, typically 0.5, determines whether the crossover is conducted or

not. A single crossover point on both parents chromosomes is selected at random. Then, all bit strings beyond that point are swapped between the two parent's chromosomes. For example, the bit strings 10000100 and 1111111 could be crossed over after the third point in each to produce the two offspring 10011111 and 11100100 for the next generation.

5. Mutation is another fundamental feature for GAs. It involves a probability that an arbitrary bit in the genetic sequence will be changed from its original state. The purpose is to allow the algorithm to avoid local minima by preventing the population from becoming too similar to each other, thus slowing or even stopping the evolution. However, mutation is not implemented in the MicroGA, and it is replaced by two new operators called Newgen and Gamicro. To maintain the elite individual, Newgen will surrogate a random offspring by the best parent if it is not replicated by one of the children. This is particularly true when a good solution contains building blocks which might be disrupted by a non-respectful crossover operator. Compared to GAs, an extremely small size of population will result in the early convergence denoted by the fact that more than 95% of the number of bits in the micro population are the same as the best member. The Gamicro operator will keep the best individual and fill the remainder of the population with new randomly generated parents for the next generation.

6. Check if the maximum number of generations is reached. If not, return to step 2.

Execution time is always a concern for the heuristic approaches. Independence of individuals within each generation facilitates the parallelization of the MicroGA code without touching upon memory and processor management. To save the actual elapsed time (the CPU running time can not be reduced.), MicroGA was parallelized for the work of this thesis with OpenMP [*Sun*, 2003], which is a shared memory multiprocessing programming model. The original and parallel do constructs for objective function call are shown in Figure 2.3. The first command on the right sets 4 threads to use for the subsequent parallel region. A do loop directive is encompassed by a parallel do directive. Two private parameters, j and funcval, are not shared among multiple processors.

Probability of crossover is set at 0.5 based on a consideration of previous experience with this MicroGA [*Carroll*, 1996]. The population size was set to 4, in order to conform

```
                                call omp_set_num_threads(4)
                                c$omp parallel do private(j, funcval)
do j=jstart,jend                    do j=jstart,jend
  call func(j,funcval)                call func(j,funcval)
  fitness(j)=funcval                  fitness(j)=funcval
end do                              end do
                                c$omp end parallel do
```

Figure 2.3: Original and parallelized codes of the objective function call with OpenMP.

the maximum number of processors sharing memory on the Sharcnet with AMD's dual-core Opteron processors. The generation limit is the maximum objective function call divided by 4 processors. Thus, OpenMP parallelization will make the calibration almost 4 times faster than before for the case study of this thesis, because function evaluation requires much more computing effort than MicroGA itself.

2.3.3 DDS

Dynamically dimensioned search (DDS), developed by *Tolson and Shoemaker* [2007], is a novel stochastic single-solution based heuristic global search algorithm. Like MicroGA, DDS is designed to find multiple good global solutions within the specified parameter range. Unlike PEST, the stopping criteria for this algorithm is the user-specified maximum number of function evaluations. It is dynamically and probabilistically adjusted from global to local search by scaling down the dimensions of decision variables for perturbations in the neighbourhood as the search proceeds. Thus, DDS is specifically designed for high dimensional calibration problems that require only good solutions rather than a global optimum. *Tolson and Shoemaker* [2007]compared DDS performance to the shuffled complex evaluation (SCE) algorithm for multiple optimization test functions and the calibration of a real case study, the SWAT2000 Cannonsville watershed model with various formulations. With respect to convergence rates and objective function values, DDS was demonstrated to be more efficient and effective than SCE in all of the formulations. The complete DDS

algorithm is summarized as follows:

1. Define the neighbourhood perturbation size, maximum number of function evaluations, m, and randomly generate an initial solution.

2. Evaluate the objective function for the initial solution.

3. For each parameter, determine whether it is selected for the following perturbation with probability P, where $P = 1 - ln(i)/ln(m)$, i is the iteration count. Generate a random number from the uniform [0,1] distribution. If the random number is less than probability P, this parameter will be selected for the perturbation.

4. Perturb the selected parameters using standard normal random variables scaled by the neighbourhood perturbation size and parameter ranges; all the parameter box constraints must be honoured.

5. Evaluate the objective function at the new solution and update the best solution if necessary.

6. Check the stoping criteria. If not, return to step 3.

In step 4, a reflecting boundary is employed to enforce the constraints of the upper and lower limits on the decision variables. For instance, if the perturbed candidate solution is 0.1 smaller than the lower bound, the new solution will be the minimum plus 0.1. The scalar neighbourhood perturbation magnitude is the only one algorithm parameter that defines the standard deviation for random perturbation size as a fraction of the feasible parameter range. By default, the neighbourhood size parameter of 0.2, recommended by *Tolson and Shoemaker* [2007], tends to produce a sampling range over halfway of the feasible parameter ranges.

Chapter 3

Development of the Groundwater Flow Model for the Reich Farm Site

The Reich Farm Superfund site is located in the Pleasant Plains area of Dover Township, Ocean County, New Jersey. A regional view of the key features near the Reich Farm site is shown in Figure 3.1. This site has attracted considerable attention from groundwater modelers as a result of environmental pollution incidents over the past 30 years. Many groundwater models have been created since 1986 [*Sykes and Normani*, 2002] (Appendix A) and [*Normani et al.*, 2003], with most of them being based on MODFLOW and the finite difference method. In this chapter, a description of the site history and physical setting are presented and used to construct the conceptual model. Subsequently, the groundwater flow investigation and contaminant transport simulation analysis using HydroGeoSphere are described with the objective being the formulation of the optimization problem for the calibration of contaminant transport parameters.

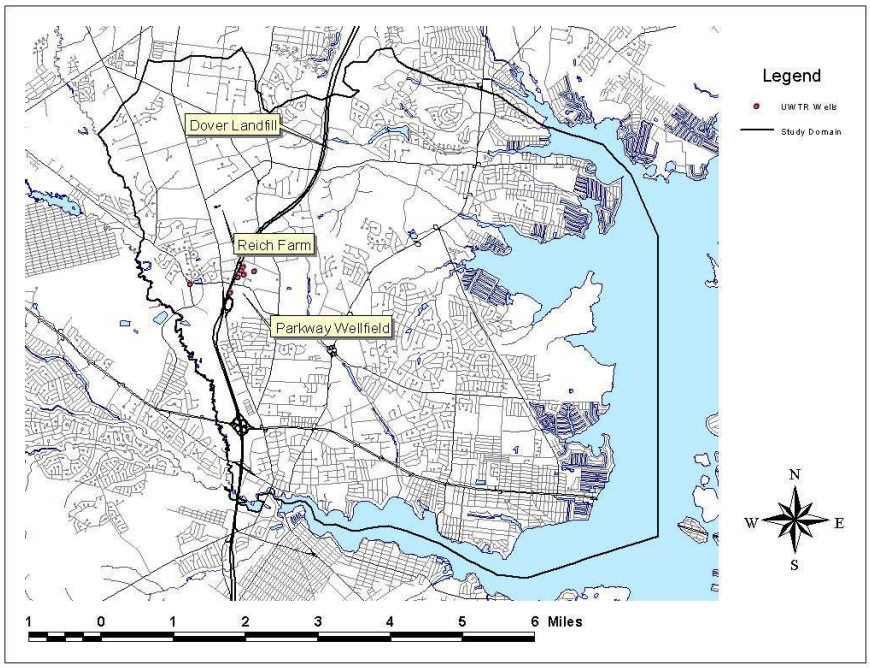

Figure 3.1: A regional view of Toms River and locations of key sites

3.1 Site History

In 1971, an independent waste hauler was hired by the Union Carbide Corporation (UCC) to transport 55 gallon drums of chemical wastes from the Bound Brook facility to the Dover Township Municipal Landfill (DTML) in Figure 3.1 for disposal. During a 5-month period in 1971, 5,000 to 6,000 drums of wastes containing aromatic hydrocarbons, phenols, halogenated aliphatic hydrocarbons, polymeric resins and unspecified petrochemicals were removed from the UCC facility by the contractor.

In August 1971, the waste removal contractor leased a part of the Reich Farm property from the owner for the use of storing the empty drums. four months later, the owner of the property found approximately 4,500 drums of chemical waste on the part of the property leased to the contractor, and trenches where the wastes were dumped. Of 4,500 drums of wastes on the Reich Farm property, approximately 10% were partially or completely dumped, indicating that part of the wastes were discharged into the soil and groundwater

at the site. This led to contamination of local groundwater, and consequently the source of drinking water for that area. The remainder of the drums and trench wastes were removed for burial or incineration. In 1974, approximately 51 additional drums and 29,700 cubic feet of contaminated soil were removed from the Reich Farm property by UCC and transferred to the Kin-Buc Landfill in Edison, New Jersey. Furthermore, 37 UCC drums were discovered in Dover Township on two parked trailer trucks of the waste hauler. The drums were removed by UCC.

Because the local groundwater system, which was serving as a major source of drinking water for Dover Township, was heavily polluted by the dumped organic wastes, the cleanup action was carried out approximately 2 years after the discovery of the dump. At the beginning of 1974, an investigation of groundwater quality confirmed that the groundwater in private wells in the area near the Reich Farm site was contaminated with organic compounds. But, most chemical testing was not able to identify individual organic compounds at that time. In 1986, additional private well and a few community wells at the Parkway well field in Figure 3.1 were discovered to contain certain VOCs and/or semivolatile organic compounds. The organic compounds were identified as predominantly trichloroethylene (TCE) and tetrachloroethylene (PCE). In 1996, another chemical released from organic wastes back in November 1971, styrene-acrylonitrile trimer, was found in some wells in the Parkway well field.

To reduce the potential exposure to site-related contaminants, many private wells affected by the Reich Farm contaminant plume were closed by the Dover Township Board of Health after the investigation indicating the presence of organic compounds. A well restriction zone in the area of Reich Farm was also established, following the recommendation by the New Jersey Department of Environmental Protection. In response to the fact that the contaminant plume originating from Reich Farm was extending toward the Parkway well field of United Water Toms River (UWTR), in 1988 UWTR installed a packed tower aeration (air stripper) treatment system for the output of two community water supply wells (well 26 and well 28, shown in Figure 3.2) to remove VOCs. The high pumping rate at the Parkway well field influenced the movement of the contaminant plume and its spatial distribution, and in fact captured the plume emanating from the Reich Farm site. The

treated groundwater was used to meet municipal demand. To eliminate contaminated soil as a long term source of groundwater contamination, in 1995, over 378,000 ft^3 of contaminated soil was further excavated and treated in an enhanced volatilization unit to meet the soil cleanup goals, and then backfilled to the excavated area.

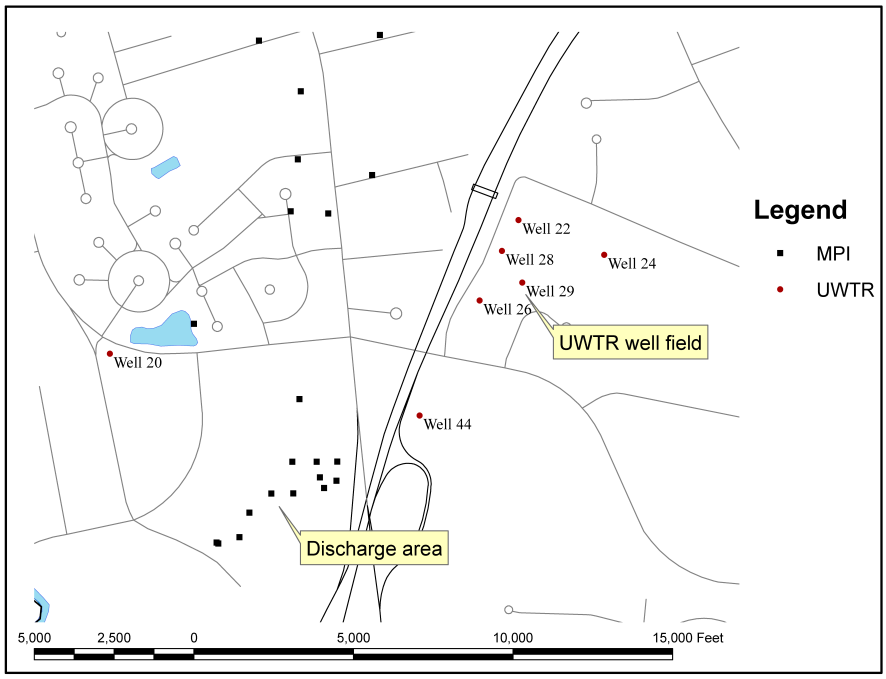

Figure 3.2: Discharge area and UWTR well names and locations

In a model study of the Reich Farm contaminant plume [*Sykes*, 1995], it was determined that the most effective and reliable method to protect uncontaminated wells was to continue pumping and treating of groundwater at the existing Parkway well field. The treated groundwater was allowed to be distributed to the community water supply if the water quality satisfied Federal and State drinking water standards. However, in November 1996, with the discovery of site-related styrene-acrylonitrile trimer in the community water supply from two wells (well 26 and well 28), activated carbon adsorption units were built to enhance the removal of organic chemicals. The treated water was discharged to the surface area shown in Figure 3.2.

Original data for this study are in US customary units. These original units have been

preserved in the analysis of this thesis.

3.2 Physical Setting

In the area of the Reich Farm site from Toms River eastward to Barnegat Bay in Figure 3.4, the surficial geology is primarily the Quaternary unit comprised of undifferentiated alluvium deposits, beach sands and gravels, and Cape May Formation, which are composed by black mud, light-coloured and pebbly sand and quartz. The Cohansey Sand that underlies the Quaternary unit outcrops in the southern part of the study area. The Cohansey Sand is predominantly a light-coloured quartz sand, containing very fine- to coarse-grained sand, silty and clayey sand, and interbedded clay units [*Zapecza*, 1989]. Along the Toms River, the Kirkwood Formation outcrops. The basal part of the Kirkwood Formation, which makes up the underlying upper part of the composite confining unit, contains regionally extensive clay layers. The upper part of the Kirkwood Formation, which is hydraulically connected to the overlying Cohansey Sand and surficial deposits, is composed primarily of dark gray to yellowish-brown fine sand to fine gravel and diatomaceous silty clay.

The Kirkwood Formation and Cohansey Sand comprise the major stratigraphic unit in the study area: the Kirkwood-Cohansey aquifer system. The aquifer system thickness is about 120 ft in the northwestern part of the study area and increases to about 220 ft near Barnegat Bay. The elevation of the base of the Kirkwood-Cohansey aquifer system, originally interpolated by *Zapecza* [1989], is shown in Figure 3.9.

3.3 Model Development History for the Reich Farm site

Many groundwater models of the Reich Farm site have been created. The U.S. Geological Survey (USGS), in cooperation with the New Jersey Department of Environmental Protection (NJDEP), created a MODFLOW [*McDonald and Harbaugh*, 1988] groundwater flow

model to study the local aquifer system and its interaction with surface water. The aquifer system's response to groundwater withdrawals [*Nicholson and Watt*, 1997] was also evaluated under both steady-state and transient scenarios. The New Jersey Geological Survey (NJGS) developed a steady-state groundwater flow model, based on MODFLOW, to determine the well head protection areas for the Parkway well field and Well 20 shown in Figure 3.2 [*Spayd*, 1997].

The Reich93 Model [*Sykes*, 1995] was developed, based on the finite difference method with both MODFLOW and SWIFT3 [*Ward et al.*, 1984], to assess the saturated groundwater flow and proposed remedial alternatives for the Reich Farm Superfund site. The model was calibrated for steady-state flow conditions with pumping at Parkway well field and Well 20. The model was also employed to estimate the transport of TCE and PCE from the Reich Farm site to determine the well field capture zones under transient groundwater flow conditions caused by temporally varying pumping rates. Subsequently, a new version of the Reich93 model, the Reich99 model [*Sykes*, 1999] was developed by extending the study domain eastward to Barnegat Bay and separating the Kirkwood Formation from the Kirkwood-Cohansey aquifer system. The study domain was vertically divided into 4 MODLOW layers. The top 3 layers correspond to the aquifer system, and the bottom layer is the basal portion of the Kirkwood Formation. Instead of a constant hydraulic conductivity for the whole domain, a spatially varying hydraulic conductivity field was generated from lithologic descriptions on drillers logs. The Reich99a model is an improved version of the Reich99 model through refining the grid in the vicinity of the Parkway well field and DTML. In addition, recharge variation both in time and space were taken into account for better representation of varying precipitation from month to month. Therefore, unlike the Reich99 model, the Reich99a model was designed to simulate transient groundwater flow.

Sykes and Normani [2002] further developed the Reich99a model to evaluate the transient saturated groundwater flow for the Reich Farm Superfund site. In the model, the temporally varying recharge distribution was calculated by an integrated analysis approach [*Jyrkama*, 2003] using the ArcView GIS environment and a hydrologic water routing model HELP3 [*Schroeder et al.*, 1994], which simulates daily water movement into the ground, and accounts for snowmelt, evapotranspiration, vegetative interception, surface runoff, and

temperature effects. The model included the most up-to-date well borings to further refine the hydraulic conductivity field. Because data quality plays a key role in model calibration, a thorough data quality check was performed. Inconsistency of well casing elevations from different source documents were found, and over 40% of the approximately 1,750 water level manual measurements were corrected by *Sykes and Normani* [2002] (Appendix A) as a result.

3.4 Conceptual Model

The regional-scale conceptual model for the Reich Farm study that was developed by *Sykes* [1999] forms the basis for the contaminant transport analysis developed in this thesis. The study of *Sykes and Normani* [2002] further developed the groundwater flow conceptual model, described the MODFLOW implementation of the conceptual model, calibrated the flow parameters for the model using an innovative method that honoured well logs obtained from the field-program of the study, and investigated the travel time for average water particles moving from the Reich Farm site to the Parkway well field. The modelling framework was used extensively to estimate the capture zones for the Parkway wells and to recommend flow rates for the wells that would control the contaminants migrating downgradient from the Reich Farm site. The recommended flow rates formed the basis of an agreement between the U.S. Environmental Protection Agency and United Water Toms River, the owner of the Parkway well field. However, this previous work involved no analysis of contaminant migration from the Reich Farm site. The modelling of *Sykes and Normani* [2002] was based on MODFLOW and did not include the unsaturated zone. Because a necessary component of the contaminant transport analysis of this thesis is migration through the unsaturated zone, the model HydroGeoSphere was selected. The switch of models from MODFLOW to HydroGeoSphere necessitated a complete redevelopment of the numerical aspects of the regional-scale groundwater system. The following sections of this chapter describe the groundwater conceptual model as developed by *Sykes* [1999]; *Sykes and Normani* [2002]; details of the conceptual model are provided in the Appendix of this thesis. This chapter contains only those elements of the conceptual model and the

calibration of the groundwater system model that are germane to the HydroGeoSphere implementation.

3.4.1 Spatial Domain and Boundary Conditions

The regional topographic elevation for the Reich Farm analysis, based on the USGS digital elevation model, is shown in Figure 3.3. The DEM was used to define domain boundaries and their features *Sykes and Normani* [2002]. The boundaries were chosen to match the topographic features such as rivers, surface water flow lines and divides. It is assumed that these boundaries could be used to describe groundwater divides, and also that the directions of groundwater flow along the boundaries are insensitive to changes in pumping rates at the UWTR Parkway well field. The spatial domain for the groundwater model as

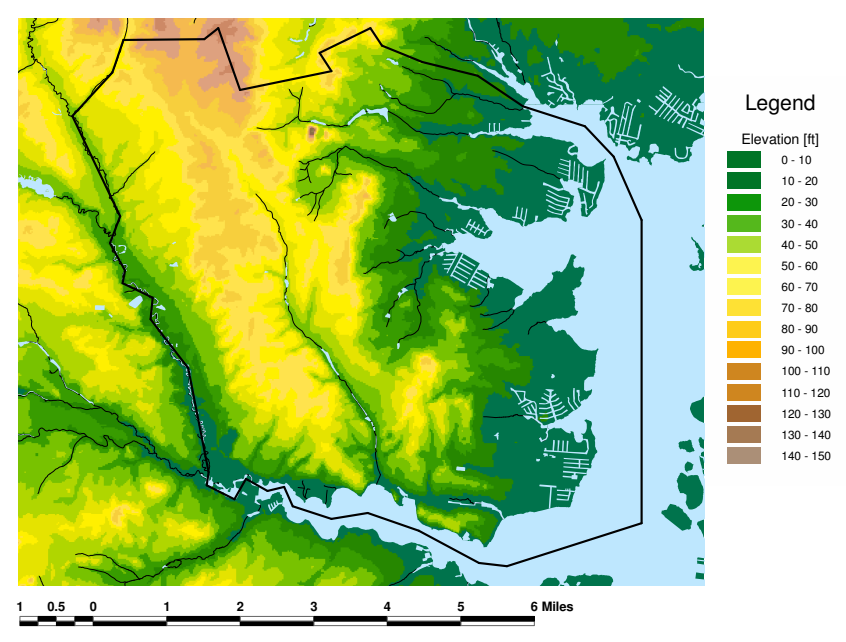

Figure 3.3: Elevation model for Toms River area

shown in Figure 3.1 extends eastward from Toms River to Barnegat Bay, and includes the Parkway well field, the Reich Farm Superfund Site, and the Dover Township Municipal

Landfill. This areal extent is considered to be large enough to reduce the possibility that the applied lateral boundary conditions will influence the migration of contaminants.

The conceptual model outline for the study domain is shown in Figure 3.4. Toms River flows through the western side of the top surface of the study domain, and it is modelled using a constant heads or Dirichlet boundary condition in HydroGeoSphere. The northern side is conceptualized as a groundwater divide, which is always orthogonal to the ground elevation contour; therefore it is the divide for the surface water system and can be regarded as a groundwater flow line (no-flow Neumann boundary condition). The Toms River estuary and Barnegat Bay are represented in HydroGeoSphere by Dirichlet boundary conditions, and the specified heads are set to zero representing mean sea level, and a high hydraulic conductivity of 150 ft/day was assigned to the associated elements beneath the estuary and Barnegat Bay. The other surface water bodies were initially described by bed elevation, bed conductivity, bed thickness, reach width, reach length and river flow depth. This representation of surface water bodies requires time steps small enough so that HydroGeoSphere was numerically convergent, which makes the inverse modelling of a transient, combined variably saturated flow and transport model too computationally intensive and thus infeasible for the analysis of this thesis with current computing resources. This fact is further developed in Section 4.3.1. To reduce the computational burden, the rivers and lakes within the domain are either perched or represented using Dirichlet boundary conditions, depending on their stage elevations. Stage elevations of surface water bodies including Toms River were estimated from the USGS 7.5 minute quadrangle maps combined with GIS layers for rivers, ponds and lakes obtained from NJDEP. The stage elevations were assumed to be constant in time.

The top boundary corresponds to the ground surface, at which a Neumann boundary condition is applied. The spatially and temporally varying inflow flux rates are equal to the recharge rates calculated by the HELP3 model at the root depth. The recharge rate depends on numerous factors that include land use/land cover, surface soils, evapotranspiration, weather data, and rainfall [*Jyrkama*, 2003]. The development of the spatially and temporally varying recharge using the HELP3 methodology is described in greater detail in the Appendix of this thesis. HELP3 model assumes that the vertical drainage is

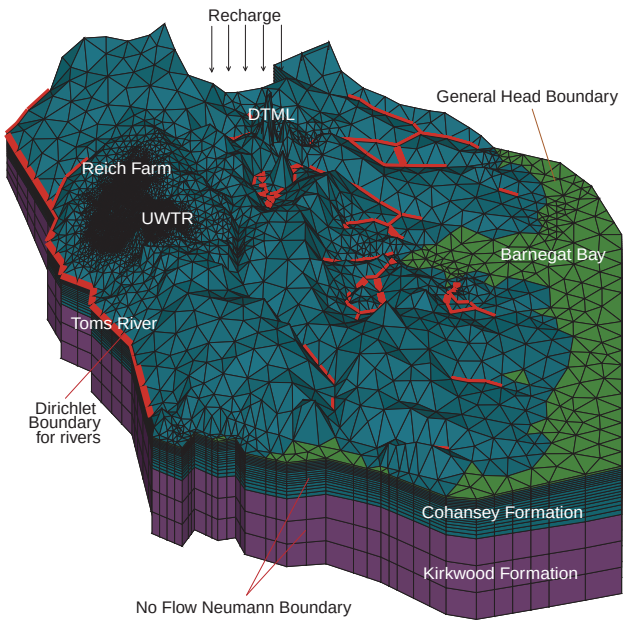

Figure 3.4: The conceptual model outline

driven by gravity alone (or unit gradient). A conclusion of *Jyrkama et al.* [2002] is that the HELP3 methodology is an important component of the calibration of the Reich Farm groundwater model; the investigation of the recharge methodology further to that developed by [*Jyrkama*, 2003] for the Reich Farm study is beyond the scope of this thesis. The water travel path within the root depth is nearly vertical and the travel time is trivial, compared to that in the whole domain. The recharge estimated with the HELP3 model at the root depth is applied to the top surface of the HydroGeoSphere model. The recharge rate for each finite element grid block is calculated by integrating the recharge rate per square feet (obtained from the *Jyrkama* [2003] database for the Reich Farm site) over the grid area. The average recharge rate for the whole study domain over 30 years was about 16.9 in/year, which is less than the value of 20 in/year used by the USGS and NJGS models. Figures 3.5 and 3.6 show typical recharge distributions for a dry (May 1981) and a wet (May 1989) month across the entire modelling domain. It is clear that the recharge varies significantly both spatially and temporally in the study area.

Below the ground surface, there is an unconfined Kirkwood-Cohansey aquifer system, which serves as the principal conduit for lateral groundwater flow in the Reich Farm re-

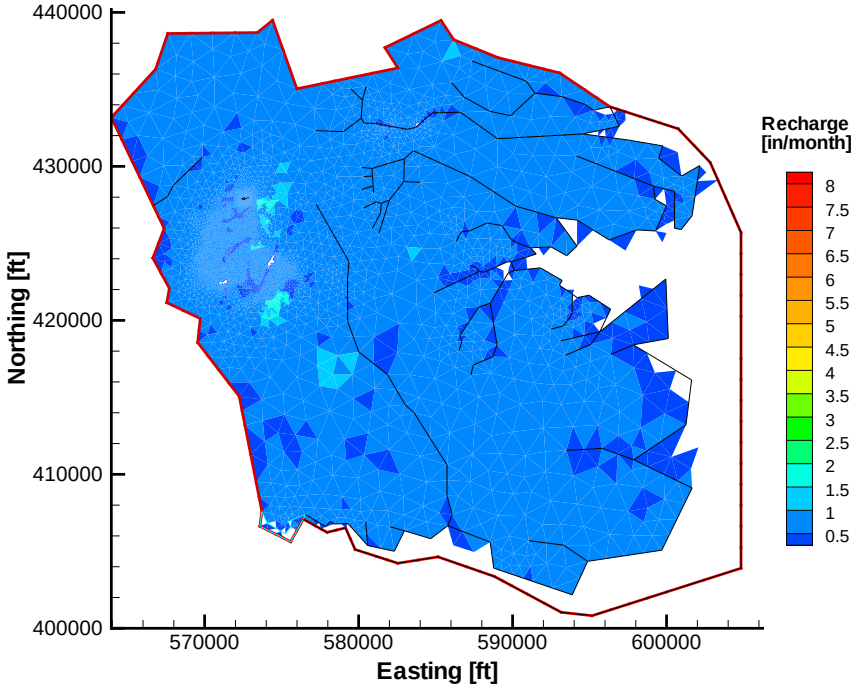

Figure 3.5: Spatially varying recharge for a dry month - May 1981

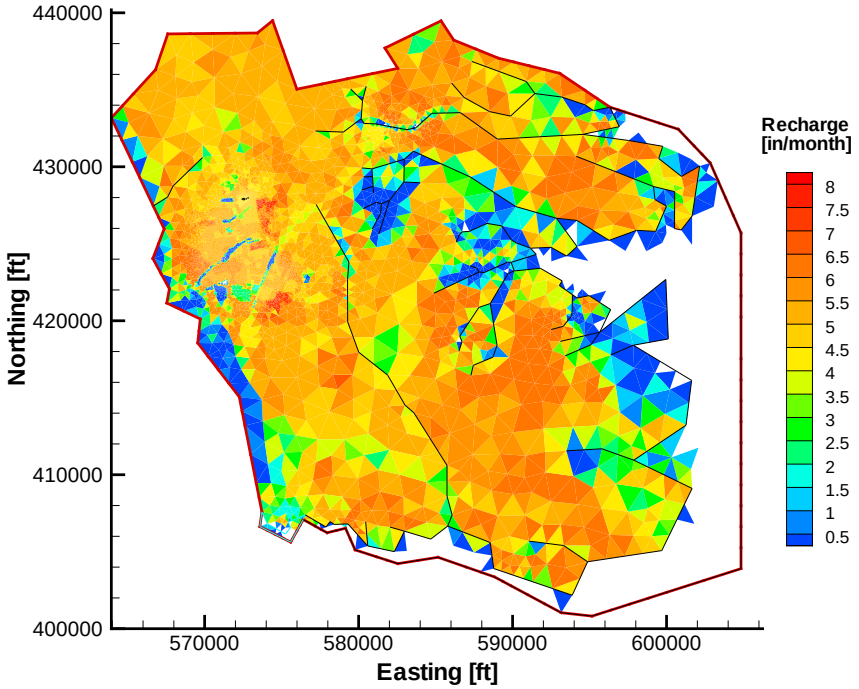

Figure 3.6: Spatially varying recharge for a wet month - May 1989

gion. The average thickness of this aquifer is approximately 200 feet. Immediately below the Kirkwood-Cohansey aquifer, there is a clay layer of low hydraulic conductivity that comprises the bottom of the domain. The bottom surface, defined by *Nicholson and Watt* [1997], is represented in modelling by a no-flow Neumann boundary condition. In addition, all vertical external edges of the study domain are represented by no-flow Neumann boundary conditions, because either flow lines or water divides are assumed at those boundaries.

3.4.2 Three-Dimensional and Variably Saturated Modelling

In Section 3.3, none of the described groundwater flow models for the Reich Farm site included the vadose zone. The source of contamination is normally located in the vadose zone. The time lag for the recharge migrating downward to the transient water table from the location of contamination is not negligible. Compared to the fully saturated model, the reduction of effective cross-section area in the vadose zone tends to increase the linear velocity and shorten the water routing. Thus, the vadose zone greatly influences the migration of the contaminant plume.

The low ratio (around 1:200) of vertical to horizontal dimensions of the study domain suggests that a two-dimensional model is sufficient for groundwater flow simulation. However it is not sufficient for contaminant transport simulation. The migration of contaminants may be significantly influenced by the vertical fluxes of groundwater flow in the Kirkwood-Cohansey aquifer system, especially around the Parkway well field area. Screen length and location is different for each monitoring and pumping well, so the knowledge of the vertical contaminant distribution is necessary in order to successfully and accurately fit the measured concentration. Therefore, a three-dimensional variably saturated groundwater flow and contaminant transport model is developed in this thesis to simulate the migration pathway of contaminants released from near ground surface at Reich Farm down through the unsaturated zone to the water table and then down gradient to the Parkway well field.

3.4.3 Domain Discretization

Figure 3.7 illustrates the HydroGeoSphere spatial domain discretized with the aid of a two-dimensional mesh generator and delaunay triangulator–Triangle [*Shewchuk*, 1997]. The generation of this two-dimensional mesh was restricted with boundary constraint and internal physical constraints such as wells and Reich Farm Superfund site locations. To satisfy the requirement of economy of computational effort, the mesh around the flow lines between the potential source zone and pumping wells was refined, as much as possible, to minimize grid Peclet number constraints, but it is coarser elsewhere. The resulting grid is significantly more refined than the MODFLOW grid developed by *Sykes* [1999]. The node number and element number for each layer are 5422 and 10726, respectively. To accurately describe the contaminant concentration distribution in the vertical direction, the domain is subdivided into 20 horizontal layers. In total, there are 113862 nodes and 214520 elements in the domain. From the top to the bottom, there are 1 top surface soil layer with 5 feet of depth representing surficial geology, 16 layers in the high hydraulically conductive part of the Kirkwood-Cohansey aquifer system and 3 layers for the basal part of the Kirkwood Formation, respectively. The interface elevation between the two aquifers, shown in Figure 3.8, and the bottom of the domain in Figure 3.9, were constructed by interpolation from well driller's logs using the kriging method. Such an interpolation is implemented using SURFER. The interpolated elevations for the sub layers in the basal part of the Kirkwood Formation are evenly distributed at intermediate distance. For the sake of delineating the water table, the vertical discretization within the Kirkwood-Cohansey aquifer system is dense for the upper layers, where the water table is most likely to be located, and sparse for the lower layers. While 20 grid layers were used for the vertical discretization of regional-scale model, this number is recognized to be insufficient for the detailed modelling of variably saturated flow and the algorithms describing it in HydroGeoSphere. The limitation of the vertical discretization is further developed in Section 3.4.5 of this thesis.

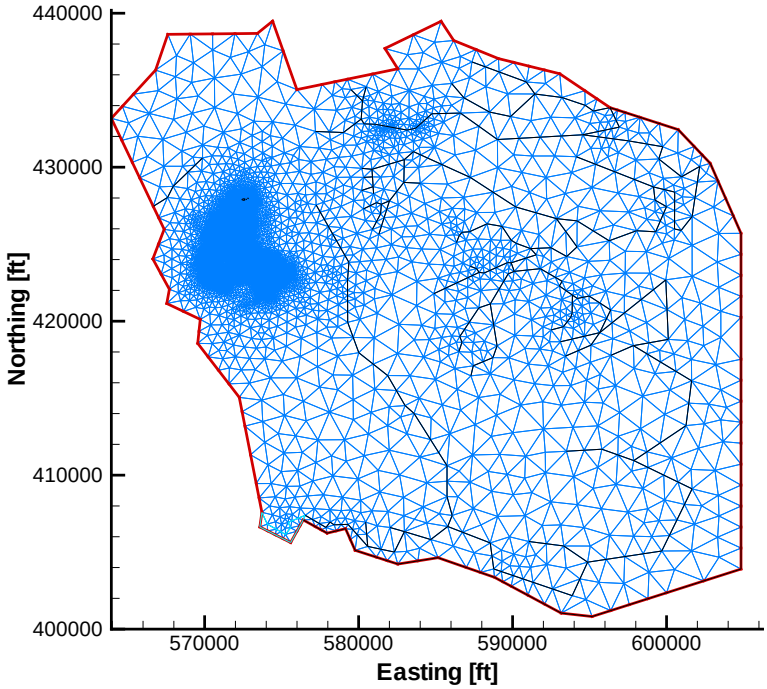

Figure 3.7: 2-D triangular finite element for HydroGeoSphere groundwater flow model

3.4.4 Modelling Parameters

Porosity is defined as the ratio of the volume of voids to the bulk volume of a porous media. It is independent of particle size but depends strongly on size distribution and packing. For this case study, the main component of the Kirkwood-Cohansey aquifer system is predominantly a light-coloured fine to coarse mixed sand. The range of value of porosity for this type of porous medium is 30-40% [*Bear*, 1988]. Effective porosity is slightly smaller than porosity, defined as the ratio of the interconnected voids to the bulk volume. In this conceptual model, effective porosity of 0.3, inherited from the Reich99a model, is suggested by *Sykes* [2008]. This value was used in the average water particle travel time analysis developed for the U.S. Environmental Protection Agency. In this thesis, the porosity is only used in the accumulation term of Equation 2.14. The accumulation term also includes the linear adsorption isotherm.

Specific storage is composed of water compressibility and bulk medium compressibility,

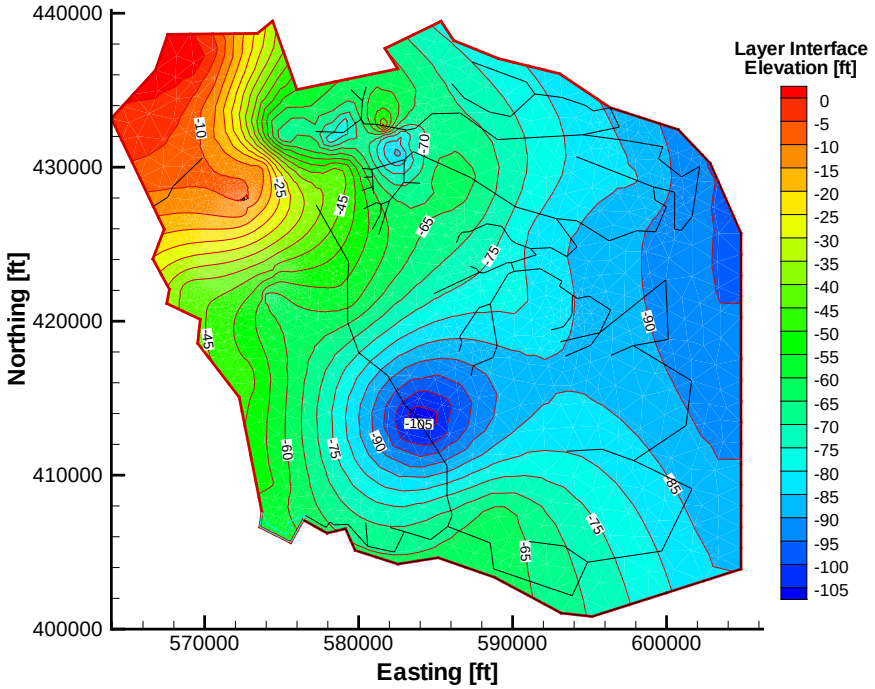

Figure 3.8: Interpolated elevation for the interface between Kirkwood and Cohansey aquifers

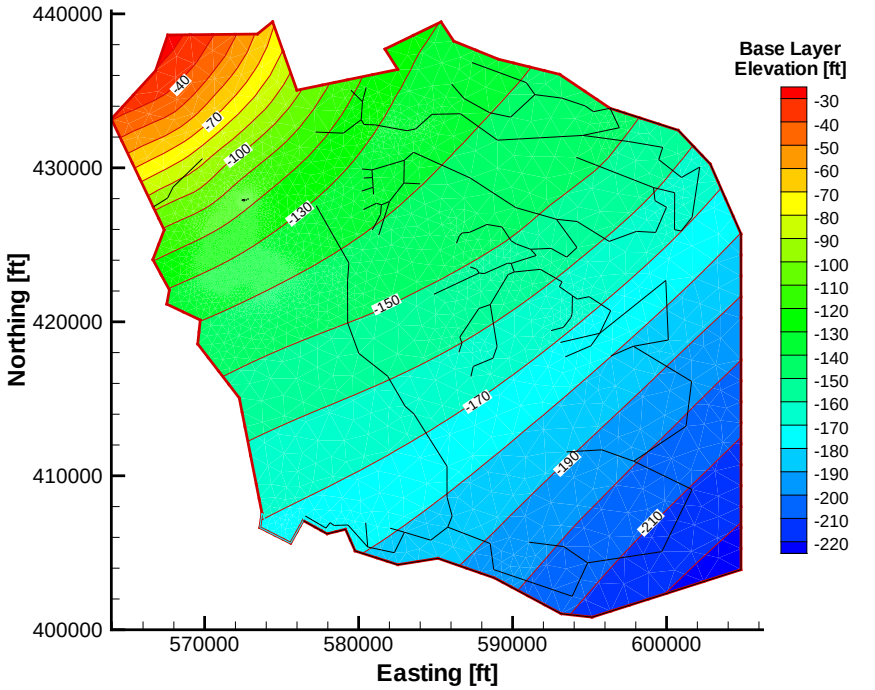

Figure 3.9: Interpolated base elevation for Kirkwood formation

and defined as:

$$S_s = \gamma \left(\beta_p + n \cdot \beta_w \right)$$

Where γ is the specific weight of water, n is the porosity of the porous media, β_p is the compressibility of the bulk aquifer, and β_w is the compressibility of water. The β_p-value of sand lies in the range $10^{-7} - 10^{-9}$ m^2/N [*Domenico and Mifflin*, 1965]. The specific weight and compressibility of water at 20°C are 9.8×10^3 N/m^3 and 3.0×10^{-6} m^2/N respectively. Substituting into the equation above, the specific storage is obtained as $3.92 \times 10^{-3} - 2.95 \times 10^{-3}$ m^{-1} or $1.19 \times 10^{-3} - 0.90 \times 10^{-3}$ ft^{-1}. In groundwater hydrology, this parameter is normally encountered when pumping takes place in a confined, completely saturated aquifer. For the unsaturated, unconfined shallow aquifer system, specific storage does not play a key role in supplying the pumping water, because specific storage is far less than the water capacity. In addition, its impact on the groundwater flow field and the coupled contaminant transport is considered to be trivial, compared to the water capacity. In this case study, a unique value of specific storage (1.0×10^{-3} ft^{-1}) is selected at no expense of the subsequent transport model calibration.

The vertical hydraulic conductivity for the top surface layer in each grid block is an areally weighted average of conductivities associated with each soil type. The specific conductivity values assigned to soil types are the same as those used in the HELP3 model to estimate the recharge rate. Based on the Reich99a model by *Sykes and Normani* [2002] by matching the simulated and measured hydraulic heads, the horizontal hydraulic conductivity in Figure 3.10 was calculated using an anisotropy ratio of 20 : 1. The grid blocks that underly surface water bodies were assigned a horizontal hydraulic conductivity of 150 ft/day. For the layers within the Kirkwood-Cohansey aquifer system, the saturated horizontal hydraulic conductivity distribution was derived by a kriging interpolation algorithm with no drift, an exponential variogram, and a nugget variance of 225(ft/day)^2 [*Sykes and Normani*, 2002; *Normani et al.*, 2003]. Known hydraulic conductivities were estimated based on the data from the logs for wells and borings. Each soil category identified in a well log was initially assigned a calibrated hydraulic conductivity value. The equivalent horizontal hydraulic conductivity for the location of each well is the weighted arithmetic mean of the soil-related horizontal K value, and the weight is proportional to the thickness of each soil unit identified in the well log. The resulting two-dimensional interpolated field

shown in Figure 3.11 was used to estimate the hydraulic conductivity for each element. A constant horizontal hydraulic conductivity of 0.283 ft/day was assigned to the 3 layers in the bottom of Kirkwood Formation. The vertical hydraulic conductivity was determined using a horizontal to vertical anisotropy ratio of 20 : 1, which was shown appropriate by the previous modelling effort.

3.4.5 Flux based method to calculate relative hydraulic conductivity

The HydroGeoSphere source code is distributed with the model executable program. The availability of the source code facilitated adapting and modifying the code so that it was suitable for this analysis of this thesis. In the process of model parameter estimation, it was observed that the model deficiency of coarse discretization could not be overcome by parameter adjustment alone. Further investigation into these deficiencies resulted in the identification of the model modifications as follows.

Modelling unsaturated groundwater flow not only requires a highly intensive computing effort, but also has numerical convergence issues. This is particularly the case in this study where a temporally varying recharge boundary condition is applied at the top surface of the model. A few options are available with HydroGeoSphere to describe the unsaturated properties of porous media. Van Genuchten and Brooks-Corey expressions are the most extensively used empirical relationships among relative hydraulic conductivity, water saturation, and capillary pressure. The coefficients are normally provided by the user. The coarse sandy soils of the Kirkwood-Cohansey aquifer of the Reich Farm study results in highly nonlinear relationships for both models. Their use requires a very fine temporal and spatial discretization, which leads to not only an extremely high computing demand, but also non-convergence issues. HydroGeoShpere also is able to deal with user-defined tabular linear relationships for the unsaturated properties. Efforts were made to implement all three methods to model unsaturated flow. However, due to the coarse discretization of the domain necessary for the field-scale model, neither the intensive computing requirement or convergence problem could be successively overcome in this case study with the model

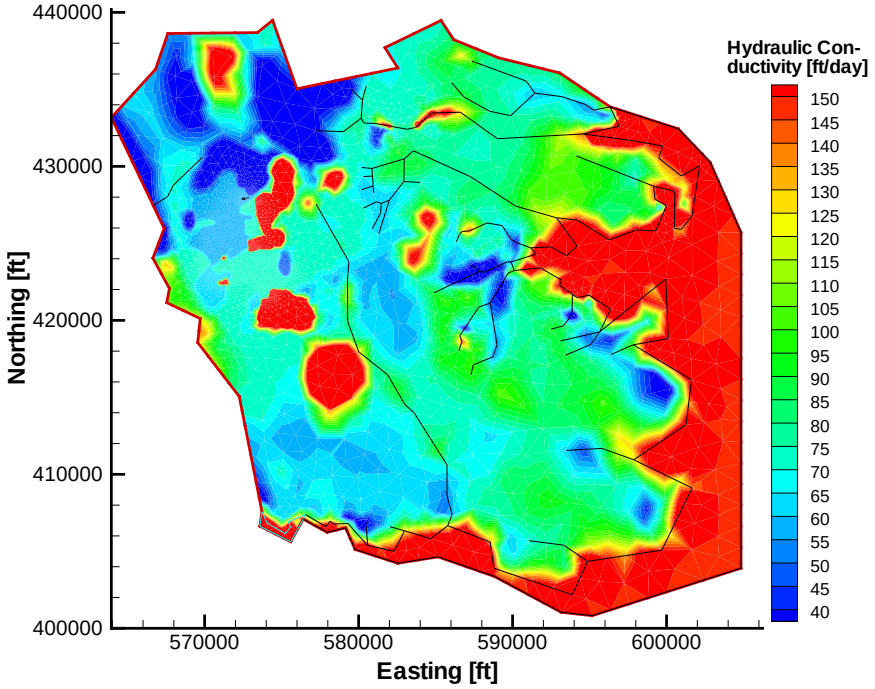

Figure 3.10: Horizontal hydraulic conductivity field for the top surface layer

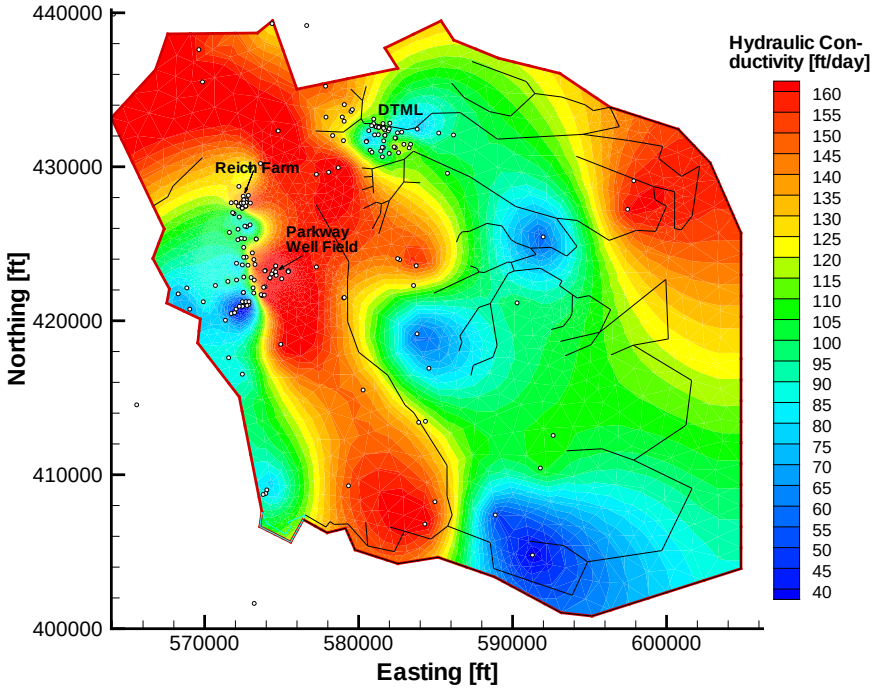

Figure 3.11: Horizontal hydraulic conductivity field for Kirkwood-Cohansey aquifer

as received. In a similar finding, *Sousa* [2009] observed that "the original van Genuchten equation with FEFLOW required sub-centimeter thick layers to achieve convergence; this is obviously too computationally expensive for a large-scale model. Therefore, a simplified representation of the pressure head versus relative permeability relationship was used" in his simulations. In a study of groundwater flow in the Grand River watershed *AquaResource Inc.* [2007] used linear relationships in FEFLOW to describe the unsaturated zone. They indicate that it is an industry standard to linearize the pressure versus saturation relationship for the vadose zone. While such an approximation may enable the simulation of unsaturated flow, if a solution can be obtained, the resulting saturations can lead to a misrepresentation of the contaminant migration.

For this case study, because the top soil at the Reich Farm site has an average hydraulic conductivity of around 140 ft/day, and is vertically penetrable, the vertical flux rate in the lower layer of the vadose zone is dominated by the flux in its adjacent upper layer. Here, a flux-based method to calculate the relative hydraulic conductivity for the vadose zone is introduced in Figure 3.12. In this method, assuming a unit hydraulic gradient, with this being consistent with the HELP3 model used to estimate recharge, the effective hydraulic conductivity is equal to the recharge rate for the top layer. Therefore, the effective unsaturated hydraulic conductivity $K_{ij,eff}$ is equal to the unsaturated Darcy velocity divided by the vertical hydraulic gradient which is subjectively set to be constant at 1.0. For the top layer, the recharge flux, calculated using the HELP3 algorithm, is used as the Darcy velocity. It must be noted that the discharge area, normally related to the surface water body, is regarded as the first type boundary condition. No discharge flux calculation is required. The relative permeability k_r can be determined by dividing $K_{ij,eff}$ by the saturated hydraulic conductivity K_{ij}. Then, the groundwater flow simulation can be conducted in the normal manner, where the water content θ is calculated using the modelled suction pressure in the vadose zone and the van Genuchten soil water retention function of equation (2.7) with the fitting parameters α, β, and the residual saturation S_{wr}. For the other layers in the vadose zone, the unsaturated hydraulic conductivity, $K_{ij,eff}$, is set using the Darcy velocity of the upper adjacent layer with a minimum of 0.03 ft/day. Based on experience, a lower value than the minimum tends to result in instability and non-convergence issues. The soil water retention curve describing water content and

pressure head relationship still holds, which is represented by the constitutive relationship equation (2.7) of *Van Genuchten* [1980]. It is worth noting that the Darcy velocity in the vadose zone used to determine $K_{ij,eff}$ is updated recursively for each time step until the convergence is reached. By using the described method for determining estimates of $K_{ij,eff}$, solutions for the unsaturated zone could be obtained for the coarse vertical discretization necessary for the regional-scale analysis of this thesis. It is re-emphasized that due to the coarse vertical discretization, a solution for the unsaturated zone could not be obtained with the traditional algorithms for the unsaturated zone that were implemented in HydroGeoSphere. It also is recognized that the numerical difficulties that were encountered in the unsaturated zone simulation could possibly be lessened by the use of a boundary condition at the domain surface as invoked in the integrated surface water and groundwater module of HydroGeoSphere. However, such an implementation requires small time steps making long-term simulations a computational burden.

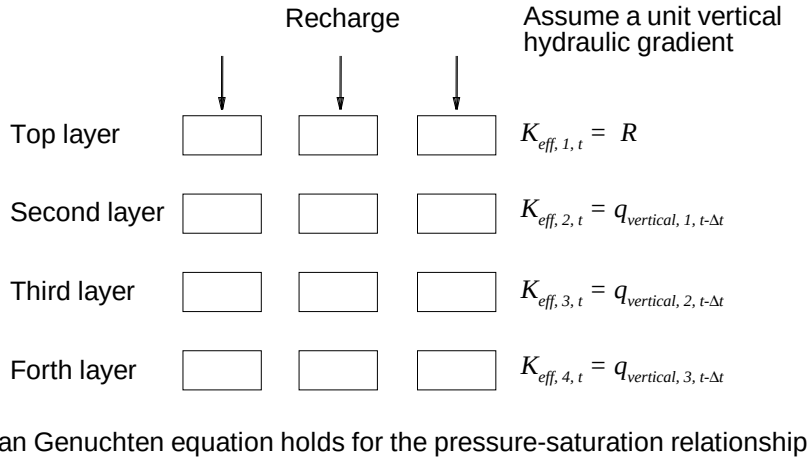

Figure 3.12: Flux based method to calculate k_r

The parameters in equation (2.7) and the saturated hydraulic conductivity, required for the variably saturated groundwater simulation, form a set of material properties that characterize hydraulic properties of a particular soil. Values of these parameters can be obtained by measurement, but such measurements are time-consuming and require specialized equipment. Laboratory based values were not available for this study. An alternative

for parameter estimation is the determination of sand, silt and clay content of the soil followed by estimation for the parameter from a catalog of soil textures [*Carsel and Parrish*, 1988]. For all the 12 Soil Conservation Service (SCS) textural classifications, *Carsel and Parrish* [1988] compiled a soil database obtained from measurements, estimated water retention parameters for the van Genuchten model using a multiple regression equation, and analyzed the limits of variation for each curve-fitting parameter. As mentioned in Section 3.2, the surficial geology is mostly comprised of alluvium deposits, sands and gravels, and the underlying Kirkwood-Cohansey sand is comprised of quartz sand, silty and clayey sand. Due to the shortage of data for the actual content of sand, silt and clay, we can roughly characterize this particular soil as sandy clay based on the soil description, and determine the proper parameter values of professional judgment as displayed in Table 3.1. For this case, the saturated volumetric water content θ_s is equal to porosity $n = 0.3$, and the inverse of the air entry pressure head α is 2.0 ft^{-1}.

Table 3.1: Limits of variation and actual values of sandy clay unsaturated hydraulic parameters for the van Genuchten model

	Lower limit	Upper limit	Selected value
α [ft^{-1}]	0.00	4.57	2.0
β	1.0	1.5	1.4
θ_r	0	0.12	0.07

The simulated hydraulic head distribution in December 1990 in Figure 3.22 was compared to the results derived from the MODFLOW free surface groundwater flow model (Reich99a) by *Sykes and Normani* [2002] (reproduced as Figure 3.13 in this thesis). At the drawdown area of the Parkway well field, the vadose zone for the MODFLOW model is deeper. The free surface model (Reich99a) tends to overestimate the hydraulic head gradient in the vicinity of the wells due to the neglect of the horizontal groundwater flow in the vadose zone. For other areas of the domain, there is good agreement between the hydraulic head contours of the two analyses. This justifies the use, in part, of the flux based method for simulating the groundwater flow in the vadose zone.

Figures 3.14 and 3.15 show the extracted slices for the water saturation and the flow lines located at Easting 572550 feet and Northing 427900 feet, respectively. There is no significant difference between the water saturation distributions for the dry and wet months, because the water saturation is calculated by the van Genuchten model based on the pressure heads, which are almost constant given the assumption of unit hydraulic gradient and the relatively stable water table at the Reich Farm site. However, the vertical Darcy fluxes at the vicinity of the Reich Farm determined by the recharge are 0.0026 ft/day for the dry month, and 0.0130 ft/day for the wet month. The flow lines verify the statement that the top soil at the Reich Farm site is vertically penetrable so that the vertical flux rate in the lower layer of the vadose zone is dominated by the flux in its adjacent upper layer. The flow lines in Figures 3.14 and 3.15 become more horizontal when approaching the water table.

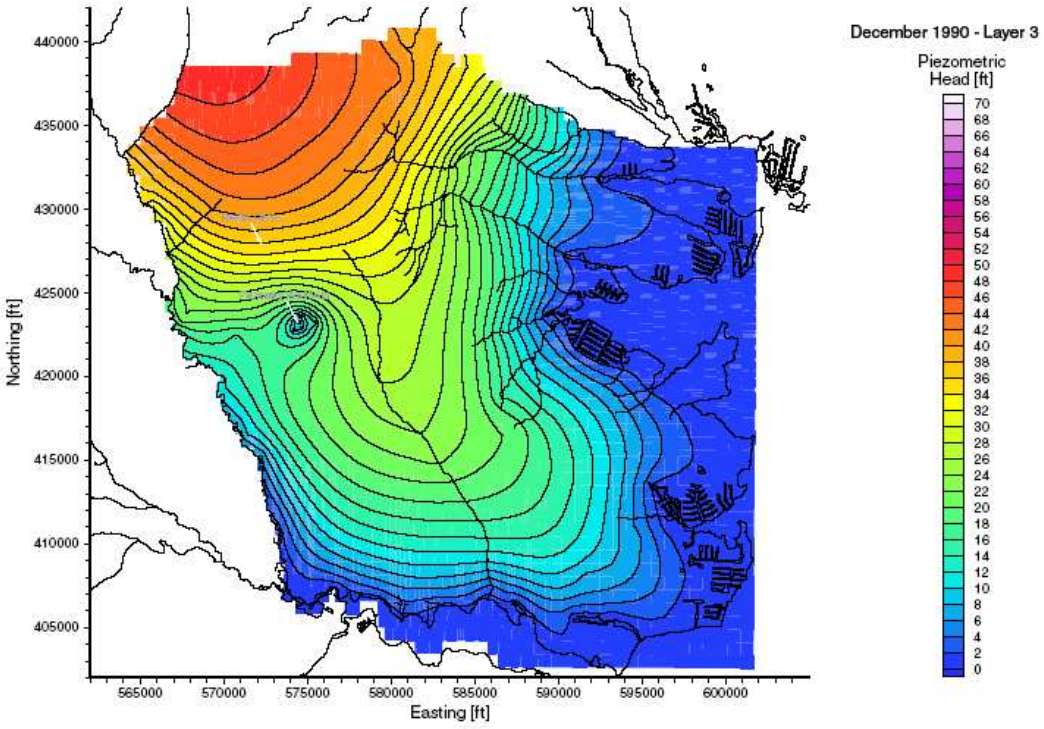

Figure 3.13: Simulated transient piezometric heads in layer 3 of the Reich99a groundwater flow model in December of 1990 from *Sykes and Normani* [2002]

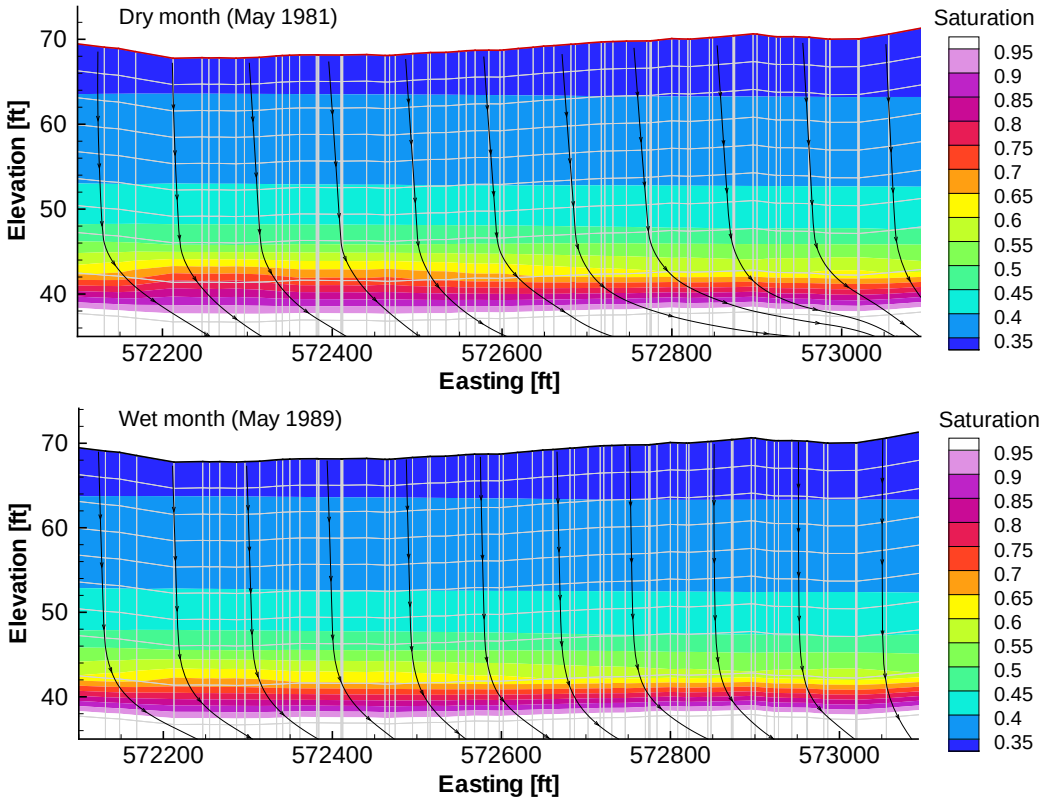

Figure 3.14: Plot of the water saturation distribution and the flow lines at Easting 572550 ft in the vadose zone

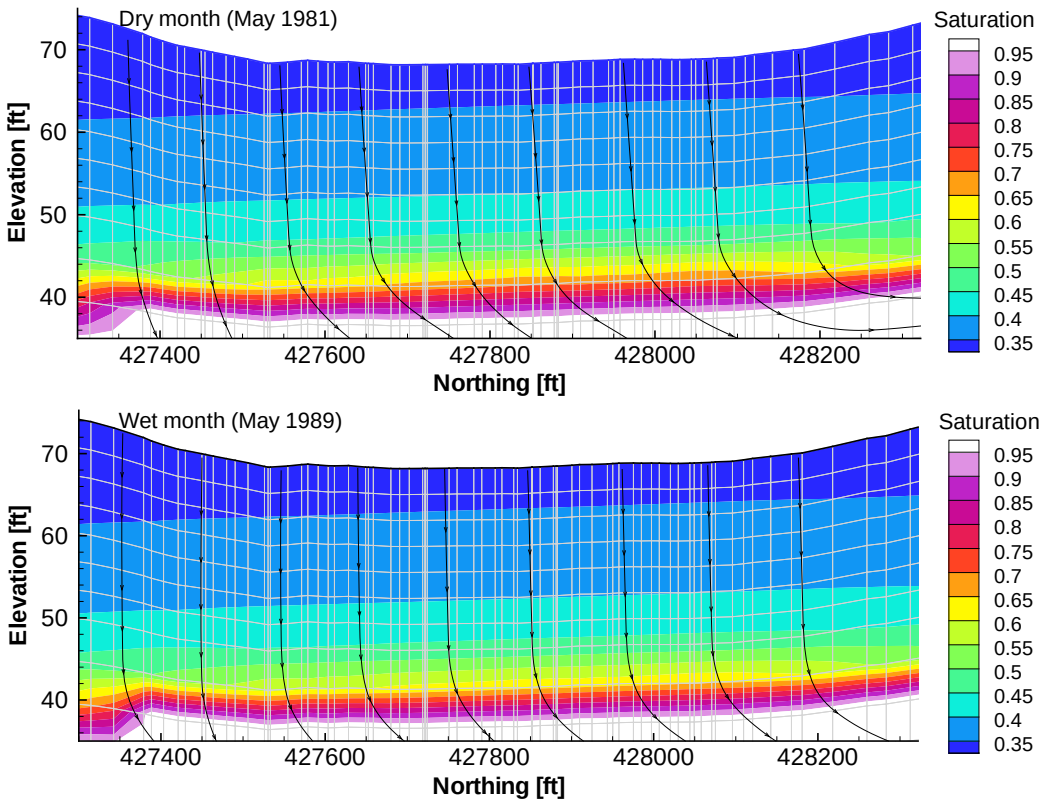

Figure 3.15: Plot of the water saturation distribution and the flow lines at Northing 427900 ft in the vadose zone

The actual travel time in the vadose zone is the key concern of incorporating the unsaturated zone. Excluding the vicinity of the pumping well field, the flow direction is straightly vertical. To verify the flux based approach in estimating the actual travel time in the vadose zone, the example consists of 1 dimensional transient infiltration in an unsaturated vertical column. The physical system is 30 ft long in the vertical direction, with the top face corresponding to the ground surface and the bottom face corresponding to the water table. The column is evenly vertically discretized into 30 elements. Initially, the pressure head is determined by a steady-state simulation with the infiltration at the rate of 16.9 in/year. The temporally varying recharge rate at the Reich Farm is then extracted from the regional-scale model and applied to this 1-d verification example for a period of 30 years. The porous medium property and the simulation control options are identically the same as those with the regional model, except that the constitutive relationships for the porous medium are given by van Genuchten model in Equation (2.7) with the parameters specified in Table 3.1. Figure 3.16 shows the evolution of particle travel time at different releasing time for the regional model and the 1-d column. The travel paths are 15 ft long for both models. In the regional model, the travel time was calculated from the location of dumping site at 55 ft above the sea level to the water table roughly at 40 ft above the sea level. In the 1-d column verification example, the travel time was estimated from the midpoint to the bottom. The evolution of travel time for both models follows the same trend. The travel time reflects the magnitude of the recharge rate with short travel time indicating high recharge rate and vice versa. The maximal values of the absolute and relative difference of travel time are 1.3 years and 0.16 in 1985. At the starting time of the simulation period, the difference of the travel time is 0.07 years, which is trivial compared to approximately 19 years of travel time from the dumping site to the pumping well field.

3.4.6 Hysteresis and Capillarity

Hysteresis and capillarity are important effects in groundwater modelling. The hysteresis behaviour of the groundwater levels can be explained by a time lag in the flux in the unsaturated zone. If the quick runoff processes, such as saturated overland flow and flow in macropores, is the main mechanism for discharge or water infiltration down to the water

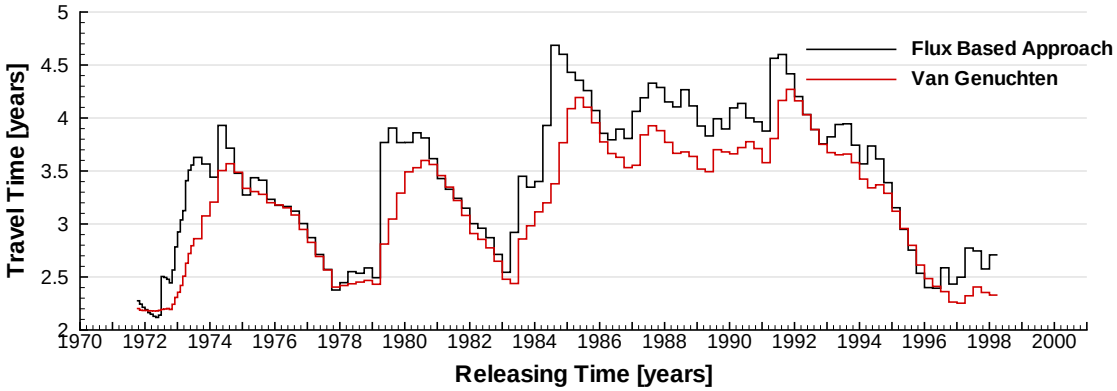

Figure 3.16: Comparison of travel time for particles releasing at the Reich Farm contaminant source (z=55 ft)

tables, the time delays are negligible and the hysteresis effect becomes less pronounced [*Myrabo*, 1997]. Capillary effect is strongly influenced by the pore size and connection of porous media. Capillary rises for coarse sand and sand are only 2-5 cm and 12-35 cm, respectively [*Bear*, 1988]. For the case study of the Reich Farm site, sand is the main component of Kirkwood-Cohansey aquifer. Besides urban areas, most of the lands are highly vegetated so that plant roots may serve as the major channel for the groundwater flow in top soil. In addition, observed fast water infiltration demonstrates that flow in macropores dominates. Therefore, hysteresis and capillary effects are minimized and are not taken into account.

3.4.7 Pumping Wells

The transient groundwater flow is simulated by monthly averaged flow rates for the UWTR Cohansey wells (Wells 20, 22, 24, 26, 26B, 28, 29, and 44). At the beginning of the simulation period in October 1971, only Well-20 was in operation. Plots of the temporally varying pumping rates for the wells, provided by George Flegal at United Water Toms River, are shown in Figures 3.17 and 3.18. The varying withdrawal rates of the municipal pumping well field result in a highly dynamic groundwater flow field, which also necessitates the use of a transient modelling scheme. Of note in Figures 3.17 and 3.18 is the shut down

of well 20 and the Parkway wells 26, 28 and 29 in 1988.

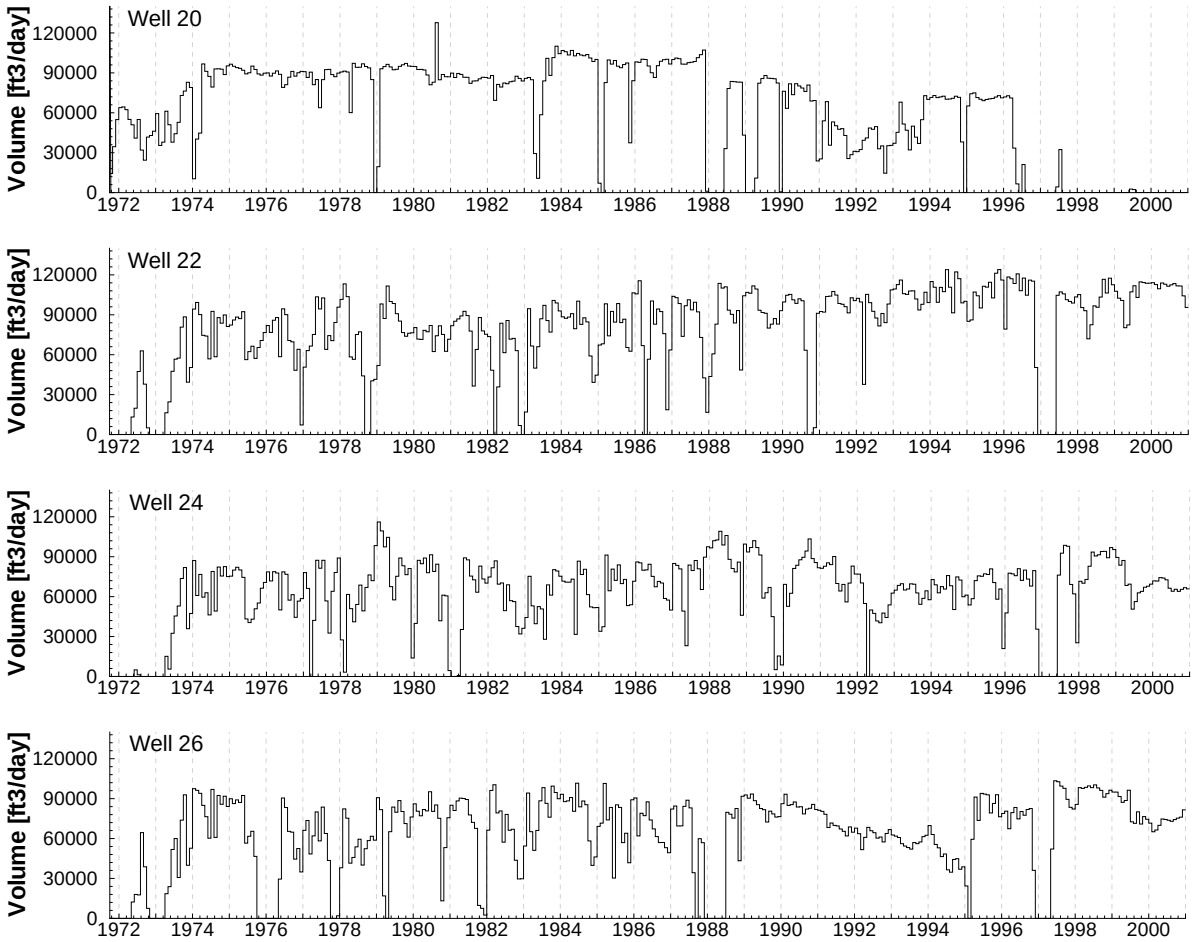

Figure 3.17: UWTR monthly production volumes for Wells No. 20, 22, 24, and 26

3.4.8 Initial Condition

A steady-state groundwater flow analysis was performed to provide the initial condition for the transient groundwater flow model. In the steady-state simulation, the spatially varying recharge distribution to the top layer was determined as the average recharge estimated using HELP3 for each HydroGeoSphere element for the period from October 1971

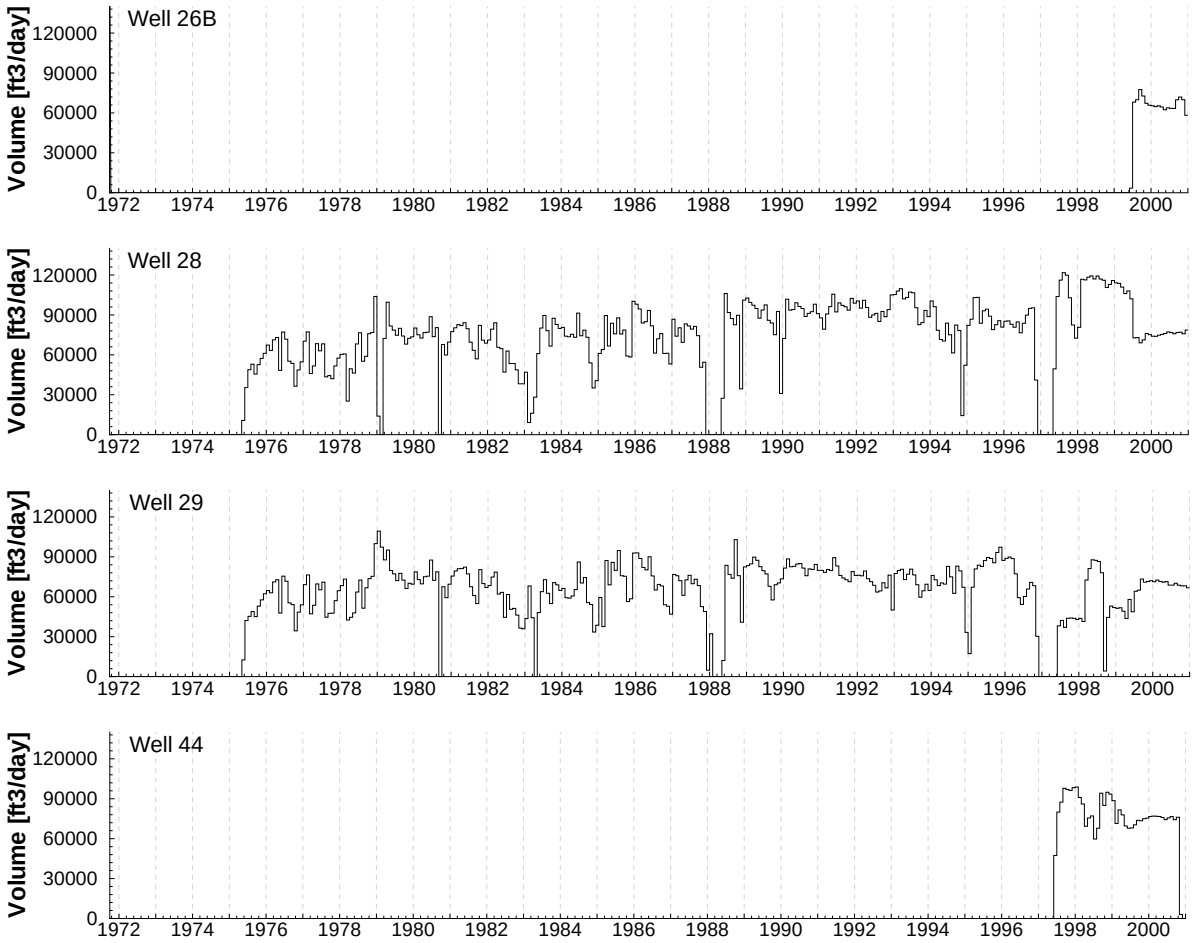

Figure 3.18: UWTR monthly production volumes for Wells No. 26B, 28, 29, and 44

to December 2000. A well-20 pumping rate of 46,420 ft^3/day was applied. The pumping rate is the average for the period from October 1971 to February 1972. The steady-state hydraulic head distribution in layer 4 of the model is shown in Figure 3.19. From the Reich Farm site, the groundwater approximately flows from northeast to the southwest towards Toms River and its estuary. The steady-state groundwater flow established above also served as the initial condition for transient groundwater modelling. The simulation period for transient groundwater flow commenced in October 1971 and ended in December 2000.

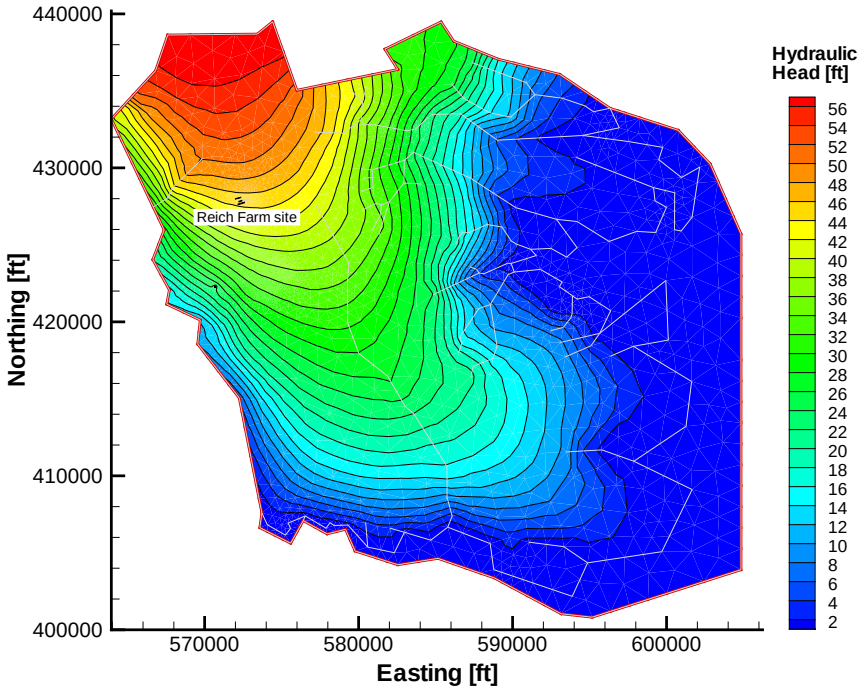

Figure 3.19: Steady-state hydraulic heads contour in Oct 1971

3.5 Groundwater Flow Simulation and Calibration

In order to utilize a predictive flow model in simulating contaminant transport and estimating the effectiveness of future potential management practices, the flow model must be

calibrated to field measured piezometric heads. Model calibration determines a reasonable parameter set which produces satisfactory performance in fitting the measured data set. The groundwater flow model calibration procedure discussed in this section follows the methodology developed by *Sykes and Normani* [2002]. The results are produced using HydroGeoSphere.

3.5.1 Observed Water Elevations

A large amount of water level measurements, both spatial and temporal, exist to facilitate the calibration of the groundwater flow model for the Reich Farm site. Hydraulic heads at 101 wells shown in Figure 3.20, were measured since 1985 by NUS, Ebasco, CH2M-Hill, and Malcolm Pirnie Inc. There are over 410,000 discrete measurements, most were from nine data loggers. However, there were a few issues related to data loggers. When the data loggers were removed from the monitoring wells for downloading, they still continued to record water levels at pre-specified times. In addition, discrepancies in water levels were not always avoidable between readings prior to logger removal and after logger return to the wells. Thus, the data from loggers were ignored in the calibration of the groundwater flow model parameters. However, the logger data can be used to validate the appropriateness of the flow model calibration (refer to Figures A.1 to A.6).

For the manual measurements, a through data quality check was performed by *Sykes and Normani* [2002]. It was found that the top of casing elevations for a large portion of the monitoring wells varied with the different source documents. Some top of casing elevations had been recorded incorrectly for nearly 10 years, due to transcription errors or using the wrong elevations. Further, the elevation benchmarks used by different contractors to calculate water levels were inconsistent with each other. For the measurements at the Dover Township Landfill wells, significant changes in water levels with time were discovered, which might be caused by incorrect measurements or clogged well screens. Due to the reasons stated above, around 20% of approximately 1,750 manual measurements were removed from the data used for the following parameter calibration. Table 3.2 adjusted from *Sykes and Normani* [2002] shows the number of data points in each year and well category.

Table 3.2: Count of scatter data points by category and year

	Category							
Year	**A**ll	**C**H2M	**D**TML	**E**BASCO	**M**PI	**N**US	**O**ther	**U**WTR
1986	151					151		
1987	48		1	24		20	3	
1988	94		5	47		42		
1989	15	15						
1990	35	5		12		18		
1991	63	6		22	14	21		
1992	44	1		13	22	8		
1993	54	6		12	22	14		
1994	30	3		8	12	7		
1997	300	12		45	177	38		28
1998	203	10		35	109	29		20
1999	157	6	48	20	53	18		12
2000	189	7	88	18	43	17		16
All Years	1383	71	142	256	452	383	3	76

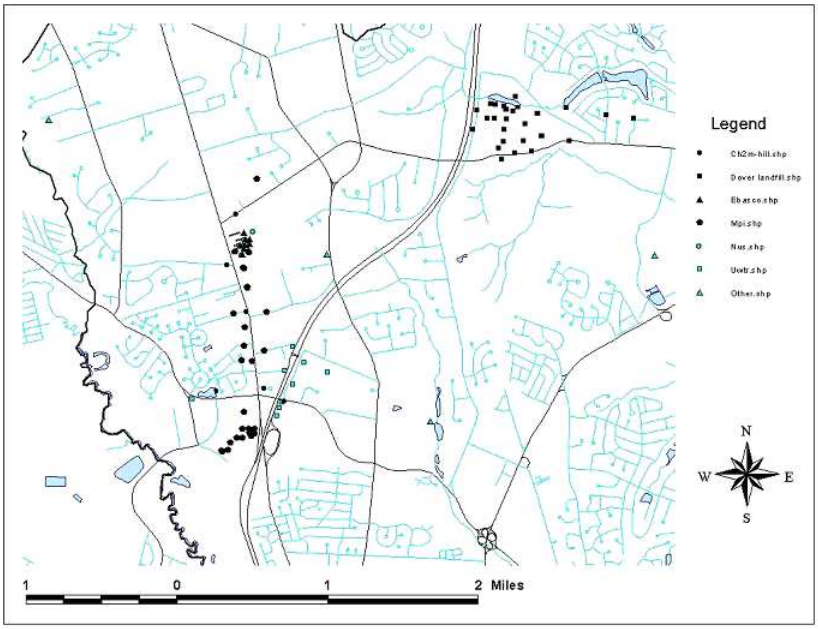

Figure 3.20: Locations of 101 domestic, pumping and monitoring wells

3.5.2 Calibration and Results

An iterative manual calibration procedure was employed to adjust model input parameters to provide a better match between simulated and measured water levels. The infiltration of water at the model surface was determined by *Jyrkama et al.* [2002]. The calculation of the recharge was based on a LULC/soil combination map, snowmelt, evapotranspiration, vegetative interception, surface runoff and temperature effects. Most of these factors were fixed. Thus, the spatially and temporally varying recharge on the top surface was assumed known, and the remaining parameters that could be perturbed were the hydraulic conductivity, porosity, and surface water elevations, especially the Toms River along the western side of the domain.

The hydraulic conductivity field was generated by *Sykes and Normani* [2002] (refer to the Appendix of this thesis) using a kriging algorithm, based on the estimated but adjustable hydraulic conductivities from the lithology in well and boring logs. The value for each element in the domain was assigned by interpolating the hydraulic conductivity

field to the 2-D triangle mesh. This process was repeated until a good match was obtained between simulated and observed water levels. However, ill-posedness due to a solution highly sensitive to small changes in data can arise when predictions are made at locations or times distant from observations clustered in space or time [*McLaughlin et al.*, 1996]. In this case study, the contaminant was released at the beginning of simulation period in October, 1971, but the water level measurements commenced from 1986. There exists the possibility that the calibrated groundwater flow model in the early stage of the simulation period could be inappropriately defined. In addition, as a result of pumping activities, the groundwater flow system in the study domain was highly variable. The future contaminant transport path is very sensitive to its early migration which in practice is dominated by advective transport or groundwater flow direction. The contaminant travel time can be roughly identified as the occurrence of the peak concentration in the concentration breakthrough curves in the monitoring or pumping wells. Therefore, a particle tracking analysis, based on the general knowledge of contaminant travel times, was conducted to assist in the flow model calibration. This analysis, performed in this thesis using HydroGeoSphere, also provided an initial estimate of the retardation factor.

The bottom layer of the high hydraulic conductive part of the Kirkwood-Cohansey aquifer system, that most wells were in hydraulic connection with, is believed to be representative. Figures 3.21, 3.22, and 3.23 show the hydraulic head contours for this layer in December of 1980, 1990, and 2000, respectively. The highly dynamic nature of the local groundwater flow system necessitated the transient model to describe the evolution of the groundwater flow pattern. Initially, the groundwater around Reich Farm and the Parkway well field flows from northeast to southwest towards the Toms River and its estuary. The flow direction gradually shifted to the Parkway well field as a result of the operation of its pumping wells. Hydraulic heads continued to decline for the same reason, especially around the Parkway well field forming a head drawdown area. After year 1997, water mounding formed in the southwest of Parkway well field was caused by recharging the pumping water from wells 26 and 28 back to the subsurface after air stripper treatment.

A comparison between the HydroGeoSphere simulated and measured water levels is presented in Figure 3.24. The agreement is considered to be good, given the fact that

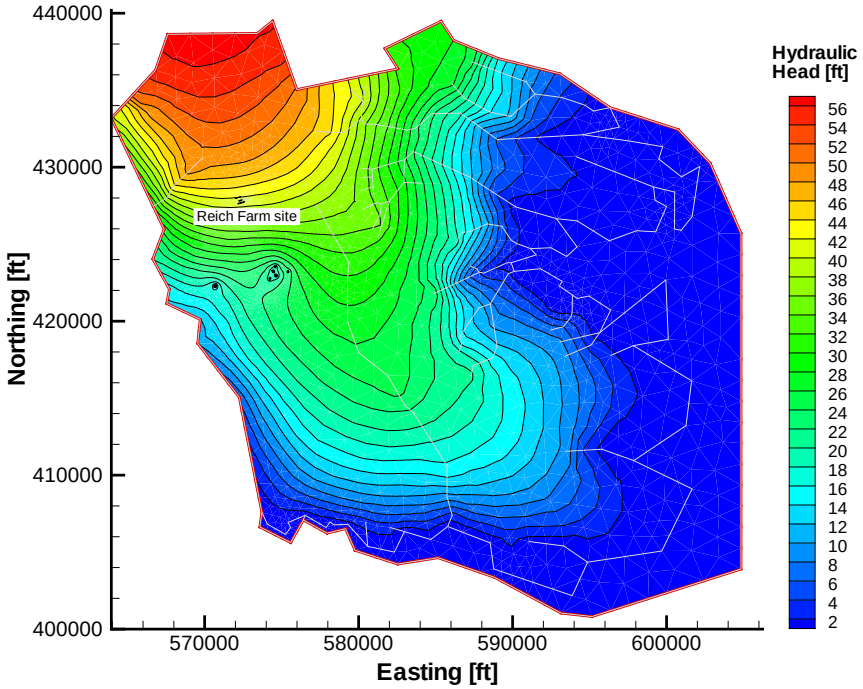

Figure 3.21: Simulated hydraulic head contour in Dec 1980

most of the scatter points lie within the narrow band of a 5 feet diagonal. Table 3.3 represents the averages of the absolute value of residuals between simulated and observed water levels for monitoring years and well categories. The overall average of the absolute value of the residuals is 1.93 feet. Except for the category with only 3 measurements, NUS, Ebasco, and DTML have relatively small average values. Fluctuations of water level are shown to be greatest near the Parkway wells and become smaller as the distance from the wells increases. From the time point of view, the largest average value occurred in 1997, even though there were 300 water level observations in that year. Residuals from the MPI observations contributed most to the average value, because the recharge activity commenced in 1997 greatly disturbed the original hydraulic system where the MPI monitoring wells are located. Early water level observations before 1991 and recent water level observations after 1997 show more favorable comparisons with the simulated water levels.

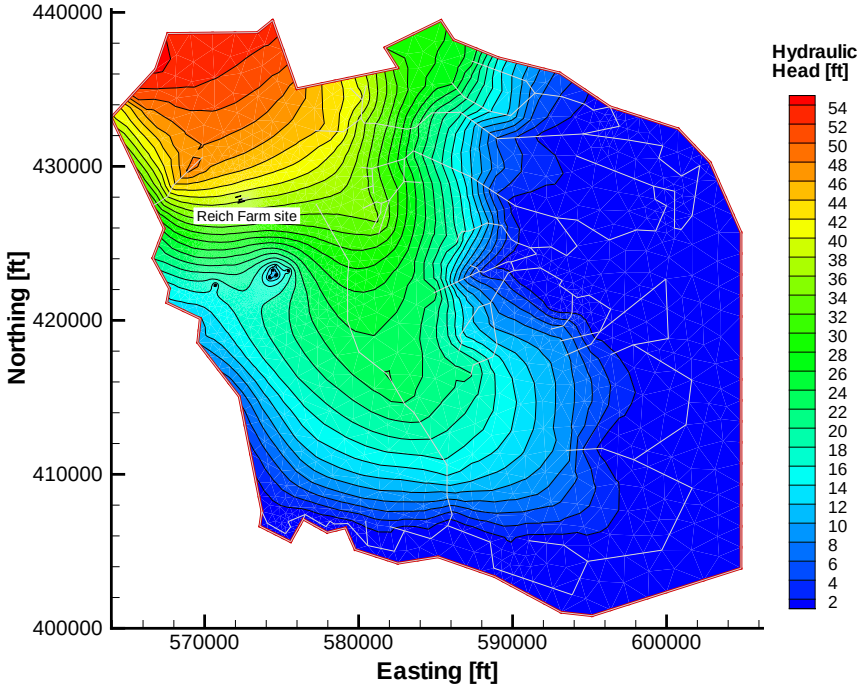

Figure 3.22: Simulated hydraulic head contour in Dec 1990

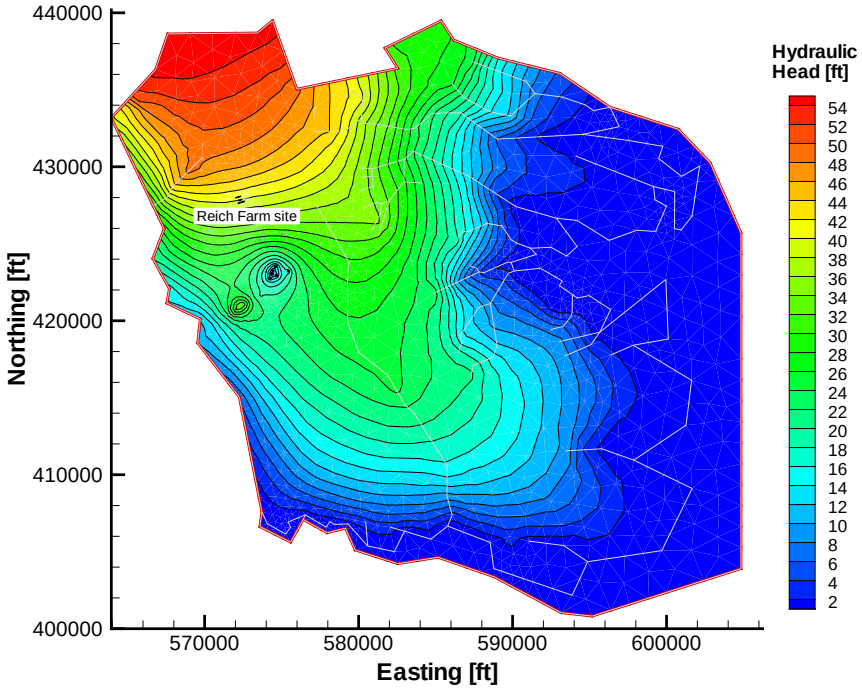

Figure 3.23: Simulated hydraulic head contour in Dec 2000

Table 3.3: Average of absolute value of head residuals [ft] by category and year

	Category							
Year	**A**ll	**C**H2M	**D**TML	**E**BASCO	**M**PI	**N**US	**O**ther	**U**WTR
1986	1.84					1.84		
1987	1.09		0.49	0.79		1.17	3.22	
1988	1.05		0.78	0.82		1.33		
1989	1.47	1.47						
1990	0.79	1.77		0.66		0.61		
1991	0.89	0.93		0.45	1.07	1.22		
1992	2.26	0.31		2.96	1.67	3.01		
1993	2.20	1.42		2.94	1.65	2.77		
1994	1.38	2.12		0.73	2.04	0.69		
1997	3.10	3.84		1.06	3.89	1.15		3.71
1998	2.04	1.86		2.39	2.09	2.09		1.16
1999	1.31	1.67	0.42	1.65	1.82	1.41		1.66
2000	1.72	2.42	0.55	3.37	2.36	3.29		2.64
All Years	1.93	2.02	0.52	1.48	2.71	1.69	3.22	2.50

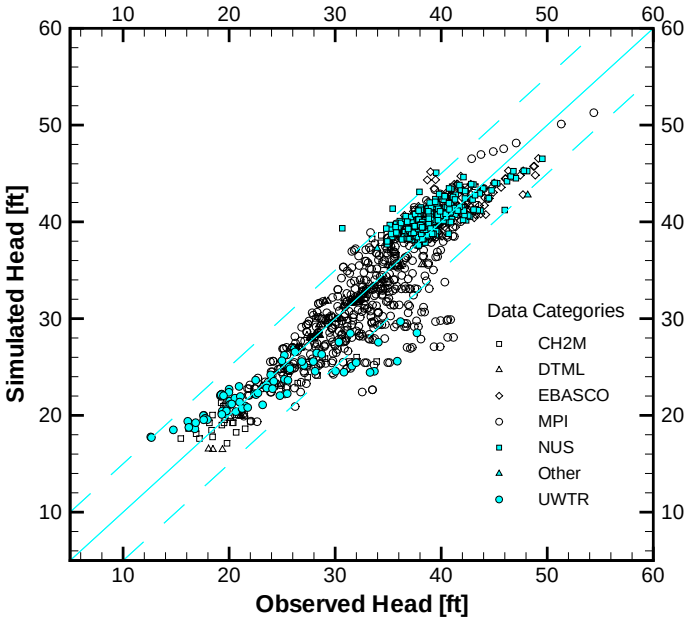

Figure 3.24: Scatter plot of observed versus simulated water levels (1383 data points)

3.6 Particle Tracking

Assuming no degradation and dispersion processes, the advective transport of a contaminant in the subsurface is identically the same as the movement of average water particles. A particle tracking analysis was conducted in this thesis by the use of DROG3D developed by *Blanton* [1995] at the University of North Carolina at Chapel Hill. DROG3D, written in Fortran 77, tracks passive particles with given harmonic velocity fields in a three-dimensional finite element mesh. A 5^{th} order Runge-Kutta algorithm was employed to integrate the path with an adaptive time step. Both forward and backward particle tracking are enabled. The code is also able to perform the computation of particle travel path within evolving seasonal mean or monthly mean velocity fields.

In this case study, particle paths were integrated over 351 monthly mean flow fields to investigate the movement of average water particles from the Reich Farm Superfund site. Originating from the source zone delineated by Malcolm Prinie Inc, the particles were

placed under the top layer with approximately 55 feet above sea level. Figures 3.25 and 3.26 show the water particle paths with varying retardation values and starting times. The time markers on the paths are spaced at 1 year intervals. Unlike the steady-state scenario, the average water particle travel paths depend on the starting times and retardation values in a transient groundwater flow field, because the flow field is significantly shifted towards the Parkway well field as the pumping activities continue. Comparing Figures 3.25 and 3.26 for average water particle paths with a retardation rate of unity, it can be observed that only the western most particles released in October 1971 do not migrate to the Parkway wells.

Based on the TCE breakthrough curve (shown in Figure 4.6) for the air-stripper influent accommodating flows from Wells 26 and 28, the peak concentration was observed between 1988 and 1991, which approximately corresponds to a travel time of 17 to 20 years with no dispersion assumption. Given an effective porosity of 0.30, Figure 3.25 shows that paths from the Reich Farm Superfund site to the Parkway well field roughly have the travel times of 12, 14, 17, and 20 years for several retardation coefficients in increasing order. Then, based on average water particle analysis, the TCE retardation coefficient in this case study is estimated to be between 1.25 and 1.75, which encompass the measured average retardation factors by EPA-certified Lancaster Laboratories [*Gillham*, 1998] (1.65 for the soil from the 12-24 ft depth, and 1.65 for the soil from the 24-33 ft depth). The lower bound of (R_{TCE}) is extended further to compensate the neglection of the dispersion process, which can possibly delay the occurrence of the peak concentration.

3.7 Summary

In this Chapter, a physically based transient three-dimensional groundwater flow model was developed using HydroGeoSphere and the methodology presented in *Sykes and Normani* [2002] (reproduced, in part, in the Appendix of this thesis). The flow model, which encompasses the Reich Farm Superfund site located in the Pleasant Plains area of Dover Township, Ocean County, New Jersey, incorporates a spatially and temporally varying recharge calculated by the ArcView GIS environment and an accepted hydrologic model

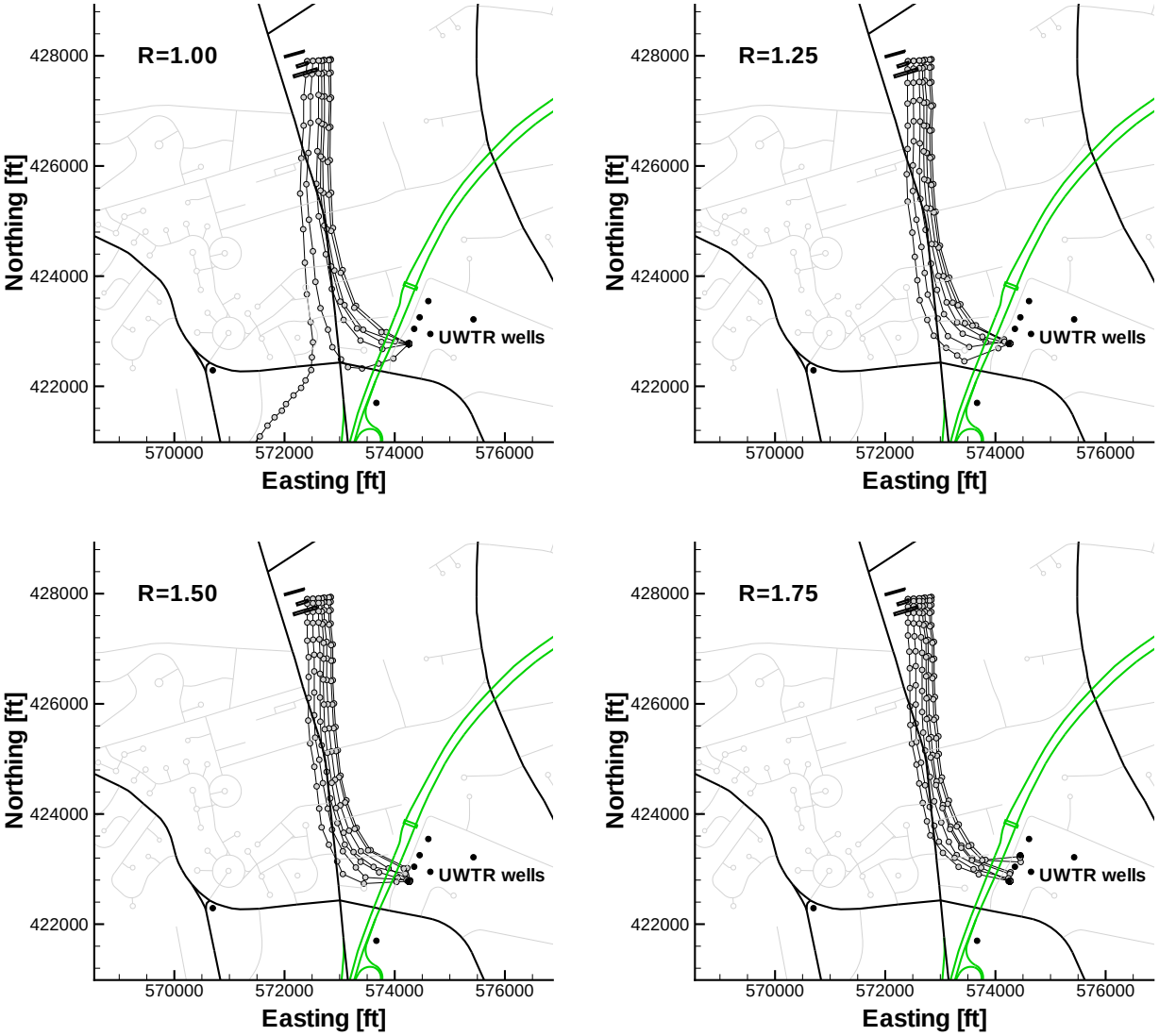

Figure 3.25: Forward average water particle tracks with varying retardation factors. Time markers are spaced at 1 year intervals, beginning in October 1971

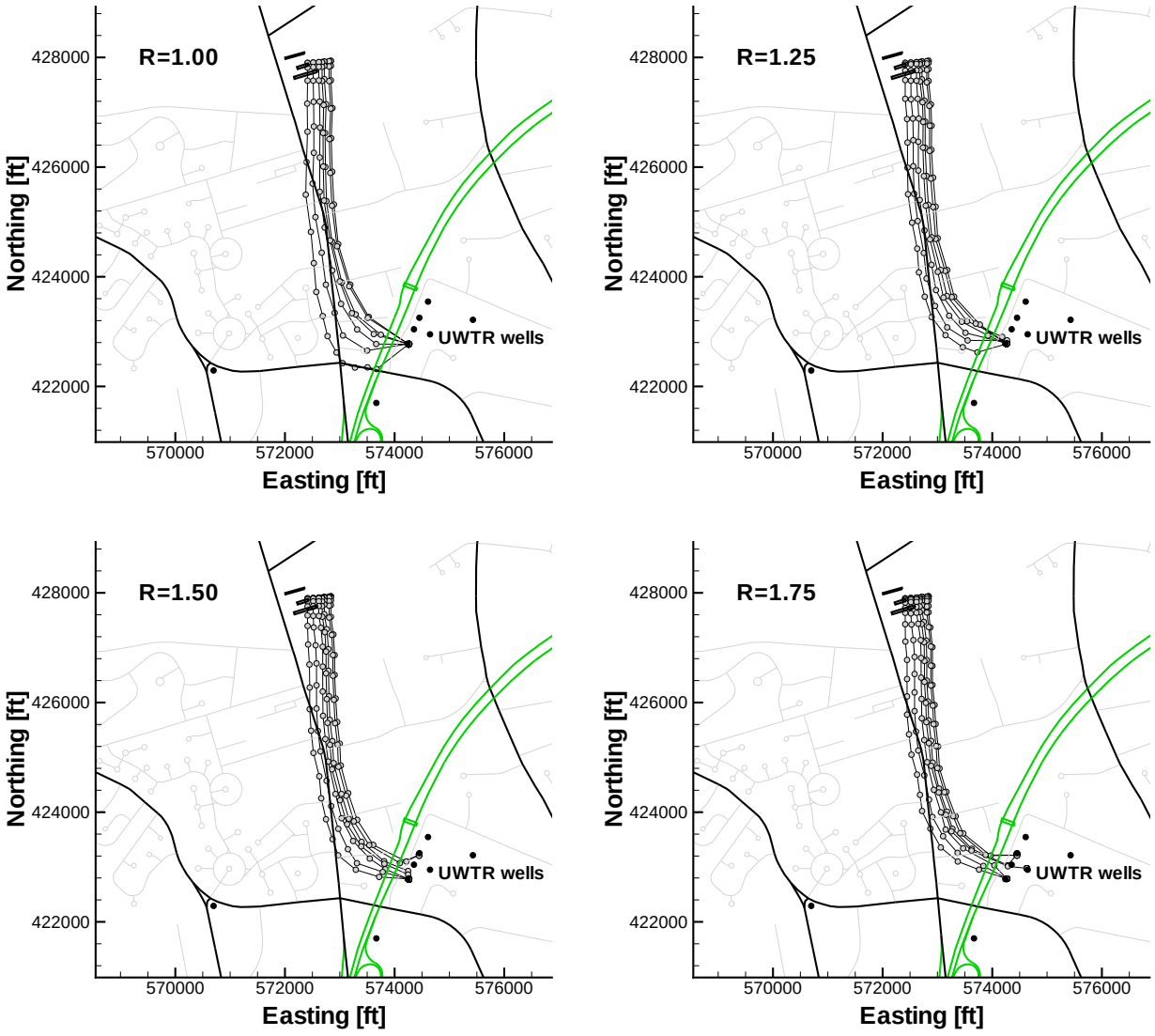

Figure 3.26: Forward average water particle tracks with varying retardation factors. Time markers are spaced at 1 year intervals, beginning in October 1972

HELP3. The hydraulic conductivity was derived by a kriging interpolation algorithm based on the estimated hydraulic conductivities from the lithology in well and boring logs. With highly changing pump rates at the 7 municipal wells, time increments over the approximately 30 year simulation period varied dynamically between several days and 3 months. Calibration of such a groundwater flow model was conducted using both approximately 9 years of head data from continuous well records and data over a period of approximately 30 years from traditional monitoring wells. A good agreement between the simulated and measured water levels was reached.

The simulation from 1971 to 2000 suggested that the groundwater flow around the Reich Farm gradually shifted to the Parkway well field as a result of the pumping activity. Then, the contaminants originating from the Reich Farm could be captured by the Parkway well field. A particle tracking analysis indicates that the retardation factor for TCE is estimated to be between 1.25 and 1.75, based on the TCE breakthrough curves from Wells 26 and 28 and the occurrence of the peak concentration.

Chapter 4

Reich Farm Contaminant Transport Model Calibration Case Study

4.1 Source Characterization

Contaminants in the aquifer from the Reich Farm site are NAPLs, including TCE, PCE and the styrene-acrylonitrile (SAN) trimer. Only trichloroethylene (TCE) that migrated from the site through the Kirkwood-Cohansey aquifer, and entered one or more of the Parkway wells, is modelled in this thesis. The mass transfer process at the interface between NAPLs and the aqueous phase has been extensively studied in recent years. The linear driving-force mass transfer model is commonly used to represent the interphase mass transfer flux, where the mass transfer coefficient can be related to the modified Sherwood number. The Sherwood number is frequently represented by a function of Reynolds number and the initial volumetric fraction of NAPLs. Sherwood number correlations were developed by *Miller et al.* [1990]; *Powers et al.* [1992]; *Imhoff et al.* [1994]; *Powers et al.* [1994b]; *Nambi and Powers* [2003]. However, their applicability is greatly limited by scale problems, and the difference of porous media or techniques used in experiments and field processes and methods respectively [*Zhu and Sykes*, 2000].

The location, strength and composition of a source play key roles in simulating the transport of contaminants. Incorrect source characteristics are likely to produce a poor match to measured concentrations or unsatisfactory predictions even if a good match is achieved. Simultaneous estimation of both source location and transport parameters can be problematic, due to the ill-posed nature of the inverse modelling problem. In this case study, the source locations are well known as a result of the soil sampling program at Reich Farm, and these locations can be used directly in the model. The contaminant barrels dumped at the Reich Farm site in 1971 were excavated and removed right after the incident. The dumping site is delineated in Figure 4.1. The three soil remediation areas, enclosed by sheetpiles in 1995 and identified as Areas 1, 2 and 3, are deemed as the location of the contaminant source. The dumping site is at a depth of 5 ft below ground surface. The representation of source location in the two-dimensional triangle mesh is shown in Figure 4.2. Vertically, the contaminant source area is considered right beneath the top HydroGeoSphere model layer, which has a depth of 5 ft. The same layer was used in the analysis of the release of average water particles. In this thesis, two different models of mass transfer from the NAPL residual to the solution phase are considered. In the first model, the NAPL residuals are assumed to be immobile, and a linear equilibrium model (refer to Equation (2.13)) can be employed to describe the mass transfer from the residual to the dissolved phase. Therefore, the contaminant source term can be approximated by an immobile phase that is in equilibrium with the dissolved phase. For this conceptual model, the source information can be represented by only two parameters: equivalent distribution coefficient at the source and the initial dissolved phase TCE concentration in Equation (2.13). A second model is developed in Section 4.7.5. The initial TCE concentration for the non-source area is 0, and all the transport boundary conditions are assigned as third-type Cauchy boundaries, where the fluxes are calculated from the flow model. Under favourable microbial conditions in groundwater, TCE concentrations have been observed to sharply decline from high levels to low levels. On the other hand, TCE also has been shown to be strongly conservative at some sites [*Benker et al.*, 1997; *McKelvie et al.*, 2007]. In the study area of interest, nutrient levels and TCE concentrations are too low to support microorganisms and there is no co-substrate or oxidant [*Providenti et al.*, 1993]. Therefore, the non-conservative process of biological or chemical decay has been neglected.

The detailed discussion on calibration with the second source characterization method will be presented in a later section.

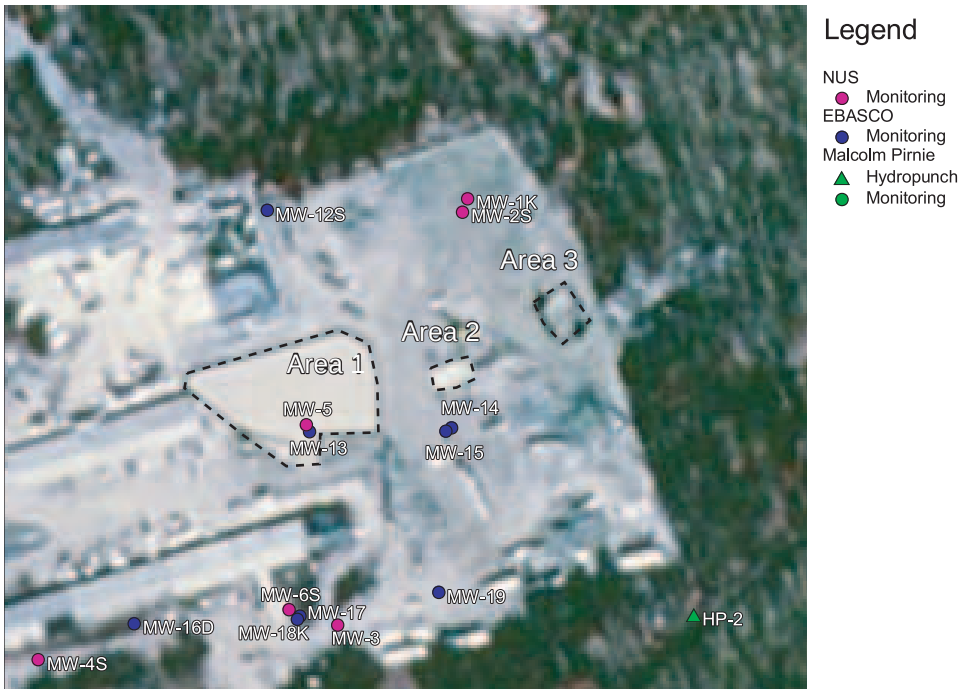

Figure 4.1: Locations of soil remediation areas, monitoring wells, and hydropunch borings on and near the Reich Farm Superfund site overlaying an aerial photo

4.2 Contaminant Characteristics

TCE is volatile, non-flammable and colourless. It is an industrial solvent that is also the most commonly found contaminant in groundwater. TCE exposure is associated with several adverse health effects, and several forms of cancer [*U.S. EPA*, 2001].

According to U.S. National Primary Drinking Water Regulations (Jun 2003), the maximum acceptable TCE concentration for drinking water in the United States is 0.005 mg/L (or 5 ppb). TCE, with an aqueous solubility ranging from 1300 to 1500 mg/L (or 1.3 to 1.5 $\times 10^6$ ppb) from 9 to 71°C (Henry's law is approximately constant for TCE between 10

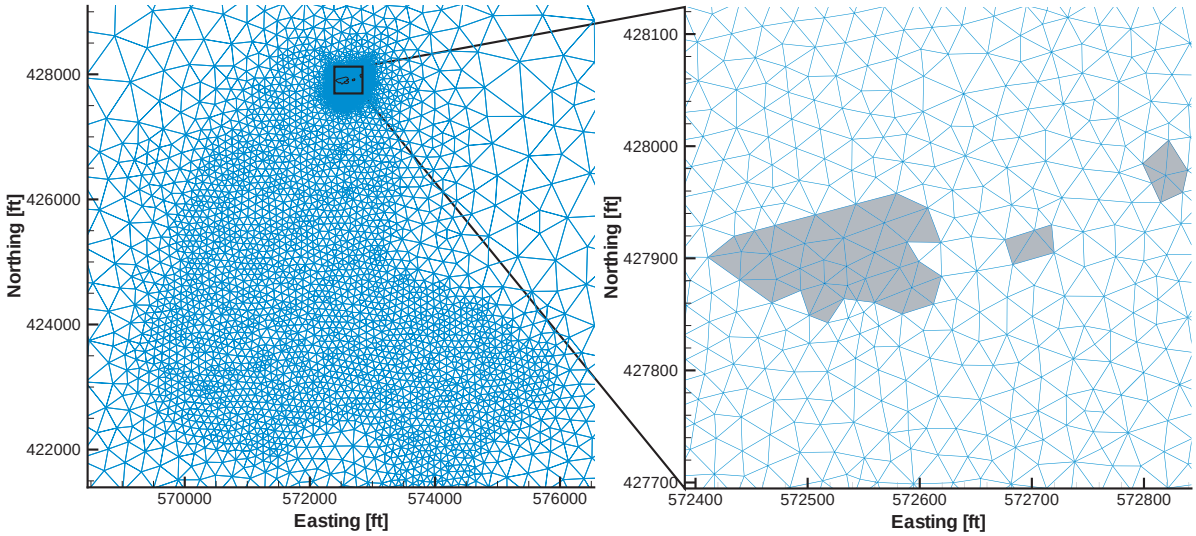

Figure 4.2: Defined source zone elements at the Reich Farm Superfund site

and 95°C), does not break down very readily in the soil, and it can pass through the soil into groundwater.

Additionally, because its density of 1.44 mg/l is higher than water, it tends to form pools of dense nonaqueous phase liquid (DNAPL) and settle at or below the lowest groundwater strata. The residence time of TCE in groundwater is much longer than in surface waters. From these subsurface pools, TCE may be slowly released over long periods of time. This may be problematic because long time exposure to even low levels of TCE can still cause health concerns. TCE present in soil or groundwater can be transformed by bacteria under anaerobic conditions into vinyl chloride, which is far more toxic than TCE. Therefore, extensive research on TCE removal by biological or chemical methods has been carried out in recent decades.

4.3 Separation of flow and transport model

Extensive investigations have been carried out to estimate the flow parameters. A review and summary of the inverse problem solution through flow modelling is given by *Yeh* [1986]. *McLaughlin et al.* [1996] reviewed flow inverse problems and provided a more

general formulation and methodology for parameter estimation. Solutions of the inverse problems for contaminant transport models have not been studied as extensively as the flow inverse problems, partly because of the ill-posed nature induced by the dispersion term. In spite of this limitation, research has been conducted on the coupled flow, transport, and/or heat inverse problems in an attempt to improve the calibration performance through the use of more data [*Woodbury and Smith*, 1988; *Sun and Yeh*, 1990; *Medina and Carrera*, 1996; *Jiang and Woodbury*, 2006; *Mayer and Huang*, 1999; *Friedel*, 2005; *Wagner*, 1992]. However, these studies have not investigated problems that are as computationally intensive as the problem of this thesis.

Woodbury and Smith [1988] employed the temperature measurements to improve the resolution of model parameters for a steady-state groundwater flow system. *Jiang and Woodbury* [2006] solved the inverse problem for a 2-dimensional steady-state groundwater flow and heat transport using a full-Bayesian approach. *Mayer and Huang* [1999] applied the maximum likelihood method to a coupled inverse problem of the saturated groundwater flow and solute transport. Calibration of the coupled modelling of vadose zone water, heat and solute transport was applied to an artificial recharge experiment [*Friedel*, 2005].

For the case study of this thesis, the field-scale groundwater flow and solute transport model is transient, variably saturated, and highly computationally intensive. To ensure the convergence of the solution, adaptive timestepping, based on changes in peizometric heads, must be used to adjust the timestep values. The result is that the timestep could be very small when the sharp changes are induced at the transient boundary conditions, such as pumping rates. More importantly, unsaturated properties of the porous media are dependent on the transient water saturation of each element, and must be calculated iteratively for every timestep. Compared to the fully saturated model, one more loop of computation, which is designed to find the proper unsaturated properties for every timestep, needs to be embedded in the solution process.

With the current model setup, it takes around 4 hours to run the 30 years groundwater flow and contaminant transport coupled HydroGeoSphere model on a single 2.2 GHz processor with 2 GB of RAM. Such a computationally intensive large scale model is normally unfeasible for most of the optimization methods, especially for the heuristic algorithms,

because at least hundreds of model calls are necessary to successively solve the optimization problem. For example, for the coupled model, 100 model calls would require approximately 17 days computation time on a single processor, while 500 model calls would require more than 83 days. It is clear that based on computational time, a more efficient modelling strategy much be developed.

Most of the computation time is attributed to the flow model. As described in Section 3.5.2, the flow model has already been well calibrated, and only the transport model is of interest in this thesis. In addition, in comparison to the measured hydraulic heads, the concentrations can vary over several orders of magnitude. The selection of the weights in an objective function or estimator for the misfit of the simulated head and concentrations as compared to the measured values of head and concentration respectively, which plays a key role in the calibration process, is subjective. In summary, for the purpose of calibration, decoupling of the flow and transport models is essential for the computationally intensive problem investigated in this thesis.

Through the separation of the transport model from the flow model, for every timestep, the transient flow field, head distribution, saturation, flux at the first type boundary conditions and pumping rates at wells are saved when running the flow model and loaded when dealing with the transport model. Then the required CPU time for each contaminant transport model run can be greatly reduced. For the coupled model, the required CPU time includes the running time for both the flow and transport models. Figure 4.3 compares the TCE concentration breakthrough curves from the coupled model and transport models with three different timestep schemes. For Parkway well 26, the break at the peak concentration occurring in 1988 was induced by the cease of pumping operation (refer to Figure 3.17). A variant timestep scheme generally means the use of smaller timesteps at the beginning and larger timesteps at the end of a simulation. In addition, it tries to honour the variation of the pumping rates at the Parkway well field. Transport models for all the three different schemes approximate the coupled model fairly well. Table 4.1 lists the total timesteps and CPU times for each employed timestep scheme. With an increase of timestep length, less CPU time is required, but the discrepancy of TCE BTCs at UWTR 26 becomes larger. The simulation with a 1 month timestep needs twice or more computa-

tion time than the other two timestep schemes, but the advantage of the improved match to the coupled model is not significant, given the purpose of the analysis of this thesis. The average simulation time for the variant timestep case is only 25% longer than the use of the 3 months timestep case, and it does not show the obvious advantage in the match with the coupled model. Therefore, in the following optimization experiments, the variant timestep scheme is used. By separating the flow model from the transport model and by using a variant timestep, the optimization of the transport parameters for 500 model calls can then be accomplished with a computation time of less than 9 hours. The CPU time for each model simulation or model call is reduced from the 4 hours required for the coupled model to around 10 minutes for the separated model.

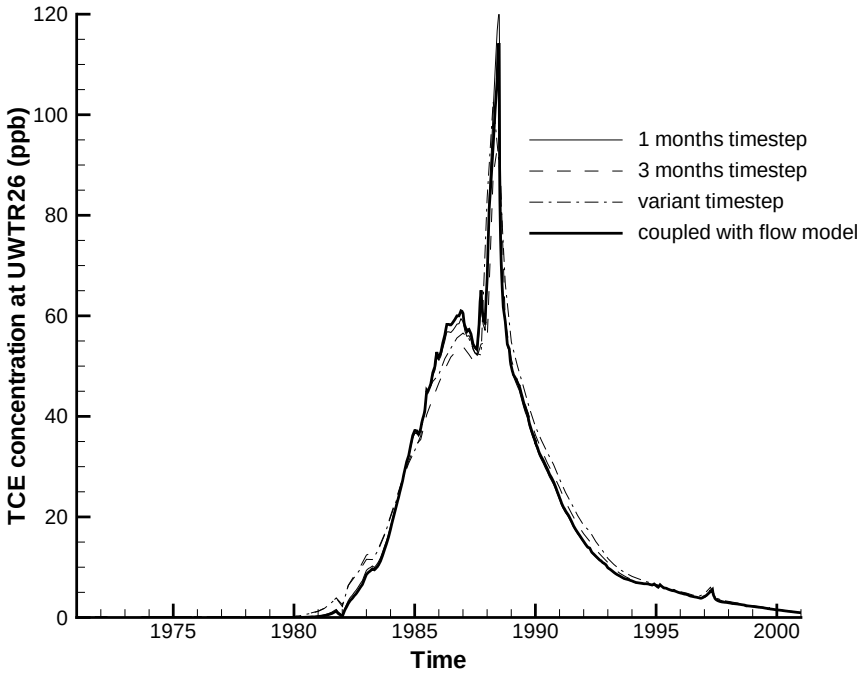

Figure 4.3: TCE concentration BTCs from coupled model and transport models with different timestep schemes

Table 4.1: Number of timesteps and computation time for different timestep schemes

	Timesteps in total	CPU time
1 month timestep	351	∼24 minutes
3 months timestep	117	∼8 minutes
Variant timestep	138	∼10 minutes
Coupled with flow model (Max 1 month)	∼700	∼4 hours

4.4 Parameterization

Spatial and temporal variations of the model parameters can lead to a large number of parameter values that must be estimated. However, the estimation of numerous parameters is generally restricted by factors such as a limited number of observations and limited computing resources for this thesis. A relatively small group of contaminant transport model parameters were selected as optimization decision variables and their values were iteratively adjusted to best match measured concentrations. Given the fact that the transient flow model has been carefully and successfully calibrated, the time varying Darcy flux distribution in the study domain has been determined. In addition, for this specific site, the flow field was driven and dominated primarily by the pumping wells. Given the fact that the pumping rates were well defined, the flow field was well determined. The unknown parameters of the transport model in Equation (2.14) are transport parameters including porosity θ, hydrodynamic dispersion tensor D_{ij}, distribution coefficient , K' (or retardation factor R), and others characterizing the contaminant source term including the equivalent distribution coefficient at the source, K'_{source} and the initial source TCE concentration, $C_{TCE,initial}$ for the first source term model investigated. For the left hand side (LHS) of Equation (2.14), the Darcy flux is fixed by the calibration of the flow model. The product of the porosity and the hydrodynamic dispersion tensor is comprised of the mechanical dispersion and the effective diffusion in Equation (2.12). For a highly dynamic groundwater system, the diffusion process is dominated by the mechanical dispersion, which only depends on the dispersivities and the Darcy flux. This statement will be justified in

Section 4.7.4. On the right hand side (RHS) of Equation (2.14), the porosity and the retardation factor are inversely related under the conditions that the Darcy flux is fixed and the diffusion process is negligible. The retardation factor is a function of porosity and distribution coefficient K' shown in Equation (2.15). Therefore, it is impossible to calibrate both the porosity and retardation factor (or distribution coefficient which still holds a one-to-one mapping relationship with porosity) simultaneously. The selected parameters of the transport model are considered as more uncertain and sensitive compared to the flow parameters. Then, the unknown transport parameters were reduced to the following: longitudinal dispersivity, a_l, transverse dispersivity/longitudinal dispersivity, a_t/a_l, transverse vertical dispersivity/longitudinal dispersivity, a_{tv}/a_l, and distribution coefficient, K'. All of the parameters are assumed to be constant across the domain over the time duration of the simulation. Upper and lower limits on parameters that constrain the estimated values are commonly suggested, because unrealistic parameter values can be obtained both through the inverse modelling due to its ill-possed nature and a possible lack of sufficient information. The box constraints listed in Table 4.2 are derived based on field experiment data [*Gelhar et al.*, 1992] and sensitivity analysis in previous manual calibration efforts by the author of this thesis.

Table 4.2: Decision variables and box constraints

Decision variables	Lower bound	Upper bound
Longitudinal dispersivity, a_l [ft]	1.0	50.0
Transverse/longitudinal dispersivity, a_t/a_l	2.0E-03	0.5
Transverse vertical/longitudinal dispersivity, a_{tv}/a_l	1.0E-03	0.2
Distribution coefficient, K' [ft^3/kg][a]	1.2E-03	6.0E-03
Equivalent K' at the source , K'_{source} [ft^3/kg]	2.0E-04	1.0E-02
Initial source TCE concentration, $C_{TCE,initial}$ [kg/ft^3][b]	2.0E-03	1.0E-02

[a] 1 kg = 2.204623 lb, 1 ft^3/kg = 28316.8 cm^3/kg

[b] 1 kg/ft^3 = 3.53×10^{-5} kg/cm^3 = 3.53×10^7 ppb

4.5 Objective Function Definition

The contaminant transport model calibration focused on replicating the measured solute concentrations for the Reich Farm TCE plume. The comparison of simulated and observed values is accomplished quantitatively using an objective function. Parameter values that produce the best fit are defined as those that give the smallest value of the objective function. Various forms of objective functions or estimators exist. Some of the estimators accentuate small residuals (defined as the difference between simulated and measured data), while others are very sensitive to large residuals or outliers. An example of the latter is the convex optimization [*Boyd and Vandenberghe*, 2003]. Both of these two cases could result in unfavourable outcomes. Thus, an appropriate selection of the objective function plays an important role in the calibration process

The objective function in PEST is formulated such that the sum of the squared residual errors (SSE, or L_2 estimator based on L_2 norm [*Xiang et al.*, 1993]) between the measured and simulated flux averaged concentrations is minimum at all times. SSE guarantees that the estimator is unbiased from a statistical point-of-view. The drawback of SSE lies in the fact that it tries to fit all of the measurements including outliers. The outliers, which are extreme measurements that do not belong with the other measurements, can arise from errors in data entry, analytical instrument failure and inherent spatial or temporal variability in concentrations [*Gibbons*, 1994]. Given the fact that SSE is equivalent to assigning more weight to outliers, it is very sensitive to the outliers [*Carrera et al.*, 2005]. SSE in equation (A.1) can be viewed as a weighted average of the absolute residuals, where the equivalent weights are dependent on the magnitude of the errors.

Minimize

$$\Phi = \sum_{i=1}^{N_{obs}} w_i \left(C_{obs,i} - C_{sim,i}\right)^2 = \sum_{i=1}^{N_{obs}} w_i' |C_{obs,i} - C_{sim,i}|, i = 1, 2, 3, \ldots, N_{obs} \tag{4.1}$$

where:

$$w_i' = w_i |C_{obs,i} - C_{sim,i}| \tag{4.2}$$

and where, $C_{obs,i}$ and $C_{sim,i}$ are measured and simulated TCE concentrations (ppb) respectively, w_i is the weighting factor associated with i^{th} observation (the default value is 1), w_i'

is the equivalent weighting factor, and N_{obs} is total number of observations. As a result, the parameter estimation might be heavily deviated by the outliers. The L1-norm, sum of the absolute residuals is often considered as a robust estimator as it treats the errors with equal equivalent weight [*Xiang et al.*, 1993]. To deemphasize the outliers during optimization and to honour the formulation of objective function definition in PEST simultaneously, a modified SSE is introduced as follows in (4.3):

Minimize

$$\Phi = \sum_{i=1}^{N_{obs}} w_i \left(C_{obs,i}^x - C_{sim,i}^x\right)^2, i = 1, 2, 3, \ldots, N_{obs} \tag{4.3}$$

where, the parameter x denotes the strength of deemphasizing factor. The transformed concentrations are assumed to obey Gaussian statistics. The equivalent weights of residuals for the newly introduced objective function can be expressed as:

$$w_i' = w_i \frac{\left(C_{obs,i}^x - C_{sim,i}^x\right)^2}{|C_{obs,i} - C_{sim,i}|} \tag{4.4}$$

The equivalent weights are dependant on both residuals and observed data themselves. Thus, the parameter x can only be selected so as to mimic the L1-norm estimator in the desired range. Figure 4.4 compares equivalent weights by varying x with an average observed TCE concentration of around 20 ppb. The dashed red line representing the weights for SSE is a straight line with a slope of 1. For this specific setting, the scenario with a power of 0.8 is superior than others in approximating the horizontal line of the L1-norm weights.

Figure 4.4 demonstrates that for a very specific range of observation data, the modified SSE with a power parameter of 0.8 can produce roughly the same equivalent weights for residuals as the L1-norm estimator. Its application to other values of observations were extended and shown in Figure 4.5, which compares SSE, sum of absolute residuals and modified SSE when the difference between measured and simulated TCE concentration is less than 10 ppb. The equivalent weights associated with outliers are largely reduced.

It should be noted that no penalty term is adopted in this objective function definition.

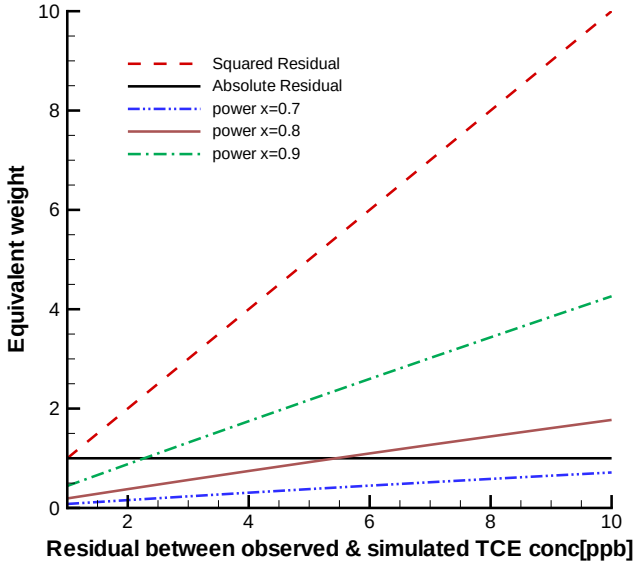

Figure 4.4: Equivalent weights for different power parameters at an observed TCE concentration of 20 ppb.

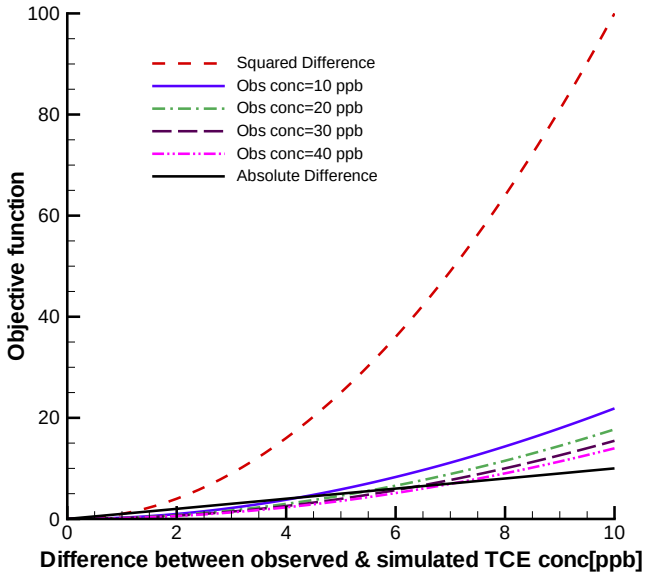

Figure 4.5: Comparison of objective function

The penalty functions are frequently motivated by the desire to use unconstrained optimization techniques to solve constrained problems. For this case study, the box constraints in Table 4.2 are relatively easy to be handled by the selected optimization algorithms. The regularization terms, by adding to the objective function, enable the ill-posed inverse problem to be a well-posed optimization problem [*Skaggs and Kabala*, 1994]. However, the regularization terms improve the stability of the inverse problem at the expense of accuracy [*Neupauer et al.*, 2000]. In addition, the relative weight assigned to the regularization terms is essential and hard to determine. Therefore, both the penalty term and the regularization term were not incorporated in the objective function definition for this thesis, and could be considered for the future work.

4.6 TCE concentration measurements

Concentration observations, which contain information about the dynamics and features of a groundwater system, are commonly used to estimate the value of model parameters by minimizing the objective function defined above. Goodness of fit, with clearly realistic parameter values, suggests appropriate model construction and parameterization.

Pumping and observation wells at 101 locations shown in Figure 3.20 were sampled and analyzed since 1985 by NUS, Ebasco, CH2M-Hill, and Malcolm Pirnie Inc. Because different laboratories were used, the detection limits of the analytical method employed to test groundwater samples for TCE vary from 0.3 to 10 ppb. Because a large portion of samples referred to as non-detectable can provide valuable information as well, based on the uniform distribution assumption between 0 and detection limits, an unbiased estimator of the TCE concentrations for non-detects were equal to one half of the associated detection limits. In fact, it is trivial of the impact of this arbitrary selection of representative concentrations for non-detects.

Of the 3947 TCE measurements in total, 135 data were sampled from the Dover Township Municipal Landfill (DTML; the location is shown in Figure 3.1) monitoring wells and most of them are below detection limits. DTML monitoring wells are far from the Reich

Farm Superfund site. Therefore, TCE detected in those wells is believed to originate from the local landfill. These measurements are not considered in the calibration process. A further 1836 measurements from the private wells could not be included, because the wells can not be precisely located. Another portion of the data came from surface water, UWTR distribution system, and the UWTR intermediate treatment and air stripper effluent measurements. These data are not directly related to the groundwater system and will not be considered. The remaining 935 TCE measurements shown in Figure 4.6, are used to facilitate the calibration of the transport model. The fact that the peak concentration appeared around 1990 demonstrates approximately 19 years travel time from the Reich Farm site to the well field. TCE observations are limited to recent times from 1989 to 2001, long after the start of substantial pumping at the Parkway well field. In a spatial sense, the data locations are densely clustered in the well field in an area of high hydraulic conductivity. Data clustering is not problematic in this case, because in the pumping areas with high hydraulic conductivity, sensitivities of observations to most parameters are inclined to be relatively small and the clustered wells will not adversely affect the calibration [*Hill and Tiedeman*, 2007]. It should be noted that multiple TCE concentration measurements from the same locations in a timestep of the transport model were not averaged, and they are still regarded as multiple measurements. Concentration measurements can vary over many orders-of-magnitude, and suffer from a variety of unexpected random errors, reading error and equipment error. The concentration measurements at the pumping wells are regarded less sensitive to the numerical dispersion than those at the monitoring wells, because of the conservation of the mass balance and the use of flux averaged concentrations, which tend to suppress the numerical oscillation, especially at the margin of the plume. In addition, multiple measurements give more confidence, and deserve more weight to be assigned to them.

Data sources fall into 6 categories listed in Table 4.3, in which air stripper influent is a mixture of pumping wells 26 and 28. Influent concentration can be calculated through the following mass balance equation.

$$C_{airstripper} = \frac{Q_{26}C_{26} + Q_{28}C_{28}}{Q_{26} + Q_{28}} \tag{4.5}$$

where, C_{26} and C_{28} are concentrations in well 26 and well 28, and Q_{26} and Q_{28} are pumping

rates of well 26 and well 28. There are 484 (268 from UWTR + 216 from air stripper influent) of the 935 TCE measurements attributed to well 26 and well 28. Thus, these two heavily monitored wells are selected to compare the simulated breakthrough curves with measured data. The approximate extent of the plume delineated in Figure 4.7 is based on a concentration contouring of the averaged concentration distribution from 1991 to 1993 inclusive. These measurements are from a variety of sampling methods: shallow and deep observation wells, pumping wells and hydropunch samples. The general trend of plume migration complies with the particle tracking path discussed in a previous section. TCE initially migrated with the groundwater flow to the southwest, then shifted towards the Parkway well field as a result of the pumping operation.

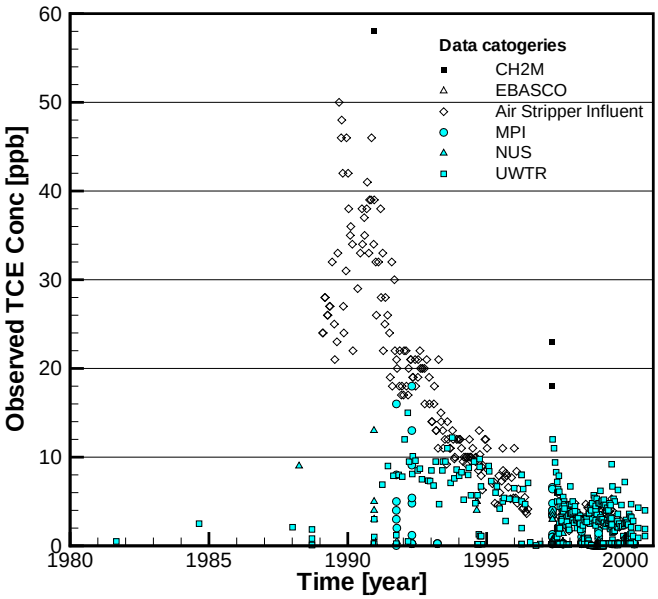

Figure 4.6: Scatter plot of observed TCE concentrations versus sampled time (935 data points)

One thing needs to be emphasized here is that measured contaminant concentrations obtained from pumping and observation wells are flux averaged values over screen intervals. Their counterparts in the transport model must be the same type of concentrations. Using

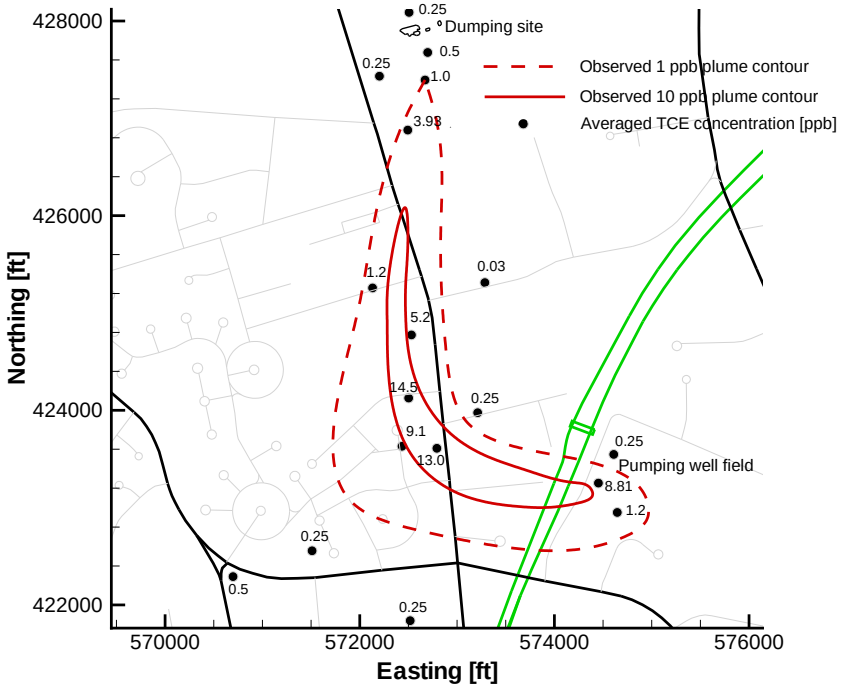

Figure 4.7: Observed TCE contaminant plume from 1991 to 1993 [*Normani*, 1998].

Table 4.3: A list of data sources and number of measurements employed in the calibration.

Category	Number of TCE measurements
CH2M	7
EBASCO	17
Air stripper influent	216
MPI	90
NUS	30
UWTR	562

flux averaged values in calibration avoids the dependence on possible vertical heterogeneity in the model, which could be problematic for point measurements.

4.7 Results and Discussion

4.7.1 Outline of algorithm comparisons

The comparisons are focused on the ability of each algorithm to optimize the objective function in calibrating a case specific transport model, in this case for the TCE plume from the Reich Farm site. Candidate algorithms for comparison include straight PEST, multi-start PEST, MicroGA, straight DDS, and multi-start DDS. As a result of the stochastic nature of algorithms compared here, their relative performance must be assessed over multiple independent optimization trials. On the basis of statistical consideration, algorithms are compared using 30 optimization trials. For each trial, the maximum number of transport model evaluations is constantly 500 for DDS and MicroGA. In other words, 500 model evaluations serve as a stopping criteria. For PEST, the number of model calls varies depending on the convergence criteria. In global optimization problems, a performance measure is frequently utilized for the purpose of algorithm comparison. Average performance measure denoted by the best solution is plotted against the number of model evaluations for each algorithm. In other words, for a particular algorithm, the average of the best objective function value found so far across all optimization trials is computed after each model call. Because the performance measure can not provide a complete picture of results, the empirical distribution of final best solutions is also investigated to assist in the comparison.

Algorithm parameters often exert a substantial influence on the effectiveness of the calibrations. Dozens of algorithm parameters could be employed to tune PEST to the particular case problem. In this thesis, default settings for all the algorithm parameters are used. The only one algorithm parameter in DDS, the neighbour size parameter, r, is set to the default value of 0.2. The individuals in MicroGA are comprised of 60 binary bits (10 binary bits for each parameter). A population size of 4 is recommended for each generation.

Thus, there are 125 generations in total. Crossover probability is set at 0.5. All the other algorithm parameters are set to the default value, unless otherwise noted. Computing work was carried out on a cluster of processors (Sharcnet), which is comprised of thousands of 2.2 GHz AMD's dual-core Opteron processors with 2 GB Of RAM each. The estimated simulation time for a single transport model evaluation is around 10 minutes.

4.7.2 Comparison of optimization algorithms

The HydroGeoSphere TCE transport calibration problem described previously is solved using various optimization algorithms with multi-start technique. Table 4.4 summarizes the average number of transport model evaluations, CPU time, and elapsed time for each optimization algorithm. For straight PEST trials, the average number of function evaluations is about 85 plus an additional 10 model evaluation calls, which were utilized by the constructive run (described in Section 2.3.1) to generate a better initial solution. Then, a 5 multi-start PEST can provide comparative results with DDS and MicroGA in terms of the number of transport model evaluations. The actual CPU time was not recorded, but could be estimated by the number of model evaluations multiplying by the CPU time for each transport model run, because the computing time assigned to optimization algorithms themselves is negligible. All the computing effort was conducted on the Sharcnet, where thousands of processors are readily available. Ideally, all the optimization trials could be run simultaneously on the Sharcnet. Thus, comparing the elapsed time would make more sense. The actual elapsed time for the multi-start scheme is independent of the number of multi-start, and equal to CPU time for just one trial. In addition, there are 4 processors on each node of Sharcnet clusters, so the elapsed time for parallelized MicroGA is approximately equal to a quarter of the CPU time for one trial. Although Table 4.4 demonstrates that straight PEST requires the least CPU time and elapsed time, its calibration performance needs to be compared further with that of the other algorithms.

Figure 4.8 illustrates the average best objective function values against the number of model evaluations for DDS and MicroGA, and the final best objective function values of 30 trials for PEST. The output of objective function values with PEST occurs at a certain

Table 4.4: Number of model calls, CPU time, and elapsed time for different algorithms

Algorithm	Number of model calls	CPU time for 1 trial	Elapsed time for 1 trial
PEST	∼85+10	15hr50min	15hr50min
3 multi-start PEST	∼255+30	47hr50min	15hr50min
5 multi-start PEST	∼425+50	79hr10min	15hr50min
MicroGA	500	83hr20min	20hr50min
DDS	500	83hr20min	83hr20min
5 multi-start DDS	500	83hr20min	16hr40min

number of model calls depending on each trial and local search iteration. No intermediate results are available. Thus, only 30 scattered dots corresponding to 30 trials of each PEST algorithm are plotted in Figure 4.8. Clearly, both straight DDS and MicroGA are converging to poor local minima with DDS performing slightly better. The 5 multi-start DDS is comprised of 5 DDSs, and each can run the transport model evaluations at most 100 times. As the number of model evaluations increases, straight DDS keeps refining solutions by searching the local neighbourhood further, while the multi-start technique is capable of eliminating poor solutions. In this case, the marginal benefit of model call increment is dominated by continued local search in the early stage, but taken over by the multi-start technique when the total model calls is more than around 100. Some of the scatter for PEST are far below DDS and MicroGA curves, which indicates that PEST is more efficiently convergent to fairly good solutions. The probability of finding these solutions is getting higher with more multi-starts or model evaluations.

Figure 4.9 provides a more complete description of algorithm performance by plotting all 30 objective function values with fixed levels of computational effort (500 model evaluations) for DDS and MicroGA, and various levels in Table 4.4 for PEST. In terms of final best solutions, multi-start PEST is superior to the heuristic algorithms (DDS and MicroGA). It is worth noting that at the probability of 0.4, straight PEST is better than DDS and MicroGA. This may indicate that 40% of the selected initial solutions for PEST are located in the area of fairly good solutions. Heuristic algorithms are designed to find

good solutions, instead of precise local or global optima. The performances of two heuristic algorithms, DDS and MicroGA, seem to be nearly indistinguishable from a practical calibration perspective, while multi-start DDS is slightly better in this case study. Overall, knowing that DDS is a more appropriate optimizer for high dimensional (10 or more) problems when total allowable model evaluations are essentially unlimited (which is not realistic for such a computationally high demanding transport model calibration) [*Tolson and Shoemaker*, 2007], the 3 multi-start PEST with default algorithm parameter setting and even less model evaluations outperforms DDS in terms of objective function values. It is also more effective than straight PEST because of substantial performance improvement, and more efficient than 5 multi-start PEST by having indistinguishable different performance measure while requiring less computing effort. However, it should be noted that the comparison conclusions are largely drawn from very specific circumstances (only one case study and the limited model evaluation of 500). To guide the selection of estimation in other inverse modelling problems, algorithm comparisons need to be further investigated.

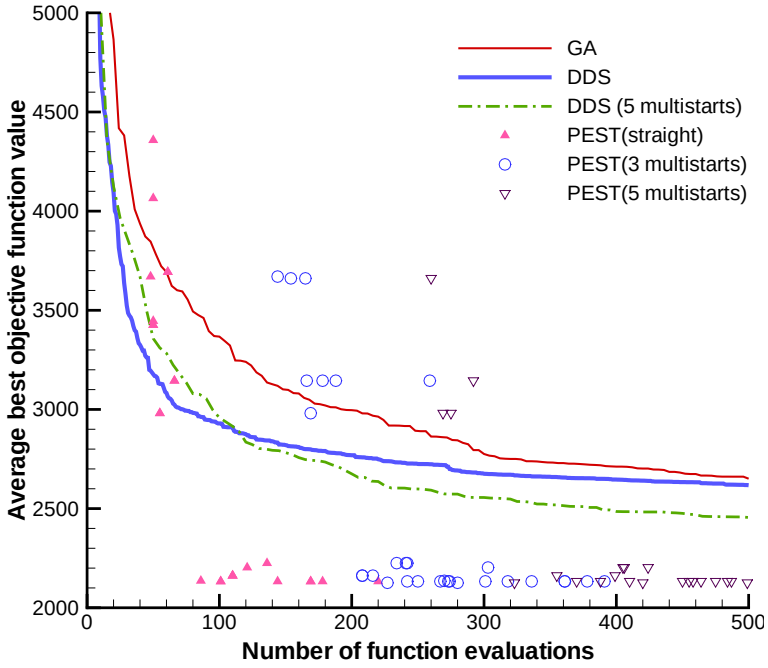

Figure 4.8: Average best objection function values versus the number of model evaluations

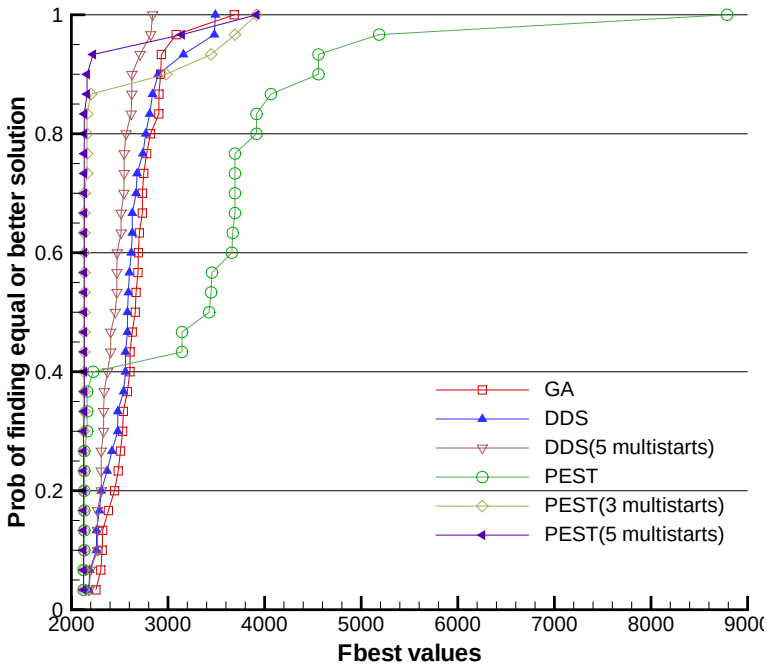

Figure 4.9: Empirical cumulative distribution function of final best solutions

4.7.3 Evaluation of contaminant transport model calibration results

Due to the ill-posed nature of inverse modelling problems, the solutions of concern are non-unique. Uncertainty analysis incorporating multiple solutions will be carried out in the next chapter. It must be noted that without exhaustive searching of the feasible space, it is arduous to determine a global minimum for the ill-posed inverse problems with complex non-convex response surfaces [*Ivanov et al.*, 2005]. Many solutions exist that are equally good candidates for the optimal solution. The solution with the best performance measure cannot 100% guarantee attainment of the global minimum. In this section, calibration results from the PEST trial with the best performance measure, as an example, are evaluated to determine whether the estimates are realistic. From the calibration process, the PEST estimated local optimal combination of parameters and associated confidence intervals are presented in Table 4.5.

The estimated retardation factor for the whole study domain is calculated using Equation (2.15):

$$R = 1 + \frac{\rho_b}{n} K' = 1 + \frac{50.97 \text{kg/ft}^3}{0.3} \times 3.4 \times 10^{-3} \text{ft}^3/\text{kg} = 1.67$$

The measured average retardation factors over 9 samples by EPA-certified Lancaster Laboratories [*Gillham*, 1998] are 2.17 for the soil collected from the 2-15 ft depth, 1.65 for the soil from the 12-24 ft depth, and 1.65 for the soil from the 24-33 ft depth. As the TCE plume mostly migrated in the lower aquifer, the value of 1.65 was selected as the representative retardation factor of this study domain. Then, a very good agreement was reached between the calibrated and measured retardation factors. Because there exists a one-to-one mapping relationship between the retardation factor and porosity, this agreement justifies the selected porosity of 0.3 in Section 3.4.4.

Table 4.5: Summary of parameter estimates and confidence intervals derived from PEST calibration trial

	Estimated	**9**5% percent confidence limits	
Parameter	value	Lower limit	Upper limit
a_l [ft]	10.1	7.8	12.4
a_t/a_l	2.0E-03	-2.8E-02	3.2E-02
a_{tv}/a_l	9.0E-03	6.5E-03	1.1E-02
K' [ft^3/kg]	3.4E-03	3.2E-03	3.6E-03
K'_{source} [ft^3/kg]	2.0E-04	-3.0E-03	3.4E-03
$C_{TCE,initial}$ [kg/ft^3]	8.1E-03	6.5E-03	9.8E-03

Two parameters, the ratio between transverse and longitudinal dispersivities, and equivalent distribution coefficient at the source, reach the lower bounds. The relatively large intervals for these two parameters indicate less confidence in the optimized parameter value. In other words, with the current settings, these parameters are not sensitive to the performance measure, and consequently not able to be effectively calibrated.

To save computation time, the mesh size is enlarged when discretizing the study domain, and the flux based concentrations are averaged temporally over 3 months and spatially across well screens. These model simplification techniques definitely contribute to a large numerical dispersion. More importantly, the variation in the local velocity caused by the spatial and temporal fluctuation of the recharge tends to increase the advective dispersion, compared to the normally used averaged recharge. To verify this phenomenon in this thesis, the 3 multi-start PEST was employed to calibrate the transport model parameters under different scenarios including spatially and temporally varying recharge and averaged recharge. As for the equivalent distribution coefficient at the source, one reason lies in the coarse discretization of the study domain and the time increments selected for the purpose of the computational efficiency. The numerical dispersion induced by the truncation error is proportional to the grid size and the timestep. Given the condition that the total mass of the contaminants is conserved, the numerical dispersion tends to suppress the sharp front of the plume, prolong the actual duration of contaminant source release, and reduce the estimated equivalent distribution coefficient as a result. The other explanation is that because the contaminated site was cleaned up three times after the dumping event; the source term might behave like a step-wise function. Therefore, for these two parameters, negative values of the lower limits of 95% confidence intervals were obtained, due to the small magnitude of the estimated values and insensitivity to the current performance measure. In the stochastic analysis of Chapter 5, a positive constraint on all the parameters will be enforced to produce reasonable estimates and reduce the associated uncertainty. Calculated confidence intervals around other parameters exhibited a relatively narrow range, suggesting a good level of precision in the final parameter estimates.

Correlation coefficients are defined as the covariance between two parameters divided by the product of their standard deviations. *Hill et al.* [1998] suggested that the absolute value of parameter correlations in the range of 0.80 to 1.0, particularly those in excess of 0.95, may be indicative of problematically strong correlations between parameter pairs. Strong correlation, dictating the dependency between the parameters, may do harm to the uniqueness of the optimized parameter values. The correlation coefficient matrix corresponding to the estimated parameters above is summarized in Table 4.6. Longitudinal and vertical transverse dispersivities have a strong negative correlation of -0.96. This high cor-

relation does not represent the actual relationship between these two parameters, because the vertical transverse dispersivity could not be properly calibrated due to the scarcity of vertically distributed data. For instance, the flux based concentrations were vertically (perpendicular to the flow direction shown in Figures 3.14, 3.15) averaged over the well screens. The other strong negative correlation exists between source concentration and the equivalent distribution coefficient, which indicates that the total mass of the contaminants at the source, expressed as the product of these two parameters, is constant but unknown.

Table 4.6: PEST parameter cross correlation matrix

	a_l	a_t/a_l	a_{tv}/a_l	K'	K'_{source}	$C_{TCE,initial}$
$a_l ft$	1.00	-0.66	-0.96	-0.53	0.18	0.21
a_t/a_l	-0.66	1.00	0.58	0.18	0.26	-0.51
a_{tv}/a_l	-0.96	0.58	1.00	0.39	-0.13	-0.24
K'	-0.53	0.18	0.39	1.00	-0.71	0.41
K'_{source}	0.18	0.26	-0.13	-0.71	1.00	-0.91
$C_{TCE,initial}$	0.21	-0.51	-0.24	0.41	-0.91	1.00

Figure 4.10 analyzes the model fit graphically by showing a scatter plot for all simulated versus measured TCE concentrations for all times; around half of the measured TCE concentrations are below detection limits, and most of the data lie within the bands of diagonal $\pm$ 5 ppb. Table 4.7 represents the averages of the absolute value of residuals between simulated and observed TCE concentrations for monitoring years and well categories. Compared to Table 3.3, the relative residuals for TCE concentrations are more pronounced than hydraulic heads, because (1) concentration measurements can vary many orders of magnitude, (2) concentrations are more likely affected by the particular representation of heterogeneity in the model, and (3) due to differences in the sampling methodology and laboratory analysis of the various contractors for the Reich Farm site.

The overall average of the absolute value of the residuals is 1.48 ppb. Except for CH2M-Hill with only 7 observations, the air stripper influent, with the highest level of

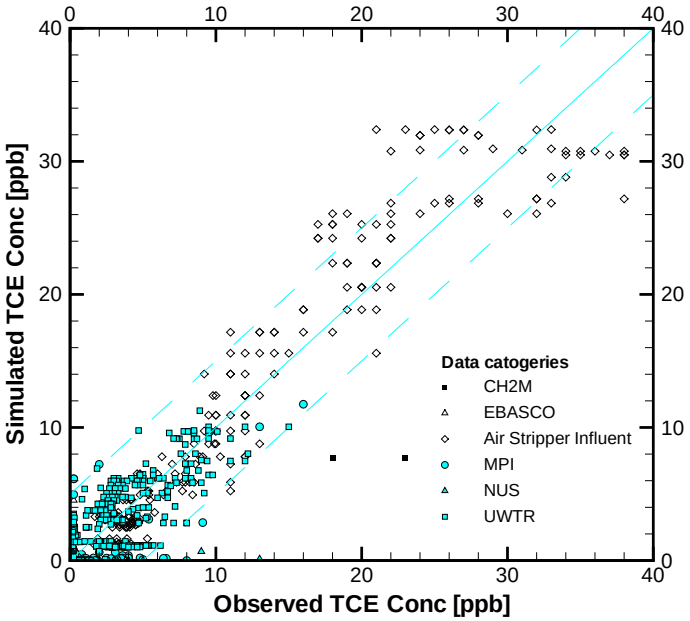

Figure 4.10: Scatter plot of observed vs. simulated TCE concentrations (935 data points)

TCE observations, has the largest average residual. In a time sense, the residual for "all" category is most variable for the peak TCE concentration observed around 1989, and decreases gradually in both directions of the time scale. No model bias calibration suggests that residuals, which often follow a normal distribution with mean 0.0, is supposed to be independent of observed values. The positive correlation between residuals and observed values here is induced by the new definition of the objective function. To deemphasize the outliers in the calibration, changing the objective function would produce biased estimates from a statistical point of view. Thus a trade-off needs to be balanced between unbiased estimation and better calibration performance.

Given the estimated transport parameters in Table 4.5, running the transport model simulation generates TCE concentration breakthrough curves for well 26, well 28, and the air stripper influent (blended effluent from well 26 and well 28), as shown in Figure 4.11. The calibrated parameters are capable of reproducing the measured TCE breakthrough curves at Wells 26 and 28 and the air stripper. Their modified SSE in Equation (4.3)

Table 4.7: Average of absolute value of TCE residuals [ppb] by category and year

	Category						
Year	**A**ll	**C**H2M	**E**BASCO	**A**ir stripper influent	**M**PI	**N**US	**U**WTR
1981	0.50						0.50
1984	2.48						2.48
1988	3.50					8.32	2.54
1989	5.45			5.45			
1990	3.30	2.82	0.25	5.12		3.66	1.36
1991	3.65			4.73	4.21	0.50	1.43
1992	2.30		0.25	2.55	2.88		1.37
1993	2.14			2.53	0.93		1.66
1994	1.49		0.45	1.32		4.49	1.50
1995	1.88			1.93			1.80
1996	1.48			0.92			1.88
1997	1.45	12.98	1.27	0.82	1.85	1.58	0.98
1998	1.09		0.25	0.83	1.00	0.45	1.15
1999	0.78			1.41	1.23	0.40	0.69
2000	0.86			2.07	0.32		0.68
All Years	1.48	6.74	0.51	2.48	1.76	2.14	0.98

with a power of 0.8 is 911 over 475 measurements. The break in the peak concentration occurring in 1988 was caused by the fact that a few pumping wells were out of operation during that period. The sharp peak concentration at Well 26 occurring in 1997, induced by the same reason, was captured by the TCE breakthrough curve simulated by the transport model. For the most densely monitored air stripper, the TCE breakthrough curve closely follows the trend of the measured concentrations.

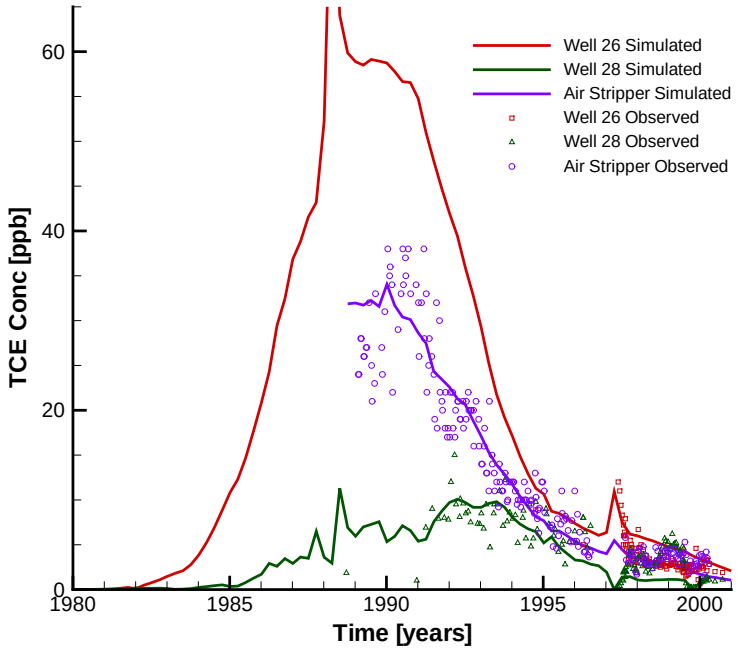

Figure 4.11: Simulated TCE concentration breakthrough curves and observed TCE concentration

To examine the TCE plume evolution, Figure 4.12 shows the extracted slice located at Easting 572775 ft for the vertical concentration contour. It suggests that near the source, the TCE in the groundwater system sank significantly in a short distance and then sinks very slowly as it spreads and moves downstream toward the Parkway well field. This situation would possibly result from the effects of areal recharge creating small downward vertical velocities especially near the surface, and the effects of advection and dispersion. The two-dimensional plan view of the TCE concentration contour at elevation

-50 ft is presented in Figure 4.13. The TCE concentration oscillation around the pumping wells is caused by the fact that a large maximum timestep of 3 months was employed in order to save the computing effort. However, the calibration performance measure, mainly determined by the TCE concentrations from the pumping wells, is not affected, because the flux averaged concentrations are integrated over the screen length and the influx flow during each timestep. Then, the oscillation of the TCE concentrations will be cancelled out. It can be verified by the smoothness of the simulated breakthrough curves for the pumping wells 26 and 28 in Figure 4.11. Compared to the hand made contour in Figure 4.7, point values from the numerical model exhibit higher concentrations than flux averaged observations over well screen intervals.

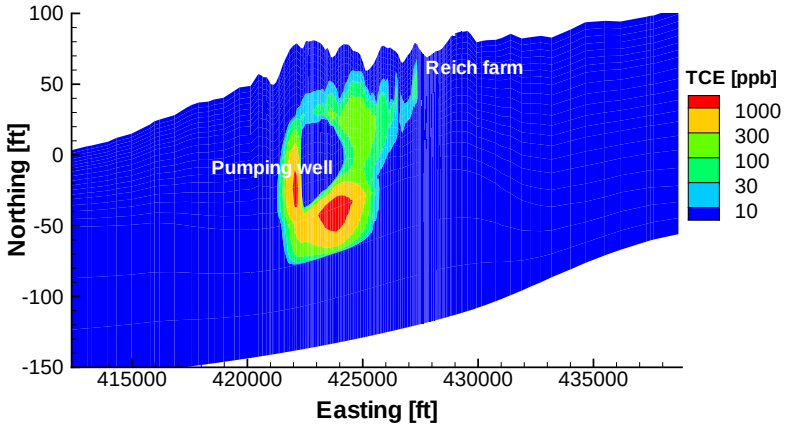

Figure 4.12: Plot of vertical TCE concentration at Easting 572775 ft in Dec 1990

4.7.4 Evaluation of the influence of recharge averaging

As discussed previously, the use of varying recharge both in space and time partly attributes to the inability of identifying transverse dispersivities. According to the classical dispersion theory, the dispersion coefficient in the advection-dispersion equation represents the spreading of the tracer by diffusion and local dispersive mixing. A critical review on field-scale dispersion [*Gelhar et al.*, 1992] pointed out that the dispersivity values in the dispersion coefficient for the field-scale problems were much larger than expected based on local dispersive mixing. The impact of spatial averaging of local velocity on the mass

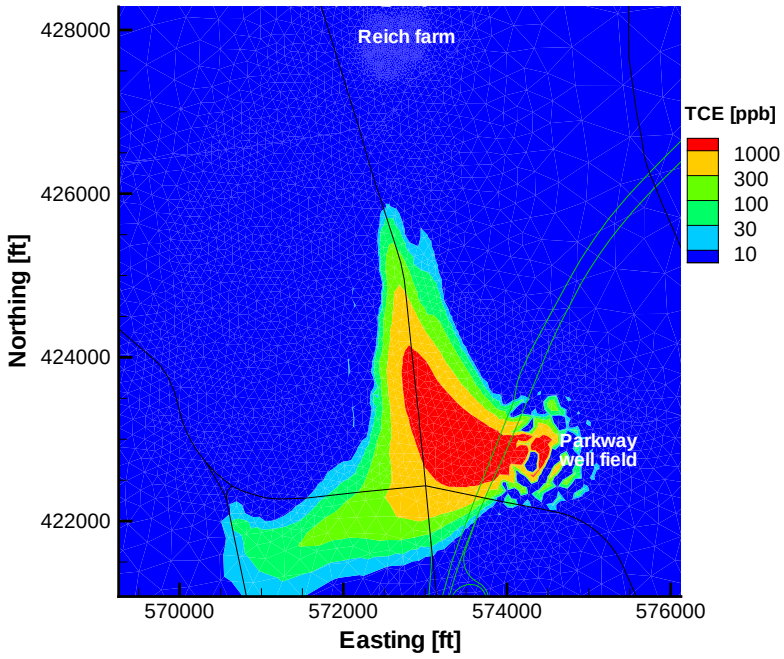

Figure 4.13: Plot of 2D TCE concentration at elevation -50 ft in Dec 1990

dispersion in heterogeneous porous media has been well addressed [*Smith and Schwartz*, 1980; *Tang et al.*, 1982; *Gelhar and Axness*, 1983; *Plumb and Whitaker*, 1988; *Rubin*, 1990; *Kitanidis*, 1992; *Rehfeldt and Gelhar*, 1992]. This work has shed light on the mechanism of macro-dispersion, has explained the large rate of tracer spreading in field scale heterogeneous formations caused by the spatial variations, and has suggested methods for the calculation of macro-dispersion coefficients. Using three-dimensional stochastic transport theory, *Gelhar and Axness* [1983] developed a method to predict the macro-dispersivities resulting from heterogeneous porous media. *Rehfeldt and Gelhar* [1992] developed a model to calculate the macro-dispersivity tensor which accounts for spatial variability caused by geological heterogeneity and temporal variability caused by transient effects. Through two special cases of unsteady flow, *Rehfeldt and Gelhar* [1992] found that hydraulic gradient magnitude variation tended to increase the longitudinal macro-dispersivity, whereas the gradient direction variation was inclined to enlarge the transverse macro-dispersivity. Most of these authors evaluated the macroscopic dispersion caused by the spatial variation in the local velocity, which, in turn, is primarily due to the variability in heterogeneous

hydraulic conductivity, while they overlooked the importance of varying recharge.

The varying recharge can induce variation in local velocity, both in magnitude and direction, especially for the upper layers where the source of the contaminant is located. This variation in local velocity will cause the initial tracer mass within the flow domain to spread and occupy an ever-increasing volume of the porous medium [*Bear*, 1988]. *Swain and Chin* [2003] developed an analytical formulation of enhanced dispersion by surficial recharge variation for two-dimensional groundwater system. A 9 % increase of groundwater dispersion was observed for the field case study. Deterministic and stochastic modelings by *Sykes et al.* [1982] also have shown that temporal averaging in local velocity may have a significant impact on mass transport. Theoretical work on the spatial-temporal averaging method for modelling mass transport was presented by *He and Sykes* [1996].

In addition to the base-case with varying recharge in time and space, transport model calibrations under three other scenarios including spatially varying recharge, temporally varying recharge, and constant recharge were carried out. Following the same procedure as the base-case analysis, for each scenario, a transient groundwater flow simulation with varying or averaging recharge was conducted first. The flow distributions for all the output timesteps were saved and loaded when dealing with the contaminant transport models. Thus, the flow fields depending on the surface recharge are specific to each scenario. Figure 4.14 illustrates that the averaging of recharge in space and time has a fundamental impact on the estimated transverse dispersivity. The three evident outliers were excluded when fitting the trend line to the data set for the case with temporally and spatially varying recharge. When the averaging scheme is employed, the calibrated macroscopic transverse dispersivity incorporates not only the inherent contribution to the mechanical dispersion by the porous media itself, but also the impact of the varying recharge on the local flow velocity. Calibration of the physically based transport model with temporally and spatially varying recharge tends to produce the smallest transverse dispersivity. The estimated transverse dispersivities for other scenarios are greatly enlarged to compensate for the loss of local deviations of the velocity that results from upscaling the recharge.

The variation reduction of the local vertical velocity for the upper layer of the Reich Farm model for both the space and time sense is shown in Figures 4.15 and 4.17 respectively

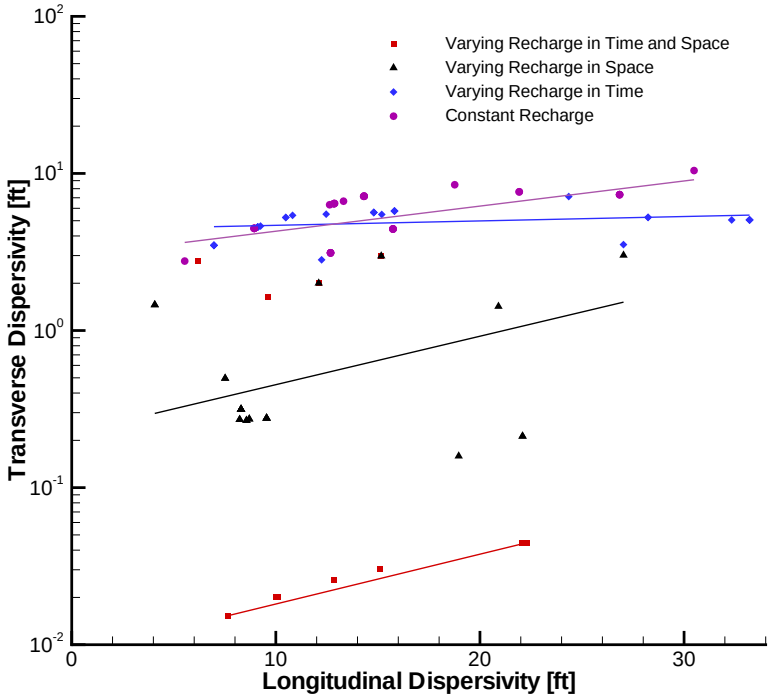

Figure 4.14: Scatter plot and trend line of estimated longitudinal vs. transverse dispersivities for the 3 multi-start PEST for various spatial and temporal averaging schemes for recharge

for various recharge averaging schemes. In accordance with the relative magnitude of the estimated transverse dispersivities, both figures support the fact that the base-case has the most varying local velocity field, and that the scenario with constant recharge has the most uniform velocity distribution. In the upper layers at the source zone area, the flow is largely influenced by the recharge. Thus, variation of the recharge tends to increase the local velocity variation and decrease the estimated transverse dispersivity as a result. The contaminant plume migrates horizontally towards the pumping well locations within the lower aquifer from layer 12 to layer 17. Its linear velocity distribution in the northing direction for two extreme cases of constant recharge, and spatially and temporally varying recharge are shown in Figure 4.16 with no obvious distinct difference. Therefore, unlike heterogeneous hydraulic conductivity which frequently has an influential impact on local velocity variation across the whole domain, varying recharge acts primarily on the shallow layers.

Figure 4.14 also suggests that the temporal averaging seems to be less important than spatial averaging. The temporally varying recharge is still a monthly averaged value, while the spatial discretization of recharge for the Reich Farm model has been refined. Thus, the relative importance of the two averaging schemes can not be determined based solely on this case study. One point that should be noted is that during the calibration process, the transverse dispersivity for the base-case reached the lower limit. The upper bound of the transverse dispersivity range was reached for the scenarios with spatially averaged recharge. However, the validity of the conclusion relating to the importance of transverse dispersivity is robust as the extension of the parameter's feasible range will widen the differences.

From Figures 4.15, 4.16, and 4.17, the magnitude of linear velocity is mostly above 0.01 ft/day, except for the Barnegat Bay, where the groundwater system is much more stable. Given the estimated longitudinal dispersivity equal to 10.1 ft, the mechanical dispersion part in Equation (2.12) is roughly:

$$\alpha_l q = \alpha_l v n = 10.1\text{ft} \times 0.01\text{ft/day} \times 0.3 = 0.03\text{ft}^2\text{/day}$$

where, v is the linear velocity. The free solution diffusion coefficient for TCE is $9.3 \times$

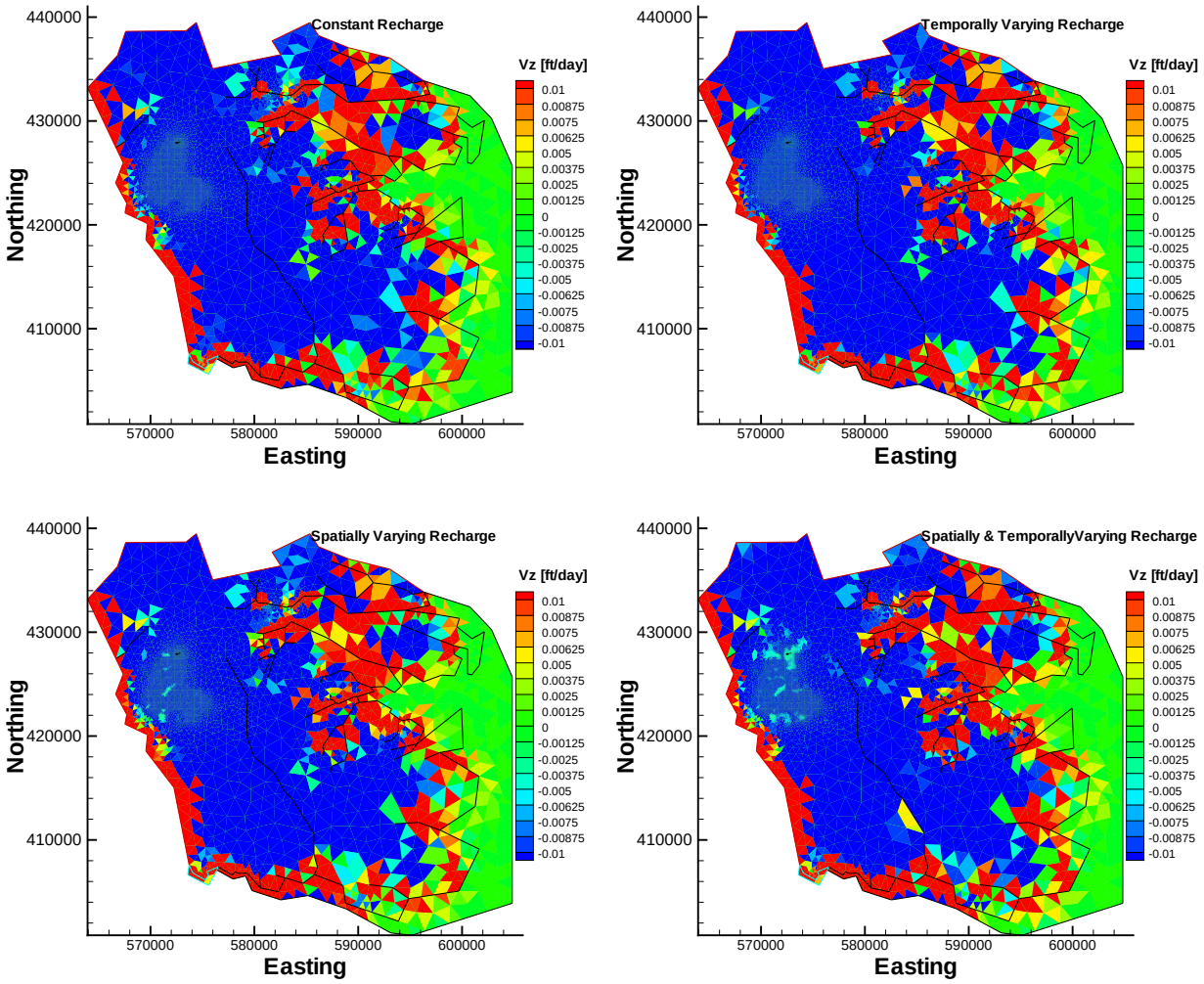

Figure 4.15: Vertical linear velocity distributions in layer 5 for scenarios with varying and averaging recharge in time and space

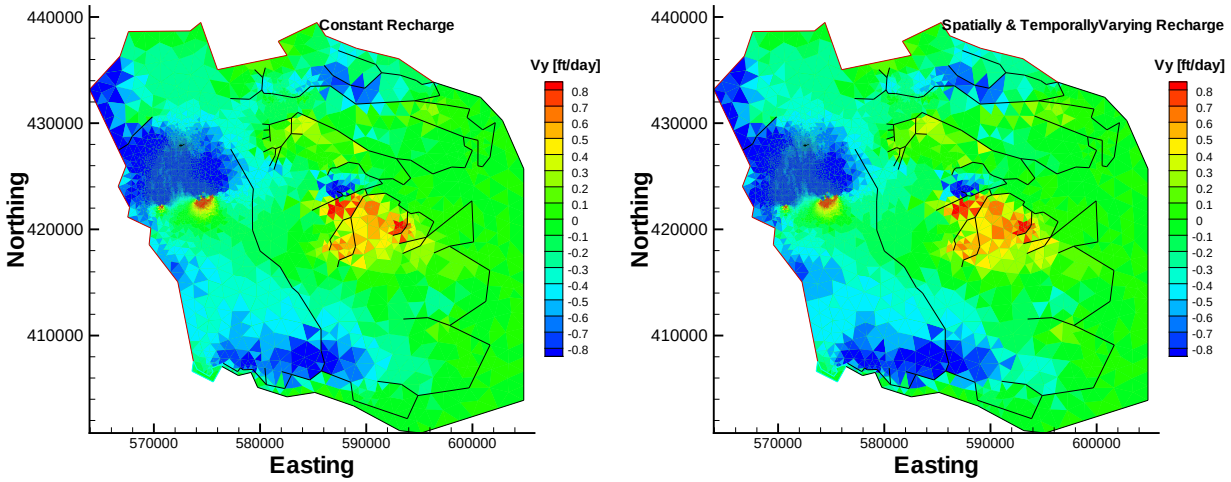

Figure 4.16: Linear velocity distributions in northing direction in layer 15 for scenarios with varying and averaging recharge in time and space

10^{-4}ft^2/day [*Pankow and Cherry*, 1996], far less than the mechanical dispersion coefficient, which verifies the assumption in Section 4.4 that the diffusion process can be neglected.

4.7.5 Calibration with different contaminant source characterization methods

As discussed in Section 4.1, NAPL dissolution at the source was simplified using the linear equilibrium model where the adsorbed phase is always in equilibrium with the dissolved phase. To evaluate the impact of the contaminant source characterization approaches on the calibration results, other dissolution model are considered. *Abriola and Pinder* [1985] attempted to quantify the interphase mass exchange in porous media by a local equilibrium assumption, which stated that the concentration of dissolved NAPL in the aqueous phase was equal to the correspondent NAPL solubility. More investigations have shown that measured organic solute concentrations, especially in heterogeneous systems, were below solubility limits, indicating non-equilibrium conditions [*Mackay et al.*, 1991; *Geller and Hunt*, 1993; *Powers et al.*, 1991, 1992, 1994b; *Nambi and Powers*, 2000, 2003]. In fact, heterogeneity increases the tendency toward preferential flow and kinetic mass transfer.

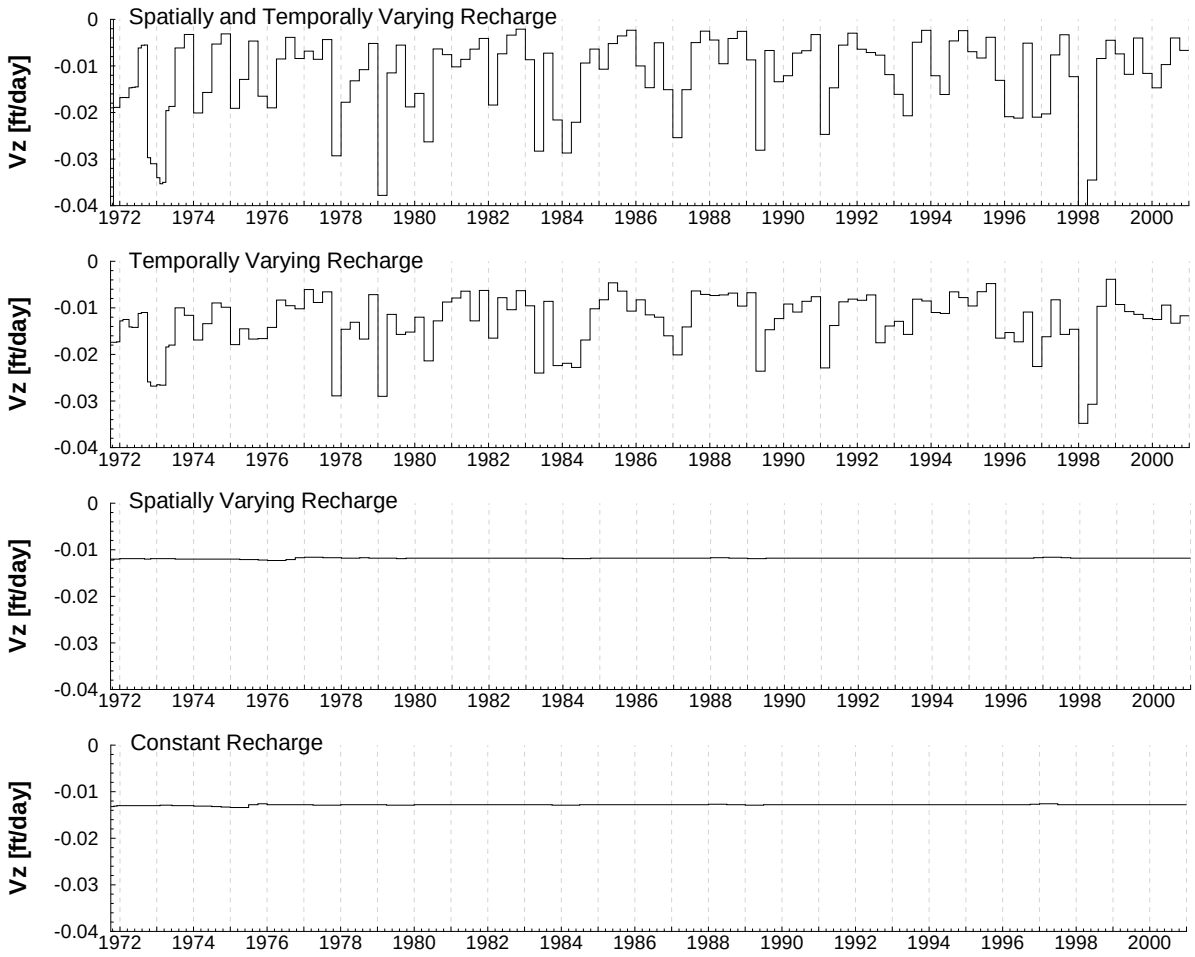

Figure 4.17: Vertical velocity vs. time at Reich Farm site for scenarios with varying and averaging recharge in time and space

Thus, the linear driving-force mass transfer model is commonly used to represent the interphase mass transfer flux (J_0):

$$J_0 = k_f \left(C_s - C\right) \tag{4.6}$$

where k_f is the mass transfer coefficient, C is the bulk aqueous phase concentration, and C_s is the effective solubility that would be in equilibrium with the organic phase. The interfacial areas between two immiscible phases are a fundamental factor affecting the mass transfer coefficient k_f [*Powers et al.*, 1992]. To account for the shrinkage of interfacial areas caused by NAPL dissolution, one simplified approach borrowed from chemical engineering is to use idealized geometry of spheres representative of actual NAPL blobs. *Powers et al.* [1994a] proposed a phenomenological model describing NAPL dissolution rates by representing irregular NAPL blobs with a distribution of effective spheres. The primary limitation of this model lies in the requirement to quantify the NAPL blob shape and size distribution. Other researchers have avoided the need to quantify the geometry of immobile NAPL blobs by using lumped mass transfer coefficients [*Miller et al.*, 1990; *Powers et al.*, 1992, 1994b; *Imhoff et al.*, 1994; *Nambi and Powers*, 2003]. The dimensionless forms of lumped mass transfer coefficients, the modified Sherwood number Sh, are frequently employed in the formulation of mass transfer correlations.

$$\widehat{k} = k_f a_0, Sh' = \widehat{k} d_{50}^2 / D_L \tag{4.7}$$

where $\widehat{k}$ is the mass transfer rate coefficient, a_0 is the interfacial area per unit volume of porous medium, d_{50} is the median grain diameter, and D_L is the free liquid diffusivity of the organic species in water. A selection of recently developed correlations is shown in Table 4.8.

Aside from the valid condition constraints for each lumped mass transfer correlation listed in Table 4.8, their applicability is also greatly restricted by scale problems and the difference of porous media or techniques used in experiments and fieldworks respectively [*Zhu and Sykes*, 2000]. Furthermore, it should be noted that all these studies were performed in saturated subsurface systems. For unsaturated conditions, the volatilization and

Table 4.8: Sherwood number correlations from Laboratory investigations.(x/d_p is the dimensionless distance into the residual saturation zone, θ_n^0 is initial volumetric fraction of NAPL in the system, Re employs the Darcy velocity, Re' employs the interstitial velocity.)

Correlation	Valid condition	Reference
$Sh = 425Re^{0.75}\theta_n^{0.60}$	$0.016 < \theta_n < 0.07$	*Miller et al.* [1990]
	$0.0015 < Re < 0.1$	(steady-state dissolution)
$Sh = 57.7Re^{0.61}d_{50}^{0.64}U_i^{0.41}$	θ_n=constant	*Powers et al.* [1992]
	$0.012 < Re < 0.2$	(steady-state dissolution)
$Sh = 340Re^{0.71}\theta_n^{0.87}\left(\frac{x}{d_p}\right)^{-0.31}$	$0 < \theta_n < 0.04$	*Imhoff et al.* [1994]
	$0.0012 < Re < 0.021$	
	$1.4 < x/d_p < 180$	
$Sh = 4.13Re'^{0.60}\delta^{0.67}U_i^{0.37}\left(\frac{\theta_n}{\theta_n^0}\right)^{\beta}$	$0.0003 < \theta_n < 0.065$	*Powers et al.* [1994b]
	$0.052 < Re' < 0.08$	
	$1.19 < U_i < 3.33$	
$Sh = 37.15Re'^{0.61}S_n^{1.24}$	$0.01 < S_n < 0.35$	*Nambi and Powers* [2003]
	$(0.0048 < \theta_n < 0.168)$	
	$0.0012 < Re < 0.021$	

water/gas partitioning of NAPL serve as other mass transfer processes and need to be taken into account in contaminant transport models. In an attempt to simply the mass transfer correlation, *Zhu and Sykes* [2004] proposed three simple models shown in Table 4.9 based on the NAPL mass conservation as well as the assumption that a linear or nonlinear power-law relationship exists between the effluent concentration and the remaining NAPL mass. The simple models lump the entire NAPL source zone, and completely neglect the spatial variation of the aqueous phase concentration.

Table 4.9: Simple models for the NAPL dissolution [*Zhu and Sykes*, 2004]

	Fundamental relationship	Model formulation
L model	$C_0(t) = C_s \dfrac{M(t)}{M_0}$	$C_0(t) = C_s e^{-\lambda t}$
N_1 model	$C_0(t) = C_s \left[\dfrac{M(t)}{M_0}\right]^{\alpha}$	$C_0(t) = C_s \left[(\alpha-1)\lambda t + 1\right]^{\alpha/(1-\alpha)},\ \alpha > 1$
N_2 model	$C_0(t) = \beta C_s \left[\dfrac{M(t)}{M_0}\right]^{\alpha}$	$C_0(t) = \beta C_s \left[\beta(\delta-1)\lambda t + 1\right]^{\delta/(1-\delta)},\ \delta = \alpha\beta > 1$

In Table 4.9, $C_0(t)$ is the residual NAPL zone solute concentration at time t, C_s is the NAPL solubility, $M(t)$ is the NAPL mass remaining at time t, M_0 is the initial NAPL residual mass, α is a power-law index indicating non-linearity of the initial solute concentration at the source, and β represents the degree of non-equilibrium at the initial stage of dissolution.

The model L, often used as a hypothetical contaminant source release function, assumes a linear relationship. The model N_1 extends the model L by containing a power-law index to control the rate of NAPL source zone concentration decrease. The model N_2 again incorporates a non-equilibrium indicator to adjust the initial aqueous phase NAPL concentration. *Zhu and Sykes* [2004] compared them with the lumped mass transfer correlation of *Imhoff et al.* [1994] through a 1-D saturated column simulation. It was demonstrated

that all the simplified models are capable of reproducing the rapid drop and tailing parts of effluent concentration breakthrough curves with appropriate model parameters. For simplicity, a modified model L in (4.8) was implemented in HydroGeoSphere, and the calibration result was compared with those obtained from the linear equilibrium model.

$$C_0(t) = \beta C_s \frac{M(t)}{M_0}, \quad C_0(t) = \beta C_s e^{-\lambda t} \tag{4.8}$$

Following the same procedure and objective function definition employed in the previous section of this chapter, 3 multi-start PEST was used to calibrate the transport model with the source term characterized by the modified model L. One source-related parameter was replaced by the mass of TCE residual per unit volume of the porous medium. Box constraints associated with each parameter are listed in Table 4.10.

Table 4.10: Decision variables and box constraints

Decision variables	Lower bound	Upper bound
Longitudinal dispersivity, a_l [ft]	1.0	50.0
Transverse/longitudinal dispersivity, a_t/a_l	2.0E-03	0.5
Transverse vertical/longitudinal dispersivity, a_{tv}/a_l	1.0E-03	0.2
Distribution coefficient, K' [ft^3/kg]	1.2E-03	6.0E-03
Initial source TCE concentration, βC_s [kg/ft^3]	1.0E-03	4.0E-02
Initial TCE residual per unit volume, M_0 [kg/ft^3]	1.0E-03	0.5

Given the model L describing the NAPL residual dissolution process, the optimal parameters and associated confidence intervals are presented in Table 4.11. Like the previous calibration attempt, the ratio between transverse and longitudinal dispersivities arrives at the lower bound. The same explanations lie in the upscaling and the incorporation of spatially and temporally varying recharge. For the two most sensitive parameters, longitudinal dispersivity and distribution coefficient, empirical cumulative distribution functions over 30 trials are shown in Figures 4.18 and 4.19. Both parameters, derived on a basis of the modified model L, tend to have wider ranges. A two-tailed non-parametric Wilcoxon-Mann-Whitney test was performed. The results demonstrate that for both parameters, the

computed p-value is lower than the significance level $\alpha = 0.05$. Thus, the alternative hypothesis that the distributions from two source characterization methods are significantly different should be accepted with a risk to reject the null hypothesis of less than 0.01%.

Table 4.11: Summary of transport parameter estimates and confidence intervals for the case with the source term characterized by the modified linear dissolution model

	Estimated	**9**5% percent confidence limits	
Parameter	value	Lower limit	Upper limit
a_l [ft]	18.4	12.0	24.8
a_t/a_l	2.0E-03	-1.7E-02	2.1E-02
a_{tv}/a_l	4.5E-03	3.5E-03	5.4E-03
K' [ft^3/kg]	3.2E-03	3.1E-03	3.3E-03
βC_s [kg/ft^3]	7.6E-03	6.2E-03	9.1E-03
M_0 [kg/ft^3]	3.5E-03	2.9E-03	4.0E-03

Substituting the optimal parameter values into the transport model, the TCE source concentration histories and breakthrough curves at the pumping wells are depicted in Figures 4.20 and 4.21, respectively. A good agreement between the TCE source concentration profiles was reached. As a result of the cleanup of the dumping site, both source concentrations dropped very sharply at the beginning. The lack of smoothness in the curves in Figure 4.20 were possibly the result of the fluctuation of temporally varying recharge. Comparing Figure 4.21 to the breakthrough curves in Figure 4.10, no significant difference is observed. Therefore, different source characterization approaches and parameter settings yield a similar response in the temporal change in the TCE source concentration, and the estimated parameters are likely to be unique, even for the commonly shared transport parameters. However, the target performance measure, TCE concentrations in this case, is not heavily affected, and can produce comparable breakthrough curves with different NAPL dissolution models.

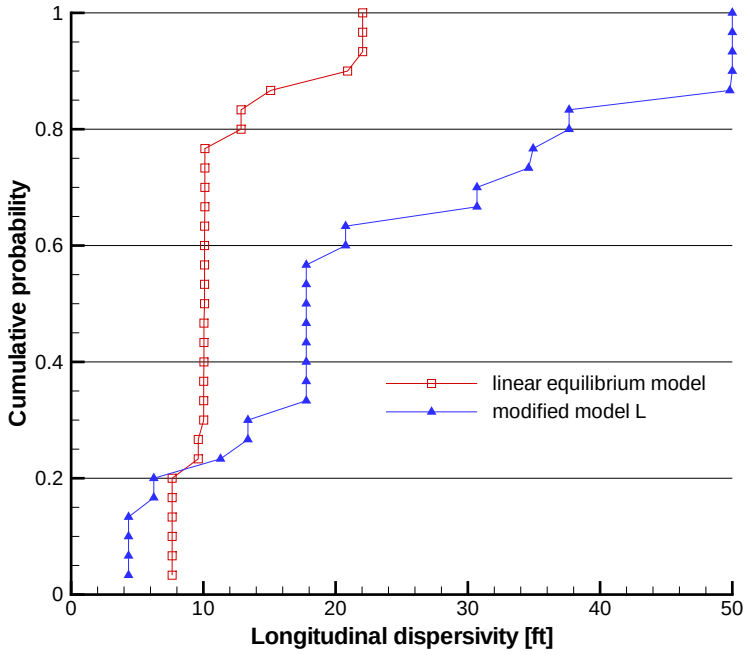

Figure 4.18: Empirical cumulative distribution function of longitudinal dispersivity

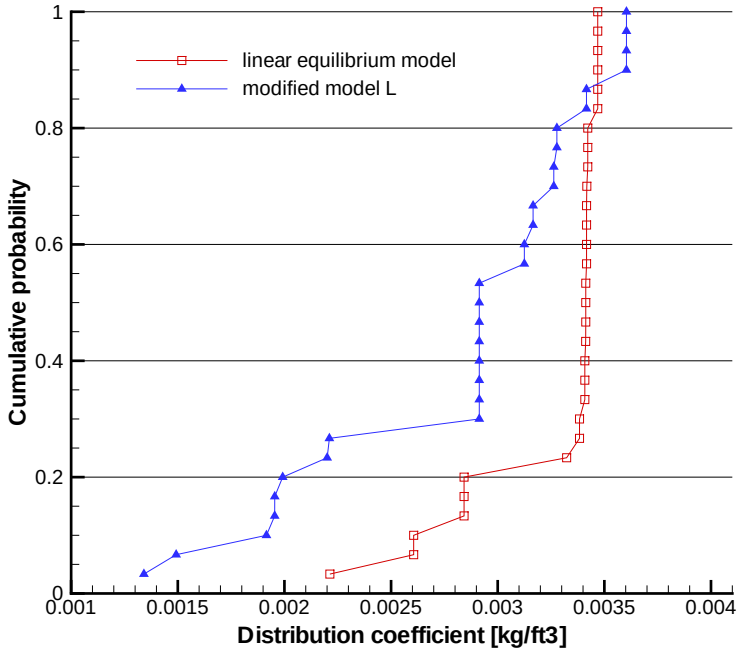

Figure 4.19: Empirical cumulative distribution function of distribution coefficient

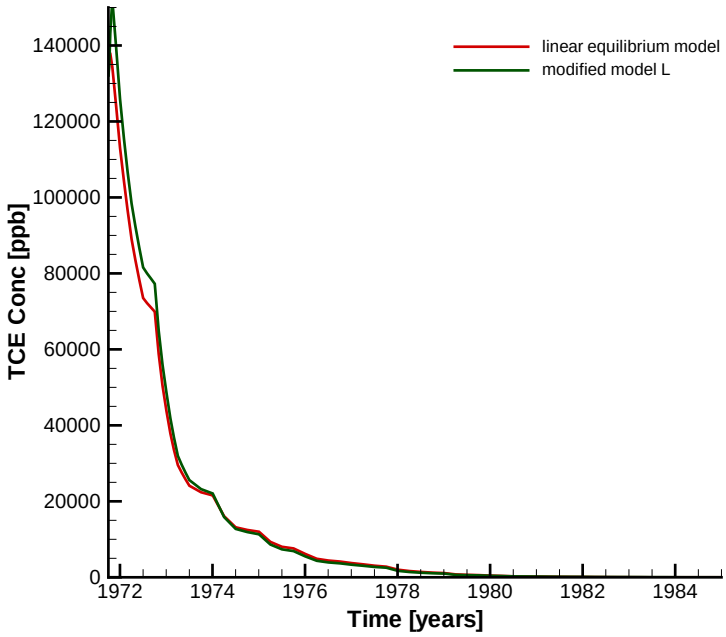

Figure 4.20: Estimated TCE release history at the source area

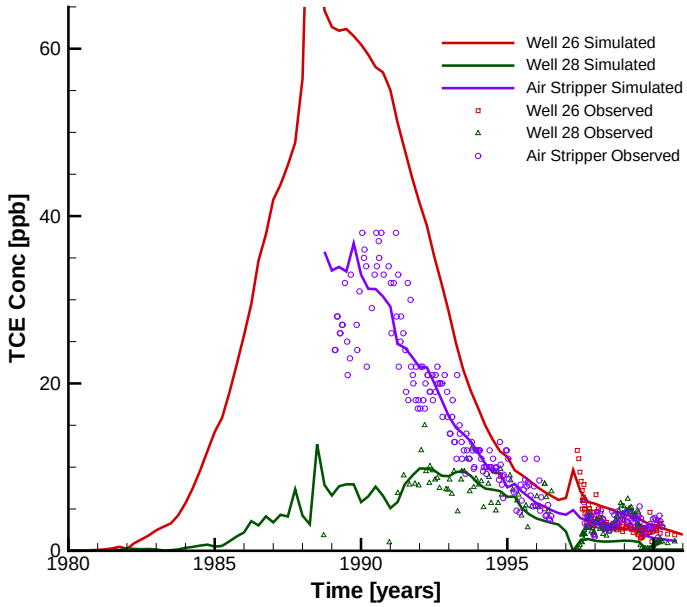

Figure 4.21: Simulated TCE concentration breakthrough curves and observed TCE concentration with the source term described by the modified linear dissolution model

4.8 Summary

Due to the limitation on the computing resource and the computationally demanding modelling of the three-dimensional variably saturated transient groundwater system, the groundwater flow and contaminant transport models were separated with a variant timestep scheme that honours the variation of the pumping rates at the Parkway well field being used. Given the flow model was well determined, three optimization algorithms, multi-start PEST, MicroGA and DDS were employed to calibrate the transport model with a modified SSE as the objective function. The modified model is less sensitive to the outliers than the standard SSE.

The calibration results indicate that overall, multi-start PEST performs best in terms of the final best objective function values with equal number of function evaluations. The actual elapsed computer time for multi-start PEST was far less than that for the two heuristic algorithms. Gradient based methods such as PEST are in general computationally faster than heuristic algorithms such as MicroGA and DDS, although they can converge to a local minimum, instead of a global one. Alternatively, heuristic algorithms can ideally converge to a global minimum with thousands of model evaluations. The objective function is sensitive to the uncertain measured concentrations. A hybrid algorithm, called multi-start PEST takes advantage of the robustness of the heuristic methods and of the fast convergence of gradient based approach. However, in order to guarantee the attainment of a global minimum, an exhaustive searching of the feasible space is necessary due to the ill-posed nature of contaminant transport problem and noisy data. Such a search is indeed impractical for this large scale computationally intensive problem. Thus, the solutions, obtained from the multi-start PEST calibration attempts, are the local minima or the candidates for the global minimum.

Multi-start PEST also was employed to identify contaminant transport and source parameters under different scenarios including spatially and temporally varying recharge and averaged recharge. For the transient model with spatially and temporally varying recharge, the estimated transverse dispersivity coefficients were estimated to be significantly less than that reported in the literature for the more traditional approach that uses steady-state flow

with averaged, less physically based recharge values.

In addition to the linear adsorption isotherm model characterizing the TCE dissolution at the source, a modified model L describing the source release, was also implemented and calibrated by multi-start PEST. The results indicate that the estimated parameters are different, even for the commonly shared transport parameters. However, the impact on the target performance measure, TCE concentrations in this case, is trivial.

Chapter 5

Uncertainty Analysis

Uncertainty is the inherent nature of physical systems due to unavoidable unpredictability. Some quantities are random in principle, such as brownian motion in the microscopic world; some vary over time or space, for example the recharge rate in hydrogeology, and are often termed "variability"; other quantities can be precisely measured, but modelled as uncertain due to a lack of knowledge. The last type of uncertainty is usually reducible by further measurement or study.

The third case of uncertainty includes parameter uncertainty, model uncertainty, and scenario uncertainty [*U.S. EPA*, 1997]. Model uncertainty arises due to the necessary simplification of real-world processes. Scenario uncertainty stems from errors in professional judgment and incomplete analysis. The groundwater flow model in this case study was well calibrated; there is a good fit between simulated and measured groundwater levels from 101 wells over a 20 year time period. The heterogeneity of porous medium and recharge both in time and space was incorporated in the physically based model, and the boundary conditions were well-defined based on surface topology. Thus, model and scenario uncertainty, which is difficult to quantify, is largely reduced. This chapter will mainly address parameter uncertainty.

Parameter uncertainty, which most likely originates from the measurement errors and

spatial variability in properties of the medium, is ubiquitous in environmental processes in subsurface systems. There has been a growing awareness of the need to quantify the uncertainty associated with outcomes produced by models. Failure to account for this uncertainty may lead to erroneous results, and may have significant environmental and economic consequences. Groundwater management, such as remediation planning for contaminant cleanup, can be meaningful only if the analysis includes quantitative estimates of parameter uncertainty, because remediation requirement can increase dramatically due to parameter uncertainty [*Wagner and Gorelick*, 1987; *James and Oldenburg*, 1997]. Over the past two decades, the significance of the uncertainty associated with parameters has been recognized, and considered in forward and backward groundwater modelling. There are relatively fewer studies on uncertainty for the inverse modelling in identifying contaminant source parameters.

Mishra and Parker [1989] examined the effects of parameter uncertainty, with known mean and variance, on predictions of unsaturated flow by the comparison of the predicted and measured results. The experimental results were shown to slightly differ from the mean of predicted values. Therefore, because the predictions are not frequently linearly related to the uncertain parameters, simply substituting the mean values of parameters into the governing equations will not lead to the proper results. Actually, even the uncertain coefficient of variation (COV) may influence the predictions. *Jyrkama* [2003] integrated FORM with MODFLOW in the analysis of the Reich Farm site to estimate the impact of hydraulic heads and pumping rates caused by the uncertainty of COV. It was shown that the larger COV results in the higher failure probability of maintaining the head at a pumping well.

Cawlfield and Wu [1993] evaluated the probability estimate for 1-D advective-dispersive transport model under four scenarios. The uncertain parameters included flow velocity, dispersivity, and distribution coefficient. Among them, flow velocity was demonstrated to be the most important variable for nonreactive transport cases with gamma sensitivity approximately equal to unity. *Tebes-Stevens et al.* [2001] examined the impact of uncertainty associated with groundwater flow, contaminant transport and reactive parameters on the outcome of the solute transport model in the subsurface. The results indicated that

the concentration is most sensitive to the hydraulic conductivity with relative sensitivity of 7.57. Further, half of the uncertain parameters of concern have the relative sensitivity larger than 0.5. *Hamed et al.* [1996b] investigated the effect of spatial random variability of the hydraulic conductivity on the failure probability to meet a target concentration level at a downgradient well. It was shown that the probabilistic events were very sensitive to hydraulic conductivities along the stream tubes and the tortuous flow path respectively.

Thus, parameter uncertainty seriously impacts the simulation results of forward groundwater modelling problems. Neglecting the uncertainty generally will reduce the credibility of model predictions. Uncertainty analysis methods developed in the past thirty years enable us to quantify the uncertainty of the estimates, and will be presented in the following section.

5.1 Stochastic and Uncertainty Modelling Methods

Parameter uncertainty can be propagated into model results and thus bring into question the utility of deterministic groundwater flow and contaminant transport models. Approaches to perform uncertainty analysis include interval mathematics, fuzzy theory, and probabilistic analysis. Interval mathematics is used to estimate the bounds on various model outputs based on the upper and lower limits of the model inputs and parameters without probability structures. This method can not provide adequate information on the nature of output uncertainty, even if the probability structure of parameters is known. Fuzzy theory's application to groundwater flow and transport model was initiated by *Dou et al.* [1995, 1997]. However, it is considered to be more suitable for qualitative reasoning rather than quantitative estimation of uncertainty [*Isukapalli*, 1999]. Probabilistic analysis is the most widely used method to quantify the uncertainty in physical processes in the subsurface. In recent years, a few stochastic modelling approaches have been developed and successfully applied to the forward groundwater modelling method. The approaches include the Monte Carlo method coupled with stratified sampling and Latin hypercube sampling, perturbation method, first-order second moment method (FOSM), first-order reliability method (FORM) and second-order reliability method (SORM).

5.1.1 Perturbation method

The perturbation method is a simple and efficient stochastic modelling technique. This approach decomposes the random parameters into mean terms and perturbation terms, and then propagates parameter uncertainty into the model outputs based on their relationships, such as partial differential equations describing groundwater flow and contaminant transport [*Tang and Pinder*, 1977; *Gelhar*, 1986; *Mantoglou and Gelhar*, 1987; *Li et al.*, 2004; *Shvidler*, 1993; *Tiedeman and Gorelick*, 1993; *Russo*, 1993; *Satish and Zhu*, 1992]. This method does not require the probability density function of the input random parameters but rather their statistical moments, for instance mean and covariance. The results are statistical moments of model outputs. It does not provide any information about probability density functions. In addition, this method is only valid when perturbation terms are relatively small in comparison to mean terms. Another limitation is the methods applicability to problems of simple geometry and boundary conditions [*Gelhar*, 1986].

5.1.2 First-order second moment method

Another uncertainty modelling approach is the first-order second moment (FOSM) method [*Dettinger and Wilson*, 1981; *Wagner and Gorelick*, 1987; *James and Oldenburg*, 1997; *Kunstmann et al.*, 2002]. By performing a first-order Taylor series expansion with respect to all the random parameters at their mean values, the model output can be expressed as the linear combination of random parameters, where the coefficients are the sensitivities to each random parameter. Then a common uncertainty propagation approach can be employed to evaluate the means and variances of model output. Despite the computational advantage of the FOSM method, it can only be applied to the cases with moderate uncertainty in parameters, because the method is inherently a linear method. Using a second-order Taylor series expansion can improve the accuracy of estimated variances. However, the calculation of second derivatives will give rise to a huge computational burden.

5.1.3 First and second-order reliability methods

First- and second-order reliability methods (FORM and SORM) with high computational efficiency were originally developed in the structural engineering field [*Madsen et al.*, 1986], but have recently been applied to both surface and subsurface problems [*Schanz and Salhotra*, 1992; *Cawlfield and Wu*, 1993; *Hamed et al.*, 1995, 1996b,a; *Hamed*, 2000; *Vasquez et al.*, 2000; *Portielje et al.*, 2000; *Skaggs and Barry*, 1997; *Jyrkama and Sykes*, 2006]. The goal is to estimate the failure probability of a specific model output exceeding the target level. So "the method estimates only one percentile of one model output at a time" [*Schanz and Salhotra*, 1992]. For the sake of obtaining the probability density function of model output, FORM or SORM has to be executed repeatedly with a series of target levels.

FORM and SORM are much more efficient than other approaches for low probability events with a great many parameters. They can also provide a measure of sensitivity of the failure probability to the random parameters at no extra computational effort with this enabling the identification of the parameters of the highest influence. The weakness of these approaches is that they are not always accurate, particularly in the vicinity of the 50^{th} percentile [*Schanz and Salhotra*, 1992]. In practice, the accuracy of FORM and SORM depends on the shape of the limit state surface (where the model output is equal to the target level). SORM is generally more accurate than FORM, because SORM employs a second-order approximation to the limit state surface. But SORM has no distinct advantage over the Monte Carlo method for the cases with a great many parameters [*Skaggs and Barry*, 1997].

5.1.4 Sampling based approaches

The Monte Carlo method is probably the most widely used uncertainty modelling method. "Any existing flow and transport model can be used with the method, provided a random number generator is available to sample from the prescribed distribution function of the basic random variables, whether independently or jointly" [*Hamed et al.*, 1996b]. So far, it has been implemented to both analytical and numerical problems in hydrology [*Schanz and*

Salhotra, 1992; *Hamed et al.*, 1995; *James and Oldenburg*, 1997; *Kunstmann et al.*, 2002; *Fenton and Griffiths*, 1993]. The Monte Carlo method is regarded as a most reliable and robust method of uncertainty analysis, because it makes no assumption or approximation in application. Therefore, it is often used as the benchmark solution for a problem to verify other uncertainty modelling methods. This approach is not appropriate for the calculation of sensitivity analysis to assess the relative importance of each random parameter. What is more, the major drawback to the classic Monte Carlo approach with random sampling is the computational cost, especially for the low probability events, because the model has to be solved for thousands of times to attain an acceptable result. Therefore, sampling approaches are needed to improve the efficiency of the Monte Carlo method. Stratified sampling (importance sampling) and Latin hypercube sampling [*Melching and Bauwens*, 2001; *Portielje et al.*, 2000; *Helton and Davis*, 2003] can provide significant computational benefits.

In stratified sampling, the probability space is divided into strata. The strata should be mutually exclusive. Then, equal number of samples are generated randomly within each stratum. It has been proved theoretically that for the same number of realizations, the variance for the percentile on the estimated distribution function with stratified sampling is less than that with random sampling [*Helton and Davis*, 2003]. Indeed, the stratified sampling is always regarded as the best sampling technique if enough information is available for its appropriate implementation [*Helton and Davis*, 2003]. The drawback associated with this technique lies in the difficulty in determining the strata and their probabilities. "Latin hypercube sampling can be viewed as a compromise procedure that incorporates many of the desirable features of random sampling and stratified sampling" [*Helton and Davis*, 2003]. In Latin hypercube sampling, the range of each parameter is divided into intervals of equal probability. Equal number of samples are generated within each interval and paired at random. Latin hypercube sampling has an advantage over random sampling in variance reduction as well. The relative magnitude of variances calculated by stratified sampling and Latin hypercube sampling is not identified theoretically. Unlike the stratified sampling, Latin hypercube sampling does not require identifying the strata and probabilities, and is easy to implement. Thus, Latin hypercube sampling is often the preferred sampling procedure in Monte Carlo analysis.

5.2 Derivation of Estimated Parameters Sampling Space

To quantify prediction uncertainty by Latin hypercube sampling, the uncertainty in parameters and the associated correlations (or parameters sampling space) must be characterized. As has been stated, the inverse modelling of a transport model is an ill-posed problem. The non-uniqueness may arise because of the effects of error and uncertainty in the modelling process, resulting from model errors, measurement errors, scale effects, and an imperfect knowledge of the system. Moreover, the optimal calibration result can not guarantee the best solution, and the selection of only one result to conduct uncertainty analysis is rather subjective. Table 5.1 shows that even with similar objective function values for the problem of this thesis, the parameter values can be very different. Therefore, the sampling space should incorporate all the calibration results, even those with bad performance. The calibrated parameter sets with good performance should be more likely to be selected in the following sampling procedure. A goodness-of-fit measure between simulated and observed concentrations, that reflects the ability of a parameter set to yield the real situation, is often considered as an adequate indicator to define the likelihood. Examples of the empirical likelihood measures used in generalized likelihood uncertainty estimation (GLUE) [*Beven and Binley*, 1992] applications are shown in Table 5.2. If the shaping factor N equals 0, each parameter set is given equal weighting. Then, the likelihood function is no longer dependent on the error variance σ_e^2 and/or the variance of the observations σ_{obs}^2, and just follows the frequency of each parameter set. As N is increased, the difference in the likelihood assigned to parameter sets is magnified with more probability given to the better optimization trials. As N approaches infinity, only the parameter set with the smallest error variance σ_e^2 has non-zero likelihood. As a result, the shaping factor N plays a significant role in the representation of uncertainty. However, the choice of the shaping factor N is subjective, which will lead to non-uniqueness of uncertainty analysis. In this section, a new statistically based, empirical approach to allocate the weights for 30 optimization trials from multi-start PEST will be presented.

It is worth noting that the clustering of the solutions (several distinct solutions are similar and located within the attraction area of the same local minimal.) might be problematic for optimization algorithms which result in incomplete convergence to the various

Table 5.1: Selected transport parameter estimates for different calibration trials

Obj Fun	a_l[ft]	a_t/a_l	a_{tv}/a_l	K'[ft^3/kg]	K'_{source}[ft^3/kg]	$C_{TCE,initial}$[kg/ft^3]
2126	10.1	2.0E-03	9.0E-03	3.4E-03	2.0E-04	8.1E-03
2133	12.9	2.0E-03	6.8E-03	3.4E-03	2.0E-04	8.7E-03
2136	15.1	2.0E-03	5.7E-03	3.3E-03	2.0E-04	9.5E-03
2162	7.6	2.0E-03	12.6E-03	3.5E-03	2.0E-04	7.6E-03
2225	27.3	2.0E-03	2.8E-03	3.3E-03	4.7E-03	9.8E-03
3426	9.6	0.17	7.6E-02	2.6E-03	1.0E-03	8.0E-03

Table 5.2: Likelihood definitions employed to determine the weight for the parameter sets, where N is the shaping factor

Likelihood function	Description	Reference
$L(\theta) = (\sigma_e^2)^{-N}$	Based on inverse error variance	*Beven and Binley* [1992]
$L(\theta) = \left(1 - \frac{\sigma_e^2}{\sigma_{obs}^2}\right)^N$	Based on Nash and Sutcliffe efficiency criterion	*Freer et al.* [1996]
$L(\theta) = \exp(-N\sigma_e^2)$	Based on exponential transformation of error variance	*Freer et al.* [1996]

$\sigma_{obs}^2 = 1/N_{obs} \sum_{i=1}^{N_{obs}} \left(C_{obs,i} - \overline{C}_{obs}\right)^2$, $\sigma_e^2 = \sum_{i=1}^{N_{obs}} \left(C_{sim,i} - C_{obs,i}\right)^2$

local minima, such as simulated annealing and genetic algorithm. *Vasco et al.* [1996] developed a methodology to de-cluster the solutions by partitioning them into groups according to the "distance" between the parameters. A representative value by averaging the parameters is employed to characterize each specific group. The number of partitions (groups) plays a key role in the de-clustering procedure. In the determination of the number of partitions, it is better to over estimate rather than underestimate, because the underestimation tends to group clusters that are truly distinct [*Vasco et al.*, 1996]. In this research, the calibration results for the 30 optimization trials (refer to Table 5.3) from the gradient-based multi-start PEST are completely convergent to the local minima, which can be partially verified by the recurrences of the solutions shown in Table 5.4. Each solution from multi-start PEST is a local minimal of a contraction area. Therefore, 30 optimization trials are deemed appropriate for the following uncertainty analysis.

Through the calibration of the Reich Farm transport model by multi-start PEST, 30 independent parameter sets and their associated parameter covariance matrixes and sum of squared errors (SSE) between the simulated and observed concentrations at different time and locations were obtained. *Xiang et al.* [1994] proposed the assumption that the parameter vector obeys a multivariate Gaussian distribution with the estimated parameter vector as the mean vector and the posterior covariance as the covariance matrix. Thus, for each parameter set, the parameters are assumed to obey multivariate normal distribution with given means and covariance matrix obtained through PEST calibration outcomes. To incorporate all the calibrated parameter sets in Latin hypercube sampling, a weighted p-multivariate normal distribution for parameter probability space is considered:

$$\mathbf{X} \sim \sum_{i=1}^{m} w_i N_p(\mu_i, \sigma_{i,j}), \sum_{i=1}^{m} w_i = 1, i, j = 1, 2, 3, \ldots, m \tag{5.1}$$

where, m is the number of parameter sets or calibration trials for each algorithm (30 in this case), p is the number of variables, μ and σ are the mean and variance-covariance matrix for each parameter set respectively, w is the weighting factor associated with parameter sets, and the p-multivariate normal distributions for the parameter sets are mutually independent. The next step is to determine the weighting factor.

Because concentration measurement suffers from a variety of unexpected random errors,

reading error, equipment error as examples, the observed concentrations are reasonably assumed to follow a normal distribution with constant variance and the mean equal to the simulated concentrations, giving:

$$C_{obs,i} \sim N(C_{sim,i}, \sigma), i = 1, 2, 3, \ldots, N_{obs} \tag{5.2}$$

where, $C_{obs,i}$ and $C_{sim,i}$ are measured and simulated TCE concentrations (ppb) respectively, σ is the standard deviation, and N_{obs} is the total number of measurements. Equation (5.2) is assumed to be true for transformation of $C_{obs,i}$ and $C_{sim,i}$. Therefore, $\mathbf{Y}$, defined in Equation (5.3), follows a standard normal distribution, and SSE divided by σ^2 is the sum of squared $N(0,1)$ or theoretically follows a Chi-squared distribution with the degree of freedom of N_{obs}.

$$Y_i = \frac{C_{obs,i} - C_{sim,i}}{\sigma} \sim N(0,1), i = 1, 2, 3, \ldots, N_{obs} \tag{5.3}$$

$$\frac{SSE}{\sigma^2} = \sum_{i=1}^{N_{obs}} \left(\frac{C_{obs,i} - C_{sim,i}}{\sigma} \right)^2 = \sum_{i=1}^{N_{obs}} Y_i^2 \sim \chi^2_{N_{obs}}, i = 1, 2, 3, \ldots, N_{obs} \tag{5.4}$$

Because we have 30 trials for each multi-start PEST algorithm, there are 30 SSEs readily available and they are assumed to be mutually exclusive. The maximum likelihood method can be employed to find the value of constant variance. Firstly, the probability density function of the Chi-squared distribution χ^2_n,

$$f(y) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2} & y > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

where, n is the degree of freedom, equal to the total number of measurements in this case, and Γ is a gamma function. Then, the likelihood function is

$$f\left(\frac{SSE_1}{\sigma^2}, \frac{SSE_2}{\sigma^2}, \ldots, \frac{SSE_m}{\sigma^2} | \sigma \right) = \prod_{i=1}^{m} f\left(\frac{SSE_i}{\sigma^2} \right), \text{by independency} \tag{5.6}$$

where, m is the number of trials for each multi-start PEST algorithm. The degree of freedom is constant, then

$$\begin{aligned}\prod_{i=1}^{m} f\left(\frac{SSE_i}{\sigma^2}\right) &\propto \prod_{i=1}^{m}\left[\left(\frac{SSE_i}{\sigma^2}\right)^{n/2-1}\exp\left(-\frac{SSE_i}{2\sigma^2}\right)\right] \\ &\propto \sigma^{m(2-n)}\exp\left(-\frac{\sum_{i=1}^{m} SSE_i}{2\sigma^2}\right)\end{aligned} \tag{5.7}$$

Applying the maximum likelihood method by taking the first derivative of the likelihood function with respect to σ to be 0 yields

$$\begin{aligned}&\frac{df\left(\frac{SSE_1}{\sigma^2}, \frac{SSE_2}{\sigma^2}, \ldots, \frac{SSE_m}{\sigma^2}|\sigma\right)}{d\sigma} \\ &\propto \exp\left(-\frac{\sum_{i=1}^{m} SSE_i}{2\sigma^2}\right)\left(m(2-n)\sigma^{m(2-n)-1} + \frac{\sum_{i=1}^{m} SSE_i}{\sigma^3}\sigma^{m(2-n)}\right) = 0\end{aligned} \tag{5.8}$$

Rearranging gives

$$m(2-n) + \sum_{i=1}^{m}\frac{SSE_i}{\sigma^2} = 0 \tag{5.9}$$

So

$$\sigma^2 = \sum_{i=1}^{m}\frac{SSE_i}{m(n-2)} \tag{5.10}$$

To ensure that the parameter sets with good performance are weighted more than others, the likelihood function is defined as the complementary cumulative distribution function in Equation (5.11). In other words, it represents the probability that the calibrated parameter sets have a performance worse than or equal to the i^{th} parameter set.

$$F_c\left(\frac{SSE_i}{\sigma^2}\right) = P\left(y > \frac{SSE_i}{\sigma^2}\right) = \int_{\frac{SSE_i}{\sigma^2}}^{\infty} f(y)\, dy, \quad i = 1, 2, 3, \ldots, m \tag{5.11}$$

The likelihood weights associated with the parameter sets are rescaled to give a cumulative sum of 1.0, and can be determined by the following probability of each SSE in the Chi-squared distribution.

$$w_i = \frac{F_c\left(\frac{SSE_i}{\sigma^2}\right)}{\sum_{i=1}^{m} F_c\left(\frac{SSE_i}{\sigma^2}\right)}, \quad i = 1, 2, 3, \ldots, m \tag{5.12}$$

The weighted p-multivariate normal distribution for the parameter probability space is given by.

$$X = \frac{1}{\sum_{i=1}^{m} F_c\left(\frac{SSE_i}{\sigma^2}\right)} \sum_{i=1}^{m} F_c\left(\frac{SSE_i}{\sigma^2}\right) N_p(\mu_i, \Sigma_i), \quad i = 1, 2, 3, \ldots, m \tag{5.13}$$

5.3 Latin Hypercube Sampling Approach

Latin hypercube sampling can be viewed as a compromise procedure that incorporates many of the desirable features of random sampling and stratified sampling. In Latin hypercube sampling, the range of each parameter is divided into intervals of equal probability. Equal number of samples are generated within each interval and paired at random. In this case, the number of samples for each sub p-multivariate normal distribution is given by,

$$n\mathbf{S}_i = \text{int}\left(w_i n\mathbf{S}\right), i = 1, 2, 3, \ldots, m \tag{5.14}$$

where, $n\mathbf{S}$ is the total number of samples. Then Latin hypercube sampling can be operated for each sub p-multivariate normal distribution, in which correlations among variables can also be specified and form part of the definition of the corresponding probability space. The method by *Iman and Conover* [1982] for inducing correlation is summarized as follows:

1. Generate $n\mathbf{S} \times k$ sample matrix $\mathbf{X}$ using straight Latin hypercube sampling of k variables at sample size $n\mathbf{S}$, in which upper and lower bounds for each variable are imposed.

2. Calculate the $\mathbf{P}$ lower triangular matrix of the target correlation matrix $\mathbf{C}$ (from the known covariance matrix $\sigma_{i,j}$) using Cholesky factorization $\mathbf{C} = \mathbf{P}\mathbf{P}'$.

3. Generate a new matrix $\mathbf{Z}$ that has the same dimension as $\mathbf{X}$, but is otherwise independent of $\mathbf{X}$. Each column of $\mathbf{Z}$ contains a random permutation of the $n\mathbf{S}$ van der Waerden scores $\phi^{-1}[i/(n\mathbf{S}+1)], i = 1, 2, 3, \ldots, n\mathbf{S}$, where ϕ^{-1} is the inverse of the standard normal distribution.

4. Calculate $\mathbf{E}$, the correlation matrix of $\mathbf{Z}$.

5. Calculate $\mathbf{Q}$ the low triangular matrix of $\mathbf{E}$ by Cholesky factorization, $\mathbf{E} = \mathbf{QQ}'$.

6. Calculate matrix $\mathbf{Z}^* = \mathbf{S}(\mathbf{Q^{-1}})'\mathbf{P}'$, which has a correlation matrix equal to $\mathbf{C}$.

7. Rearrange the values of each variable in $\mathbf{X}$ so they have the same rank as the target matrix $\mathbf{Z}^*$.

8. Check whether the correlation matrix of rank $\mathbf{X}$ is close enough to the target correlation matrix $\mathbf{C}$. If not, repeat the construction procedure for $\mathbf{Z}^*$ (step 3 to 7) until a suitable approximation to $\mathbf{C}$ is obtained.

McKay et al. [1979] proved that Latin hypercube sampling has an advantage over random sampling in variance reduction for monotonic functions. For non-monotonic functions, *Helton and Davis* [2003] illustrated an example in which Latin hypercube sampling produced more stable CDF estimates than produced by random sampling. Therefore, Latin hypercube sampling from each sub p-multivariate normal distribution has a smaller variance than random sampling:

$$Var\left(\overline{Y}_L|N_p\left(\mu_i,\sigma_{i,j}\right)\right) \leq Var\left(\overline{Y}_R|N_p\left(\mu_i,\sigma_{i,j}\right)\right), i,j = 1,2,3,\ldots,m \tag{5.15}$$

where, $\overline{Y}_L$ and $\overline{Y}_R$ represent the means of model output obtained with equal random sample size and Latin hypercube sample size. *McKay et al.* [1979] also show that both Latin hypercube sampling and random sampling yield unbiased estimates for Y:

$$E\left[\overline{Y}_L|N_p\left(\mu_i,\sigma_{i,j}\right)\right] = E\left[\overline{Y}_R|N_p\left(\mu_i,\sigma_{i,j}\right)\right], i,j = 1,2,3,\ldots,m \tag{5.16}$$

From the conditional variance formula,

$$\begin{aligned} Var\left(\overline{Y}_L\right) &= E\left[Var\left(\overline{Y}_L|N_p\left(\mu_i,\sigma_{i,j}\right)\right)\right] + Var\left(E\left[\overline{Y}_L|N_p\left(\mu_i,\sigma_{i,j}\right)\right]\right) \\ &= E\left[Var\left(\overline{Y}_L|N_p\left(\mu_i,\sigma_{i,j}\right)\right)\right] + Var\left(E\left[\overline{Y}_R|N_p\left(\mu_i,\sigma_{i,j}\right)\right]\right) \\ &\leq E\left[Var\left(\overline{Y}_R|N_p\left(\mu_i,\sigma_{i,j}\right)\right)\right] + Var\left(E\left[\overline{Y}_R|N_p\left(\mu_i,\sigma_{i,j}\right)\right]\right) \\ &= Var\left(\overline{Y}_R\right) \end{aligned} \tag{5.17}$$

Therefore, Latin hypercube sampling from weighted p-multivariate normal distribution produces more stable estimates than random sampling. A computer module based on the Latin hypercube sampling methodology with correlation was developed and used to investigate parameter uncertainty in the Reich Farm contaminant transport study of this thesis. The computer module was written in Matlab and generated parameter sets for the Latin hypercube sampling. The generation of these samples is described in the following section.

5.4 Generation of Latin Hypercube Samples

The Latin hypercube sampling approach requires that for each input parameter that has associated uncertainty or variability, a probability distribution be provided. This method involves the repeated generation of pseudo-random values of the uncertain input variables drawn from each interval of equal probability by the selected probability distribution and within the feasible range. In this study, the multivariate probability distribution for each parameter is obtained through the previous calibration procedure. In order to combine all of the estimated parameter sets and honour the associated correlation among parameters simultaneously, a modified Latin hypercube sampling scheme is introduced as follows.

Although density distribution and correlation are readily available for each parameter set with known likelihood weight, it is not easy to determine the unified correlation over multiple parameter sets. Instead of making one comprehensive distribution with correlation over 30 calibration trials, the number of samples for each parameter set is calculated first by multiplying its likelihood weight with the total sample size. Because the probability distributions associated with their own parameter sets are mutually independent, Latin hypercube sampling which honours parameter correlation can be conducted in a normal manner for each parameter set with given but different sample size. And the statistical analysis can be conducted by directly grouping Latin hypercube simulations from different parameter sets, given that each Latin hypercube sample represents an equal probability in our likelihood definition. The number of Latin hypercube samples plays a key role in the balance of accuracy and economy of this technique. To avoid the imperfection or

non-representation caused by a shortage of samples, a number of 300 Latin hypercube simulations for the 6 random parameters are used in the first attempt for evaluation of the model uncertainty. When a calibration trial is repeated, the Latin hypercube simulations assigned to the trial are summed with those of the other similar trials. If the 30 calibration trials have equal weights, then 10 Latin hypercube simulations would be performed for each trial, yielding 300 simulations in total. The impact of different sample sizes on the convergence of the approximation to the actual probability density will be investigated in a later section. The Chi square likelihood weights assigned to each parameter set for straight PEST, 3 multi-start PEST, and 5 multi-start PEST are given in Table 5.3. It is obvious that more weight, and hence Latin hypercube simulations, is allocated to realizations that accurately predict the observations through the definition of the likelihood measure, and unfavourable realizations with poor likelihood measure are reasonably removed from the sampling process. This trend becomes more pronounced in relation to the number of multi-starts, because the calibrated solution tends to converge to better local optima. In addition, to compare the variance reduction with the Chi square weighting method, the uncertainty analysis with the equal likelihood approach was also conducted. Tables 5.4, 5.5 and 5.6 show the number of Latin hypercube samples allocated to each optimization trial for straight PEST, 3 multi-start PEST, and 5 multi-start PEST estimations. The parameter sets with bad performance are successfully screened out by the Chi square weighting method.

Given the likelihood measure, by adding the weighted probability density function over the different parameter sets, the unified PDF for each parameter can be obtained and illustrated through Figure 5.1 to Figure 5.6. It should be noted that generation of these PDFs does not show the correlations between parameters. All the 6 parameter distributions based on 3 and 5 multi-start PESTs estimates result in very good agreement. However, PDFs derived from a straight PEST solution differs with the other multi-start PESTs in terms of probable sampling intervals and the mean values. Straight PEST estimates, which can be identified as flat distributions with weaker peaks, exhibit more variability than 3 and 5 multi-start PESTs. This can be explained by the fact that an increase in the number of multi-starts is more likely to lead to the recurrence of good solutions. Figure 5.2 and Figure 5.5 shows the PDFs for transverse dispersivity and equivalent distribution coefficient

Table 5.3: Summary of objective function values by PEST estimation and the assigned weights in constructing the parameter probability space

	Straight PEST		3 multi-start PEST		5 multi-start PEST	
Trial #	Obj Fun	Weight	Obj Fun	Weight	Obj Fun	Weight
1	2133	0.056	2133	0.042	2133	0.039
2	4687	0.000	2133	0.042	2126	0.039
3	3145	0.054	3145	0.000	3145	0.000
4	2225	0.056	2162	0.041	2162	0.037
5	3145	0.054	2133	0.042	2133	0.039
6	3426	0.025	3426	0.000	2133	0.039
7	2133	0.056	2133	0.042	2133	0.039
8	2126	0.056	2126	0.042	2126	0.039
9	3426	0.025	2133	0.042	2133	0.039
10	3661	0.003	2133	0.042	2133	0.039
11	6311	0.000	2162	0.041	2126	0.039
12	2133	0.056	2133	0.042	2126	0.039
13	3693	0.002	2126	0.042	2126	0.039
14	2133	0.056	2133	0.042	2133	0.039
15	4558	0.000	2133	0.042	2133	0.039
16	3661	0.003	2162	0.041	2162	0.037
17	3446	0.022	2133	0.042	2133	0.039
18	2162	0.056	2162	0.041	2133	0.039
19	2133	0.056	2133	0.042	2133	0.039
20	2133	0.056	2133	0.042	2133	0.039
21	4687	0.000	3145	0.000	3145	0.000
22	8787	0.000	2162	0.041	2162	0.037
23	2136	0.056	2136	0.042	2136	0.038
24	7383	0.000	3145	0.000	2126	0.039
25	3426	0.025	3426	0.000	3426	0.000
26	3145	0.054	2133	0.042	2133	0.039
27	2133	0.056	2133	0.042	2133	0.039
28	2126	0.056	2162	0.041	2133	0.039
29	2133	0.056	2133	0.042	2133	0.039
30	3693	0.002	3670	0.000	3661	0.000

Table 5.4: Objective function values and the number of samples by the equal likelihood and the Chi square methods for straight PEST estimation

Trial #	Obj Fun	Number of recurrences	Chi square	Equal likelihood
1	2126	2	34	20
2	2133	8	135	80
3	2136	1	17	10
4	2162	1	17	10
5	2225	1	17	10
6	3145	3	48	30
7	3426	3	22	30
8	3446	1	7	10
9	3661	2	2	20
10	3693	2	1	20
11	4558	1	0	10
12	4687	2	0	20
13	6311	1	0	10
14	7383	1	0	10
15	8787	1	0	10
Total	—	30	300	300

Table 5.5: Objective function values and the number of samples by the equal likelihood and the Chi square methods for 3 multi-start PEST estimation

Trial #	Obj Fun	Number of recurrences	Chi square	Equal likelihood
1	2126	2	25	20
2	2133	15	188	150
3	2136	1	13	10
4	2162	6	74	60
5	3145	3	0	30
6	3426	2	0	20
7	3670	1	0	10
Total	—	30	300	300

Table 5.6: Objective function values and the number of samples by the equal likelihood and the Chi square methods for 5 multi-start PEST estimation

Trial #	Obj Fun	Number of recurrences	Chi square	Equal likelihood
1	2126	6	70	60
2	2133	16	185	160
3	2136	1	12	10
4	2162	3	33	30
5	3145	2	0	20
6	3426	1	0	10
7	3661	1	0	10
Total	—	30	300	300

(or decay) at the source. The distribution functions differ from the others, where one or multiple rather complete normal distributions are presented. This mainly resulted from the incorporation of spatial and temporal varying recharge and the insensitivity of the parameters to the performance measure.

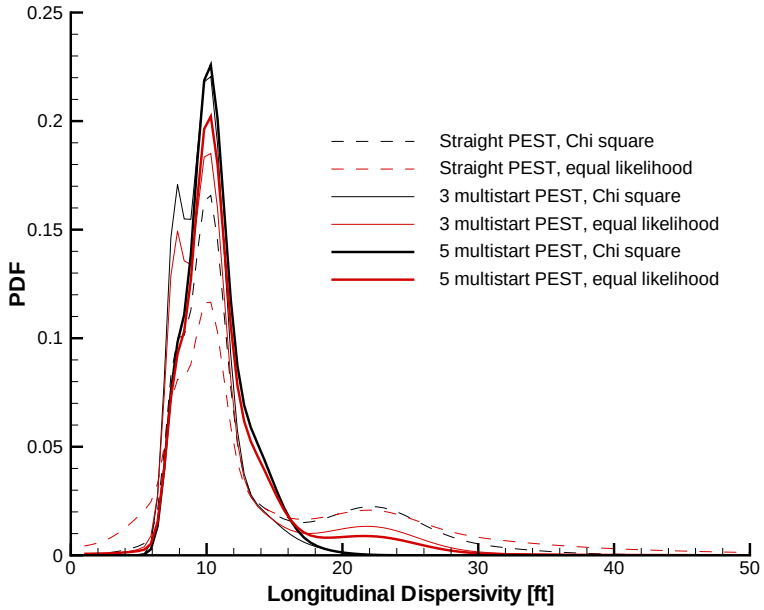

Figure 5.1: Unified probability density function for longitudinal dispersivity over 30 calibration trials

The unified probability density functions shown above can not exhibit correlations among parameters. Sampling accounting for correlations plays a key role in reducing uncertainty and producing stable numerical estimates. Because it is too tedious to show 2-D sampling scatter plots for every combination of any two parameters, only the two most sensitive parameters, longitudinal dispersivity and distribution coefficient, were selected and plotted in Figures 5.7, 5.8, and 5.9 for straight, 3 multi-start and 5 multi-start PESTs respectively. The mean values represent the calibrated parameter sets or local optima, which are surrounded by the randomly generated Latin hypercube samples following specific correlation patterns. Compared to Figures 5.8 and 5.9, sampling with straight PEST calibration results (refer to Figure 5.7) incorporates more local optima with different correlation coefficients, and samples are more sparsely distributed within the feasible region.

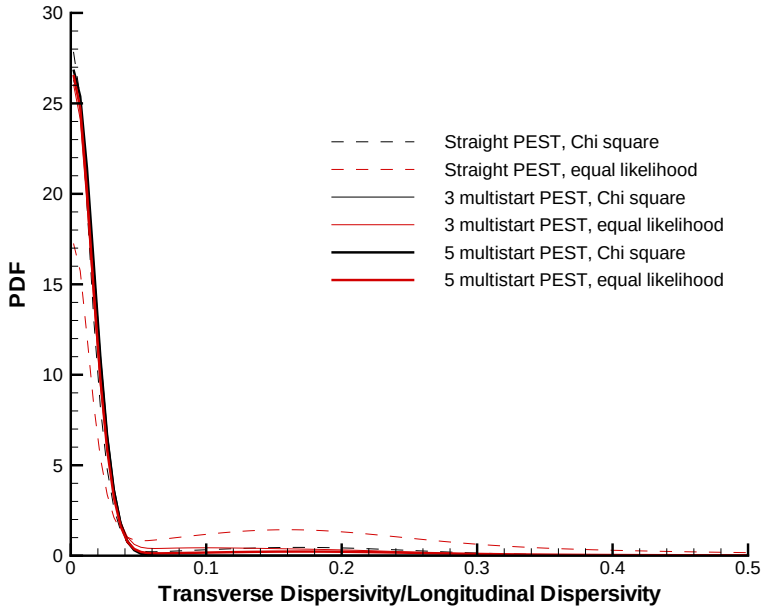

Figure 5.2: Unified probability density function for the ratio of transverse dispersivity to longitudinal dispersivity over 30 calibration trials

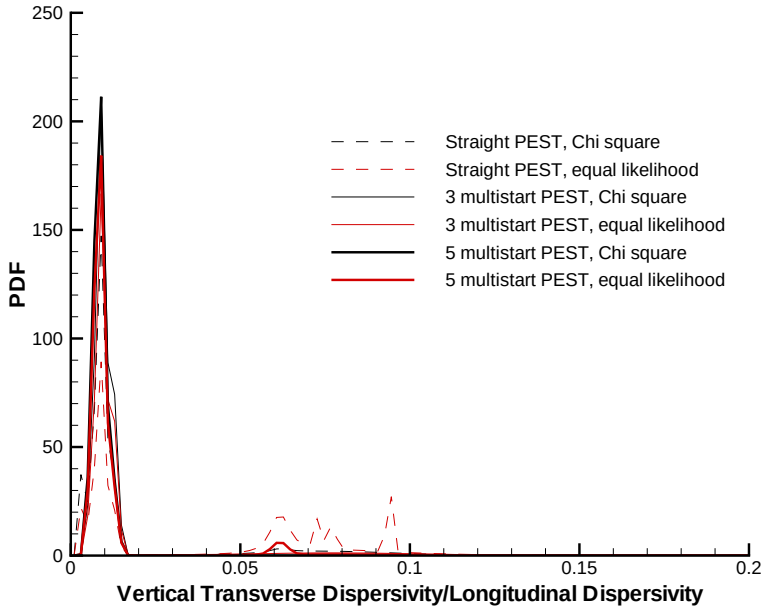

Figure 5.3: Unified probability density function for the ratio of vertical transverse dispersivity to longitudinal dispersivity over 30 calibration trials

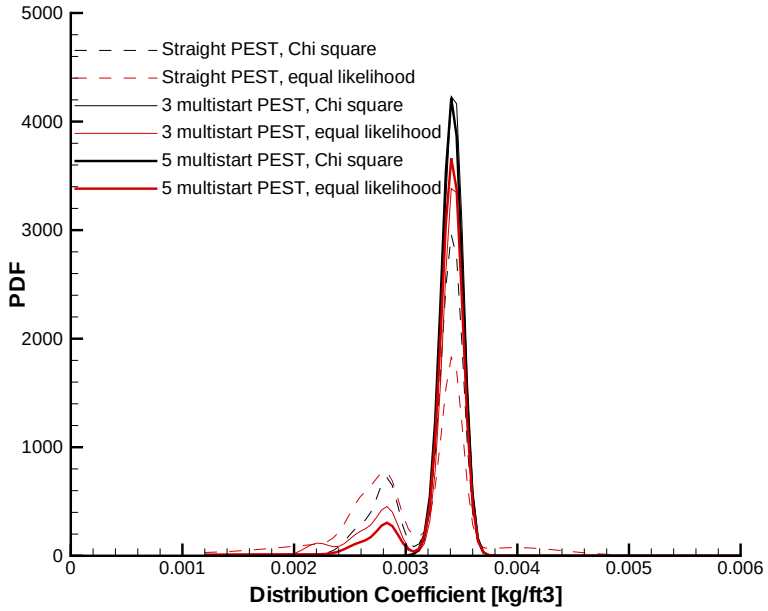

Figure 5.4: Unified probability density function for distribution coefficient over 30 calibration trials

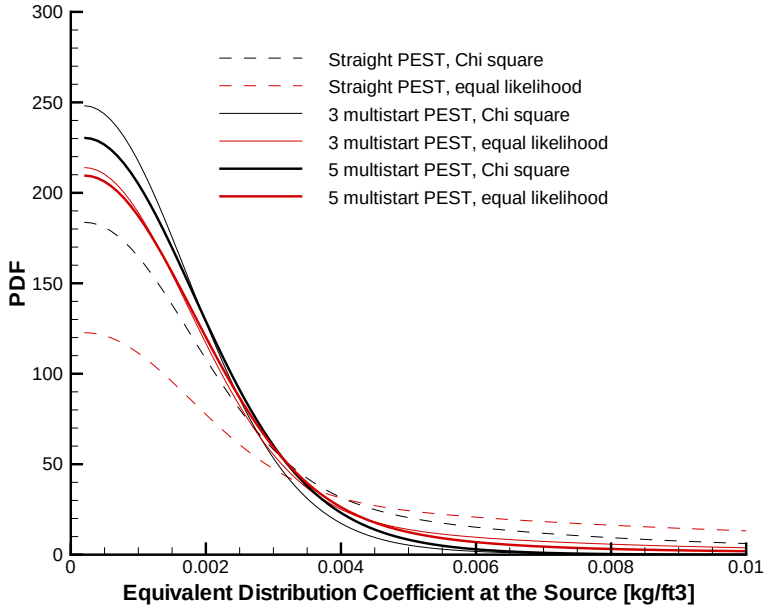

Figure 5.5: Unified probability density function for the equivalent distribution coefficient at the source over 30 calibration trials

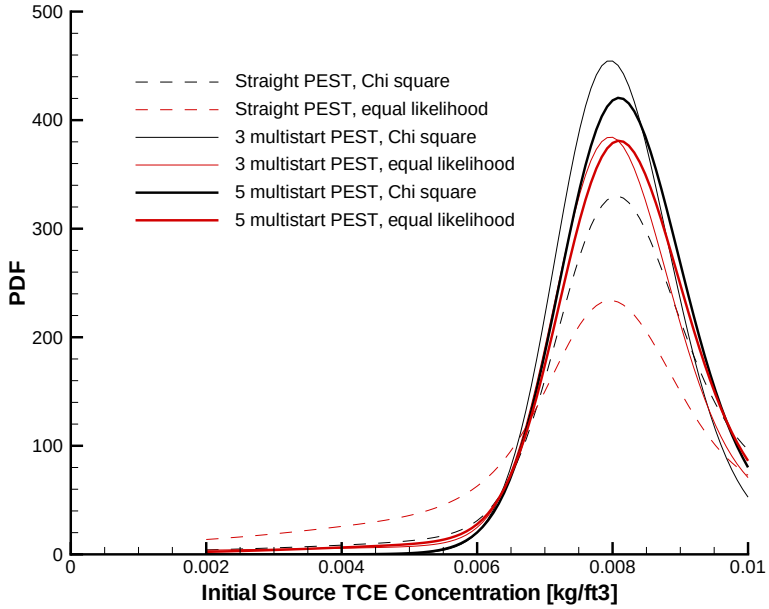

Figure 5.6: Unified probability density function for the initial TCE concentration at the source over 30 calibration trials

Most of the samples of the uncertainty analysis for 3 and 5 multi-start PESTs emphasize a few fairly good realizations. Therefore, the overall correlation, which can be deemed as the weighted averaged value over 30 parameter sets, is much weaker for straight PEST. The dependency between these two parameters can be easily identified for 3 and 5 multi-start PESTs, and the coefficients are -0.305 and -0.304, respectively.

5.5 Concentration Breakthrough Curves

The concentration breakthrough curves are generated at 101 wells including all the observation wells and pumping wells for each of the 300 Latin hypercube trials. Common across all the trials, 351 time steps were required to complete the simulation from October 1971 to the end of 2000. Then, for each well, 300 concentration values are readily available at every output time, and can be averaged and sorted to produce the mean, 5th and 95th concentration percentiles for the concentration breakthrough curves.

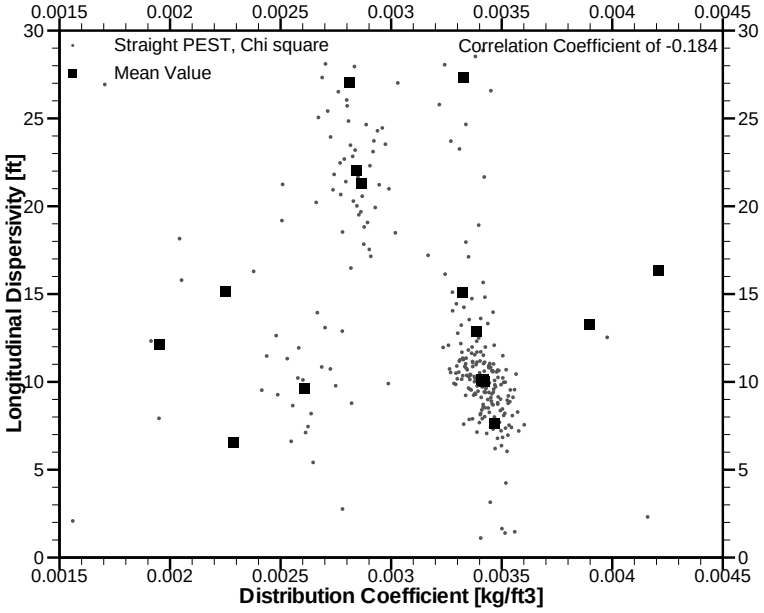

Figure 5.7: Scatter plot produced in a Latin hypercube sampling of size $nS = 300$ and parameter sets of size 30 for straight PEST

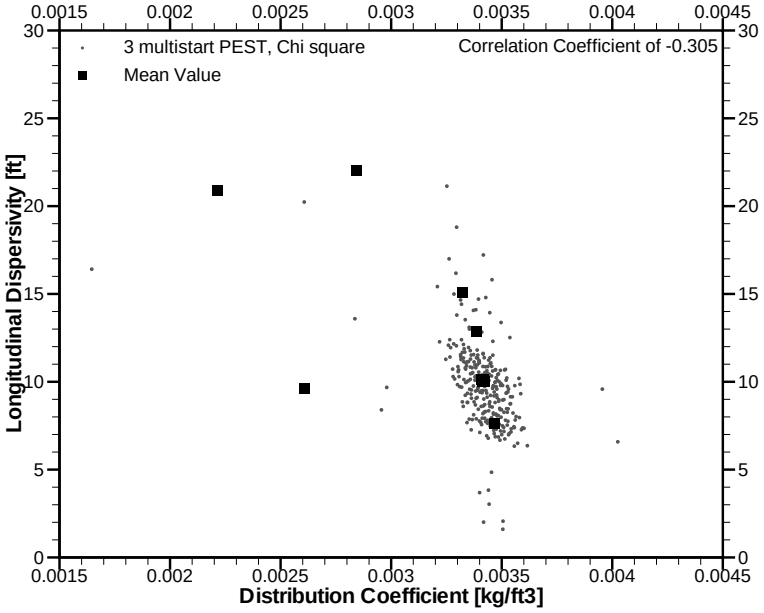

Figure 5.8: Scatter plot produced in a Latin hypercube sampling of size $nS = 300$ and parameter sets of size 30 for 3 multi-start PEST

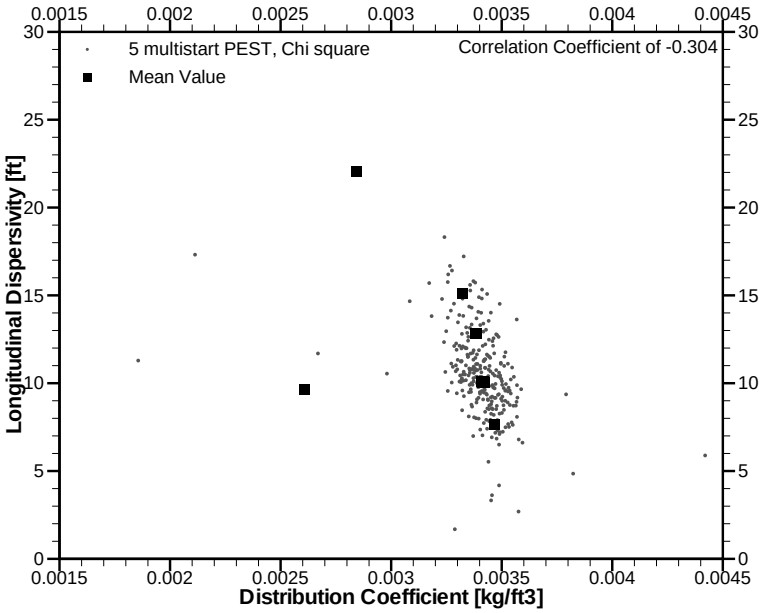

Figure 5.9: Scatter plot produced in a Latin hypercube sampling of size $nS = 300$ and parameter sets of size 30 for 5 multi-start PEST

It is redundant to show the uncertainty analysis for every well. Wells 26 and 28, and the influent to the air stripper were chosen and their uncertain breakthrough curves were plotted in Figures 5.10, 5.11, and 5.12, respectively. Wells 26 and 28 are used to serve as drinking water supply wells for the local community. The concentration for the influent of the air stripper treating the confluent flow from wells 26 and 28 was calculated on a flux averaging basis. What's more, both wells and the air striper were relatively densely monitored, and comprised 51% of the total TCE observations.

In uncertainty analysis, a confidence interval is normally used to measure the precision of the estimated value. The interval represents the range of the values, consistent with the data, which is believed to encompass the "true" value with high probability. As shown in Figures 5.10, 5.11, and 5.12, percentile based confidence intervals are bounded by the 5^{th} and 95^{th} percentile curves. It seems that the confidence intervals have a linear relationship with the mean values, which implies that the relative confidence intervals defined by the

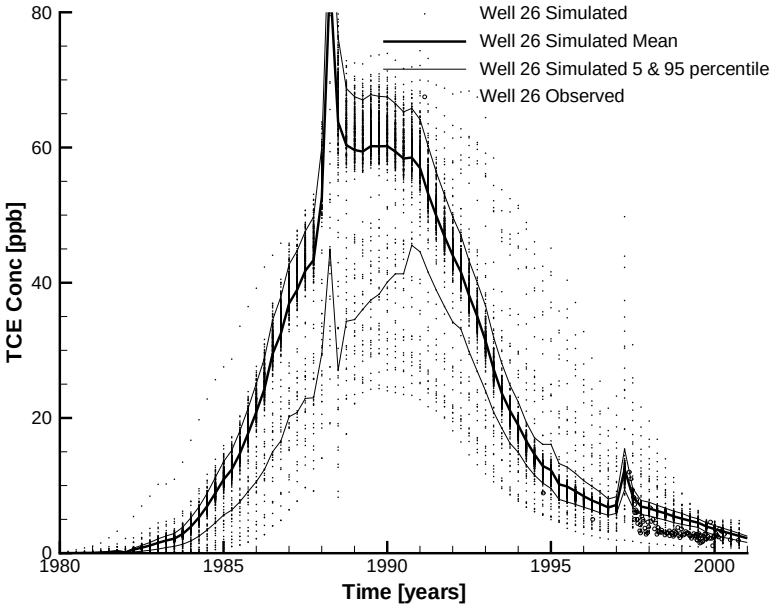

Figure 5.10: Comparison of measured and simulated uncertain TCE concentration breakthrough curves for well 26 with 3 multi-start PEST and the Chi square weighting method

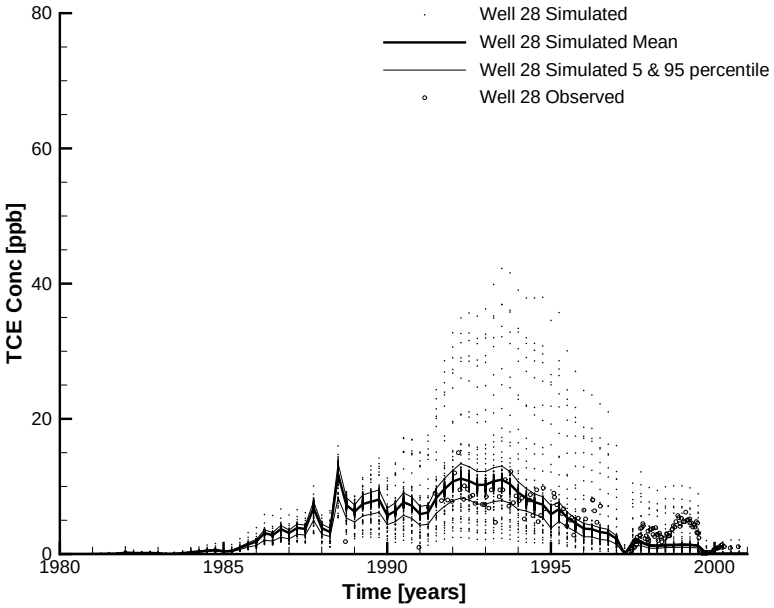

Figure 5.11: Comparison of measured and simulated uncertain TCE concentration breakthrough curves for well 28 with 3 multi-start PEST and the Chi square weighting method

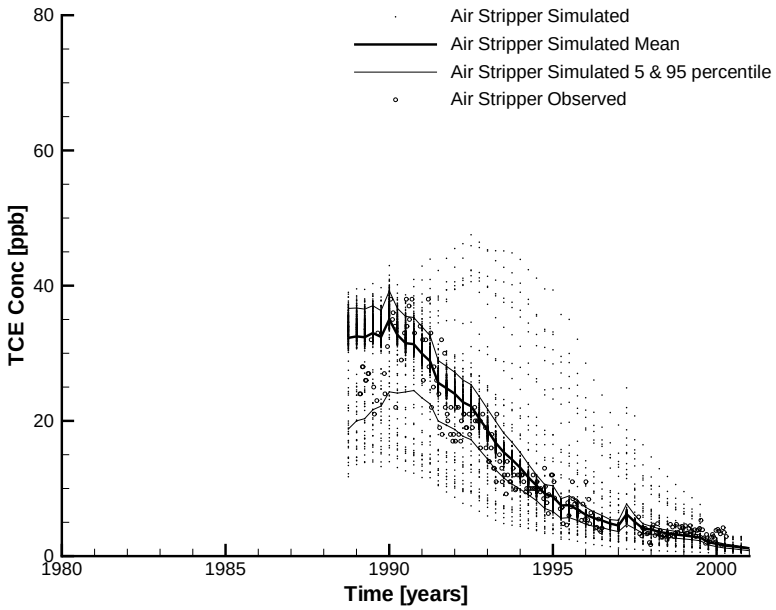

Figure 5.12: Comparison of measured and simulated uncertain TCE concentration breakthrough curves for the air stripper with 3 multi-start PEST and the Chi square weighting method

following equation are rather constant:

$$CI_{relative} = \left(\frac{Bound_{up} - Bound_{low}}{Mean} \right) \tag{5.18}$$

where, $Bound_{up}$ and $Bound_{low}$ represent upper and lower limits for the confidence interval respectively. The simulated results depicted in Figure 5.13 clearly show that the relative uncertainty roughly maintains around 0.4 after 1985 for well 26 and the air stripper. For well 26 and 28, the relative uncertainty is high before 1985, because the numerical dispersion, proportional to the squared Darcy velocity and timestep ($D_{num} \sim v^2 \Delta t/2$), will contribute to the instability which is relatively large for small concentrations in the margin of the contaminant plume. With given models, the uncertainty caused by numerical error is unavoidable. However, part of uncertainty depends on the sensitivity of parameters, and normally has a monotonic increased relationship with perturbation magnitude. When the observation points are located at either the margin of the contaminant plume or out of the contaminant plume, the simulated TCE concentration is extremely low, and the numerical error becomes a dominant factor. The oscillation of relative confidence interval for well 28 in the late of 1990s also is attributed to this explanation. While the use of a smaller timestep size would reduce the impact of numerical dispersion, the results would not alter the given explanation.

5.6 Impact of Calibration and Weighting Approaches on Uncertainty Reduction

As discussed in a previous section, weighting schemes and parameter sets generated by different calibration algorithms tend to have various definitions of the likelihood measures and the sampling spaces, with this having a key role in the evaluation of uncertainty. With the same parameter sets, the Chi square weighting method is more efficient in uncertainty reduction with narrower 90% confidence intervals in Figure 5.12 than the equal likelihood approach shown in Figure 5.14.

The uncertain breakthrough curves and related 90% confidence interval at the influ-

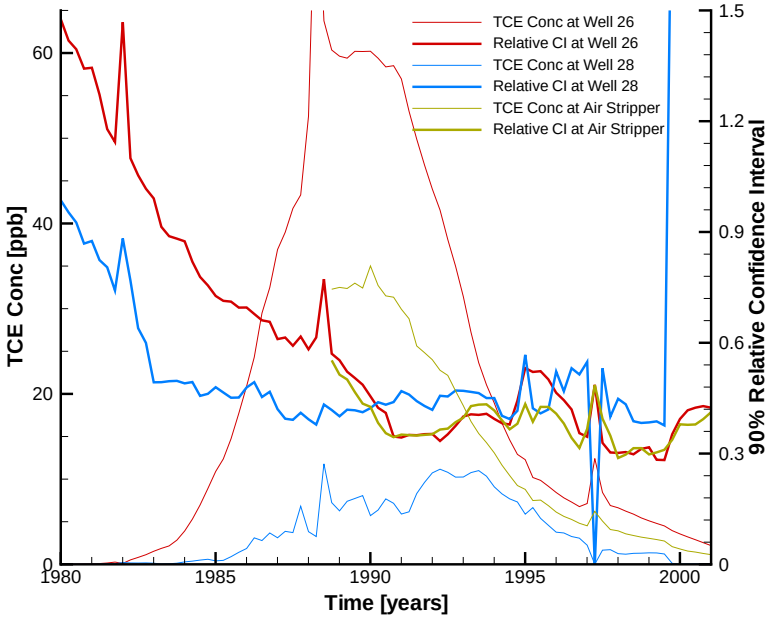

Figure 5.13: Simulated TCE concentration breakthrough curves and the associated relative confidence intervals for wells 26, 28, and the air stripper with 3 multi-start PEST

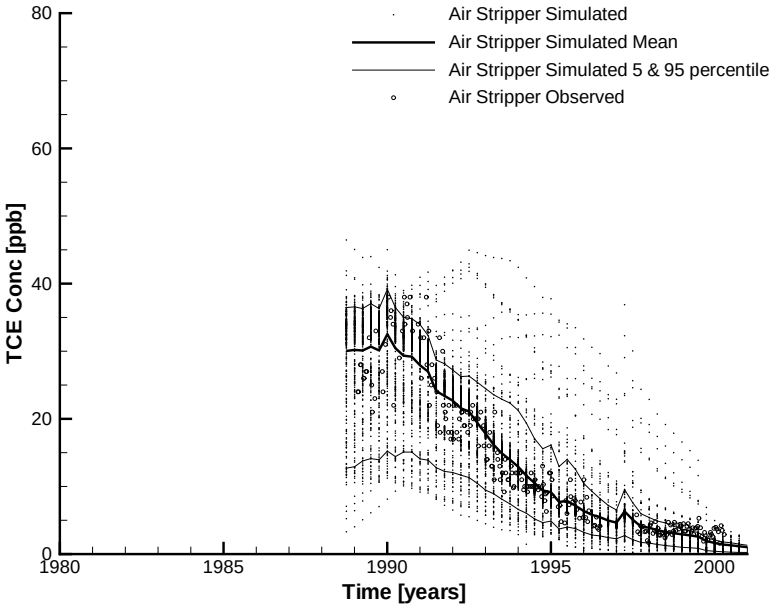

Figure 5.14: Comparison of measured and simulated uncertain TCE concentration breakthrough curves for the air stripper with 3 multi-start PEST and the equal likelihood method

ent to the air stripper for straight and 5 multi-start PESTs are presented in Figures 5.15 and 5.16 for the Chi square and equal likelihood weighting schemes respectively. Compared to Figures 5.12 and 5.14 with a 3 multi-start PEST, TCE concentration breakthrough curves with straight PEST show more variability, and are less efficient in uncertainty reduction. No obvious difference between 3 and 5 multi-start PESTs was observed in terms of the mean and 90% confidence intervals of breakthrough curves. This result can be inferred by the fact that estimated parameter sets with straight PEST contain more diversified local optima with more evenly assigned likelihood weights, as shown in Table 5.3, and subsequently, randomly generated Latin hypercube samples are more sparsely distributed in Figure 5.7 than those in Figures 5.8, and 5.9.

5.7 Impact of the Sample Size on Convergence

Sample size is regarded as an influential factor in the evaluation of sampling based methods. The unbiased uncertainty estimation will converge to reality as the sample size approaches infinity. However, with limited computational resources, a trade-off between the efficiency and the accuracy always arises. To ensure that the numerical estimates are accurate approximations of the actual distribution, the statistical quantities obtained by ensemble analysis must converge. Therefore the minimum number R_{min} of realizations generated per randomly sampled parameter set must be identified so that the numerical estimates do not depend on the number of realizations.

Uncertainty estimation with 300 samples was conducted in a previous section, and a sample size of 30 [*David et al.*, 1991] is usually deemed as the least number of samples to apply the central limit theorem in order to obtain the variance-based confidence intervals. A series of sample sizes is evenly distributed between 30 and 1000 on a logarithmic basis to yield 30, 100, 300, and 1000. The uncertain breakthrough curves with means and percentile-based 90% confidence intervals are shown in Figure 5.17. The result shows that not only the uncertainty decreases considerately as the sample size increased, but also the mean breakthrough curves are affected by the small sample size. Although the improvement is evident by comparison of means and confidence intervals with 300 and 1000 samples, the

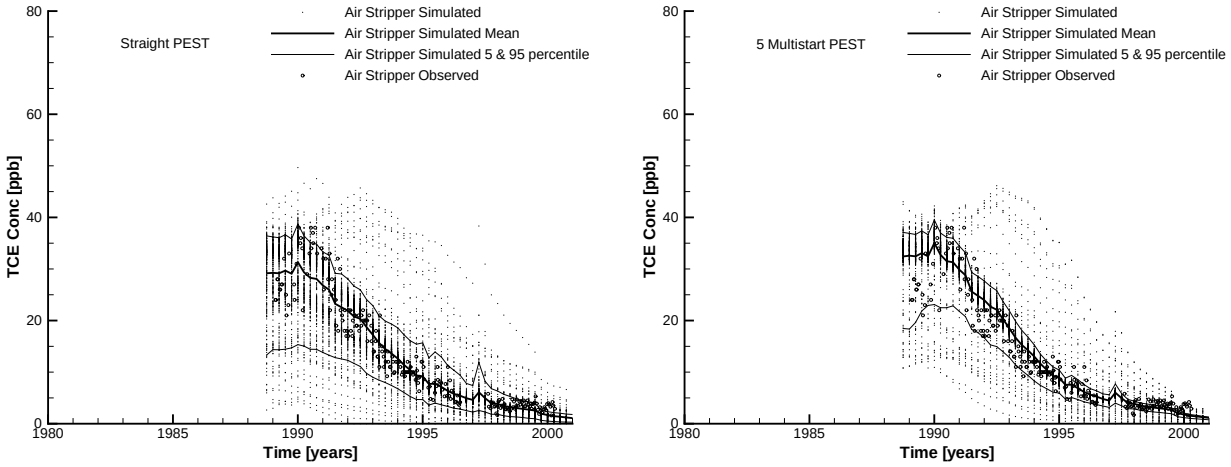

Figure 5.15: Simulated means and percentiles of uncertain TCE concentration breakthrough curves for the air stripper with straight and 5 multi-start PEST and the Chi square weighting method

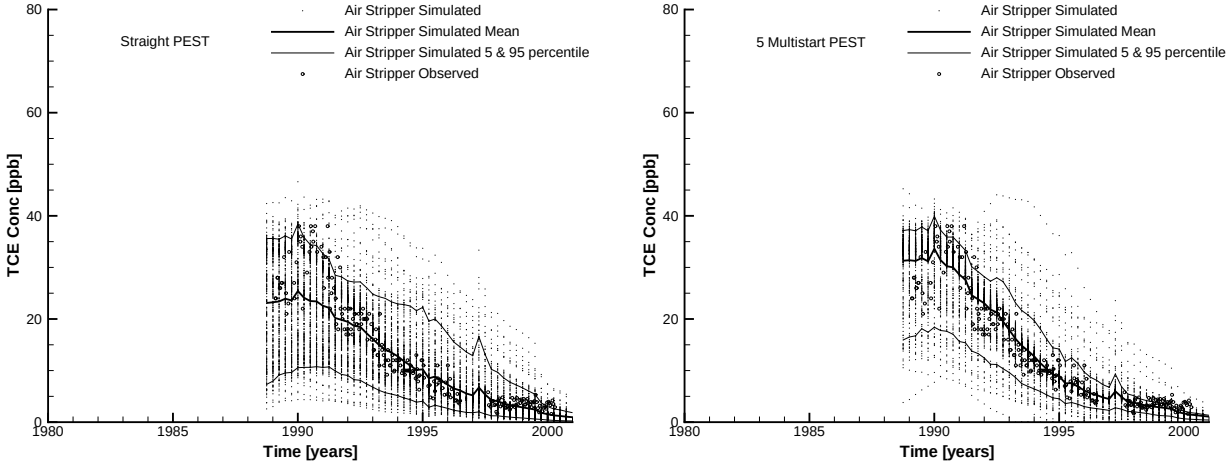

Figure 5.16: Simulated means and percentiles of uncertain TCE concentration breakthrough curves for the air stripper with straight and 5 multi-start PEST and the equal likelihood method

Latin hypercube sample size of 300 is considered enough to yield satisfactory results in terms of the efficiency and the accuracy.

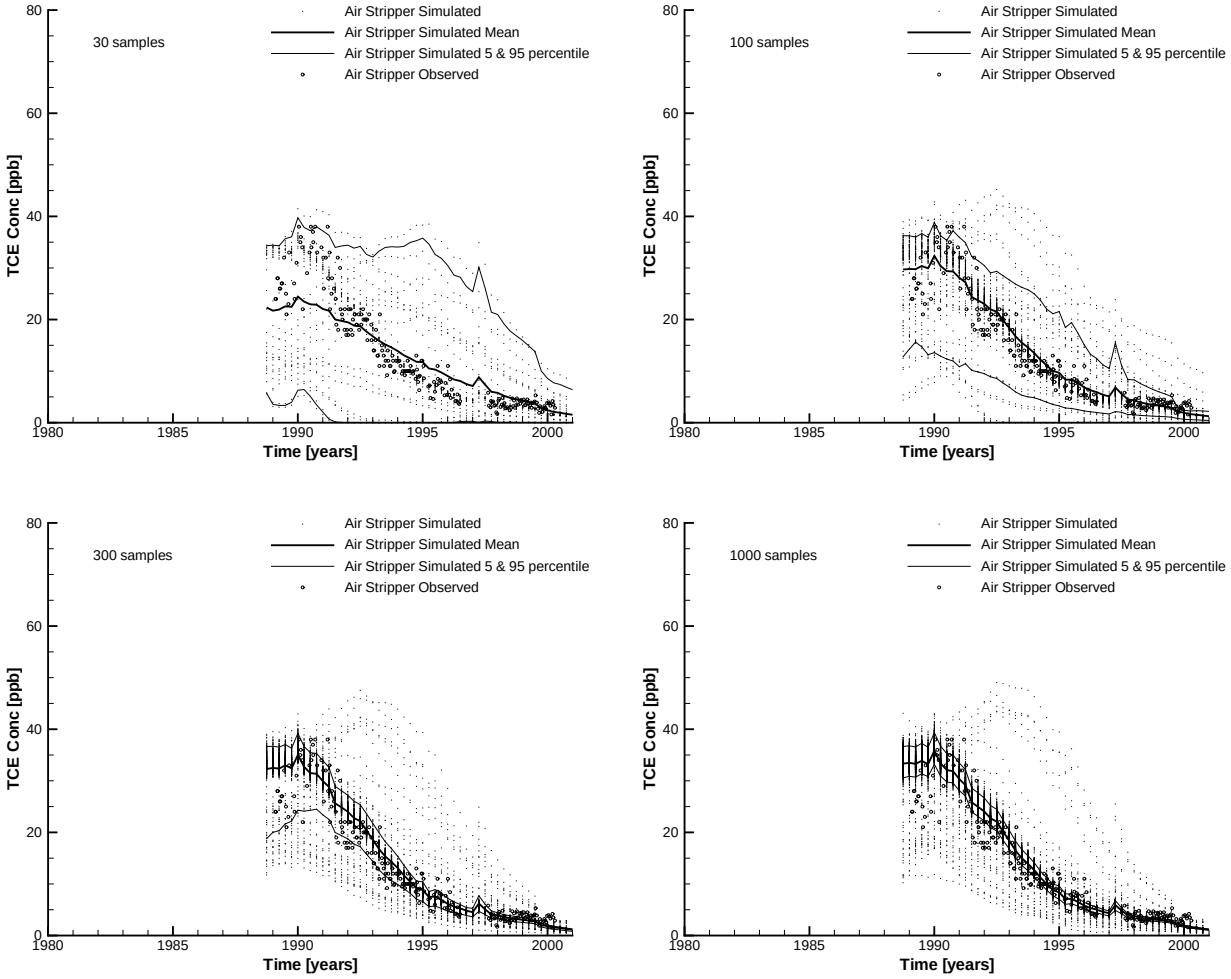

Figure 5.17: Simulated means and percentiles of uncertain TCE concentration breakthrough curves for the air stripper with 3 multi-start PEST and the Chi square weighting method

Figure 5.18 depicts the means and variance-based 90% confidence intervals of TCE concentration breakthrough curves for the air stripper with different sample size. In spite of the fact that the variance-based confidence interval estimates do not depend on the magnitude of the mean values, and are more stable than those of the percentile-based

approach, the impact of population size on the CI estimates is still evident. Large sample size leads to narrow confidence interval curves, and marginal improvement of CI curves dwindles as the population increases. Both the percentile and variance based uncertainty analysis with 300 samples provides comparable results to scenarios with more population. Therefore, for this case study, when the number of samples is larger than 300, Latin hypercube sampling with 3 multi-start PEST can consistently guarantee convergence of the approximation towards the actual, but unknown probability density. This conclusion may be problem specific.

5.8 Summary

Propagating the parameter uncertainty into the model outcome is an important step to apply the model to the future planning of the well field. Based on multi-start PEST calibration results, a modified Latin Hypercube sampling approach accounting for correlation between parameters was employed to conduct uncertainty analysis for contaminant concentration breakthrough curves in pumping wells. Due to the non-uniqueness nature for ill-posed inverse problems, multiple feasible parameter sets and covariance matrices (instead of selecting the global optimal for the well-posed inverse problems), which were generated using the multi-start PEST algorithm, should be employed to form the solution space. Assuming the observed concentrations follow a normal distribution with constant variance and the mean equal to the simulated concentrations, the likelihood for each parameter set, expressed as the normalized CDF of a Chi square distribution, was determined by a newly developed Chi square based weighting scheme. The Chi square weighting method is more efficient than the equal likelihood method in uncertainty reduction.

Impact Evaluation of the impact of sampling size on the estimated uncertainty indicates that the least number of the sample size to yield satisfactory results in this case study was 300 distributed over 30 feasible parameter sets. The mean and confidence intervals of TCE breakthrough curves at the municipal wells and influent to air stripper were estimated using the LHC sampling approach. The results demonstrate that the relative confidence interval

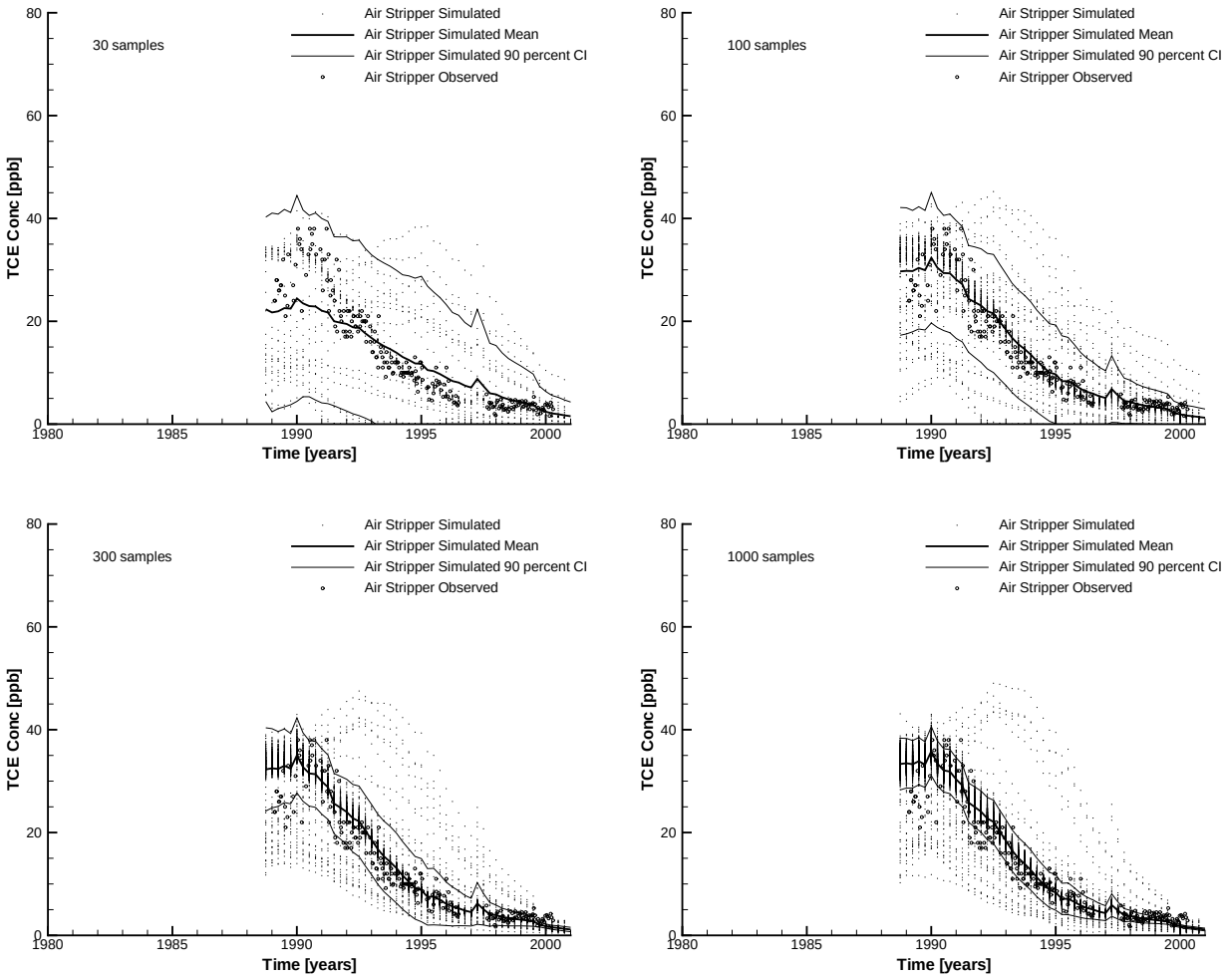

Figure 5.18: Simulated means and 90 percent confidence intervals of uncertain TCE concentration breakthrough curves for the air stripper with 3 multi-start PEST and the Chi square weighting method

is considered large at the margin of the contaminant plume as a result of the dominance of the numerical dispersion.

Chapter 6

Conclusions and Recommendations

6.1 Conclusions

The emphasis of this thesis is the investigation of a computationally intensive groundwater contaminant transport problem. This thesis has described the development of a three-dimensional, physically based, transient, computationally intensive groundwater flow and contaminant transport model for the Reich Farm Superfund site near Toms River, New Jersey. The flow system was manually calibrated by *Sykes and Normani* [2002] using the method of trial-and-error. The hydraulic conductivity distribution and the spatially and temporally varying recharge developed in their analysis were adopted for the contaminant transport analysis of this thesis. The transport model was efficiently calibrated by a methodology presented in the preceding chapters; the calibration methods included multi-start PEST and two heuristic algorithms. Data for the contaminant transport parameter estimation include measured TCE concentrations at monitoring wells and measured concentrations at Toms River Parkway wells 26 and 28. Measured TCE concentrations were also available for the influent to an air stripper at the Parkway well field. A comparison of the transport model calibration under four different scenarios was conducted to evaluate the importance of spatially and temporally varying recharge on the estimation of contaminant parameters. In addition, two different forms of contaminant source char-

acterization, representing the NAPL dissolution process at the source area, were used in the calibration to investigate the impact on the parameter estimates and the breakthrough curves at the pumping wells. In the end, a Latin hypercube simulation comprised of 300 trials distributed in 30 solutions sets was performed on the Sharcnet. Each contaminant transport trial yielded a simulated TCE concentration breakthrough curve at the pumping wells. The resulting 300 breakthrough curves for each well were used to generate an uncertain concentration confidence interval. More concretely, the following conclusions may be drawn from this study:

- Due to the convergence issue of implementing saturation versus permeability curves (van Genuchten equation) for the large scale models with coarse discretization, a flux based method, with an assumption of unit hydraulic gradient in the vadose zone, was developed to calculate relative hydraulic conductivity in the simulation of the unsaturated groundwater flow. The flux based approach was compared in a 1-d column verification example. The difference of actual travel time from the dumping site to the water table is trivial, compared to the travel time in the whole study domain.

- The highly varying withdrawal rates of the municipal pumping well field necessitate a small adaptive time step and large computational effort for the simulation of the coupled transient groundwater flow and contaminant transport. With current computing resources, it is unrealistic if not impossible to identify the contaminant source release history and contaminant transport parameters through an optimization approach with the coupled system. Through the separation of the transport model from the flow model, the required computational time for one model run can be greatly reduced to 10 minutes from 4 hours by using a variant timestep scheme honouring the variation of the pumping rates. This approach was validated by the good agreement of the TCE breakthrough curves at well 26 between the separated and originally coupled model simulation.

- The objective function in PEST is formulated as the sum of the squared residual errors between the measured and simulated concentrations, which is extremely sensitive to the outliers. To deemphasize the outliers in the optimization and conform

to the formulation of the objective function definition in PEST, a robust estimator of the modified SSE is introduced in equation (4.3).

- A three-dimensional, physically based, field-scale groundwater flow and contaminant model for the Reich Farm Superfund site was developed and implemented in HydroGeoSphere. The model incorporates the vadose zone, highly varying pumping rates, and spatially and temporally varying recharge. The TCE residuals serving as the contaminant source are considered to be immobile, and for the base-case analysis, a linear equilibrium model was employed to describe the mass transfer from the residual to the dissolved phase. Using an optimization algorithm, the transport model was adequately calibrated against the observed data (935 TCE concentration measurements, and most of them are at later time). The decision variables include four transport parameters (longitudinal dispersivity, transverse dispersivity, vertical transverse dispersivity, distribution coefficient), and two source term parameters (NAPL source distribution coefficient and initial concentration at the source). The calibration results indicate that the performance measure is sensitive to the longitudinal dispersivity and distribution coefficient, and that these parameters have relatively narrow confidence intervals.

- The inclusion of spatially and temporally varying recharge in the model tends to induce the variation in local velocity in the shallow aquifer, and reduce the macroscopic dispersion. Then, the estimated transverse dispersivity will be largely reduced, and its values are significantly less than that reported in the literature for the more traditional approach that uses steady-state flow with averaged, less physically based recharge values. Compared to the transport model calibration under three other scenarios, the base-case has the most varying local velocity field and the least transverse dispersivity. Undoubtedly, the scenario with constant recharge has the most uniform velocity distribution, and the estimated transverse dispersivity reached the upper bound.

- Two NAPL residual dissolution models, the linear equilibrium at the interphase and the modified model L from *Zhu and Sykes* [2004], were implemented and calibrated in the case study for the Reich Farm Superfund site. The results demonstrate that the

estimated parameters are unique, even for the commonly shared transport parameters. However, the impact on the target performance measure, TCE concentrations in this case, is trivial. Therefore, calibration can only be used to estimate a feasible range of the parameter values, instead of the accurate values.

- Given the groundwater flow model well calibrated, the contaminant transport parameters were estimated using optimization with two heuristic search algorithms and a gradient based multi-start PEST algorithm. In other words, the impact of flow model on the contaminant transport was not addressed in this thesis. The results for the calibration of the contaminant transport model demonstrate that all three approaches are effective and give satisfactory performance. However, the solution sets are sensitive to the uncertain measured concentrations in the definition of the objective function. Even with the measured concentrations being certain, thousands of model evaluations are normally required to achieve a global minimum for the ill-posed non-convex (or flat response surface) inverse problems. Such an exhaustive search is indeed impractical, if not impossible, for this large scale computationally intensive problem. Thus, the solutions are regarded as the local minima or the candidates for the global minimum.

- The performances of DDS and MicroGA are almost indistinguishable from a practical calibration perspective. 40% of the selected initial solutions for straight PEST converge to good solutions. Multi-start PEST, which takes advantage of the robustness of the heuristic methods and of the fast convergence of gradient based approach, performs best in terms of the final best objective function values with equal number of function evaluations. As a result of the inherent parallelization of the multi-start technique, the actual elapsed time for multi-start PEST for one trial is less than 1/5 of those for both DDS and MicroGA when running on the Sharcnet.

- Based on the Latin Hypercube sampling, a methodology for comprehensive uncertainty analysis, which accounts for multiple parameter sets and the associated correlations, was developed and applied to the case study. Two assumptions are necessary in the derivation of this methodology. For each parameter set, the parameters are assumed to obey multivariate normal distribution with given means and covariance

matrix obtained through calibration outcomes. The observed concentrations are assumed to follow a normal distribution with constant variance and the mean equal to the simulated concentrations. In the method, the likelihood related to each parameter set was determined by the performance measure (objective function value). Given the uncertainties associated with the contaminant transport parameters for TCE and the source zone, a 300 Latin hypercube simulation distributed in 30 solution sets was selected to calculate mean and confidence intervals of TCE breakthrough curves at the municipal wells and influent to air stripper. The relative confidence interval is considered large at the margin of the contaminant plume. The main limitation in performing a sampling based uncertainty analysis is the associated cost and effort. The uncertainty was propagated into TCE breakthrough curves with series sample sizes of 30, 100, 300, and 1000. 300 is considered the least number of the sample size to yield satisfactory results in this case study.

6.2 Recommendations

During the present study, some areas were revealed of interest for future research. They are listed as follows:

- This study estimated the transport parameters and the contaminant source term in a transient, but deterministic flow field. However, the flow field usually suffers from uncertainties, to some extent, because of the imperfect and sparse measurements. As an important parameter affecting the flow distribution, hydraulic conductivity is deemed crucial for the modelling of the fate and transport of contaminants in subsurface systems. To date, its impact on the inverse problems for finding contaminant source has not been identified. It is foreseeable that this kind of impact will most likely be even worse due to the ill-posedness nature of inverse problems.
- To conform to the formulation of the objective function definition in PEST, a modified SSE was used in the parameter estimation processes. Although the sum of the squared residuals is normally considered, the definition of the objective function

controls the relative influence of measurements on the parameter estimates. Except for the L2-norm estimator, many others are available, including L1-norm estimator, Huber's M-estimator, and Cauchy's estimator. Because the robustness of an estimator is problem-dependent, there is a need to evaluate the performance of different estimators in the application of this case study.

- This research reconstructed the contaminant source release history with known source location. However, in some cases, the location and the extent of the contamination source are not readily available. It would be beneficial to develop a methodology capable of identifying the contaminant source concentration, duration, and the location simultaneously for a complex, field scale models.

- Mass transfer, taking place on the interfaces between NAPL residual and aqueous phases, depends on the size and geometry of NAPL blobs, the history of aqueous phase movement, and capillary forces. In this research, two simplified source characterization approaches were implemented and calibrated. To provide more accurate prediction of mass transfer and ongoing contaminant plume migration, physically based dissolution models could be investigated.

- An important step in the procedure of uncertainty analysis is the calculation of a likelihood measure for the local optimal parameter sets. Because the choice of the shaping factor in the conventional likelihood functions is rather subjective, a statistically based, empirical likelihood definition was developed to determine the probability measure for each parameter set, and compared with the equal likelihood method. However, a thorough comparison between the newly introduced and traditional likelihood definitions needs to be investigated.

6.3 Contributions

- Developed a flux based method to calculate relative hydraulic conductivity in the simulation of the unsaturated groundwater flow;

- Through the separation of the transport model from the highly varying transient flow model, the required computational time for the transport model can be greatly reduced using a variant timestep scheme;
- Compared the performance of different optimization algorithms through the application to the case study: the Reich Farm Superfund site;
- Evaluated the impact of the simplified representation of the hydrologic processes on the estimated parameter values;
- Evaluated the impact of different contaminant source characterization on the calibration results;
- Based on the Latin Hypercube sampling, a methodology for comprehensive uncertainty analysis, which accounts for multiple parameter sets and the associated correlations, was developed and applied to the case study.

Appendix A

Calibration of the Reich Farm Groundwater Flow Model by Sykes and Normani (2002)

This appendix presents the calibration of the *Sykes and Normani* [2002] Reich Farm groundwater flow model that forms the basis of the contaminant transport analysis developed in this thesis; the presentation is included in the thesis as the report by *Sykes and Normani* [2002] is not widely available. The text of this appendix is taken directly from *Sykes and Normani* [2002] with their permission. Their calibration methodology and report were reviewed by the U.S. Environmental Protection Agency, the New Jersey Geological Survey, United Water Toms River and the citizens groups of Toms River including their attorneys. The groundwater flow conceptual model and calibration methodology developed by *Sykes and Normani* [2002] using MODFLOW was followed by the author of this thesis using a FRAC3DVS/HydroGeosphere framework. This framework is described in Chapter 3 of this thesis; selected figures cited in this appendix also are included in Chapter 3.

A.1 Conceptual Model

A.1.1 overview

The Reich99a model and the conceptual model on which it is based encompasses the area east of the Toms River to Barnegat Bay and encompasses the Reich Farm Superfund site and the Dover Township Landfill (see Figure 3.1 of this thesis). The northern boundary of the model domain is assumed to be a groundwater flow line and the boundary is oriented with a flow line in the surface water system. It is assumed that the direction of groundwater flow along the northern boundary is insensitive to changes in pumping at the UWTR wells. Vertical leakage is permitted at streams within the modeled domain. Long Swamp Creek is an example of such a stream. The top of the first layer of the 4-layer MODFLOW model corresponds to the ground surface as defined by a digital elevation model. The 4th model layer corresponds to the deeper, less hydraulically conductive portion of the Cohansey- Kirkwood aquifer and is assigned a horizontal hydraulic conductivity of 0.283 ft/day. The top of this layer is interpolated from well drillers logs; at the UWTR Parkway Wellfield, the top of the layer corresponds to the top of the clay layer where the screens of the pumping wells terminate. The bottom of the 4th layer corresponds to the bottom of the Cohansey-Kirkwood aquifer as defined by Nicholson and Watt in the USGS report (Figure 5, 1997).

The hydraulic conductivity for MODFLOW layers 1 to 3 varies spatially with the distribution being developed from well drillers logs. The primary aquifer in the vicinity of the Reich Farm Site is the Cohansey Formation and is comprised of sand with clay and gravel lenses. The database for the determination of the lithology and the spatial extent of the Cohansey formation includes:

- remedial investigation wells and borings associated with the Reich Farm Superfund Site
- wells and borings for the Dover Township Municipal Landfill site investigation
- borings associated with the drilling of production wells for United Water Toms River

- other domestic, industrial, or production wells and/or borings as obtained from the Bureau of

Water Allocation at the New Jersey Department of Environmental Protection. For the study area, the database includes approximately 1,900 unique well records and their associated documents, of which only 1,383 records contain a lithologic description. A summary of the range in depths for those records which contained a lithologic description is presented in Table A.1. The lithology was determined using the Reich Farm, Dover Landfill and UWTR well records, plus the domestic, industrial, and production wells/borings that had a depth greater than 100 ft.

Table A.1: Number of well logs per depth interval

Depth Range [ft]	Number of Logs	Depth Range [ft]	Number of Logs
0 to 40	288	150 to 200	44
40 to 60	387	200 to 250	29
60 to 80	300	250 to 500	32
80 to 100	150	500 to 1000	16
100 to 125	77	1000 +	22
125 to 150	38		

Details of the development of the hydraulic conductivity field for the Cohansey aquifer are presented in Section 3.4 of this report.

A.1.2 Spatial Domain

For the *Sykes and Normani* [2002] MODFLOW model, groundwater flow for the Reich Farm area was investigated by subdividing the spatial domain into 41,600 finite difference grid blocks. The discretization consists of 200 grid blocks in the westerly to easterly direction, 208 in the southern to northern direction and 4 layers in the vertical. The bottom boundary of the domain corresponds to the top of a basal clay layer, immediately below the Cohansey-Kirkwood formation. The elevation for the top of the basal clay layer was originally developed by Zapecza (1989) and is presented in Figure 3.9 of this thesis. It is assumed that there is no flow across this boundary.

The top of model layer 1 conforms to the ground surface and was determined using a Digital Elevation Model (DEM) as developed by the U.S. Geological Survey, and subsequently the New Jersey Geological Survey (see Figure 3.3 of this thesis). The top of model layer 4 was determined using only those wells that were drilled into the Kirkwood formation, where the Kirkwood formation is identified by fine gray sands and silts and gray clays. The locations of these wells are shown in Figure 10 of *Sykes and Normani* [2002] (not reproduced herein). These locations were determined from a combination of information provided on the well records, well permits, tax maps of Dover Township, a tax database containing property owners, property locations by street address and lot / block number, as well as the 1995/97 aerial digital orthophotography from the NJDEP. ArcView GIS was integral to the management of the data and determination of the well coordinates. The interpolated elevation for the top of the Kirkwood formation is shown in Figure 3.8 of this thesis. A kriging algorithm was utilized with a nugget variance of 4 ft^2 and a linear drift. The elevations for the tops of MODFLOW layers 2 and 3 were distributed at intermediate distances between the ground surface elevation and the elevation of the top of the Kirkwood formation.

A.1.3 Boundary Conditions

An important component of a groundwater flow conceptual model are the boundary conditions. At the boundary blocks of the spatial domain, either the heads must be specified (Dirichlet or Type I boundary condition) or the flux normal to the boundary must be specified (Neumann or Type II boundary condition). The most common form of the latter is the no-flow or groundwater divide boundary condition. For the Reich99a model, all vertical external edges of the modelling domain are described using a no-flow Type II Neumann boundary condition. The bottom of layer 4 is also described using a no-flow Neumann boundary condition.

Boundary conditions for the top surface of the model domain were represented by prescribed Type I Dirichlet boundary conditions (general head boundary (GHB) condition as implemented in MODFLOW), leakance from surface water bodies such as rivers, lakes,

and ponds, and recharge from precipitation. The development of these boundary conditions is presented in the following two sections.

Surface Water Bodies

The Toms River estuary and Barnegat Bay were represented using a general head boundary condition. Head elevations for these grid blocks were set to zero feet AMSL, and a high conductance was selected for these blocks. The River and Surface Water categories are represented in the Reich99a model using the MODFLOW river package. The surface water bodies in the Reich99a model were described using river width and length in a grid block, conductance, stage elevation, and water depth. Stage elevations of rivers and lakes or ponds were estimated from the USGS 7.5 minute quadrangle maps. The DEM was not sufficiently accurate to determine the elevation of surface water bodies. GIS layers for rivers, ponds and lakes were obtained from the NJDEP. Some of these GIS layers were corrected to reflect the actual locations of water features as shown in the USGS quadrangle maps for the area. ArcView GIS was used to facilitate the calculation of the grid block properties and to create the Visual MODFLOW boundaries input file. The stage elevations, surface water body areas, and conductances were assumed to be temporally invariant.

Recharge

This section describes the methodology used to obtain a temporally varying recharge distribution, from October 1971 to December 2000, for each finite difference grid block in the modelling domain and then integrating it into the Reich99a model. The NJGS has developed A Method for Evaluating Ground-Water-Recharge Areas in New Jersey (Charles et al., 1993). This methodology results in annual groundwater recharge rates for land areas five acres or larger. In order to obtain a more spatially and temporally detailed recharge distribution, an integrated analysis approach was developed using the ArcView GIS environment and an Agency accepted hydrologic model HELP3.

The HELP (version 3) model was developed for the USEPA Risk Reduction Engi-

neering Laboratory by the US Army Engineer Waterways Experiment Station. It is a quasi-two-dimensional, deterministic, water-routing model for determining water balances. It simulates daily water movement into the ground, and accounts for snowmelt, evapotranspiration, vegetative interception, surface runoff and temperature effects.

Both the NJGS recharge method and the integrated method developed for implementation in the Reich99a model, utilize information from land-use/land-cover (LULC) and soil maps in a soil water budget analysis. However, not only does the integrated approach allow for a much smaller spatial and temporal resolution, but it also performs a more sophisticated water budget analysis by including the effects of snowmelt, vegetative interception, surface runoff and variable temperature and recharge distributions. The integrated analysis is primarily based on land-use/land-surface cover and soils data as well as the actual daily precipitation and temperature records for Toms River, New Jersey.

Procedure The New Jersey Department of Environmental Protection (NJDEP) 1995/97 LULC ArcView coverages contain a detailed classification of the land surface cover based on the Anderson et al. (1976) Classification System. HELP3 uses the Natural Resources Conservation Service (NRCS) (formerly Soil Conservation Service) curve number (CN) method in calculating quantities of surface runoff, therefore, the Anderson codes were matched to the NRCS land cover categories by grouping them into 11 different LULC group codes, as presented in Table A.2. These group codes, along with percent imperviousness (IS) information, were then combined in ArcView GIS with the USDA soil maps (not reproduced herein) to obtain a LULC/Soil combination map (not reproduced herein). Combining the 11 LULC groups with 27 soil types resulted in a total of 694 unique LULC/soil combinations. Details of the procedure used to determine the temporal recharge distribution for each of the combinations are described in the following paragraphs.

NRCS Curve Numbers A curve number for each of the LULC/Soil combinations was determined using the methodology presented in Chapter 2 of the NRCA TR-55 (USDA, 1986). This methodology was developed specifically for small watersheds in urbanized areas and is therefore well suited for the Toms River watershed.

Table A.2: Land Use and Land Cover group codes

LULC Group Code	General Description	NRCS Cover Description	Detail
1	Urban / residential areas	Open space – good condition	Adjust for imperviousness using Figure 2-3*
2	Altered lands and transitional areas	Open space – fair condition	Adjust for imperviousness using Figure 2-3*
3	Barren lands	Open space – poor condition	Adjust for imperviousness using Figure 2-3*
4	Agricultural areas	Agriculture	Assume an average CN
5	Old fields	Brush – fair condition	
6	Brush and shrubland	Brush – good condition	
7	Orchards, vineyards, and nurseries	Woods/grass combination (good condition)	
8	Forests and wooded areas	Woods – good condition	
9	Wetlands	Hydric soils	Assume CN = 98 (zero recharge potential)
10	Beaches and mines	Dry soils	Assume CN = 30 (maximum recharge potential)
11	Water	Water	Excluded from the analysis

The major factors affecting the curve numbers are the hydrologic soil group (HSG), cover type, treatment, and antecedent runoff condition (ARC). Assuming an average ARC, and that all impervious areas flow directly into the drainage system, Tables 2.2a, b, and c in the NRCA TR-55 (USDA, 1986) were used to estimate the CN for each LULC/soil combination. The hydrologic soil group (HSG) classification for each soil was obtained from the USDA soil maps, and the cover type classification was based on the groupings seen in Table A.2. Since groups 1-3 contained impervious areas, their CNs were adjusted based on Figure 2-3 in TR-55. Due to the lack of information regarding the type of cover and treatment in agricultural areas, an average CN was used for group 4 soils (i.e. by averaging all the CNs in Table 2-2b of TR-55 for each HSG). For hydric soils in group 9, a CN of 98 was assumed to indicate zero recharge potential, whereas a minimum CN of 30 was assumed for dry soils in group 10 to indicate maximum recharge potential. Group 11 was not included in the analysis.

Soils Data The surface soil layering information was derived from the USDA soils database. In addition to physical and chemical details, the database contains informa-

tion on soil type, number of layers, layer depths, and soil texture classifications. A HELP3 soil code was assigned to each soil type based on the soil texture classification, which was then entered into the HELP3 model along with the layer information. For the recharge analysis, the total depth of the layering was assumed to be 10 ft, with the bottom layer consisting of coarse to medium sand.

Evapotranspiration The evapotranspiration data for the model is summarized in Table A.3. The average evaporative zone depths for the given soils and cover types were estimated by taking 50 per cent of the maximum root depths given in Table 2 in Appendix 7 of NJGS GSR-32 (Charles et al., 1993). The other evapotranspiration data was obtained from HELP3 for Edison, New Jersey. A latitude adjustment to 40.00 degrees was also applied to correspond with the location of the modelling domain.

Table A.3: Evapotranspiration data used in HELP3

Parameter	Value
Average wind speed *	10.2 MPH
Growing season start day *	109 (April 19th)
Growing season end day *	299 (October 26th)
First quarter relative humidity *	64.00%
Second quarter relative humidity *	61.00%
Third quarter relative humidity *	66.00%
Fourth quarter relative humidity *	68.00%
Latitude adjustment to 40 degrees	---

* values were generated by HELP3 for Edison, New Jersey.

Leaf Area Index HELP3 requires a value for the maximum leaf area index (LAI) to calculate transpiration rates for the vegetative cover. These values range from 0 for bare ground to 5 for maximum vegetal leaf coverage and were assigned based on the LULC group codes. For example, the LAI for agricultural areas was assumed to be 2.

Weather Data A range of weather data was also essential for the HELP3 model. The solar radiation values were generated synthetically by HELP3 for Edison, New Jersey

(with a latitude adjustment to 40.00), whereas actual daily precipitation and daily average temperature values from January 1970 to December 2000 at Toms River were used in the HELP3 model. It was assumed that each precipitation event was spatially constant across the modelling domain.

Results The daily recharge values from HELP3 were summed to obtain a temporally varying monthly recharge distribution for each LULC/soil combination. Based on the areal contribution of each combination within a finite difference grid block, an areally weighted average was then used to calculate the actual recharge distribution. For grid blocks containing water, the recharge contribution from water was always assumed to be zero thereby lowering the overall recharge in the block.

Figures 3.5 and 3.6 of this thesis show typical recharge distributions for a dry (May 1981) and a wet (May 1989) month, respectively, across the entire modelling domain. It is clear that the recharge varies significantly both spatially, and temporally in the study area. The overall spatial and temporal recharge average for the 31 year time period across the entire domain was approximately 17.9 in/year.

A.1.4 Hydraulic Conductivity Distribution

The location of the wells and borings used in the estimation of the hydraulic conductivity distribution are indicated by the circular markers in Figure 3.11 of this thesis. The wells shown in this figure include the wells and borings for the Reich Farm Superfund Site, the wells/borings for the Dover Township Municipal Landfill, UWTR wells/borings as well as domestic, industrial and production wells in the NJDEP database that had a depth greater than 100 ft. Many logs were not used as they were either located outside the modelling domain, or there was insufficient information to locate the wells.

The hydraulic conductivity at the location of well j was determined using the following formula:

$$K_j = \frac{\sum_{i=1}^{n} b_{ij} K_{ij}}{\sum_{i=1}^{n} b_{ij}} \tag{A.1}$$

where K_j is the horizontal hydraulic conductivity at location j, K_{ij} is the hydraulic conductivity of unit i in a well j log, and b_{ij} is the thickness of unit i in a well j log. The qualitative description of lithology as provided on the well log was assigned a material category. Each material category was then assigned a hydraulic conductivity value as shown in Table A.4. Calibration involved adjusting the assigned hydraulic conductivity for the material categories. Trial 7 in Table A.4 presents the values for hydraulic conductivity which were used in this report.

The point values for K_j were used to estimate a hydraulic conductivity distribution using a kriging interpolation algorithm. The kriging parameters included: no drift, an exponential variogram, and a nugget variance of 225 $(\text{ft/day})^2$. The resulting two-dimensional interpolated field was used to assign a hydraulic conductivity to each grid block within a layer, where MODFLOW model layers 1 through 3 are identical, with a distribution as shown in Figure 3.11 of this thesis. The hydraulic conductivity tends to decrease eastward as one approaches the coast. Layer 4 was assigned a constant horizontal hydraulic conductivity of 0.283 ft/day. The vertical hydraulic conductivity was determined using an anisotropy ratio of 20:1, horizontal:vertical.

* Coarse to Fine Gravel and Silty GRAVEL categories were only applied to Trial 7 as Trials 1 through 6 predate the inclusion of these two categories within the lithology database.

A.1.5 Pumping Wells

The transient groundwater flow was simulated using monthly averaged flow rates for the UWTR Cohansey wells (Wells 20, 22, 24, 26, 26B, 28, 29, and 44). Plots of the temporal variation in pumping rates for each of the wells are shown in Figure 3.17 and Figure 3.18 of this thesis.

Table A.4: Material categories and assigned values for horizontal hydraulic conductivity

Code	Description	Hydraulic Conductivity [ft/day]						
		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7*
c_g	Coarse GRAVEL	600	600	600	275	400	250	350
cm_g	Coarse to Medium GRAVEL	475	475	475	250	325	250	350
m_g	Medium GRAVEL	425	425	425	225	250	250	350
g	GRAVEL	300	300	425	225	250	200	275
cf_g	Coarse to Fine GRAVEL	NA	NA	NA	NA	NA	NA	275
mf_g	Medium to Fine GRAVEL	250	250	325	200	225	200	275
f_g	Fine GRAVEL	200	200	250	175	200	200	275
gs	Sandy GRAVEL	100	125	175	120	175	150	200
c_s	Coarse SAND	150	150	200	150	150	150	200
cm_s	Coarse to Medium SAND	64	125	175	90	125	150	200
m_s	Medium SAND	32	100	150	32	100	100	150
s	SAND	32	100	150	32	100	100	150
cmf_s	Coarse to Medium to Fine SAND	22	75	110	22	75	100	75
cf_s	Coarse to Fine SAND	16	50	60	16	50	25	75
mf_s	Medium to Fine SAND	8	25	45	8	25	25	25
f_s	Fine SAND	4	10	15	4	10	25	25
gm	Silty GRAVEL	NA	NA	NA	NA	NA	NA	6.11e-01
sm	Silty SAND	6.11e-01	6.11e-01	6.11e-01	6.11e-01	6.11e-01	6.11e-01	6.11e-01
ms	Sandy SILT	1.32e-01	1.32e-01	1.32e-01	1.32e-01	1.32e-01	1.32e-01	1.32e-01
m	SILT	2.83e-02	2.83e-02	2.83e-02	2.83e-02	2.83e-02	2.83e-02	2.83e-02
gc	Clayey GRAVEL	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03
sc	Clayey SAND	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03
mc	Clayey SILT	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03
cg	Gravely CLAY	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04
cs	Sandy CLAY	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04
cm	Silty CLAY	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04
c	CLAY	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04	2.83e-04
r	rock	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03
h	hardpan	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03	2.84e-03
p	peat	2.83e-01	2.83e-01	2.83e-01	2.83e-01	2.83e-01	2.83e-01	2.83e-01

A.1.6 Initial Condition

The transient simulation commences in October 1971. The steady-state recharge distribution was determined as the average recharge for each cell for the period from October 1971 to December 2000. The initial condition was established based on steady-state groundwater flow (see Figure 3.19 of this thesis) with only UWTR Well 20 pumping at a rate of 48,988 ft^3/day, which represents the average production rate for the period October 1971 to March 1972.

A.2 Calibration

Water level measurements for 101 wells were used in the *Sykes and Normani* [2002] MODFLOW calibration process. A database was created to facilitate the addition of future water level measurements. The database includes both manual measurements, and water elevation readings from nine data-loggers installed in monitoring wells between the Reich Farm Superfund Site and the UWTR Parkway Wellfield. The database includes over 410,000 discrete measurements, mostly from the data logger measurements.

A model calibration requires sound observed water elevations. As part of a data quality check, a large portion of monitoring wells installed during the mid 1980s (NUS, Ebasco, CH2M-Hill) were found to have different top of casing elevations, depending on the source documents. Over 40% of the approximately 1,750 manual measurements were corrected as a result. Some top of casing elevations had been incorrect for nearly 10 years as the result of using the wrong elevations or as the result of transcription errors. In some cases, well casings had been modified, but not resurveyed until several years later.

A comparison of water levels recorded at neighboring monitoring wells at Reich Farm showed large vertical gradients. A review of the lithology associated with those wells did not indicate a geological means for the existence of those gradients. A similar situation was noticed in various well pairs. A comparison of water levels based on when the wells were installed (NUS, Ebasco, and Malcolm Pirnie monitoring well generations) indicated

a consistent trend. Both NUS and Ebasco were EPAs contractors for the Reich Farm site. Due to the discrepancies identified and the resulting confusion as to which recorded data was considered true, it was decided to resurvey all the monitoring wells at Reich Farm as well as a few other monitoring wells that formed part of the network.

It was discovered that the NUS and Ebasco wells at Reich Farm had been surveyed to differing elevation benchmarks. Neither of these elevation benchmarks is equivalent to the present day benchmark. The top of casing for the NUS installed wells (MW-1K to MW-10) were too low by an average of 0.55 ft, while the Ebasco installed wells (MW-12 to MW-23) were too low by an average of 1.37 ft.

Coordinate locations of monitoring wells were also checked. In some cases the wells were missurveyed and found to be several hundred feet from their true locations, while in other cases, coordinates in NAD27 were confused with NAD83, thereby resulting in a horizontal difference of over 100 ft. Transcription or typographical errors on the part of the surveyor were also found and corrected.

Time series plots helped to find anomalous data (refer to *Sykes and Normani* [2002]). For example, data loggers continued to record water levels at pre-specified times, even though they had been removed from the monitoring well for downloading. Abrupt changes in recorded temperature and/or water levels are a key indication. Other issues related to data loggers included discrepancies in water levels between readings prior to logger removal and after logger return to the well.

Water level measurements are inconsistent and / or nonrepresentative for some monitoring wells at the Dover Landfill. This effect can be seen by reviewing the water level hydrographs for the Dover Landfill wells (not reproduced in this thesis, refer to *Sykes and Normani* [2002]). For example, wells MW-2 and MW-6 both show significant changes in water levels over time. These inconsistencies may be due to incorrect measurements and / or clogged well screens. Early water level measurements (within a year or two of the installation date of the wells) and recent water level measurements (June 1999) show more favorable comparisons with simulated water levels. Water level measurements for MW-1, MW-2, MW-3, MW-4, MW-5s, and MW-6 that predate May 1999 are not included in the

scatter plot in Figure 3.24 of this thesis. Water level measurements after May 1999 were collected and provided by Dan Raviv Associates.

Several model input parameters were adjusted during the calibration phase to provide a better match between simulated and observed water elevations. Conductivities for surface water bodies were adjusted to account for the topographic control of the water table near those surface water bodies.

The hydraulic conductivity distribution for MODFLOW model layers 1 through 3 were adjusted by first estimating the hydraulic conductivity for each material category as shown in Table A.4. The weighted average hydraulic conductivity at each well or boring location is calculated. A hydraulic conductivity field is then interpolated using a kriging algorithm, and then each grid block in the domain is assigned a hydraulic conductivity based on the interpolated field. This process was repeated several times until a good match was obtained between simulated and observed heads as shown on the scatter plot.

Figure 3.24 of this thesis shows a scatter plot for all observed water level measurements for all time; the results are similar to that obtained for the MODFLOW calibration. Multiple water level measurements in a month were averaged to provide one reading per month. The upper and lower bands represent plus or minus 5 ft. Table A.5 , with data from the calibration of *Sykes and Normani* [2002], represents the number of data points that are in each year and well category. A similar table was developed for the verification of the groundwater flow model calibration of this thesis (refer to Table 3.2 and Figure 3.24). Table A.6 represents the averages of the residuals (simulated heads minus observed heads) for various years and monitoring well categories. A positive number indicates that the simulated heads are (in an average sense) greater than the observed heads, while a number near zero indicates that the simulated heads are (in an average sense) nearly equal to the observed heads. Table A.7 is similar to Table A.6, except that the averages of the absolute value of the residuals are calculated and are used to illustrate the variation in the residuals. Numbers closer to zero indicate less scatter about the match line, and a better fit between simulated and observed values. The overall average residual of all 1,483 scatter points is 0.044 ft, while the average of the absolute value of the residuals is 1.724 ft.

Table A.5: Count of scatter plot residuals by category and year

Year	Category							
	All	CH2M	DTML	EBASCO	MPI	NUS	Other	UWTR
1986	15					15		
1987	49		1	24		20	4	
1988	48		5	23		20		
1989	15	15						
1990	20	5		6		9		
1991	56	5		22	9	20		
1992	44	1		13	22	8		
1993	54	6		12	22	14		
1994	30	3		8	12	7		
1997	346	22		45	186	49		44
1998	299	22		35	157	41		44
1999	221	18	16	20	101	30		36
2000	286	19	88	18	91	29		41
All Years	1483	116	110	226	600	262	4	165

Table A.6: Average of scatter plot residuals in feet by category and year

Year	Category							
	All	CH2M	DTML	EBASCO	MPI	NUS	Other	UWTR
1986	-1.53					-1.53		
1987	-1.68		0.00	-1.84		-1.61	-1.50	
1988	-1.98		-0.80	-2.31		-1.91		
1989	-1.28	-1.28						
1990	-4.16	-4.87		-3.83		-3.97		
1991	-1.52	-1.72		-1.67	-1.78	-1.18		
1992	-0.78	-2.51		-0.43	-1.13	-0.19		
1993	0.29	-0.21		0.88	-0.30	0.93		
1994	-3.99	-4.15		-3.74	-4.45	-3.43		
1997	-1.49	-1.46		-0.54	-2.04	-0.53		-1.26
1998	2.09	2.54		3.08	1.15	3.14		3.42
1999	0.61	0.84	0.74	0.89	0.03	1.02		1.56
2000	1.22	1.47	0.87	1.95	0.69	2.01		2.17
All Years	0.044	-0.014	0.771	-0.201	-0.390	0.102	-1.496	1.456

Table A.7: Average of absolute value of scatter plot residuals in feet by category and year

Year	Category							
	All	CH2M	DTML	EBASCO	MPI	NUS	Other	UWTR
1986	1.53					1.53		
1987	1.85		0.00	1.84		1.61	3.53	
1988	1.98		0.80	2.31		1.91		
1989	1.31	1.31						
1990	4.16	4.87		3.83		3.97		
1991	1.75	1.72		1.67	1.78	1.82		
1992	0.83	2.51		0.43	1.17	0.34		
1993	1.01	0.88		0.94	1.04	1.07		
1994	3.99	4.15		3.74	4.45	3.43		
1997	1.75	1.67		0.93	2.13	1.01		1.85
1998	2.39	2.54		3.08	1.73	3.14		3.42
1999	1.03	0.87	0.83	0.91	0.93	1.02		1.57
2000	1.36	1.47	0.91	1.95	1.08	2.01		2.17
All Years	1.724	1.803	0.884	1.800	1.629	1.810	3.530	2.288

As can be seen in Table A.6 and A.7, some years provide a better fit to the data than in other years. For example, the years 1999 and 2000 provide the best fit to the observed data in the mid to late 1990s. To facilitate the calibration of the groundwater flow model, data loggers were installed in selected wells of the Reich Farm study. Figures A.1 to A.6 present a comparison of the transient MODFLOW simulation results to the observed water levels from data loggers.

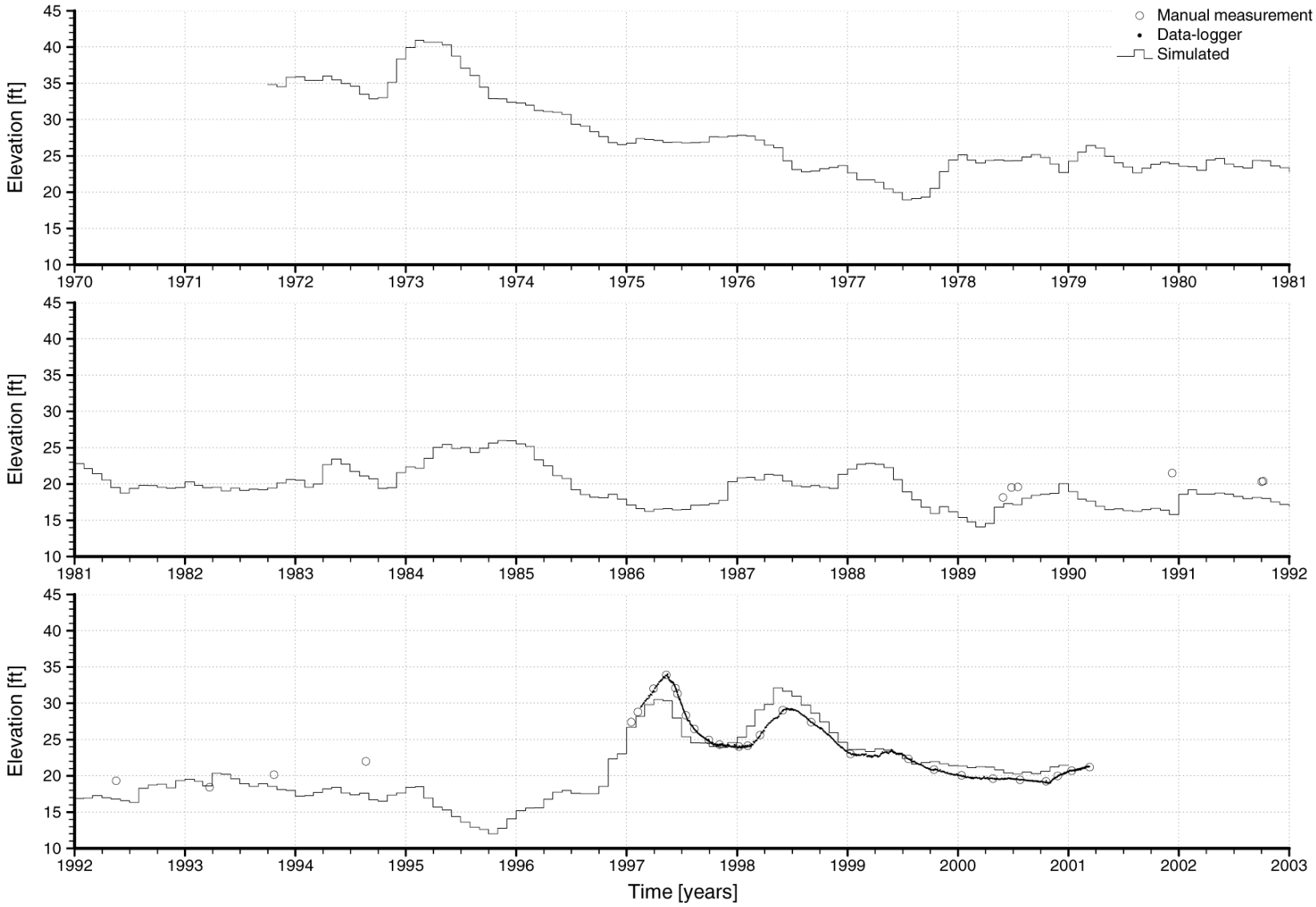

Figure A.1: Observed and simulated water levels for CHMW-4 (CH2M).

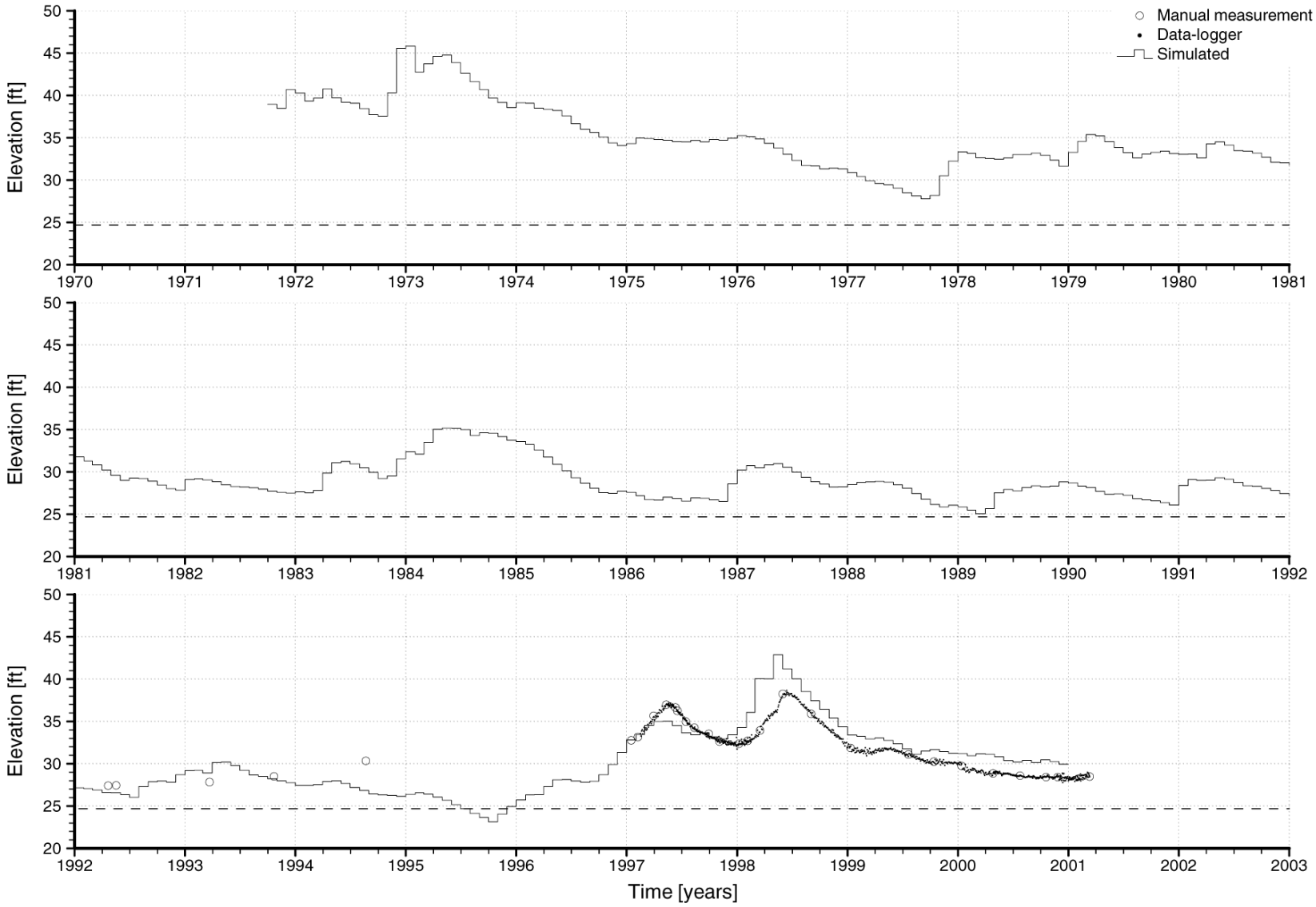

Figure A.2: Observed and simulated water levels for MP-7 (MPI).

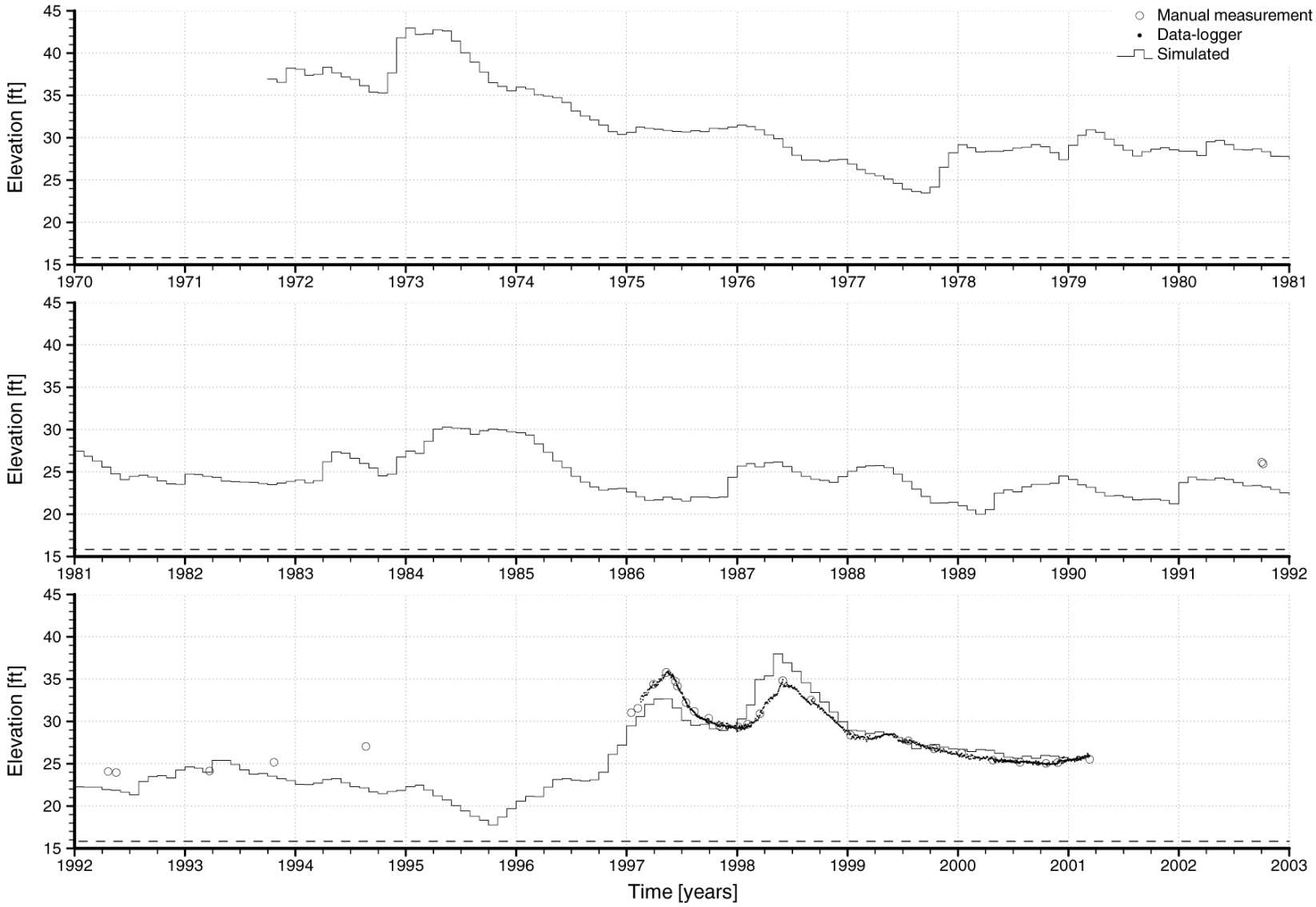

Figure A.3: Observed and simulated water levels for MP-8 (MPI).

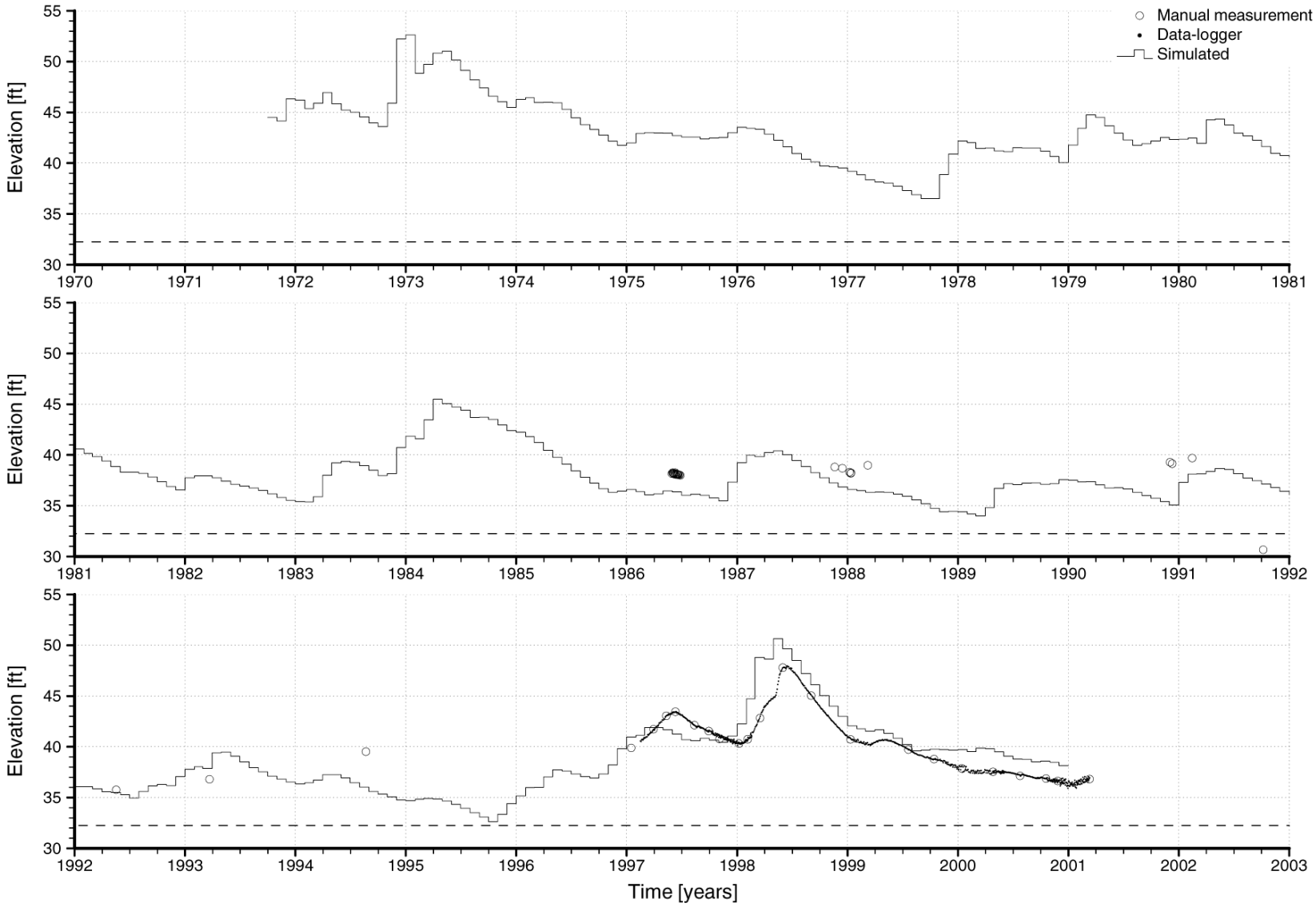

Figure A.4: Observed and simulated water levels for MW-4S (NUS).

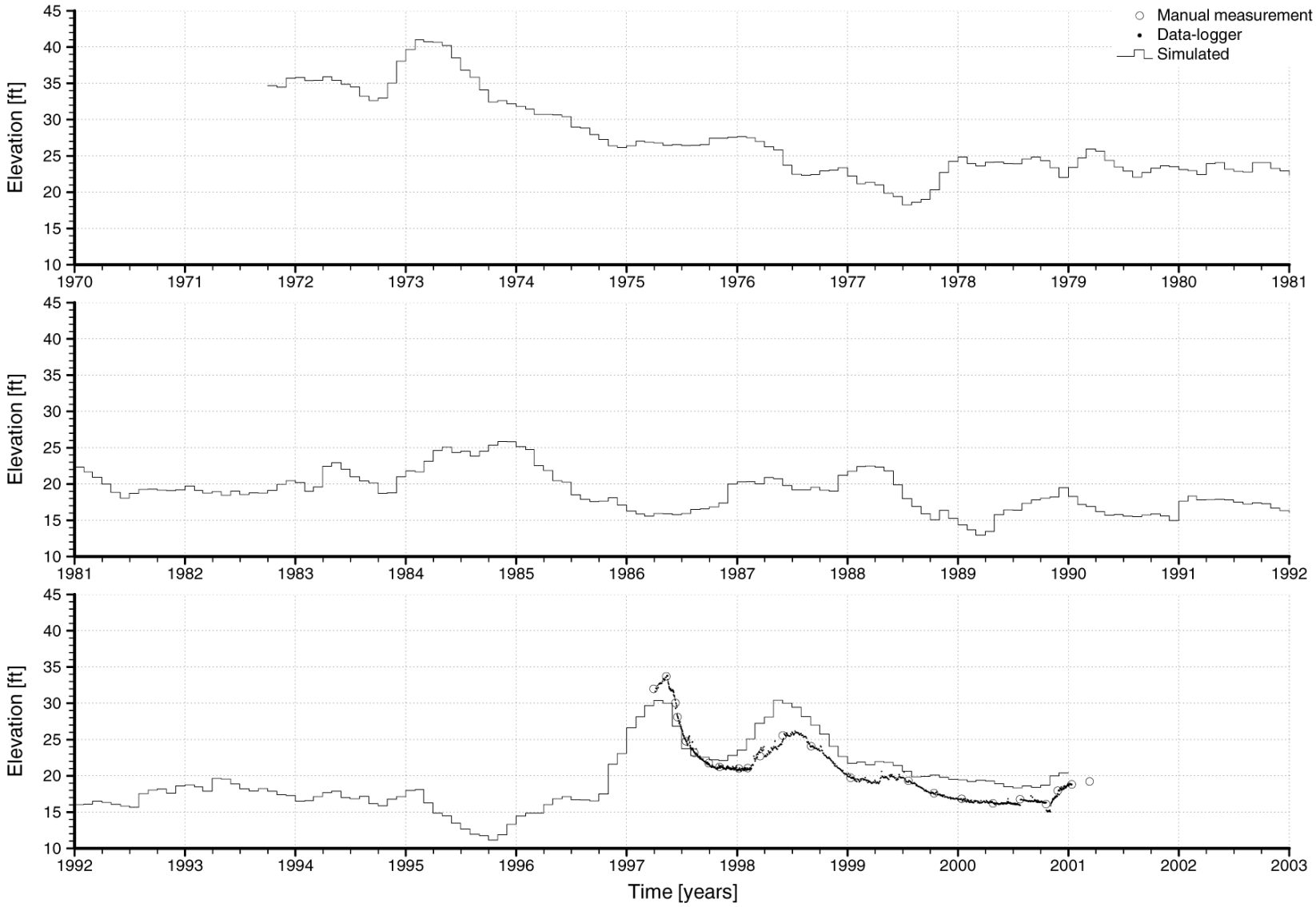

Figure A.5: Observed and simulated water levels for OW-2 (UWTR).

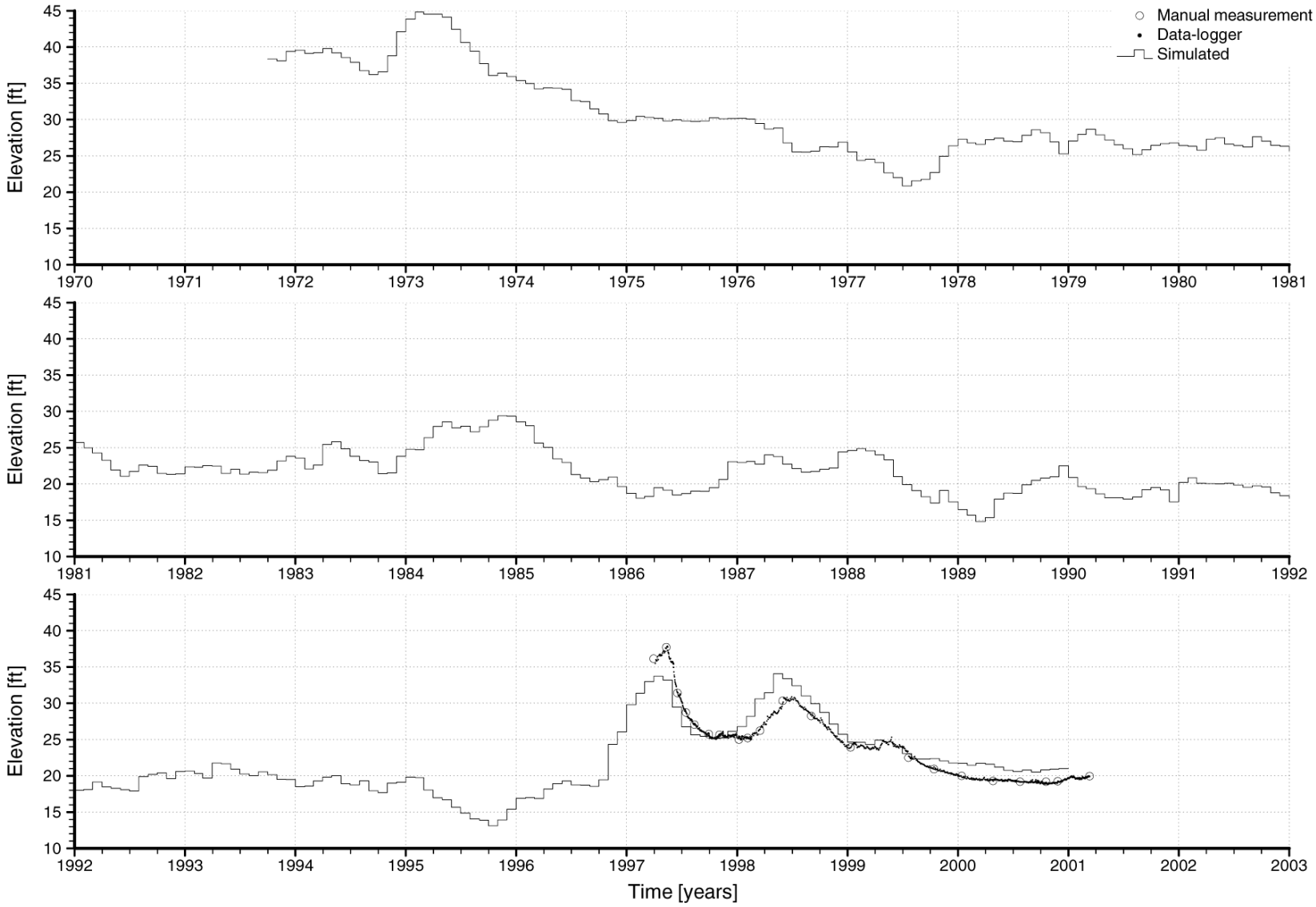

Figure A.6: Observed and simulated water levels for TRW-MWD (UWTR).

A.3 Conclusions

The groundwater flow model described in the *Sykes and Normani* [2002] report and the data upon which is it based have been updated in the following ways:

- The simulation period has been extended to the end of 2000, thereby requiring additional well pumping rates, temperature and precipitation records, water level readings and data logger measurements.
- Approximately 100 well records were added to the wells database and 19 wells were added to the water levels database.
- Software was written for the lithologic visualization of well boring logs that were used in the development of the spatially variable hydraulic conductivity field.
- Interpolation of the 3-dimensional groundwater flow field to represent the physical location of a well screen and the water level that would result at that location. This approach was required to account for the vertical flow gradients that exist at the Recharge Area.

The *Sykes and Normani* [2002] report presents the Reich99a calibrated groundwater flow model for the Cohansey-Kirkwood aquifer system with a domain that extends from Barnegat Bay to the east, the Toms River to the west, the Toms River estuary to the south and a surface water divide to the north. The modelling domain includes both the Reich Farm Superfund Site and the Dover Township Landfill site. The Reich99a model represents a refinement of the Reich99 model. Results from the Reich99a model are consistent with those of earlier modelling efforts which used the Reich99 and Reich93 models. Variations between the simulated and observed water levels can be due to the conceptual model and inconsistencies or errors in the observed data. Within the model domain, several data inconsistencies or errors were discovered and rectified:

- Checking and correcting numerous data entry, surveying, and transcription errors related to water levels, well locations, and well casing elevations.

- Checking and correcting data entry errors associated with the daily temperature and precipitation records obtained from the National Climatic Data Centre.
- After resolution of data errors and other inconsistencies, the overall average residual between simulated and observed water levels for the 1,483 scatter points is 0.044 feet, while the average of the absolute value of the residuals is 1.724 feet.

The highly dynamic behaviour of the groundwater flow system necessitated the development of a transient groundwater flow model. A steady-state groundwater flow model of the study area is inadequate for describing the evolution and behaviour of the groundwater flow system from the early 1970s to the end of 2000. A spatially and temporally varying recharge methodology was developed which uses actual daily temperature and precipitation data from 1970 to the end of 2000. The use of this data was essential, for example, in capturing the warmer than normal winter of early 1998, which lead to higher than normal recharge, as evidenced by measured water levels. Estimated travel times from Reich Farm based on average water particle paths must be based on a transient groundwater flow system, while a steady-state flow system can be used to estimate pumping well capture zones. The use of a steady-state groundwater flow system with present day pumping rates at the UWTR wells will result in the underestimation of average water particle travel times from Reich Farm as compared to those obtained using a transient model that captures the temporal evolution of the groundwater flow system.

Bibliography

Abriola, L., and G. Pinder (1985), A multiphase approach to the modelling of porous media contamination by organic compounds,1. equation development, *Water Resour. Res.*, *21*(1), 11–18.

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

AquaResource Inc. (2007), Integrated water budget report - grand river watershed, prepared for the grand river conservation authority, *Tech. rep.*, Ontario.

Aral, M., J. Guan, and M. Malia (2001), Identification of contaminant source location and release history in aquifers, *J. Hydrol. Eng.*, *6*(3), 225–234, doi:10.1061/(ASCE)1084-0699(2001)6:3(225).

Atmadja, J., and A. Bagtzoglou (2001a), Pollution source identification in heterogeneous porous media, *Water Resour. Res.*, *37*(8), 2113–2125.

Atmadja, J., and A. Bagtzoglou (2001b), State of the art report on mathematical methods for groundwater pollution source identification, *2*(3), 205–214, doi:10.1006/enfo.2001.0055.

Babbar, M., and B. Minsker (2006), Groundwater remediation design using multiscale genetic algorithms, *J. Water Resour. Plng. and Mgmt.*, *132*(5), 341–350.

Bagtzoglou, A., and J. Atmadja (2003), Marching-jury backward beam equation and quasi-reversibility methods for hydrologic inversion: Application to contaminant plume spatial distribution recovery, *Water Resour. Res.*, *39*(2), 1038, doi:10.1029/2001WR001021.

Bagtzoglou, A., A. Tompson, and D. Dougherty (1991), Probabilistic simulation for reliable solute source identification in heterogeneous porous media, in *Water Resources Engineering Risk Assessment*, edited by J. Ganoulis, pp. 189–201, Springer-Verlag, New York.

Bagtzoglou, A., D. Dougherty, and A. Tompson (1992), Application of particle methods to reliable identification of groundwater pollution sources, *Water Resour. Manage.*, *6*, 15–23.

Bear, J. (1988), *Dynamics of Fluids in Porous Media*, dover ed., Dover Publications Inc.

Benker, E., G. Davis, and D. Barry (1997), Factors controlling the distribution and transport of trichloroethene in a sand aquifer: hydrogeology and results of an in situ transport experiment, *J. Hydrol.*

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, *6*, 279–298.

Blanton, B. (1995), *DROG3D: User's Manual for 3-Dimensional Drogue Tracking on a Finite Element Grid with Linear Finite Elements*, Program in Marine Sciences, University of North Carolina, Chapel Hill, NC.

Boyd, S., and L. Vandenberghe (2003), *Convex Optimization*, Cambridge University Press, New York.

Brooks, R. J., and A. Corey (1964), Hydraulic properties of porous media, *Hydrology paper 3*.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol J.*, *13*, 206–222, doi:10.1007/s10040-004-0404-7.

Carroll, D. (1996), Genetic algorithms and optimizing chemical oxygen-iodine lasers, in *Developments in Theoretical and Applied Mechanics, Pasadena, Vol. XVIII, eds*, pp. 411–424, School of Engineering, The University of Alabama.

Carsel, R., and R. Parrish (1988), Developing joint probability distributions of soil water retention characteristics, *Water Resour. Res.*, *24*(5), 755–769.

Cawlfield, J., and M.-C. Wu (1993), Probabilistic sensitivity analysis for one-dimensional reactive transport in porous media, *Water Resour. Res.*, *29*(3), 661–672.

Cooley, R. (1971), A finite difference method for unsteady flow in variably saturated porous media: appliaction to a single pumping well, *Water Resour. Res.*, *7*(6), 1607–1625.

David, R., D. Sweeney, and T. Williams (1991), *Introduction to Statistics: Concepts and Applications*, West Pub. Co.

Dettinger, M., and J. Wilson (1981), First order analysis of uncertainty in numerical models of groundwater flow part 1. mathematical development, *Water Resour. Res.*, *17*(1), 149–161.

Doherty, J., and J. Johnston (2003), Methodologies for calibration and predictive analysis of a watershed model, *J. Am. Water Resour. Assoc.*, *39*(2), 251–265.

Domenico, P., and M. Mifflin (1965), Water from low-permeability sediments and land subsidence, *Water Resour. Res.*, *1*(4), 563–576.

Dou, C., W. Woldt, I. Bogardi, and M. Dahab (1995), Steady state groundwater flow simulation with imprecise parameters, *Water Resour. Res.*, *31*(11), 2709–2719.

Dou, C., W. Woldt, M. Dahab, and I. Bogardi (1997), Transient ground-water flow simulation using a fuzzy set approach, *Ground Water*, *35*(2), 205–215.

Fenton, G., and D. Griffiths (1993), Statistics of block conductivity through a simple bounded stochastic medium, *Water Resour. Res.*, *29*(6), 1825–1830.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the glue approach, *Water Resour. Res.*, *32*(7), 2161–2173.

Freeze, R., and J. Cherry (1979), *Groundwater*, Prentice-Hall Inc., New Jersey.

Friedel, M. (2005), Coupled inverse modeling of vadose zone water, heat, and solute transport: calibration constraints, parameter nonuniqueness, and predictive uncertainty, *J. Hydrol.*, *312*, 148–175.

Gelhar, L. (1986), Stochastic subsurface hydrology: From theory to practice, *Water Resour. Res.*, *22*(9), 135S–145S.

Gelhar, L., and C. Axness (1983), Three-dimensional stochastic analysis of macrodispersion in aquifers, *Water Resour. Res.*, *19*(1), 161–180.

Gelhar, L., C. Welty, and K. Rehfeldt (1992), A critical review of data on fieldscale dispersion in aquifers, *Water Resour. Res.*, *28*(7), 1955–1974.

Geller, J., and J. Hunt (1993), Mass transfer from nonaqueous phase organic liquids in water-saturated porous media, *Water Resour. Res.*, *29*(4), 833–845.

Gibbons, R. D. (1994), *Statistical Methods for Groundwater Monitoring*, illustrated ed., Wiley-IEEE.

Gillham, R. (1998), Sorption Tests for TCE and Trimer on Soil from the Vicinity of the Reich Farm Toms River, NJ, *Tech. rep.*, Institute for Groundwater Research, Waterloo, Ontario.

Goldberg, D. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley.

Gorelick, S., B. Evans, and I. Remson (1983), Identifying sources of groundwater pollution: an optimization approach, *Water Resour. Res.*, *19*(3), 779–790.

Hamed, M. (2000), Stochastic modelling concepts in groundwater and risk assessment: Potential application to marine problems, *6*(2), 125–132.

Hamed, M., J. Conte, and P. Bedient (1995), Probabilistic screening tool for groundwater contamination assessment, *J. Environ. Eng.*, *121*(11), 767–775.

Hamed, M., P. Bedient, and J. Conte (1996a), Numerical stochastic analysis of groundwater contaminant transport and plume containment, *J. Contam. Hydrol.*, *24*(1), 1–24.

Hamed, M., P. Bedient, and C. Dawson (1996b), Probabilistic modelling of aquifer heterogeneity using reliability methods, *Adv. Water Resour.*, *19*(5), 277–295.

He, P., and J. Sykes (1996), On the spatial-temporal averaging method for modeling transport in porous media, *Transport in Porous Media*, *21*, 1–51.

Helton, J., and F. Davis (2003), Latin hypercube sampling and the propagation of uncertainty in analysis of complex systems, *Reliab. Engng. Syst. Saf.*, *81*(1), 23–69.

Hill, M., and C. Tiedeman (2007), *Effective Groundwater Model Calibration, with Analysis of Sensitivities, Predictions, and Uncertainty*, Wiley, New York.

Hill, M., R. Cooley, and D. Pollock (1998), A controlled experiment in ground water flow model calibration, *Ground Water*, *36*(3), 520–535.

Iman, R., and W. Conover (1982), A distribution-free approach to inducing rank correlation among input variables, *Commun. Statist.-Simula. Computa.*, *11*, 311–334.

Imhoff, P., P. Jaffe, and G. Pinder (1994), An experimental study of complete dissolution of a nonaqueous phase liquid in saturated porous media, *Water Resour. Res.*, *30*(2), 321–332.

Isukapalli, S. (1999), Uncertainty analysis of transport-transformation models, Ph.D. thesis, Chemical and Biochemical Engineering. The State University of New Jersey, New Brunswick, NJ.

Ivanov, J., R. Miller, J. Xia, D. Steeples, and C. Park (2005), The inverse problem of refraction travel times, part I:types of geophysical nonuniqueness through minimization, *Pure appl. geophys.*, *19*, 447–459, doi:10.1007/s00024-004-2615-1.

James, A., and C. Oldenburg (1997), Linear and monte carlo uncertainty analysis for subsurface contaminant transport simulation, *Water Resour. Res.*, *33*(11), 2495–2508.

Jiang, Y., and A. Woodbury (2006), A full-bayesian approach to the inverse problem for steady-state groundwater flow and heat transport, *Geophys. J. Int.*, *167*, 1501–1512, doi:10.1111/j.1365-246X.2006.03145.x.

Jyrkama, M. (2003), A methodology for estimating groundwater recharge, Ph.D. thesis, Department of Civil Engineering, University of Waterloo, Canada.

Jyrkama, M., and J. Sykes (2006), Sensitivity and uncertainty analysis of the recharge boundary condition, *Water Resour. Res.*, *42, W01404*, doi:10.1029/2005WR004408.

Jyrkama, M., J. Sykes, and S. Normani (2002), Recharge estimation for transient ground water modelling, *Ground Water*, *40*(6), 638–648.

Kitanidis, P. (1992), An analysis of macrodispersion through volume-averaging: Moments equations, *Stochastic Hydrol. Hydraul.*, *6*, 5–25.

Kitanidis, P. (1995), Quasi-linear geostatistical theory for inversing, *Water Resour. Res.*, *31*(10), 2411–2419.

Kitanidis, P. (1997), *Introduction to geostatistics: applications to hydrogeology*, Cambridge Univ. Press, New York.

Kitanidis, P., and K. Shen (1996), Geostatistical interpolation of chemical concentration, *Adv. Water Resour.*, *19*(6), 369–378.

Krishnakumar, K. (1989), Micro-genetic algorithms for stationary and non-stationary function optimization, in *SPIE: Intelligent Control and Adaptive Systems, Vol. 1196*, Philadelphia, PA.

Kullback, S. (1959), *Information Theory and Statistics*, John Wiley, New York.

Kunstmann, H., W. Kinzelbach, and T. Siegfried (2002), Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modelling, *Water Resour. Res.*, *38*(4), doi:10.1029/2000WR000022.

Lattes, R., and J. Lions (1969), *The method of quasi-reversibility; applications to partial differential equations*, American Elsevier, New York.

Li, S., D. Mclaughlin, and H. Liao (2004), The accuracy of stochastic perturbation solutions to subsurface transport problems, *Adv. Water Resour.*, *27*(1), 47–56.

Liu, C., and W. Ball (1999), Application of inverse methods to contaminant source identification from aquitard diffusion profiles at dover afb, delaware, *Water Resour. Res.*, *35*(7), 1975–1985.

Mackay, D., W. Shiu, A. Maijanen, and S. Feenstra (1991), Dissolution of non-aqueous phase liquids in groundwater, *J. Contam. Hydrol.*, *8*, 23–42.

Madadi, R., and C. Balaji (2008), Optimization of the location of multiple discrete heat sources in a ventilated cavity using artificial neural networks and micro genetic algorithm, *International Journal of Heat and Mass Transfer*, *51*, 22992312, doi: 10.1016/j.ijheatmasstransfer.2007.08.033.

Madsen, H., S. Krenk, and N. Lind (1986), *Methods of structural safety*, Prentice-Hall, Englewood Cliffs, New Jersy.

Mahar, P., and B. Datta (1997), Optimal monitoring network and ground-water-pollution source identification, *123*, 199–207.

Mahar, P., and B. Datta (2000), Identification of pollution sources in transient groundwater systems, *14*(3), 209–227.

Mahar, P., and B. Datta (2001), Optimal identification of ground-water pollution sources and parameter identification., *J. Water Resour. Plann. Manage.*, *127*, 20–29.

Mantoglou, A., and L. Gelhar (1987), Stochastic modelling of large-scale transient unsaturated flow systems, *Water Resour. Res.*, *23*(1), 37–46.

Mayer, A., and C. Huang (1999), Development and application of a coupled-process parameter inversion model based on the maximum likelihood estimation method, *Adv. Water Resour.*, *22*(8), 841–853.

McDonald, M., and A. Harbaugh (1988), A modular three-dimensional finite-difference ground-water flow model, techniques of water resources investigation, *Tech. rep.*, Washington, DC.

McKay, M., W. Conover, and R. Beckman (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, *21*(2), 239–245.

McKelvie, J., S. Hirschorn, G. Lacrampe-Couloume, J. Lindstrom, J. Braddock, K. Finneran, D. Trego, and B. Lollar (2007), Evaluation of tce and mtbe in situ biodegradation: Integrating stable isotope, metabolic intermediate, and microbial lines of evidence, *Ground Water Monitoring & Remediation*.

McLaughlin, D., D. Llyod, and L. Townley (1996), A reassessment of the groundwater inverse problem, *Water Resour. Res.*, *32*(5).

Medina, A., and J. Carrera (1996), Coupled estimation of flow and solute transport parameters, *Water Resour. Res.*, *32*(10), 3063–3076.

Melching, C., and W. Bauwens (2001), Uncertainty in coupled nonpoint source and stream water-quality models, *J. Water Resour. Plann. Manage.*, *127*(6), 403–413.

Michalak, A. (2001), *Feasibility of contaminant source identification for property rights enforcement, in The Technology of Property Rights*, pp. 123–145, Rowman and Littlefield Publishers, Lanham, Maryland.

Michalak, A., and P. Kitanidis (2002), Application of bayesian inference methods to inverse modeling for contaminant source identification at gloucester landfill, canada, in *Computational Methods in Water Resources XIV*, edited by H. S.M., R. Schotting, W. Gray, and G. Pinder, pp. 1259–1266, Amsterdam, The Netherlands.

Michalak, A., and P. Kitanidis (2003), A method for enforcing parameter nonnegativity in bayesian inverse problems with an application to contaminant source identification, *Water Resour. Res.*, *39*(2), doi:10.1029/2002WR001480.

Michalak, A., and P. Kitanidis (2004a), Application of geostatistical inverse modelling to contaminant source identification at dover afb, delaware, *Journal of Hydraulic Research*, *42*(SPI), 9–18.

Michalak, A., and P. Kitanidis (2004b), Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modelling, *Water Resour. Res.*, *40*(W08302), doi:10.1029/2004WR003214.

Miller, C., M. Poirier-Mcneill, and A. Mayer (1990), Dissolution of trapped nonaqueous phase liquids: Mass transfer characteristics, *Water Resour. Res.*, *26*(11), 2783–2796.

Mishra, S., and J. Parker (1989), Effects of parameter uncertainty on predictions of unsaturated flow, *J. Hydrol.*, *108*, 19–33.

Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Advances in Water Resources*, *29*(4), 605–623.

Morrison, R. (2000a), Critical review of environmental forensic techniques: Part i, *1*(4), 157–173.

Morrison, R. (2000b), Critical review of environmental forensic techniques: Part ii, *1*(4), 175–195.

Myrabo, S. (1997), Temporal and spatial scale of response area and groundwater variation in till, *Hydrol. Process.*, *11*(14), 1861–1880.

Nambi, I., and S. Powers (2000), Napl dissolution in heterogeneous systems: An experimental study in a simplified heterogeneous system, *J. Contam. Hydrol.*, *44*(2), 161–184.

Nambi, I., and S. Powers (2003), Mass transfer correlations for nonaqueous phase liquid dissolution from regions with high initial saturations, *Water Resour. Res.*, *39*(2), doi: 10.1029/2001WR000667.

Neumann, S. (1973), Saturated-unsaturated seepage by finite elements, *ASCE J. Hydraul. Div.*, *99(HY12)*, 2233–2251.

Neupauer, R., and J. Wilson (1999), Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant, *Water Resour. Res.*, *35*(11), 3389–3398.

Neupauer, R., and J. Wilson (2001), Adjoint-derived location and travel time probabilities for a multidimensional groundwater system, *Water Resour. Res.*, *37*(6), 1657–1668.

Neupauer, R., and J. Wilson (2002), Backward probabilistic model of groundwater contamination in non-uniform and transient flow, *Adv. Water Resour.*, *25*(7), 733–746.

Neupauer, R., and J. Wilson (2003), Backward location and travel probabilities for a decaying contaminant in an aquifer, *J. Contam. Hydrol.*, *66*, 39–58.

Neupauer, R., and J. Wilson (2004), Forward and backward location probabilities for sorbing solutes in groundwater, *Adv. Water Resour.*, *27*, 689–705.

Neupauer, R., B. Borchers, and J. Wilson (2000), Comparison of inverse methods for reconstructing the release history of a groundwater contamination source, *Water Resour. Res.*, *36*(9), 2469–2475.

Nicholson, R., and M. Watt (1997), Simulation of ground-water flow in the unconfined aquifer system of the Toms River, Metedeconk River, and Kettle Creek Basins, New Jersey, *Tech. rep.*, United States Geological Survey, West Treton NJ, water-Resources Investigations Report 97-4066.

Normani, S. (1998), Methodology to assess risk and uncertainty of drinking water contaminant exposure, Master's thesis, Department of Civil Engineering, University of Waterloo, Canada.

Normani, S., J. Sykes, and M. Jyrkama (2003), Challenges and solutions in the development of a forensic transient groundwater flow model: the Toms River N.J. Cancer Cluster, in *Proceedings of MODFLOW and MORE 2003: Understanding through Modelling*.

Pankow, J., and J. Cherry (1996), *Chlorinated Solvents and Other DNAPLs in Groundwater*, eds ed., Portland, Oregon: Waterloo Press.

Plumb, O., and S. Whitaker (1988), Dispersion in heterogeneous porous media, 1. local volume averaging and large-scale averaging, *Water Resour. Res.*, *24*(7), 913–926.

Portielje, R., T. Hvitved-Jacobsen, and K. Schaarup-Jensen (2000), Risk analysis using stochastic reliability methods applied to two cases of deterministic water quality models, *34*(1), 153–170.

Powers, S., C. Loureiro, L. Abriola, and W. Weber (1991), Theoretical study of the significance of nonequilibrium dissolution of nanaqueous phase liquids in subsurface systems, *Water Resour. Res.*, *27*(4), 463–477.

Powers, S., L. Abriola, and W. Weber (1992), An experimental investigation of nonaqueous phase liquid dissolution in saturated subsurface systems: Steady state mass transfer rates, *Water Resour. Res.*, *28*(10), 2691–2705.

Powers, S., L. Abriola, J. Dunkin, and W. Weber (1994a), Phenomenological models for transient napl-water mass-transfer processes, *J. Contam. Hydrol.*, *16*(1), 1–33.

Powers, S., L. Abriola, and W. Weber (1994b), An experimental investigation of nonaqueous phase liquid dissolution in saturated subsurface systems: Transient mass transfer rates, *Water Resour. Res.*, *30*(2), 307–320.

Provencher, S. (1982), A constrained regularization method for inverting data represented by linear algebraic or integral equations, *27*(3), 213–227.

Providenti, M., H. Lee, and J. Trevors (1993), Selected factors limiting the microbial degradation of recalcitrant compounds, *J. Ind. Microbiol.*, *12*, 379–395.

Rehfeldt, K., and L. Gelhar (1992), Stochastic analysis of dispersion in unsteady flow in heterogeneous aquifers, *Water Resour. Res.*, *28*(8).

Reis, L., R. Porto, and F. Chaudhry (1997), Optimal location of control valves in pipe networks by genetic algorithm, *J. Water Resour. Plng. and Mgmt.*, *123*(2), 317–320.

Rubin, Y. (1990), Stochastic modeling of macrodispersion in heterogeneous porous media, *Water Resour. Res.*, *26*(1), 133–141.

Russo, D. (1993), Statistics modelling of macrodispersion for solute transport in a heterogeneous unsaturated porous formation, *Water Resour. Res.*, *29*(2), 383–397.

Satish, M., and J. Zhu (1992), Stochastic approach for groundwater flow in semiconfined aquifer subject to random boundary conditions, *Adv. Water Resour.*, *15*(6), 329–339.

Savic, D., and J. Walters (1997), Genetic algorithms for least-cost design of water distribution networks, *J. Water Resour. Plng. and Mgmt.*, *123*(2), 67–77.

Schanz, R., and A. Salhotra (1992), Evaluation of the rackwitz-fiessler uncertainty analysis method for environmental fate and transport models, *Water Resour. Res.*, *28*(4), 1071–1079.

Schroeder, P., T. Dozier, P. Zappi, B. McEnroe, J. Sjostrom, and R. Peyton (1994), The hydrologic evaluation of landfill performance (help) model: Engineering documentation for version 3.0, *Tech. rep.*, Cincinnati, OH.

Sciortino, A., T. Harmon, and W.-G. Yeh (2000), Inverse modelling for locating dense nonaqueous pools in groundwater under steady flow conditions, *Water Resour. Res.*, *36*(7), 1723–1735.

Shewchuk, J. (1997), Delaunay refinement mesh generation, Ph.D. thesis, School of Computer Science, Computer Science Department, Carnegie Mellon University, Pittsburgh, PA.

Shvidler, M. (1993), Correlation model of transport in random fields, *Water Resour. Res.*, *29*(9), 3189–3199.

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

Skaggs, T., and D. Barry (1997), The first-order reliability method of predicting cumulative mass flux in heterogeneous porous formation, *Water Resour. Res.*, *33*(6), 1485–1494.

Skaggs, T., and A. Kabala (1994), Recovering the release history of a groundwater contaminant, *Water Resour. Res.*, *30*(1), 71–79.

Skaggs, T., and A. Kabala (1995), Recovering the history of a groundwater contaminant plume: Method of quasi-reversibility, *Water Resour. Res.*, *31*(11), 2669–2673.

Skaggs, T., and A. Kabala (1998), Comment on minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant, *Water Resour. Res.*, *34*(8), 2077–2079.

Smith, L., and F. Schwartz (1980), Mass transport 1. a stochastic analysis of macroscopic dispersion, *Water Resour. Res.*, *16*(2), 303–313, doi:10.1029/WR016i002p00303.

Snodgrass, M., and P. Kitanidis (1997), A geostatistical approach to contaminant source identification, *Water Resour. Res.*, *33*(4), 537–546.

Sonnenborg, T., P. Engesgaard, and R. D. (1996), Contaminant transport at a waste residue deposit: 1. inverse flow and nonreactive transport modelling, *Water Resour. Res.*, *32*(4), 925–938.

Sousa, M. (2009), Personal communication.

Spayd, S. (1997), Parkway well field ground-water flow modeling, *Tech. rep.*, New Jersey Geological Survey, Trenton, NJ.

Sun (2003), *OpenMP API User's Guide*, Sun Microsystem Inc., Santa Clara, CA, U.S.A.

Sun, N.-Z., and W. Yeh (1990), Coupled inverse problems in groundwater modelling 1. sensitivity analysis and parameter identification, *Water Resour. Res.*, *26*(10), 2507–2525.

Swain, E., and D. Chin (2003), An analytical formulation of two-dimensional groundwater dispersion induced by surficial recharge variability, *Water Resour. Res.*, *39*(9), doi:10.1029/2002WR001821.

Sykes, J. (2008), Personal communication.

Sykes, J., S. Pahwa, R. Lantz, and D. Ward (1982), Numerical simulation of flow and contaminant migration at an extensive monitored landfill, *Water Resour. Res.*, *18*(6), 1687–1704.

Sykes, J. F. (1995), Final gound water supplement to the phase ii pre-design report-evaluation of groundwater remedial design scenarios (june 18, 1993), *Tech. rep.*, Union Carbide Corporation, South Charleston WV.

Sykes, J. F. (1999), A comparison of groundwater flow models of the reich farm superfund site [draft], *Tech. rep.*, Union Carbide Corporation, South Charleston WV.

Sykes, J. F., and S. D. Normani (2002), An evaluation of groundwater flow using a calibrated model of the reich farm superfund site pleasant plains, new jersey, *Tech. rep.*, J.F. Sykes and Associates Limited.

Tang, D., and G. Pinder (1977), Simulation of groundwater flow and mass transport under uncertainty, *Adv. Water Resour.*, *1*(1), 25–30.

Tang, D., F. Schwartz, and L. Smith (1982), Stochastic modeling of mass transport in a random velocity field, *Water Resour. Res.*, *18*(2), 231–244, doi:10.1029/WR018i002p00231.

Tebes-Stevens, C., F. Espinoza, and A. Valocchi (2001), Evaluating the sensitivity of a subsurface multicomponent reactive transport model with respect to transport and reaction parameters, *J. Contam. Hydrol.*, *52*, 3–27.

Therrien, R., E. A. Sudicky, and R. G. McLaren (2004a), *FRAC3DVS: An Efficient Simulator for Three-dimensional, Saturated-Unsaturated Groundwater Flow and Density-dependent, Chain-Decay Solute Transport in Porous, Discretely-Fractured Porous or Dual-porosity Formations. User's Guide*, Groundwater Simulations Group, University of Waterloo, Waterloo, Ontario, Canada.

Therrien, R., E. A. Sudicky, and R. G. McLaren (2004b), *HydroSphere: A Three-dimensional Numerical Model Describing Fully-integrated Subsurface and Surface Flow and Solute Transport. User's Guide*, Groundwater Simulations Group, University of Waterloo, Waterloo, Ontario, Canada.

Tiedeman, C., and S. Gorelick (1993), Analysis of uncertainty in optimal groundwater contaminant capture design, *Water Resour. Res.*, *29*(7), 2139–2153.

Tolson, B., and C. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, *43*(W01413), doi:10.1029/2005WR004723.

Ulrych, T., and A. Woodbury (2003), Extensions to minimum relative entropy inversion for noisy data, *J. Contam. Hydrol.*, *67*, 13–25.

U.S. EPA (1997), Guiding principles for monte carlo analysis, *Tech. Rep. EPA-630-R-97-001*, Washington, DC.

U.S. EPA (1999), Safe drinking water act, section 1429. groundwater report to congress, *Tech. Rep. EPA-816-R-99-016*, Office of Water, Washington, DC.

U.S. EPA (2001), Trichloroethylene health risk assessment: Synthesis and characterization, *Tech. Rep. EPA-600-P-01-002A*, Office of Research and Development, Washington, DC.

U.S. EPA (2002), 2000 national water quality inventory, ground water chapters, *Tech. Rep. EPA-841-F-02-003*, Office of Water, Washington, DC.

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.

Vasco, D., J. Jr, and E. Majer (1996), Nonuniqueness in traveltime tomography: Ensemble inference and cluster analysis, *Geophysics*, *61*(4), 1209–1227.

Vasquez, J., H. Maier, B. Lence, and B. Tolson (2000), Achieving water quality system reliability using genetic algorithms, *J. Environ. Eng.*, *126*(10), 954–962.

Wagner, B. (1992), Simultaneous parameter estimation and contaminant source characterization for coupled groundwater flow and contaminant transport modelling, *J. Hydrol.*, *135*, 275–300.

Wagner, B., and S. Gorelick (1987), Optimal groundwater quality management under parameter uncertainty, *Water Resour. Res.*, *23*(7), 1162–1174.

Ward, D., M. Reeves, and L. Duda (1984), Verification and field comparison of the sandia waste-isolation flow and transport model (swift),, *Tech. rep.*, Albuquerque, NM.

Wilson, J., and J. Liu (1994), Backward tracking to find the source of pollution, in *Waste-management: From Risk to Remediation*, edited by R. B. et.al., pp. 181–199, ECM Press, Albuquerque, NM.

Woodbury, A., and L. Smith (1988), Simultaneous inversion of hydrogeologic and thermal data 2. incorporation of thermal data, *Water Resour. Res.*, *24*(3), 356–372.

Woodbury, A., and T. Ulrych (1993), Minimum relative entropy: Forward probabilistic modelling, *Water Resour. Res.*, *29*(8), 2847–2860.

Woodbury, A., and T. Ulrych (1996), Minimum relative entropy: Theory and application to recovering the release history of a groundwater contaminant, *Water Resour. Res.*, *32*(9), 2671–2681.

Woodbury, A., E. Sudicky, T. Ulrych, and R. Ludwig (1998), Three-dimensional plume source reconstruction using minimum relative entropy inversion, *J. Contam. Hydrol.*, *32*, 131–158.

Xiang, Y., J. Sykes, and N. Thomson (1993), A composite l1 parameter estimator for model-fitting in groundwater-flow and solute transport simulation, *Water Resour. Res.*, *29*(6), 1661–1673.

Xiang, Y., N. Thomson, and J. Sykes (1994), Parameter identification for groundwater flow and solute transport simulation: a comparison of l1 and l2 estimators, *Hydrol. Sci. J.*, *39*(1), 65–79.

Yeh, W.-G. (1986), Review of parameter identification procedures in groundwater hydrology: the inverse problem, *Water Resour. Res.*, *22*(2), 95–108.

Zapecza, O. (1989), Hydrogeologic framework of the new jersey coastal plain: U.S. Geological Survey Professional Paper 1404-B, 49p., p. 24.

Zhou, Z., and K. Harris (2008), Counteracting stagnation in genetic algorithm calculations by implementation of a micro genetic algorithm strategy, *Phys. Chem. Chem. Phys.*, *10*, 7262–7269, doi:10.1039/b807326k.

Zhu, J., and J. Sykes (2000), The influence of napl dissolution characteristics on field-scale contaminant transport in subsurface, *J. Contam. Hydrol.*, *41*, 133–154.

Zhu, J., and J. Sykes (2004), Simple screening models of napl dissolution in the subsurface, *J. Contam. Hydrol.*, *72*, 245–258.