

Parameter Estimation and Uncertainty
Analysis of Contaminant First Arrival
Times at Household Drinking Water Wells

by

Mary Kang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2007

©Mary Kang, 2007

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

Exposure assessment, which is an investigation of the extent of human exposure to a specific contaminant, must include estimates of the duration and frequency of exposure. For a groundwater system, the duration of exposure is controlled largely by the arrival time of the contaminant of concern at a drinking water well. This arrival time, which is normally estimated by using groundwater flow and transport models, can have a range of possible values due to the uncertainties that are typically present in real problems. Earlier arrival times generally represent low likelihood events, but play a crucial role in the decision-making process that must be conservative and precautionary, especially when evaluating the potential for adverse health impacts. Therefore, an emphasis must be placed on the accuracy of the leading tail region in the likelihood distribution of possible arrival times.

To demonstrate an approach to quantify the uncertainty of arrival times, a real contaminant transport problem which involves TCE contamination due to releases from the Lockformer Company Facility in Lisle, Illinois is used. The approach used in this research consists of two major components: inverse modelling or parameter estimation, and uncertainty analysis.

The parameter estimation process for this case study was selected based on insufficiencies in the model and observational data due to errors, biases, and limitations. A consideration of its purpose, which is to aid in characterising uncertainty, was also made in the process by including many possible variations in attempts to minimize assumptions. A preliminary investigation was conducted using a well-accepted parameter estimation method, PEST, and the corresponding findings were used to define characteristics of the parameter estimation process applied to this case study. Numerous objective functions, which include the well-known L_2 -estimator, robust estimators (L_1 -estimators and M-estimators), penalty functions, and deadzones, were incorporated in the parameter estimation process to treat specific insufficiencies. The concept of equifinality was adopted and multiple maximum likelihood parameter sets were accepted if pre-defined physical criteria were met. For each objective function, three procedures were implemented as a part of the parameter estimation approach for the given case study: a multistart procedure, a stochastic search using the Dynamically-Dimensioned Search (DDS), and a test for acceptance based on predefined physical criteria. The best performance in terms of the ability of parameter sets to satisfy the physical criteria was achieved using a Cauchy's M-estimator that was modified for this study and designated as the LRS1 M-estimator. Due to uncertainties, multiple parameter sets obtained with the LRS1 M-estimator, the L_1 -

estimator, and the L_2 -estimator are recommended for use in uncertainty analysis. Penalty functions had to be incorporated into the objective function definitions to generate a sufficient number of acceptable parameter sets; in contrast, deadzones proved to produce negligible benefits. The characteristics for parameter sets were examined in terms of frequency histograms and plots of parameter value versus objective function value to infer the nature of the likelihood distributions of parameters. The correlation structure was estimated using Pearson's product-moment correlation coefficient. The parameters are generally distributed uniformly or appear to follow a random nature with few correlations in the parameter space that results after the implementation of the multistart procedure. The execution of the search procedure results in the introduction of many correlations and in parameter distributions that appear to follow lognormal, normal, or uniform distributions. The application of the physical criteria refines the parameter characteristics in the parameter space resulting from the search procedure by reducing anomalies. The combined effect of optimization and the application of the physical criteria performs the function of behavioural thresholds by removing parameter sets with high objective function values.

Uncertainty analysis is performed with parameter sets obtained through two different sampling methodologies: the Monte Carlo sampling methodology, which randomly and independently samples from user-defined distributions, and the physically-based DDS-AU (P-DDS-AU) sampling methodology, which is developed based on the multiple parameter sets acquired during the parameter estimation process. Monte Carlo samples are found to be inadequate for uncertainty analysis of this case study due to its inability to find parameter sets that meet the predefined physical criteria. Successful results are achieved using the P-DDS-AU sampling methodology that inherently accounts for parameter correlations and does not require assumptions regarding parameter distributions. For the P-DDS-AU samples, uncertainty representation is performed using four definitions based on pseudo-likelihoods: two based on the Nash and Sutcliffe efficiency criterion, and two based on inverse error or residual variance. The definitions consist of shaping factors that strongly affect the resulting likelihood distribution. In addition, some definitions are affected by the objective function definition. Therefore, all variations are considered in the development of likelihood distribution envelopes, which are designed to maximize the amount of information available to decision-makers. The considerations that are important to the creation of an uncertainty envelope are outlined in this thesis. In general, greater uncertainty appears to be present at the tails of the distribution. For a refinement of the uncertainty envelopes, the application of additional physical criteria is recommended.

The selection of likelihood and objective function definitions and their properties are made based on the needs of the problem; therefore, preliminary investigations should always be conducted to provide a basis for selecting appropriate methods and definitions. It is imperative to remember that the communication of assumptions and definitions used in both parameter estimation and uncertainty analysis is crucial in decision-making scenarios.

Acknowledgements

I would like to express my sincere gratitude and appreciation to the following people for helping me during the course of this research:

My supervisors, J.F. Sykes and N.R. Thomson, for their guidance and support; as well as for providing me with an interesting topic

Bryan Tolson for providing me with the DDS method and many insightful comments.

Stefano Normani for his technical guidance; as well as for making time for many interesting discussions.

Natural Sciences and Engineering Research Council of Canada (NSERC) for awarding me the Postgraduate Scholarship (Master's).

The University of Waterloo for awarding me the President's Graduate Scholarship.

My family for their constant love and support.

Finally, all my friends, old and new, for all the fun times, love, and support.

Table of Contents

AUTHOR'S DECLARATION	ii
Abstract	iii
Acknowledgements	vi
Table of Contents	vii
List of Figures	ix
List of Tables	xii
Chapter 1 Introduction	1
Chapter 2 Problem Formulation	5
2.1 Site History	5
2.2 Contaminant Characteristics	7
2.3 Physical Setting	8
2.3.1 Geology	8
2.3.2 Hydrogeology	8
2.4 Conceptual Model	9
2.5 Prior Knowledge of Model Parameter Values	13
2.5.1 Parameter Values	14
2.5.2 Parameter Value Distribution	16
2.6 Observations	17
2.6.1 General Trends	17
2.6.2 Local Trends	20
2.6.3 Insufficiencies in Data	29
Chapter 3 Parameter Estimation	35
3.1 Objective Function Definition	40
3.1.1 Estimators	41
3.1.2 Add-ons to Estimators	46
3.2 Parameter Estimation Methods	47
3.2.1 PEST	47
3.2.2 DDS-AU	48
3.3 Parameter Estimation	49
3.3.1 PEST Results	50
3.3.2 DDS Results	60

3.4 Parameter Characteristics	67
3.4.1 Parameter Values	68
3.4.2 Parameter Correlations	89
3.5 Summary	93
Chapter 4 Uncertainty Analysis: First Arrival Estimates	95
4.1 Steps to Uncertainty Analysis	99
4.1.1 Sampling Approaches	99
4.1.2 Quantitative Representation of Uncertainty	101
4.2 Uncertainty Analysis Using Monte Carlo Sampling	109
4.2.1 Monte Carlo Sampling with PEST Results	110
4.2.2 Monte Carlo Sampling with DDS Results.....	116
4.2.3 Shortcoming of the Traditional Monte Carlo Analysis	120
4.3 Uncertainty Analysis Using Physically-Based Dynamically-Dimensioned Search Approximation of Uncertainty (P-DDS-AU) Sampling	120
4.3.1 Factors Affecting Likelihood Distributions.....	121
4.3.2 Steps to a Conservative Uncertainty Envelope.....	128
4.4 Summary	130
Chapter 5 Conclusions and Recommendations.....	131
5.1 On Parameter Estimation	131
5.2 On Uncertainty Analysis.....	134
Appendix A Analytical Solution to the One-Dimensional Advection-Dispersion Equation.....	136
Appendix B Green’s Function Used in the Solution for the Three-Dimensional Advection-Dispersion Equation.....	137
Appendix C Algorithm Parameters for Pest	138
Appendix D DDS Results.....	143
References	147

List of Figures

Figure 2.1. The extent of plume based on the 2001/2002 measured TCE.....	6
Figure 2.2. General schematic of the conceptual model	10
Figure 2.3. Plume centerline and the measured TCE concentrations in ppb	18
Figure 2.4. Location of measurement points for TCE in ppb in transformed coordinates	19
Figure 2.5. Representative locations for each area	20
Figure 2.6. Measured TCE concentrations in and around the Leclerq Class Area in ppb.....	21
Figure 2.7. Histogram of TCE concentration levels in and around the Leclerq Class Area (a) in ppb (b) in $\log_{10}(\text{ppb})$	22
Figure 2.8. Histogram of TCE concentration levels in and around the Leclerq Class Area in ppb with smaller intervals	23
Figure 2.9. Measured TCE concentrations in and around Mejdrech Area B in ppb	25
Figure 2.10. Histogram of TCE concentration levels in and around Mejdrech Area B (a) in ppb (b) in $\log_{10}(\text{ppb})$	25
Figure 2.11. Measured TCE concentrations in and around Mejdrech Area C in ppb	26
Figure 2.12. Histogram of TCE concentration levels in and around Mejdrech Area C (a) in ppb (b) in $\log_{10}(\text{ppb})$	27
Figure 2.13. Measured TCE concentrations in and around Mejdrech Area D in ppb	28
Figure 2.14. Histogram of TCE concentration levels in and around Mejdrech Area D (a) in ppb (b) in $\log_{10}(\text{ppb})$	28
Figure 2.15. Location in transformed coordinates and time interval of observations	33
Figure 3.1. Comparison of Huber's M-estimator with L-estimators.....	43
Figure 3.2. Comparison of Cauchy's M-estimator with L-estimators	44
Figure 3.3. Comparison of Cauchy's M-estimator with c value of 1 with revised Cauchy's M- estimator with c value of 1	45
Figure 3.4. Breakthrough concentration profiles at the four locations using the acceptable parameter set presented in Table 3.4	53
Figure 3.5. Plan view of the TCE plume in ppb after 33 years since the opening of the Metcoil facility at a depth of (a) 50 ft (b) 100 ft from the defined bottom of the limestone using the acceptable parameter set presented in Table 3.4	54
Figure 3.6. The relationship between the sum of squares value and the number of function evaluations for 22 estimated parameter sets with and without logarithmic transformations	58

Figure 3.7. Scatter plot of simulated versus measured TCE (ppb) for Parameter Set 21 with no logarithmic transformation.....	59
Figure 3.8. Scatter plot of simulated versus measured TCE (ppb) for Parameter Set 1 with no logarithmic transformation.....	60
Figure 3.9. Breakthrough concentration after 33 years since the opening of the Metcoil site at the four locations using the acceptable parameter set with the lowest objective function value using the LRS1 M-estimator.....	65
Figure 3.10. Plan view of the plume at a depth of 50 ft after 33 years since the opening of the Metcoil site using the acceptable parameter set with the lowest objective function value using the LRS1 M-estimator.....	67
Figure 3.11. Parameter values in terms of frequency histograms and objective function values defined by the L_2 -estimator using parameter sets obtained using the multistart procedure	71
Figure 3.12. Histogram of objective function values defined by the L_2 -estimator for parameter sets obtained using the multistart procedure.....	74
Figure 3.13. Relationship between parameter values and objective function values defined by the L_2 -estimator for parameter sets obtained using the multistart procedure	75
Figure 3.14. Val-Obj plots for parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function.....	76
Figure 3.15. Histograms of parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function	80
Figure 3.16. Val-Obj plot for entire range of objective function values for the parameter, c_0	84
Figure 3.17. Parameter values that were generated using the L_1 -estimator with a penalty of 10,000 in terms of performance index values and frequency for parameter sets that satisfy the physical criteria.....	86
Figure 4.1. Effect of N value on $L_{c/d} = P_k^{-N}$	105
Figure 4.2. Effect of N value on $L_{a/b} = (1 - P_k/\sigma_0^2)^N$	106
Figure 4.3. Histogram of sum of squares value for Monte Carlo samples generated using posterior distributions defined by PEST results and prior knowledge	113
Figure 4.4. Deterministic, mean, and standard deviation of TCE concentrations over time at the four locations of concern using Monte Carlo samples based on PEST results.....	114
Figure 4.5. Cumulative probability distribution using equally weighted Monte Carlo samples based on PEST results.....	114

Figure 4.6. Comparison of empirical and approximated lognormal distribution for the horizontal velocity in limestone	117
Figure 4.7. Deterministic, mean, and standard deviation of TCE concentrations over time at the four locations of concern using Monte Carlo samples based on DDS results.....	118
Figure 4.8. Cumulative probability distribution using equally weighted Monte Carlo samples based on DDS results.....	119
Figure 4.9. The tail-end of the cumulative normalized likelihood distribution plot for Mejdrech Area B using the L_1 -estimator with a penalty factor of 10,000	122
Figure 4.10. The tail-end of the cumulative normalized likelihood distribution plot for Mejdrech Area B using the L_2 -estimator with a penalty factor of 10,000	122
Figure 4.11. The tail-end of the cumulative normalized likelihood distribution plot for Mejdrech Area B using the LRS1-estimator with a penalty factor of 10,000.....	123
Figure 4.12. Cumulative normalized likelihood for the parameter sets generated using the L_1 -estimator with a penalty factor of 10,000 calculated with L_c and $L_{a/b}$ with $N=1$	124
Figure 4.13. Cumulative normalized likelihood plot for Mejdrech Area B using $L_{a/b}$ for various values of N using parameter sets generated using the L_1 -estimator with a penalty factor of 10,000....	125
Figure 4.14. Normalized likelihood distribution for Mejdrech Area B using $L_{a/b}$ using (a) $N = 0$ and (b) $N = 10$	126
Figure 4.15. Cumulative normalized likelihood plot at Mejdrech Area B using various likelihood definitions with an N -value of 0, where applicable, and 1000 and 1614 parameter sets generated using the LRS1 M-estimator with a penalty factor of 10,000.....	127
Figure 4.16. Cumulative normalized likelihood plot at Mejdrech Area B using $L_{c/d}$ with an N -value of 0 and 1000 parameter sets generated using various objective function definitions with a penalty factor of 10,000.....	128
Figure 4.17. Uncertainty envelopes for the four locations of concern.....	130

List of Tables

Table 2.1. Prior information regarding parameter values	15
Table 2.2. Lognormal standard deviations for uncertain parameters	16
Table 2.3. Representative locations for areas of concern	20
Table 2.4. Numbers and percentages of ND observations according to its type.....	30
Table 2.5. Number and percentage of observations found in each location within the 2001/2002 plume boundary.....	32
Table 3.1. Estimators used in parameter estimation process	46
Table 3.2. Adjustable model parameters not estimated using a parameter estimation method	50
Table 3.3. Adjustable model parameters estimated using a parameter estimation method	50
Table 3.4. Estimated parameter set that satisfies both constraints	52
Table 3.5. Random initial parameter values	55
Table 3.6. Optimized parameter estimates without logarithmic transformations	56
Table 3.7. Optimized parameter estimates using logarithmic transformations.....	56
Table 3.8. Comparison of PEST and DDS implementation	61
Table 3.9. A comparison of the objective function values prior to the application of criteria.....	61
Table 3.10. A comparison of the acceptance rates	63
Table 3.11. Acceptance rate for L_2 -estimator, L_1 -estimator, and the LRS1 M-estimator with respect to changes in the scaling factor for the penalty function	66
Table 3.12. Acceptance rates for objective function definitions with deadzones	66
Table 3.13. Trials to be used in uncertainty analysis	67
Table 3.14. Qualitative descriptions of optimized parameter space.....	82
Table 3.15. Qualitative descriptions of the acceptable parameter space based on the L_1 -estimator with a scaling factor of 10,000 for the penalty function.....	84
Table 3.16. Interpretation of correlation coefficient values.....	90
Table 3.17. Pearson product-moment correlation coefficient for random parameter sets.....	91
Table 3.18. Spearman's rank correlation coefficient for random parameter sets.....	91
Table 3.19. Summary of qualitative interpretation of Pearson product-moment correlation coefficients of the parameter sets generated using the LRS1 M-estimator, the L_1 -estimator, and the L_2 - estimator with a scaling factor of 10,000 for the penalty function.....	92

Table 3.20. Summary of qualitative interpretation of Pearson product-moment correlation coefficients of the parameter sets generated using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function.....	93
Table 4.1. Likelihood definitions to be used to describe the likelihood of a parameter set	108
Table 4.2. Result of Parameter Estimation using PEST	111
Table 4.3. Posterior parameter distribution to be used in Monte Carlo sampling.....	112
Table 4.4. Mean and standard deviation of parameter values generated using the DDS method.....	117
Table 4.5. Number of acceptable parameter sets used in developing the likelihood function.....	121

Chapter 1

Introduction

Chlorinated solvents have been found to be “among the most frequently detected” volatile organic compounds (VOCs) from groundwater samples taken throughout the United States (Moran et al., 2006). Chlorinated solvents commonly associated with subsurface contamination, such as perchloroethylene (PCE) and trichloroethylene (TCE), are linked to life-threatening diseases such as cancer and have consequently been assigned low Maximum Contaminant Levels (MCLs) by the U.S. Environmental Protection Agency (USEPA) (USEPA, 2007). The need to quantify the excess health risk imposed on the general public by exposure to contaminated groundwater is acknowledged by government bodies and frameworks for health risk assessments have been designed (USEPA, 1989, 1991). According to the USEPA’s risk assessment guidelines (USEPA, 1989, 1991), the quantification of the potential for adverse health impacts requires the identification of health impacts linked to the contaminant, an assessment of exposure, and an establishment of the dose-response relationship (USEPA, 2000). Since many chlorinated solvents have already been linked to adverse health effects through toxicology and epidemiology studies, exposure assessment becomes a critical step. Exposure assessment, which is an investigation of the extent of human exposure to a specific contaminant, must include the duration and frequency of exposure; and it depends on both population dynamics and aspects of contaminant hydrogeology including the nature of the contaminant source, groundwater flow and transport processes, and the location of the drinking water supply wells. Physically-based models that simulate the processes of groundwater flow and contaminant transport can be used in conjunction with site-specific observations to provide an estimate of the maximum duration of exposure at a given location and the corresponding contaminant concentration profile.

The maximum duration of exposure represents the time period during which a contaminant concentration is above some threshold value and is controlled largely by the estimate of the first arrival time of the contaminant in the water supply system. Uncertainty associated with the modelling process leads to a range of possible arrival times, and the first arrival time represents the earliest of these times. The earlier arrival times typically represent low likelihood events, but play a crucial role in the decision-making process that must be conservative and precautionary, especially when evaluating the potential for adverse health impacts. In fact, decisions made in environmental liability cases in the United States show that evidence suggesting contamination *may* have happened can be sufficient for the court to rule in favour of the claimant (Green, 2006). Therefore, an emphasis

must be placed on the accuracy of the leading tail region, which represents events of low likelihood, in the distribution of possible arrival times.

In general, a defensible conceptual model can be developed for a specific contamination problem using information on site history and geology, the contaminant's physical, chemical, and biological characteristics, and physical laws. Site-specific information is used along with the conceptual model to identify parameters that represent the hydrogeology, external system forces, and initial and boundary conditions of the system. The acceptance of a model as a descriptor of reality is dependent on its ability to incorporate all these different types of information, which are complicated by errors, biases, and limitations. Error and biases can be introduced during the model development and calibration process known as inverse modelling (Carrera et al., 2005). Efforts to reduce errors, biases, and limitations should be made for all problems by gathering more information but this is not always feasible in real problems due to time and economic constraints, and available measurement technology (Kirchner, 2006). Alternate approaches to deal with errors and biases include robust estimation techniques, which have been shown to be successful in treating some types of errors and biases (Finsterle and Najita, 1998). The impossibility of treating all sources of errors and biases and the existence of various objective function definitions in parameter calibration methods leads to multiple acceptable models at the conclusion of the inverse modelling process.

Families of common inverse algorithm approaches for both linear and nonlinear groundwater problems originate from different perspectives such as the maximum a posteriori methods, the maximum likelihood methods, and the pilot point method, which can all be viewed to be based on the maximum a posteriori estimation framework (McLaughlin and Townley, 1996). Traditionally, methods based on the maximum a posteriori estimation framework are designed to accept the single "best" model of large-scale trends while rejecting all others. The rejection of a model based on the fact that it does not have the maximum a posteriori performance index value is not sufficient in risk analysis scenarios or in legal settings. Confidence bounds derived for the "best" model using the popular variance-based approaches are not recommended for use in risk analysis since they only indicate the performance of the inverse procedure (McLaughlin and Townley, 1996) and do not represent low probability or likelihood models of large-scale trends. Large-scale trends are defined as "the components of the point variables that we can expect to estimate from available measurements" while small-scale fluctuations are defined as "unidentifiable deviations from the large-scale trends" by McLaughlin and Townley (1996). The assumption of normality typically made regarding the nature of small-scale fluctuations, which represent the cumulative effect of errors and biases, is not

valid for many real problems, and variance-based confidence bounds cannot be used directly as a measure of small-scale fluctuations. The quantification of small-scale fluctuations in addition to large-scale trends is necessary due to the importance of low likelihood events. Thus, many possible models must be examined using probabilistic analysis.

Given a mathematical model, the acceptance of only a single set of parameter values in assessing first arrival times which vary in space is essentially flawed. For example, one parameter set may produce an earlier arrival time at one location compared to another parameter set while producing a later arrival time at a different location. The recognition of the potential for multiple acceptable parameterizations, which is referred to as equifinality (Beven, 2006), is made by Beven and Binley (1992), who proposed the Generalized Likelihood Uncertainty Estimation (GLUE) method. The GLUE methodology produces a likelihood distribution based on a subjective definition of likelihood using all behavioural models and is designed to include a range of likelihood models rather than simply the “best” or maximum likelihood models. No formal definition of error structure is made in the GLUE method (Beven and Binley, 1992), and thus, it inherently captures both large-scale trends and small-scale fluctuations. The subjectivity of GLUE’s likelihood function definition is not problematic in risk analysis scenarios as long as risk communication is performed effectively.

GLUE’s likelihood distribution is a function of the sampling strategy and the definition of what constitutes a behavioural model. Models are considered to be behavioural if they can create simulations that resemble observations within some subjective threshold level (Beven and Freer, 2001). The models that are not considered behavioural are not included in the likelihood distribution. In essence, the GLUE method uses a multiple maximum a posteriori estimate framework by rejecting non-behavioural models. Multiple maximum likelihood estimates can be achieved with any existing maximum likelihood method applied in multiple optimization trials, which are designed to terminate in alternative high likelihood solutions. The sampling approach in the Dynamically-Dimensioned Search – Approximation of Uncertainty (DDS-AU) method (Tolson, 2005) exploits this characteristic by using the Dynamically-Dimensioned Search (DDS) (Tolson, 2006). Since the DDS-AU sampling approach maximizes each model’s ability to create simulations that resemble observations, the application of a behavioural threshold appears to be redundant. Also, the definition of what constitutes a behavioural model should be based on physical, chemical, and/or biological constraints.

In this thesis, this new concept of behavioural is adopted and used in conjunction with a revised DDS-AU sampling method and the GLUE framework to assess the first arrival of contamination at various locations for a real problem. The case study, which serves as a representative problem, is a

real groundwater contamination problem involving TCE, which has been detected in drinking water wells down-gradient of an industrial facility. The developed conceptual and mathematical model contains a large number of parameters that are considered uncertain due to limited information. The resulting high dimensionality in conjunction the lack of site-specific information causes limitations commonly encountered in real world problems.

In Chapter 2, a conceptual and mathematical model is developed for the case study, and the corresponding observation data is presented with a review of potential for contribution to ill-posedness.

In Chapter 3, the definition of what constitutes a behavioural model is established and multiple acceptable and behavioural models are generated using the parameter estimation algorithm, DDS. PEST, a popular gradient-based search algorithm, is also evaluated to test the abilities of the least squares method, which is a widely-used maximum likelihood method. Attempts to treat data insufficiencies and increase the number of behavioural solutions by altering the objective function definition are investigated.

In Chapter 4, uncertainty analysis using various subjective likelihood definitions is performed on models obtained with two different sampling strategies: Monte Carlo sampling and a physically-based DDS-AU sampling methodology (P-DDS-AU). P-DDS-AU introduces revisions on the DDS-AU sampling method to accommodate the new definition of behavioural. The various likelihood distributions are compared and a conclusion is drawn regarding the uncertainty of the arrival time.

A summary of the thesis and a general discussion of the relevance of the work in terms of current and future research are provided in Chapter 5.

Chapter 2

Problem Formulation

Any parameter estimation process and uncertainty analysis is meaningless if the conceptual model is not defensible. In this chapter, a description of the site history, properties of trichloroethylene (TCE), and the physical setting are provided and used to develop a conceptual model. A solution to the mathematical model is presented along with general information regarding the corresponding parameters.

2.1 Site History

Trichloroethylene (TCE) was released into the subsurface at the Metcoil site (formerly Lockformer) in Lisle, Illinois over a period of approximately 32 years since the opening of the plant in March 1969 (Clayton Group Services, Volume 1, May 2002). TCE was used as a degreasing agent in the process of manufacturing metal products and was initially stored in a 500-gallon roof-mounted tank. This tank was removed from service in June 1999 and replaced with a double-walled, secondarily contained 250-gallon TCE tank inside the Lockformer building. The use of TCE was discontinued at the facility in February 2001. Known, suspected, or potential sources of contamination are delivery spills at the former TCE fill pipe and tank area, releases in the former vapour degreaser area, releases from the sanitary sewer system, and releases to the existing and former drainage ways. The 500-gallon roof-mounted TCE tank had no gauge but two sight glasses until approximately 1980, and accidental spillage during delivery is assumed to be a major source of TCE release. Procedures to reduce accidental spillage were continually implemented from the early 1980s to approximately 1984 or 1985 when the possibility of surface spills was practically eliminated by the adoption of new operating practices. Ball (2003) estimates that overflow releases from the roof tank via the TCE tank vent pipe caused losses of approximately 6000 gallons of TCE from 1969 to at least 1985. There is considerable debate in terms of the frequency and significance of delivery spills and the overall mass of TCE released.

Residential properties located down-gradient of the Metcoil site have used household wells that are completed in the potentially contaminated aquifer. Current and past residents potentially exposed to TCE originating from the Metcoil site have taken legal action to be compensated for their losses resulting from adverse health and mental effects, and decreases in property values. The first arrival

time of the contaminant at a well plays a crucial role in determining whether or not a claimant is eligible for compensation.

The residential properties are grouped into four areas as a part of the legal proceedings (Clayton Group Services Inc., 2002): the Leclercq Class Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D. A map of the Metcoil site and the four areas are presented in Figure 2.1 along with groundwater sample locations. Figure 2.1 shows that there are ground water samples indicating TCE contamination has occurred in each of the four areas. Although anomalies in the sampling processes can take place, it is difficult to claim contamination has not occurred in locations where observed concentrations have been reported in legal proceedings. Therefore, an acceptable model must simulate TCE contamination in each of the four areas that originated from the Metcoil facility.

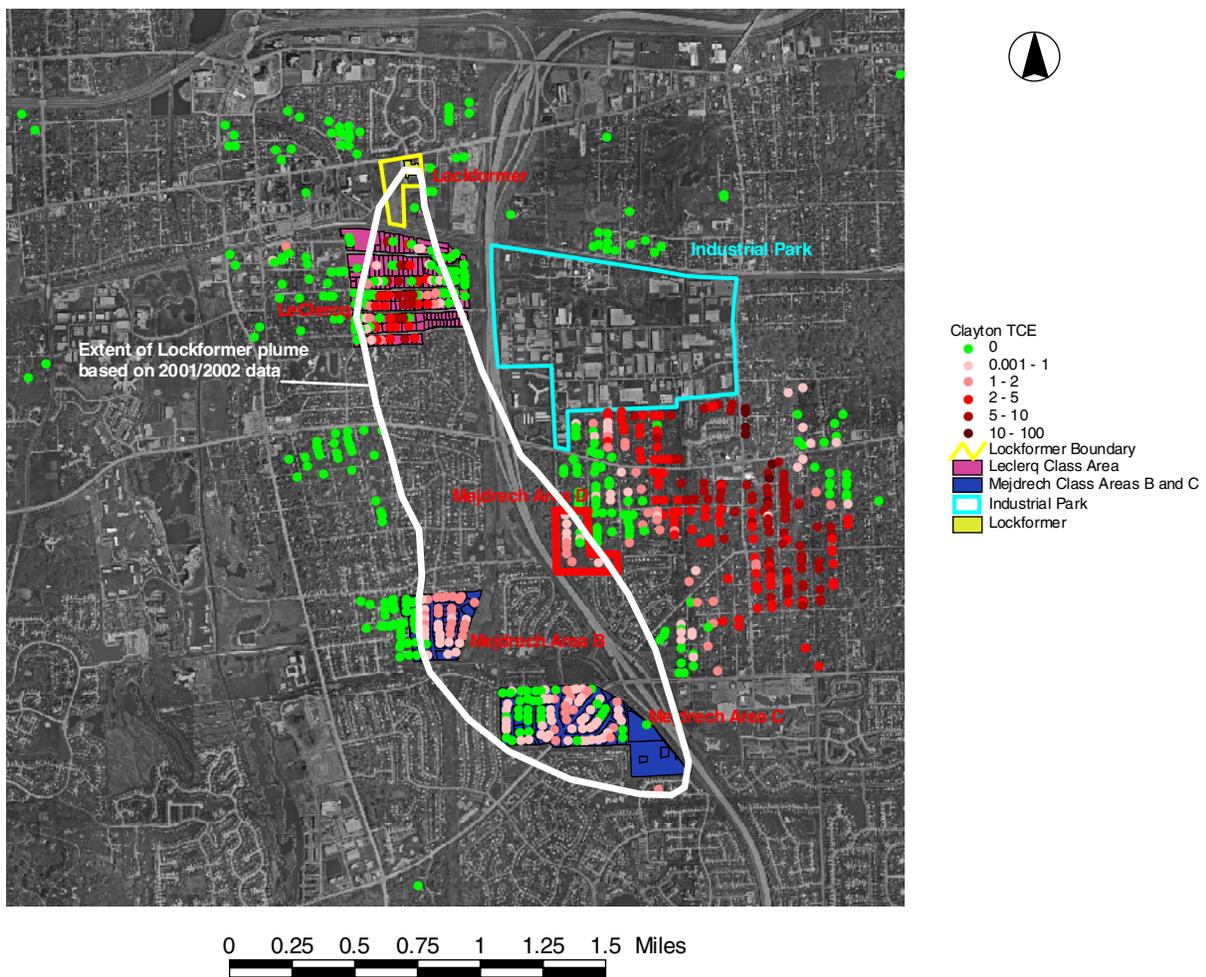


Figure 2.1. The extent of plume based on the 2001/2002 measured TCE

2.2 Contaminant Characteristics

Trichloroethylene (TCE) is an organic contaminant commonly used as an industrial solvent in the automotive and metals industry (Federal-Provincial-Territorial Committee on Drinking Water, 2004). TCE is classified as a probable carcinogen that has been linked to various cancers including liver and kidney cancer. Therefore, a low Maximum Contaminant Level (MCL) of 5 ppb with a Maximum Contaminant Level Goal (MCLG) of 0 ppb is assigned by the U.S. Environmental Protection Agency (USEPA, 2006).

TCE contamination is primarily of concern in groundwater systems since it is highly volatile in the atmosphere. TCE is a dense non-aqueous phase liquid (DNAPL) with a solubility ranging from 1100 to 1400 mg/L (Federal-Provincial-Territorial Committee on Drinking Water, 2004). Biodegradation of TCE is limited; and field and laboratory experiments have determined first-order decay coefficients with a mean and standard deviation of 0.2 day^{-1} and 0.5 day^{-1} respectively under aerobic conditions (Suarez and Rifai, 1999). TCE is persistent in groundwater because a DNAPL source can exist for a long period of time due to its low solubility and typically slow degradation rates. TCE also undergoes sorption which acts to retard plume migration. This is problematic since TCE creates health concerns even at low concentrations.

The arrival of contamination can be defined as the first moment in time that a contaminant concentration greater than zero appeared. In practice, there exists a detection limit (DL) below which concentrations cannot be detected using analytical methods. The DL varies with the selection of the analytical method and can range from 0.01–3.0 $\mu\text{g/L}$ (or ppb) for the determination of TCE in water (Federal-Provincial-Territorial Committee on Drinking Water, 2004). The usage of the DL to define the arrival of a contaminant is somewhat arbitrary since the value is determined by current technology. Ideally, a threshold concentration above which exposure can cause adverse health impacts should be employed but such information is typically unavailable. Therefore, the DL is selected to define the arrival of contamination in this thesis even though this threshold concentration can be below the MCL for TCE. If future research leads to a reduction in the DL or the identification of a threshold concentration, the methods and approach developed in this thesis are not affected.

2.3 Physical Setting

2.3.1 Geology

The Metcoil facility is located in the Wheaton Morainal section of the Great Lakes physiographic province. The uppermost surficial glacial unit is comprised of undifferentiated Valparaiso Moraine deposits, which can consist of yellow-grey silty till, sand and gravel, and dune sand (Woller et al., 1986). These morainal deposits are generally overlain by a thin Richland Loess or modern soil (Clayton Group Services Inc., 2002). According to the site reports documented by Clayton Group Services Inc. (2002), the major stratigraphic units at the Metcoil site are composed of a silty clay/till unit, a mass waste & gravel unit, a lower till unit, and the bedrock. The existence and thicknesses of the units in the overlying the bedrock are variable in space. Site investigative work has shown varying amounts of TCE in some of these units (Clayton Group Services Inc., 2002).

The Paleozoic bedrock underlying the glacial deposits in the Wheaton Morainal section comprises of about 1,100 m of lithified, stratified, sedimentary rocks of Silurian, Ordovician, and Cambrian ages respectively (Woller et al., 1986). The Silurian system primarily consists of dolomite with minor amounts of shale. Varying grades of sandstone are found in the lower units of the Ordovician system in addition to dolomites and shale. The Cambrian system underlying the Ordovician system and overlying granitic rocks consists of dolomite, sandstone, shale, and siltstone.

2.3.2 Hydrogeology

Two major hydrostratigraphic units can be defined to represent the possible areal extents of the TCE plume: the till unit and the dolomite aquifer.

Site investigations of geologic and hydrogeologic conditions at the Metcoil site by Clayton Group Services Inc. (2002) indicate the formation of a shallow water table in the mass waste and gravel unit, which underlies a relatively thin silty clay/till unit. Therefore, both hydrostratigraphic units can generally be assumed to be saturated. The hydraulic connectivity in the unsaturated surficial silty clay till and the mass waste and gravel unit are controlled by the occurrence of fractures and coarse-grained lithologies. The hydraulic conductivity in the mass waste and gravel unit is estimated by Clayton Group Services Inc. (2002) to range from 5×10^{-4} to 5×10^{-3} cm/sec using grain size analysis of soil samples. No hydraulic conductivity testing was performed at the site or for the region. An effective porosity value of approximately 0.20 can be assumed to represent this mass waste and gravel unit (Freeze and Cherry, 1979). For an effective porosity of 0.20, seepage velocities of 2.3 to

65 feet per year can be expected in this unit at the Metcoil site. The till units are hydraulically connected to the underlying Silurian dolomite bedrock at the Metcoil site and a predominantly vertical downward gradient between the till units and the dolomite has been observed.

The Silurian dolomite bedrock is highly weathered and fractured in the top approximately 10 to 20 feet zone in the Lisle and Downer's Grove areas. Site investigations at the Metcoil site show a slightly smaller thickness of 4 to 6 feet for the weathered zone. The dolomite below this region can be viewed as representative of typical fractured rock environments where groundwater exists primarily in bedding planes and individual fractures. These joints and fractures are characterized as "open, abundant, and interconnected" and thus "give a resultant regional effect equivalent to a radially homogeneous aquifer" (GSI, 2003). The groundwater velocity in the weathered dolomite has been estimated to be up to 1275 feet/year (Ball, 2003). Dolomite can have porosities that range from 0 to 0.2 (Freeze and Cherry, 1979), and the effective porosity of the Silurian dolomite has been estimated to be 0.017 (Clayton Group Services Inc., 2002). Sufficient production is typically available in the upper portions of the aquifer, and wells are primarily completed in the upper 60 feet of the Silurian dolomite (GSI, 2003). Approximately 30% of the wells in DuPage County penetrate only 0 to 20 feet into the dolomite while 34% penetrate 20 to 40 feet into the dolomite (Clayton Group Services, 2002).

The available information indicates the general direction of flow to be in the south to southeast direction throughout the possible extents of the TCE plume (GSI, 2003). The average recharge for the eastern two-thirds of DuPage County has been determined to be approximately 2.94 inch/year (Clayton Group Services, 2002). Recharge through the overburden to the Silurian dolomite is highly variable.

2.4 Conceptual Model

Based on the stratigraphy and the hydrogeology of the site described in Section 2.3, a TCE transport model representing two saturated units, the till and the bedrock, as shown in Figure 2.2 was developed. The contamination source is assumed to exist in the till unit above the dolomite aquifer as residual saturation or immiscible pools trapped on capillary barriers. This source zone is assumed to be uniformly distributed over a rectangular area; this inherently averages the effect of multiple releases at the Metcoil site occurring at different times and locations both of which are unknown. Aqueous phase TCE is assumed to dissolve from this source zone and migrate vertically downward into the bedrock aquifer. Primarily horizontal movement is assumed in the saturated aquifer such that

most of the contaminant mass remains near the top of the aquifer, where the highest hydraulic conductivities are known to exist. The contamination of this aquifer, in which residential wells are assumed to be completed, is the primary concern in this case study.

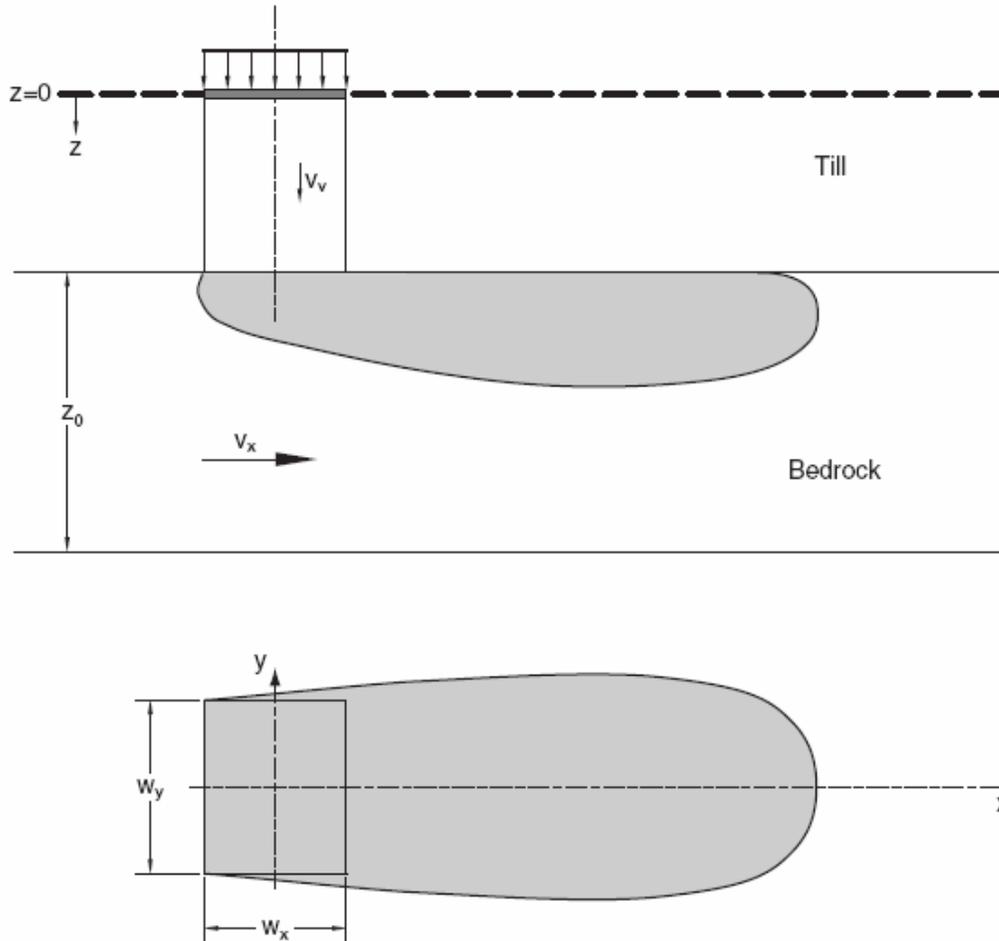


Figure 2.2. General schematic of the conceptual model

For this study, complexity is not increased because of the lack of available information and limitations in computational resources. Insufficient information regarding spatial variation in the physical setting is available to develop a distributed model, and the added complexity in distributed models increases computational effort, which is undesirable in uncertainty analysis based on large numbers of simulations. As a result, a steady-state flow field and spatially constant properties were assumed for each hydrostratigraphic unit.

The nature of the contamination source is characterized by the operating practices at the Metcoil facility and the degradation processes of TCE. The TCE releases that took place during facility operations is expected to be the dominating factor due to the characteristics of TCE and the nature of possible sources. Quantitative information regarding TCE releases originating from multiple sources, some of which are unknown, are limited, and the rate of TCE release into the surface is highly uncertain. The implementation of alternative operating practices to reduce TCE releases is expected to have reduced this rate; however, the level of this reduction and the exact time at which these measures were put into action are uncertain.

Since an exponential source model can produce a trend of higher releases at earlier times with decreases in the releases at a later time, the exponential model is assumed to be able to adequately represent source concentrations over time along with the definition of an initial source concentration, c_0 . Therefore, the TCE source concentration is assumed to follow a first-order decay model as follows

$$c(0, t) = \begin{cases} c_0 e^{-\lambda t} & 0 \leq t \leq t_r \\ 0 & t_r < t \leq \infty \end{cases} \quad (1)$$

where λ accounts for the reduction of the dissolved TCE source concentration, $c(0, t)$ represents TCE concentration at the top of the till unit at time, t , and t_r represents the time at which the source was removed. Note that the usage of the exponential model leads inherently to the assumption that the maximum dissolved TCE concentration immediately below the source zone occurs at the initial time (i.e. $t = 0$).

The one-dimensional transport from the source through the overburden is captured by the mass conservation equation

$$R_t \frac{\partial c}{\partial t} = \frac{\partial}{\partial z} \left(D_v \frac{\partial c}{\partial z} \right) - v_v \frac{\partial c}{\partial z} - \mu_t c \quad (2)$$

where c is the dissolved concentration at a point z in the overburden [M/L^3]; R_t is the retardation factor for the dissolved TCE in the overburden; D_v is the longitudinal dispersion coefficient [L^2/T]; v_v is the linear velocity in the overburden [L/T]; and μ_t represents the decay of both solute and sorbed phase of the TCE in the overburden. The initial condition for dissolved TCE transport through the till layer is

$$c(z, 0) = 0 \quad (3)$$

A semi-infinite domain is assumed and the corresponding lower boundary condition is

$$\frac{\partial c}{\partial z}(\infty, t) = 0 \quad (4)$$

where $z = 0$ represents the bottom of the TCE source zone. The analytical solution of (2) subject to (1), (3), and (4) was taken from van Genuchten and Alves (1982) (Appendix A).

The underlying dolomite aquifer is assumed to be of finite thickness that spans infinitely in the horizontal directions x and y . The corresponding three-dimensional mass conservation equation for TCE in this aquifer is

$$R_r \frac{\partial c}{\partial t} = D_x \frac{\partial^2 c}{\partial x^2} + D_y \frac{\partial^2 c}{\partial y^2} + D_z \frac{\partial^2 c}{\partial z^2} - v_x \frac{\partial c}{\partial x} - \mu_r c + S(x, y, z, t) \quad (5)$$

where R_r is the retardation factor for the TCE in the dolomite aquifer; D_x , D_y and D_z are the dispersion coefficients in the x , y , and z directions respectively; v_x is the horizontal linear velocity in the x -direction in the dolomite aquifer; μ_r accounts for the decay of TCE; $S(x, y, z, t)$ describes the source of TCE at the top of the aquifer. The initial condition for the dissolved TCE in the dolomite aquifer is

$$c(x, y, z, 0) = 0 \quad (6)$$

The x and y direction boundary conditions are

$$c(\pm \infty, y, z, t) = c(x, \pm \infty, z, t) = 0 \quad (7)$$

while the z direction boundary condition is

$$\frac{\partial c(x, y, z, t)}{\partial z} = 0 \text{ at } z = 0 \text{ and } z = z_0 \quad (8)$$

where $z = z_0$ represents the top of the aquifer, and $z = 0$ represents the bottom of the aquifer. The concentration at the top of the dolomite aquifer is obtained using the one-dimensional migration equation described by (1) to (4). This point concentration is assumed to represent the source of TCE contamination in the aquifer in terms of mass and thus, is converted to a uniform concentration throughout the rectangular source zone as described by

$$S(x, y, z, t) = \frac{c_{\text{till}} n_{\text{till}} v_v}{n_{\text{rock}}} \left[H(y - w_y / 2) - H(y + w_y / 2) \right] \cdot \left[H(x - w_x / 2) - H(x + w_x / 2) \right] \cdot \delta(z - z_0) \quad (9)$$

where w_x and w_y represent the horizontal dimensions; H is the Heaviside step function; n_{till} and n_{rock} are the porosity of the till and the dolomite respectively; and c_{till} is the concentration at the bottom of the till obtained the solution to (2). The vertical velocity in the till unit, v_v , defined in (2) is assumed to be constant throughout the unit and therefore, is also assumed to be constant beneath the areal

extent of the source zone. The source term represented by (9) assumes that the till in the rectangular source zone has a uniform thickness, and that the concentration of dissolved TCE in the till at all locations immediately beneath the source zone is determined using (1) to (4). This definition of the source term leads to discontinuities in concentration levels found at the bottom of the till or the top of the aquifer since only the mass is conserved.

The solution to (5) to (9) is obtained by using the Green's function of (5) and integrating yields

$$c(x, y, z, t) = \int_0^{z_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_0^t h(x, x', y, y', z, z', t, t') S(x', y', z', t') dt' dx' dy' dz' \quad (10)$$

where x' , y' , z' , and t' are integration variables. The Green's function in (10) is presented in Appendix B. The integral is solved using an adaptive 160 point Gauss Legendre Quadrature, which is a numerical integration method appropriate in approximating definite integrals that are difficult to solve in closed-form. The high number of Gauss Legendre points indicates the complexity of the solution surface.

The solution allows for the estimation of the concentration level in the Silurian dolomite aquifer in 3-dimensional space over time. The breakthrough curves at locations of concern determined with the model are used to estimate the arrival time of a detectable TCE concentration. For conservative estimates, a location of concern should consider its proximity to the source and the centerline since the model is designed to produce higher concentrations at locations closer to these areas.

2.5 Prior Knowledge of Model Parameter Values

Prior information such as the site history, the physical setting, and the contaminant characteristics are not sufficient to determine the appropriate values for all parameters; and quantitative observations are used with parameter estimation methods to estimate uncertain parameters. Each of the 17 parameters associated with the mathematical representation of the conceptual model presented in Section 2.4 are uncertain to differing degrees. For some parameters, a sufficient amount of information is available to select a suitable value; for other parameters, only a range of possible values can be determined. In this section, prior knowledge is used to provide quantitative information regarding each parameter. For uncertain parameters, a discussion on the possible nature of parameter values, which is of critical importance for some uncertainty analysis methods, is also provided.

2.5.1 Parameter Values

2.5.1.1 Source Zone

There are 4 parameters associated with the source zone: the initial source concentration, c_0 , the source reduction rate, λ , the source length, w_x , and the source width, w_y .

The parameters, c_0 and λ , control the highly uncertain source concentration over time described by (1). The initial source concentration, c_0 , should be less than the solubility of TCE and greater than zero; the values are selected to range from 100 mg/L to 2000 mg/L to conservatively represent the large degree of uncertainty associated with this parameter. A value significantly higher than the solubility is selected as the upper bound for c_0 due to the sharp decrease near initial times possible with (1). The source reduction rate, λ , is made to range from 0 to 1 based on the history of the site. A λ value of 0 represents TCE released uniformly over time while a value of 1 represents the concentration of releases to early times.

Locations of the multiple possible sources of release are assumed to be spread out through the Metcoil site shown in Figure 2.1. A rectangular area of 150 ft by 150 ft that appears to capture the areal extent of the site is used to represent the source zone. Therefore, the values for the parameters, w_x and w_y , are assumed to be fixable.

2.5.1.2 Till Unit

There are 6 parameters associated with the one-dimensional transport through the till unit: the vertical velocity, v_v , the dispersivity, D_v , the retardation factor, R , the decay, μ_t , the porosity, n_o , and the depth to top of the aquifer unit, b_t . The available site investigations that primarily characterize the soil properties and groundwater flow provide sufficient information to select a porosity of 0.2. The retardation factor is correlated to the porosity and approximate values can be estimated (Freeze and Cherry, 1979). A low retardation factor of 1.5 is selected to be conservative since TCE is known to be retarded.

The remaining parameters are assumed to be uncertain and conservative ranges are assigned and presented in Table 2.1. Site investigative work shows that the thickness of the till unit is highly variable; thus, the parameter is considered uncertain since the optimal approach to finding an equivalent uniform thickness is unknown.

Table 2.1. Prior information regarding parameter values

Zone / Unit	Parameter Abbreviations	Parameter Description	Units	Prior Estimate or Range	
Source	c_0 (or c0)	Initial source concentration	mg/L	100.0	2000.0
	λ (or gam)	Source reduction rate	1/years	0.00001	1.0
	w_x (or wx)	Source length	ft	150	
	w_y (or wy)	Source width	ft	150	
Till	v_v (or vv)	Vertical velocity in overburden	ft/yr	1.0	40.0
	D_v (or dispv)	Dispersivity in overburden	ft	0.001	20.0
	R_t (or rv)	Retardation in overburden	n/a	1.5	
	μ_t (or rlam)	Decay coefficient	1/years	0.00001	1.0
	b_t (or xv)	Depth to top of limestone or thickness of overburden	ft	0.0	80.0
	n_o (or porv)	Porosity in overburden	n/a	0.2	
Aquifer	v_x (or v)	Horizontal velocity in limestone	ft/yr	100.0	1000.0
	D_x (or dispx)	Longitudinal dispersivity in limestone	ft	10.0	500.0
	D_y (or dispy)	Transverse dispersivity in limestone	ft	1.0	50.0
	D_z (or dispz)	Vertical dispersivity in limestone	ft	0.05	50.0
	D^* (or dmol)	Molecular diffusion times tortuosity	ft ² /yr	0	
	R_l (or rlim)	Retardation in limestone	n/a	1.5	
	n_a (or porlim)	Porosity in limestone	n/a	0.05	

2.5.1.3 Aquifer Unit

The dolomite aquifer unit is characterized by 7 parameters: the horizontal velocity, v_x , the longitudinal, transverse, and vertical dispersivities, D_x , D_y , and D_z , the molecular diffusion times tortuosity, D^* , the retardation factor, R_t , and the porosity, n_a . The porosity and the retardation factor are selected to be 0.05 and 1.5 respectively and fixed based on the same reasoning presented for the till unit in Section 2.5.1.2. The remaining parameters are assumed to be uncertain and conservative ranges are assigned and presented in Table 2.1.

It is important to note that many of these model parameters are correlated but their exact relationship is unknown. For example, high values for both v_v and D_v , which are not directly dependent, results in arrival times that are unrealistically early. Approaches to consider high correlations in parameter estimation and uncertainty methods, which typically assume independence, must be taken to ensure sound results. Some treatment approaches are regularization and post-processing during the parameter estimation process.

2.5.2 Parameter Value Distribution

Typically, parameter values are assumed to be normally-distributed using the central limit theorem or lognormally-distributed assuming it is a multiplicative product of numerous independent and identically distributed factors. Some studies have shown that groundwater flow and transport parameters are more likely to be distributed lognormally. Although the evidence is not definitive, lognormal standard deviations estimated using expert opinion (Sykes, 2004) are presented for uncertain parameters in Table 2.2.

Table 2.2. Lognormal standard deviations for uncertain parameters

Parameter	Lognormal Standard Deviation
c_0	0.1
v_x	0.025
D_x	0.2
D_y	0.2
D_z	0.2
μ_t	0.05
λ	0.01
v_v	0.1
D_v	0.1
b_t	0.2

It is important to note that the nature of parameter value distributions is highly uncertain and the capability of standard parametric distributions to represent the “true” nature is questionable.

2.6 Observations

Automatic calibration methods can be used with quantitative observations to estimate parameter values given the model and prior knowledge regarding parameters. Residential wells at 695 locations shown in Figure 2.1 were sampled for TCE in 2001 and 2002 by Clayton Group Services. The detection limit (DL) of the analytical method employed to test ground water samples for TCE is 0.5 ppb for the majority of the samples. Observations with concentration levels below the DL are referred to as non-detects (NDs).

These residential wells are assumed to be completed in the dolomite aquifer based on the hydrogeology described in Section 2.3.2. Although the screened intervals of these wells are unknown, they are likely to be designed to produce ample yield for household use; thus, the interval is likely to be of sufficient length to cause averaging of concentrations in the interval. The top of the aquifer is selected to represent this averaged concentration since TCE transport is assumed to be primarily horizontal in the aquifer unit.

The concentration distribution of all observations is used to identify general trends and manipulations are made to facilitate the use of the model presented in Section 2.4. Further insight into the possible effects of the observation on parameter estimation process is gained by examining local trends. The insufficiencies of observations that are likely to cause problems in the calibration process are discussed and possible treatments are recommended.

2.6.1 General Trends

The plume boundary or the extent of the plume is defined in this thesis to represent the area in which observations have detectable concentration levels. The approximate extent of the Metcoil TCE plume delineated in Figure 2.1 is based on the direction of groundwater flow and the concentration distribution obtained from the 2001/2002 well survey data. The region of high concentration to the east of the delineated plume boundary is known to be caused by the industrial park located to the north to northeast of the region. This is validated by the row of NDs present between the Metcoil TCE plume and the industrial park plume. Therefore, the concentrations located outside of the delineated 2001/2002 plume boundary appear to be unrelated to TCE releases at the Metcoil site; and, only the observations shown in Figure 2.1 that lie inside of the boundary along with NDs surrounding

the plume are considered to represent the problem. The resulting number of observations, which are to be used in the model calibration process in Chapter 3, is 320.

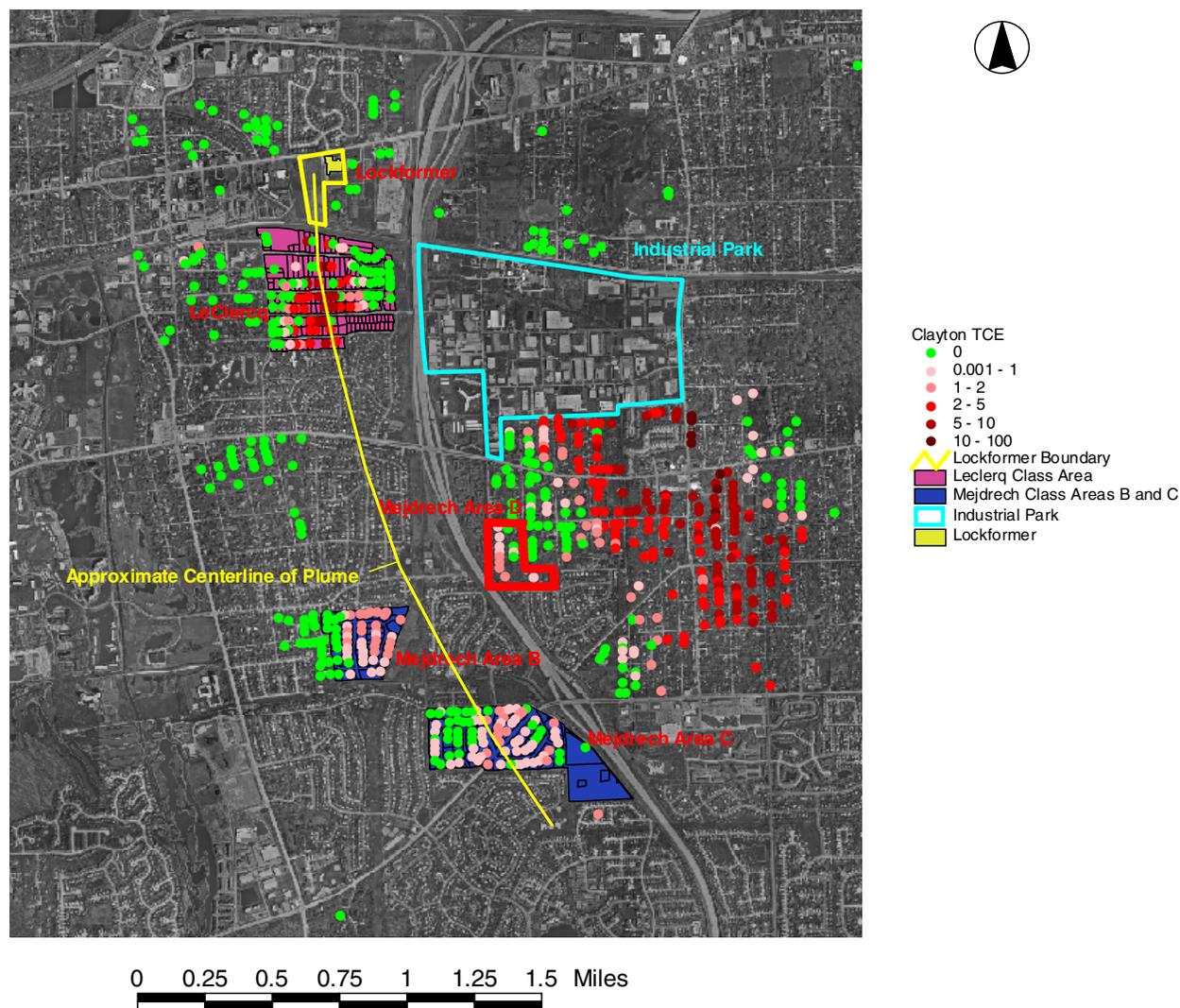


Figure 2.3. Plume centerline and the measured TCE concentrations in ppb

The centerline of the plume is approximated manually using the concentration distribution and the shape of the plume boundary as shown in Figure 2.3. The coordinates of observation locations are transformed such that the plume centerline forms a straight line and the source is located at the origin as shown in Figure 2.4. This transformation inherently includes heterogeneities and anisotropies into the model and facilitates the use of the analytical solution. An approximate extent of the 2001/2002 plume is re-delineated using the transformed distribution of concentration levels in such a way that symmetry about the centerline is achieved. The 2001/2002 plume boundary based on the transformed

coordinates is delineated to correspond to the conceptual model; this boundary, which is simply referred to as the plume boundary, is used to analyse local trends in the observation data. Note that the new plume boundary, shown in Figure 2.4, may not necessarily coincide with the extent of plume delineated in Figure 2.1 since both boundaries are delineated independently using observation locations in their respective coordinates.

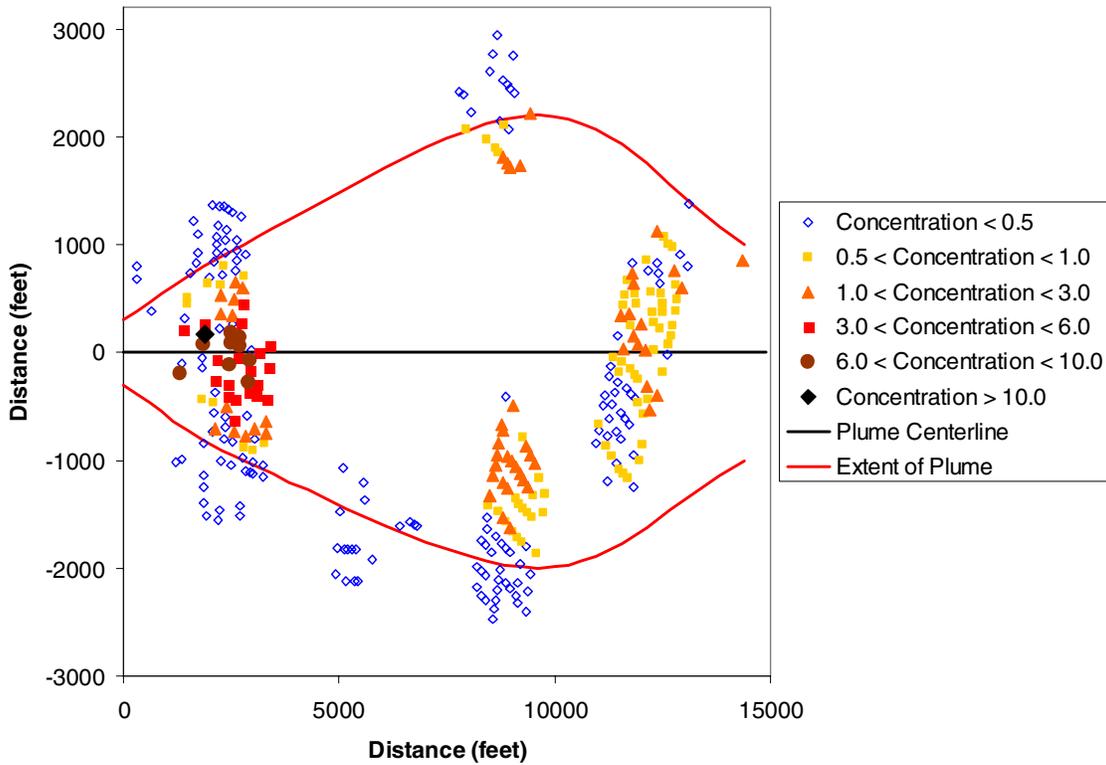


Figure 2.4. Location of measurement points for TCE in ppb in transformed coordinates

Locations close to the centerline of the plume and the source are designed to have higher concentrations, and are ideal for conservative representation of an area of concern. Although arrival times can be defined for each well location, the variation in arrival times within an area is small relative to the uncertainties associated with the problem. Therefore, the representative location for each of the four areas of concern as grouped during legal proceedings are selected as shown in Figure 2.5, and the corresponding coordinates are presented in Table 2.3.

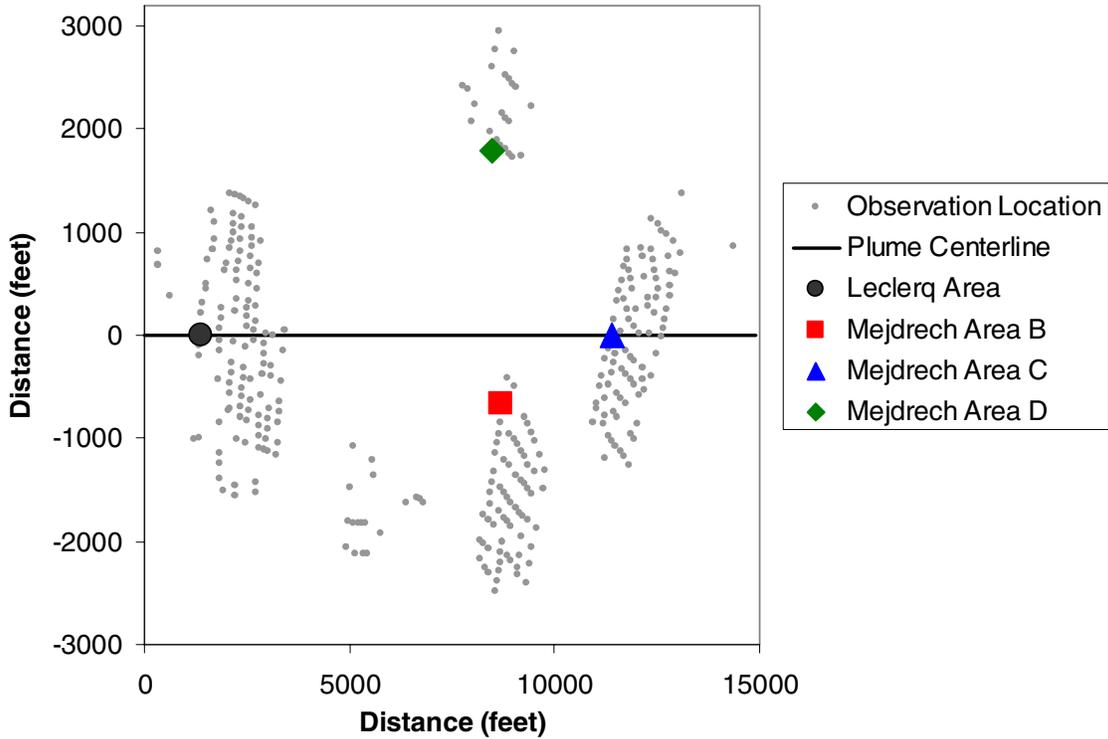


Figure 2.5. Representative locations for each area

Table 2.3. Representative locations for areas of concern

Area of Concern	x-coordinate (transformed)	y-coordinate (transformed)	z-coordinate (transformed)
Leclerq Class Area	1390	0	100
Mejdrech Area B	8707	659	100
Mejdrech Area C	11400	0	100
Mejdrech Area D	8500	1790	100

2.6.2 Local Trends

The observations are non-stationary in both time and space; thus descriptive measures such as means, medians, or standard deviations do not capture the potential for difficulties in the calibration process. However, valuable insight can be gained by examining local trends in terms of the magnitude and the distribution of observations in time and space. In Figure 2.4, the locations of the observations are primarily clustered in the four areas of concern: Leclerq Class Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D. Another cluster of observations, which are all NDs, is

located between the Leclerq Area and Mejdrech Area B near the plume boundary. Local trends are studied by examining the observations in each of these areas.

2.6.2.1 Leclerq Class Area

There are 118 observations of which 55% are non-detects (ND) in the vicinity of the Leclerq area. The corresponding concentrations range from non-detectable levels to 19.5 ppb. Five or 4% of the observations are below the previously-defined DL of 0.5 ppb with a range from 0.24 to 0.44 ppb. Only 9% of all observations associated with the Leclerq Class Area have concentrations below 1.0 ppb, and 45% of these observations are below 0.5 ppb. Thus, these observations are not likely to be anomalies and the DL of 0.5 ppb is likely to be incorrect for the observations in the Leclerq Class Area. The spatial distribution and a histogram of the TCE concentration levels in this area are presented in Figure 2.6 and Figure 2.7 respectively.

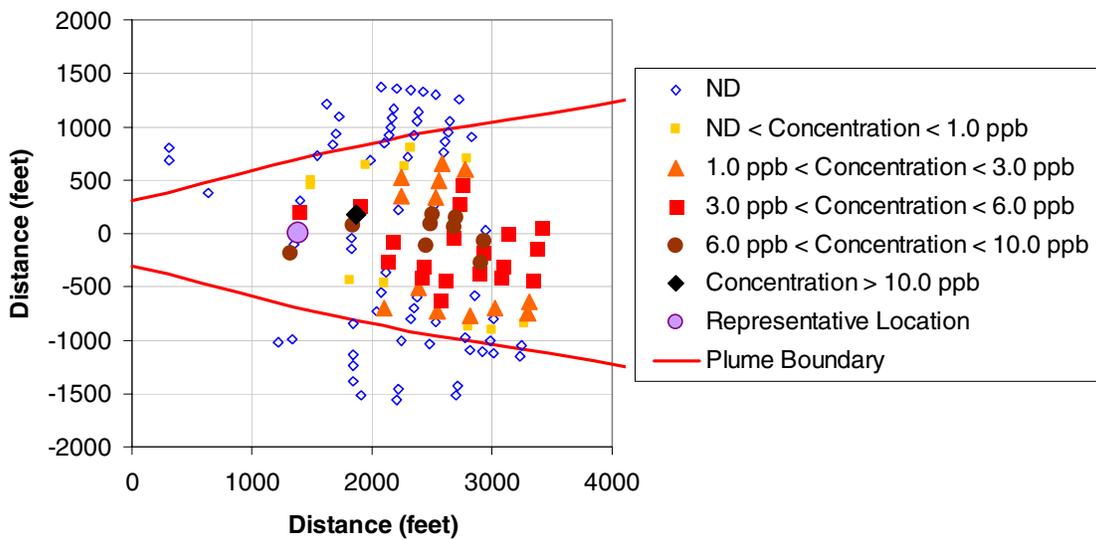


Figure 2.6. Measured TCE concentrations in and around the Leclerq Class Area in ppb

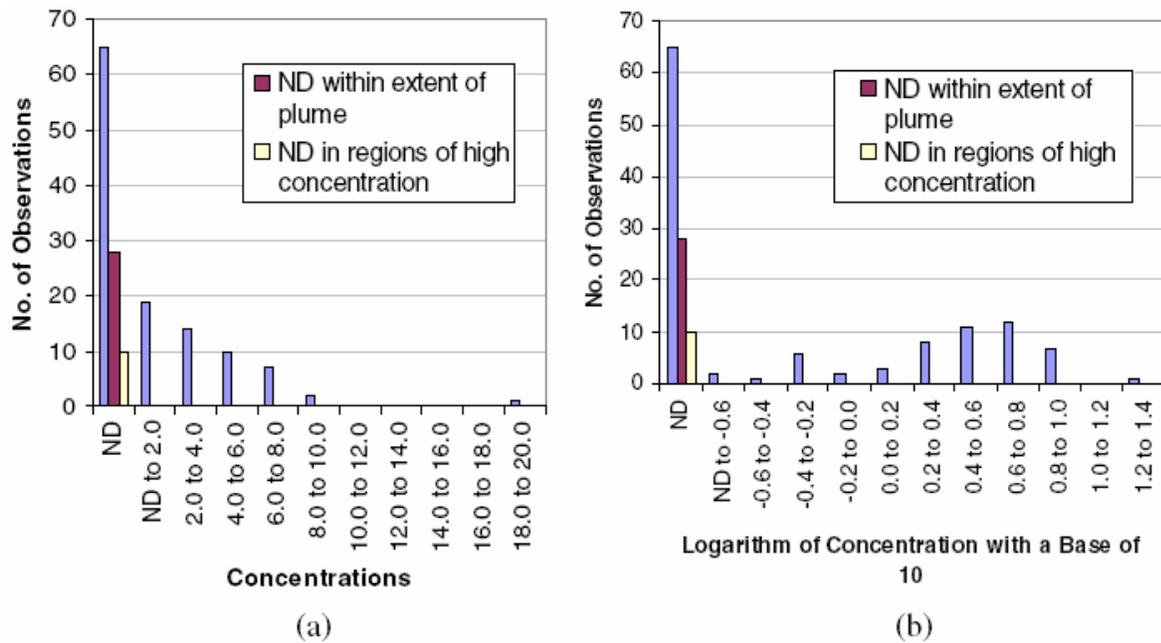


Figure 2.7. Histogram of TCE concentration levels in and around the Leclercq Class Area (a) in ppb (b) in $\log_{10}(\text{ppb})$

The histograms are created both with and without logarithmic transformations in Figure 2.7 since the effect on the estimation process is unknown. Misleading conclusions can be drawn by only examining the trends in the histogram due to the non-stationary nature of observations; thus, histograms are examined in conjunction with concentration distributions. Approximately 40% of NDs, which represents approximately 25% of the total number observations in the area, are located within the plume boundary as shown in Figure 2.6. Figure 2.7(a) shows that the number of NDs is significantly reduced but still high. Note that the intervals of the histogram are not consistent since the first and second intervals have ranges less than 2.0 ppb; the significance of the number of NDs are more pronounced in histograms with smaller interval widths, as shown in Figure 2.8. Of the 60% of NDs located within the plume boundary, the NDs of greater concern are those located in regions of high concentration. If we assume that regions of high concentration are areas with concentration greater than 3.0 ppb, only 15% of the NDs, which is only 8% of the total number of observations in the area, may be problematic. In fact, the number of problematic NDs is small in Figure 2.8 where only the NDs in regions of high concentration are considered. However, there are two NDs in the vicinity of the representative location, where high concentrations are expected; thus, these NDs may be a source of bias in arrival estimates based on the representative location. In addition, when logarithms are

applied to concentration values, the number of NDs in the plume is high relative to other intervals in Figure 2.7(b) even if only the observations within the plume boundary in regions of high concentrations are considered.

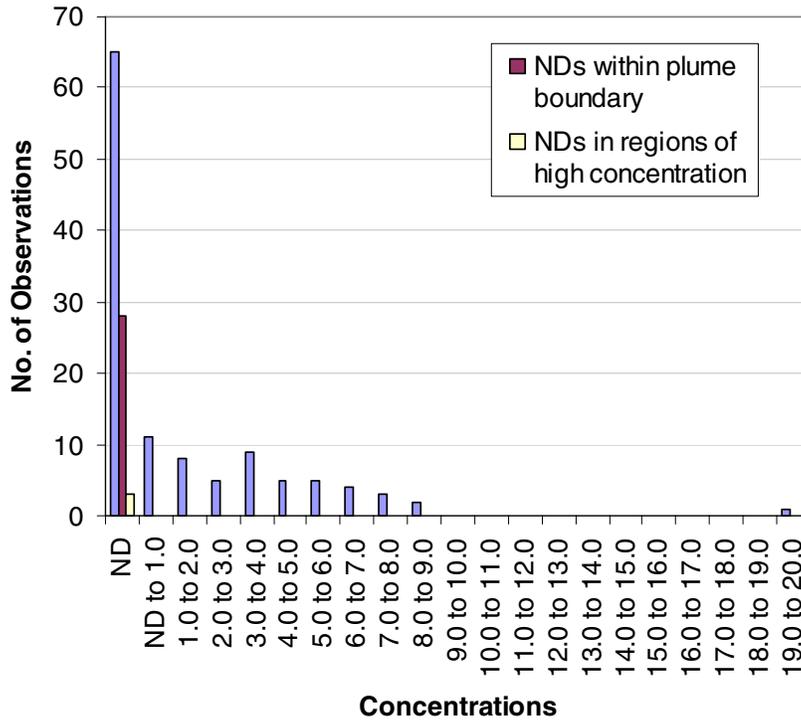


Figure 2.8. Histogram of TCE concentration levels in and around the Leclerq Class Area in ppb with smaller intervals

In addition to the concern caused by NDs, a possible outlier, caused by an observation with a concentration level of 19.5 ppb, is notable in Figure 2.7(a). The next highest concentration level is 8.3 ppb, which corresponds to a difference of 11.2 ppb from the possible outlier. The two closest observations to the possible outlier have concentrations of 3.1 and 7.4 ppb, and are separated by approximately 200 ft. The difference in concentrations typically found between observations with detectable concentration levels are approximately 1 ppb and 2 ppb for separation distances of approximately 100 ft and 200 ft respectively. Thus, the variance in observations in the proximity of the possible outlier appears to be higher, and the possible outlier may not be erroneous but a product of heterogeneity. This possible outlier is also located near the source and the centerline of the plume among observations of high concentrations. In addition, the possibility of this observation being an

outlier is significantly reduced if the logarithms of the concentrations are considered as shown in Figure 2.7(b). Therefore, the omission of this observation is not recommended. However, the possibility of this observation being an outlier still exists since the true distribution of the concentration levels is unknown.

Another concern is the proximity of the observation to the source zone since discontinuities in concentration exists in the developed model. However, the observation closest to the source zone is located at a distance of 644 ft from the edge of the source zone. Therefore, the observations are assumed to be at sufficient distances from the source zone to be insensitive to the discontinuity at the till-aquifer interface. This is not a concern in other locations which are located at greater distances from the source.

2.6.2.2 Mejdrech Area B

The spatial distribution and a histogram of the TCE concentration levels in Mejdrech Area B are presented in Figure 2.9 and Figure 2.10 respectively. There are 70 observations of which 46% are NDs in the vicinity of Mejdrech Area B. The corresponding concentrations range from non-detectable values to 1.5 ppb, which is much less than majority of concentrations observed in the Leclerq Class Area. There are no observations with concentrations less than 0.5 ppb other than those specified as NDs, and the DL of 0.5 ppb appears to be valid.

The local distribution of concentration levels in Mejdrech Area B presented in Figure 2.9 appears to disagree with the delineated TCE plume boundary. The plume boundary, also shown in Figure 2.4, is delineated to take into account the concentration distribution of all the observations. Twelve or 17% of the NDs in this area are located within the plume boundary; yet, only one ND is completely surrounded by observations with concentrations greater than the DL. This may indicate that the plume boundary or centerline needs to be revised. However, the plume boundary does not necessarily represent a DL of 0.5 ppb since the DL in the Leclerq Class Area is lower than 0.5 ppb. In addition, a smaller range in concentration levels exists, and a greater proportion of the NDs are likely to be close to the DL. Furthermore, a revision in the plume boundary can cause problems in other areas, and cannot be justified given the relatively small number of observations that are possibly problematic.

There are no obvious outliers and the difference between concentrations is relatively small in the vicinity of the representative location.

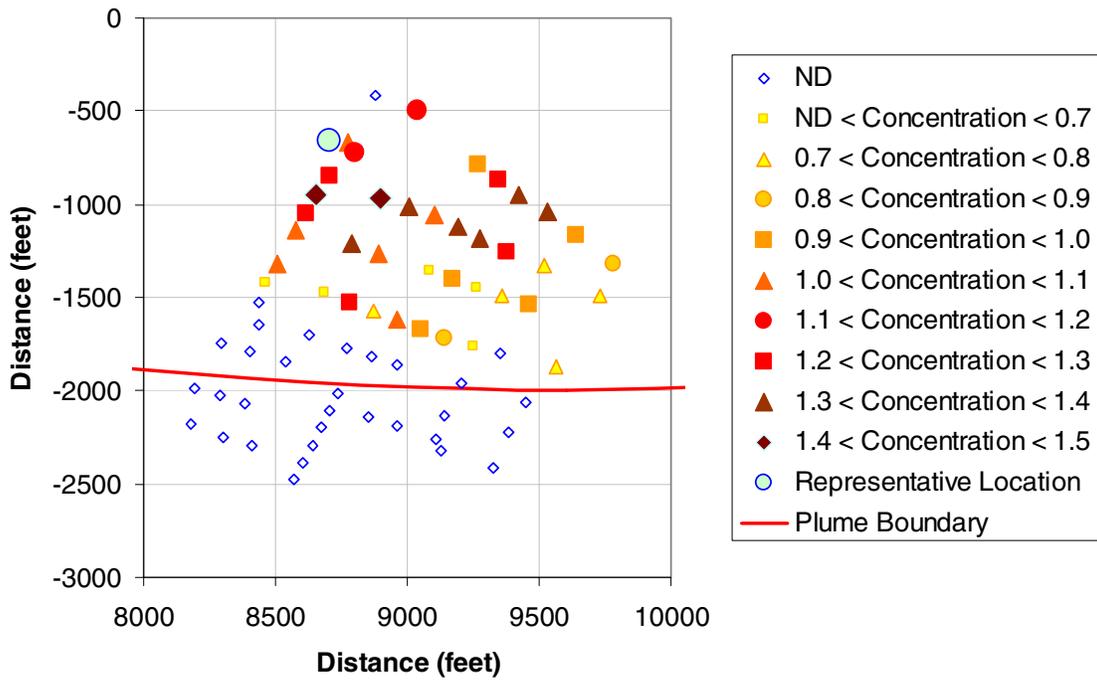


Figure 2.9. Measured TCE concentrations in and around Mejdrech Area B in ppb

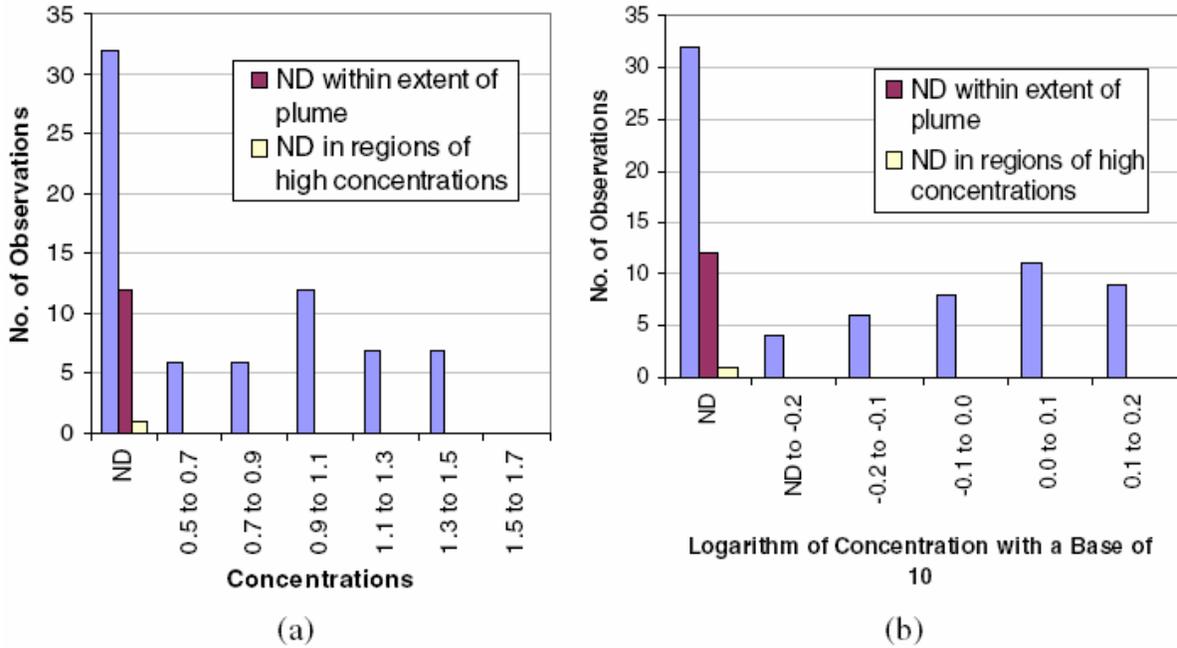


Figure 2.10. Histogram of TCE concentration levels in and around Mejdrech Area B (a) in ppb (b) in $\log_{10}(\text{ppb})$

2.6.2.3 Mejdrech Area C

The spatial distribution and a histogram of the TCE concentration levels in this area are presented in Figures 2.11 and 2.12 respectively. There are 91 observations of which 36% are NDs in the vicinity of Mejdrech Area C. The corresponding concentrations range from non-detectable values to 1.5 ppb, which is similar to Mejdrech Area B. There are also no observations with concentrations less than 0.5 ppb other than those specified as NDs, and the DL of 0.5 ppb appears to be valid.

Figure 2.11 shows that no obvious outliers exist and the number of NDs is significantly higher than the number of observations in any of the other concentration interval. Only one ND measurement lies outside of the extent of the plume. Although the DL corresponding to plume delineation is uncertain, the observations within the plume boundary should have concentration, at the least, greater than zero. No improvements are noticed with the application of logarithms.

There are many NDs below the representative location while there are many observations with detectable concentrations above the representative location. This pattern can be attributed to heterogeneities in the subsurface or an incorrect delineation of the plume's centerline. However, the range in detectable concentrations is small and these NDs may represent locations within the plume with concentrations lower than 0.5 ppb. In fact, there is a row of detectable concentrations below the area of NDs with low concentration values.

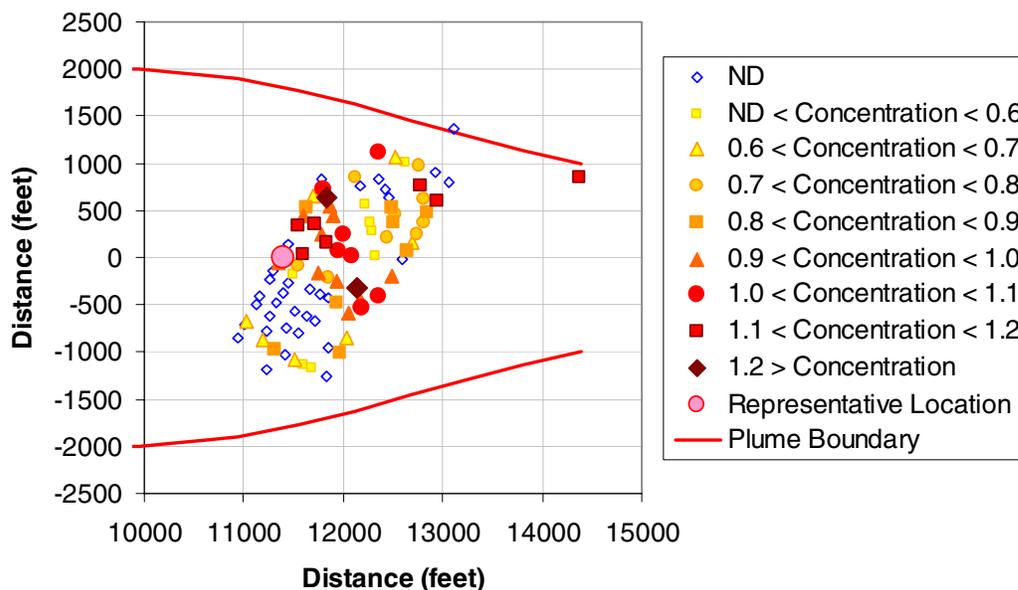


Figure 2.11. Measured TCE concentrations in and around Mejdrech Area C in ppb

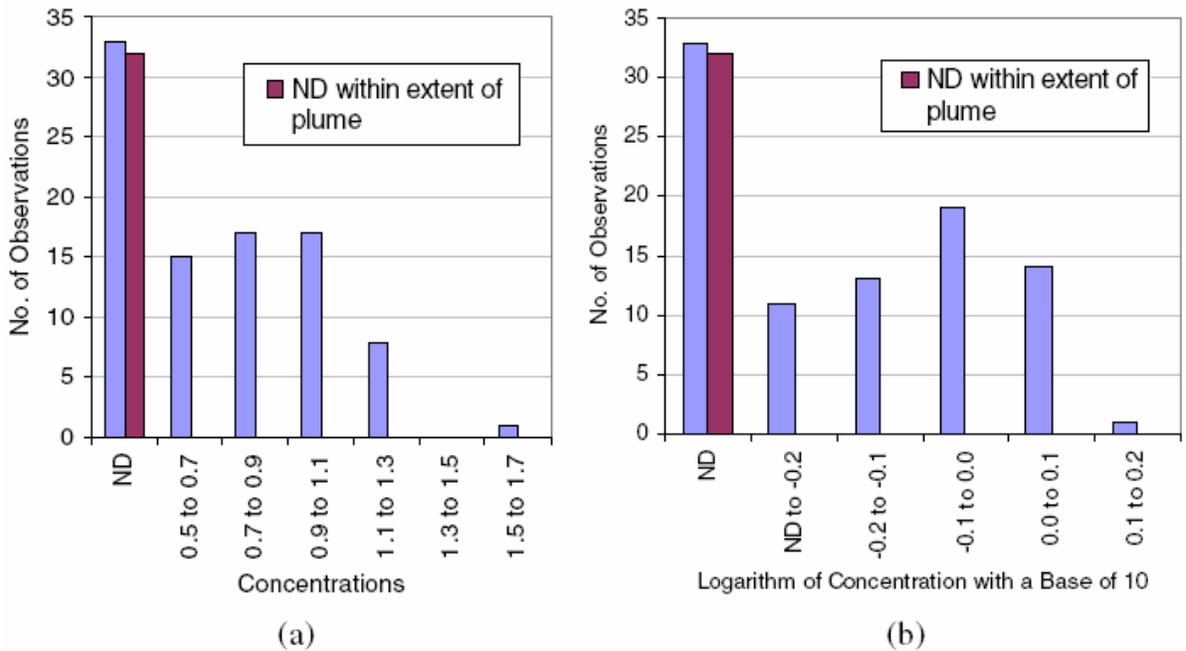


Figure 2.12. Histogram of TCE concentration levels in and around Mejdrech Area C (a) in ppb (b) in log₁₀(ppb)

2.6.2.4 Mejdrech Area D

The spatial distribution and a histogram of the TCE concentration levels in this area are presented in Figures 2.13 and 2.14 respectively. There are 23 observations of which 57% are NDs in the vicinity of Mejdrech Area D. The corresponding concentrations range from non-detectable values to 1.2 ppb, which is slightly smaller compared to Mejdrech Areas B and C. There are also no observations with concentrations less than 0.5 ppb other than those specified as NDs, and the DL of 0.5 ppb appears to be valid.

Most of the NDs are located outside of the extent of the plume. Only 2 out of 13 NDs, which correspond to 15% of ND measurements, are located within the plume. The NDs do not appear to be problematic in both distributions presented in Figure 2.14 if only the NDs in regions of high concentrations are considered. In addition, there are no obvious outliers.

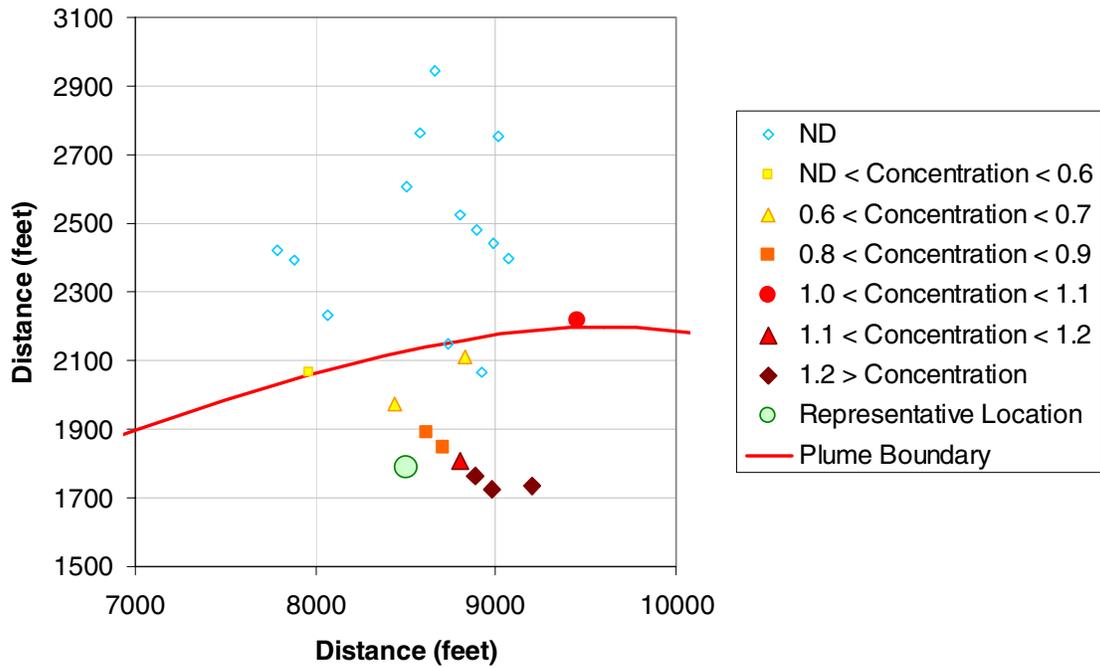


Figure 2.13. Measured TCE concentrations in and around Mejdrech Area D in ppb

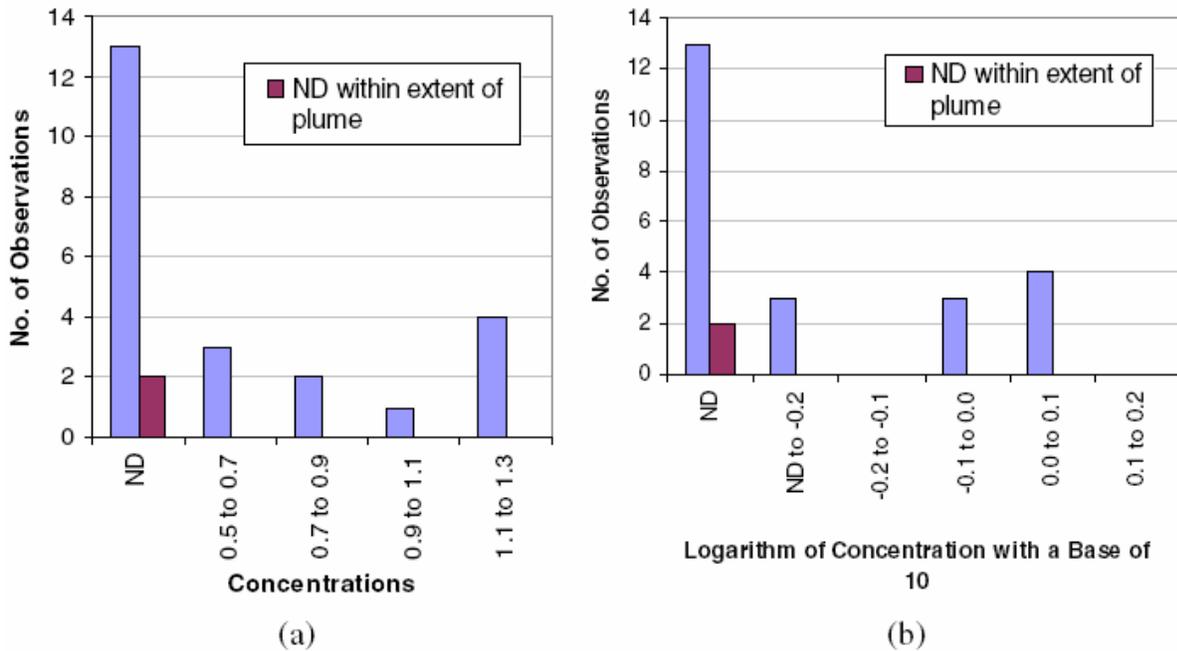


Figure 2.14. Histogram of TCE concentration levels in and around Mejdrech Area D (a) in ppb (b) in $\log_{10}(\text{ppb})$

2.6.2.5 Other Observations

There are 18 observations that are all NDs between the Leclerq Class Area and Mejdrech Area B as shown in Figure 2.4. Seven of these observations, which represent 39% of the observations in this cluster, are located within the plume boundary.

2.6.3 Insufficiencies in Data

Automatic calibration methods are generally designed to account for errors that are not strongly biased. However, five sources of bias that may affect the calibration process are present in the given data:

1. The presence of NDs
2. The presence of an extreme value (i.e., outlier)
3. The clustering of observations in space and time
4. The limitations in the type of observations
5. The limitations in the number of observations

Each of these sources is discussed below and possible treatments are recommended.

2.6.3.1 Presence and Treatment of ND Observations

Approximately 50% of the 320 observations are NDs; however, this percentage is below 75%, which USEPA (2000) declares high. Nonetheless, a significant proportion of the observations are NDs and their treatment during the calibration process is a concern. Applicable treatment approaches suggested by the USEPA (2002) are: partitioning of the data based on the conceptual model, the simple substitution method, and distributional methods. A simple substitution method is implemented by selecting a value from zero to the DL and replacing all NDs with this value. Typical values used are zero, half of the DL, and the DL. To better represent the range of possible values of NDs, a shape of the distribution based on the distribution of observations with detectable concentrations is assumed in distributional methods. If partitioning of the data based on the conceptual model is performed, a simple substitution method or distributional methods can be applied to each partition. Although histograms for clusters of observations have been created, they cannot be used to define the shape of the distribution since the data are non-stationary in both time and space; thus, the simple substitution method with partitioned data appears to be the only feasible approach.

To be consistent with the conceptual model, observations located outside of the plume boundary should have concentration values of zero or close to zero while observations inside the plume boundary should primarily consist of detectable concentration values. Based on this premise and the

discussions in Section 2.6.2, three types of ND observations can be established and are categorized as follows:

Type 1: The ND is surrounded by NDs.

The NDs of this type represent locations outside of the plume and the concentration level should be zero or very close to zero.

Type 2: The ND is surrounded by observations with concentration levels greater than the detectable limit but below 3.0 ppb.

The NDs of this type can be located inside or outside of the plume and can take on values between zero and the DL. The upper limit of 3.0 ppb is subjectively chosen based on the range of concentration levels to distinguish between high and low levels.

Type 3: The ND is surrounded by observations with concentration levels greater than 3.0 ppb.

The NDs of this type are most likely located inside of the plume and should be greater than zero.

Since the treatment of NDs should vary with type, a summary of the ND types for each locations are presented in Table 2.4.

Table 2.4. Numbers and percentages of ND observations according to its type

Type of ND Observations	Number of Observations						Percentage of Total Number of ND Observations	Percentage of Total Number of Observations
	Leclerq Area	Mejdrech Area B	Mejdrech Area C	Mejdrech Area D	Other	Total		
1	37	20	1	11	11	80	50%	25%
2	18	11	32	2	7	70	43%	22%
3	10	1	0	0	0	11	7%	3%
Total	65	32	33	13	18	161		

Table 2.4 shows that half of the NDs are of Type 1, which indicates that they are located outside of the plume and should have a concentration level of zero or very close to zero. Unnecessary biases can result if these observations are set to the DL or even half of the DL using the simple substitution method. A significant portion of NDs is categorized as Type 2 that can take on values between zero and the detection limit. For these observations, half of the DL appears to be a suitable substitution value. The NDs of Type 3, which only represents 7% of all NDs, can be substituted by either the DL or half the DL.

Although partitioning of the data may improve parameter estimations, the subjectivity of the partitioning strategy must be considered. The strategy used in this section, which was used to

generate the numbers in Table 2.4, are simply characteristics of this particular data set and may not be reproducible using observations with different times and/or locations. For example, many of the Type 2 NDs in Mejdrech Area B and the “other” or non-specified area may be of Type 1 since the delineation of the plume boundary is subjective. Therefore, a substitution of all NDs by a value of half the DL or the DL is likely to cause strong misleading biases, which will amplify problems caused by spatial clustering discussed in a subsequent section. A simple substitution of all NDs by zero is not guaranteed to create a strong systematic bias in the calibration process when there are only a few NDs within the plume boundary and even fewer NDs in regions where high concentrations are expected. If most of the NDs can be categorized as Type 1, the adoption of the simple substitution method using a concentration level of zero may be as justifiable as the implementation of a partitioning strategy and the simple substitution method.

2.6.3.2 Presence and Treatment of Extreme Value

The extremity of the observation with the concentration level of 19.5 ppb is debatable and its removal is not recommended. Nonetheless, the residual associated with this observation will have a significant impact on the performance indices of calibration methods, especially the popular least squares method. Therefore, its effect should be minimized using alternate performance indices. Since the logarithm of the value appeared to reduce the effects in Figure 2.7, performance indices based on logarithms should be considered.

2.6.3.3 Clustering Observations in Space and Time

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 and Townley, 1996). The observations provided for the case study are clustered in space and time.

2.6.3.3.1 Spatial Clustering

The locations of observations are primarily clustered in and around the four areas of concern. Since a larger amount of information in a certain spatial region biases the overall fit to observations in that region, this can be viewed positively since the purpose of modeling is to determine arrival times at the areas of concern. However, more than one area of concern exists, and problems arise from the fact that the numbers of observations in the areas are not consistent as shown in Table 2.5.

Table 2.5. Number and percentage of observations found in each location within the 2001/2002 plume boundary

Area of Concern	Regardless of Plume Boundary	
	Number of Observations	% of Total Observations
Leclerq Class Area	118	37%
Mejdrech Area B	70	22%
Mejdrech Area C	91	28%
Mejdrech Area D	23	7%

A significantly greater number of observations are provided for the Leclerq Area while a much smaller number of observations are provided for Mejdrech Area D. The number of observations is an indication of the certainty associated with the estimate for the location of concern. Therefore, it can be expected that uncertainty analysis will yield larger ranges of arrival times for Mejdrech Area D. Calibration approaches will be biased towards increasing the fit in the Leclerq area at the cost of decreasing the fit in the remaining areas, especially Mejdrech Area D. In addition, the distribution of parameters below the centerline (i.e. observations with negative y-coordinates) is more dispersed than above the centerline as shown in Figure 2.4; and, 61% of the observations are located below the centerline. Therefore, difficulties are likely to arise when attempting to model TCE transport to Mejdrech Area D.

An approach to eliminate this effect is to average the observations in such a way that each region has the same number of observations. The presumed accuracy of an averaged observation is related to the number of observations that the average is based on. Application of weights based on the accuracy of the observation is recommended if such information is available (Watermark Numerical Computing, 2004). However, such application can bring back the original bias that was removed by averaging. In addition the selection of an averaging strategy is subjective. An alternate treatment approach is the addition of soft data, the selection of which is highly subjective. Another option is to use a multi-objective optimization approach that attempts to minimize the deviation from observations for each location.

2.6.3.3.2 Temporal Clustering

The data available for this case study with three-dimensions is limited to observations made between 30.38 and 32.63 years, which correspond to 11,089 and 11,910 days, since the assumed first release of TCE. Only one measurement is made for a given location and the ranges in observation

times are small for observations in the same vicinity as shown in Figure 2.15.

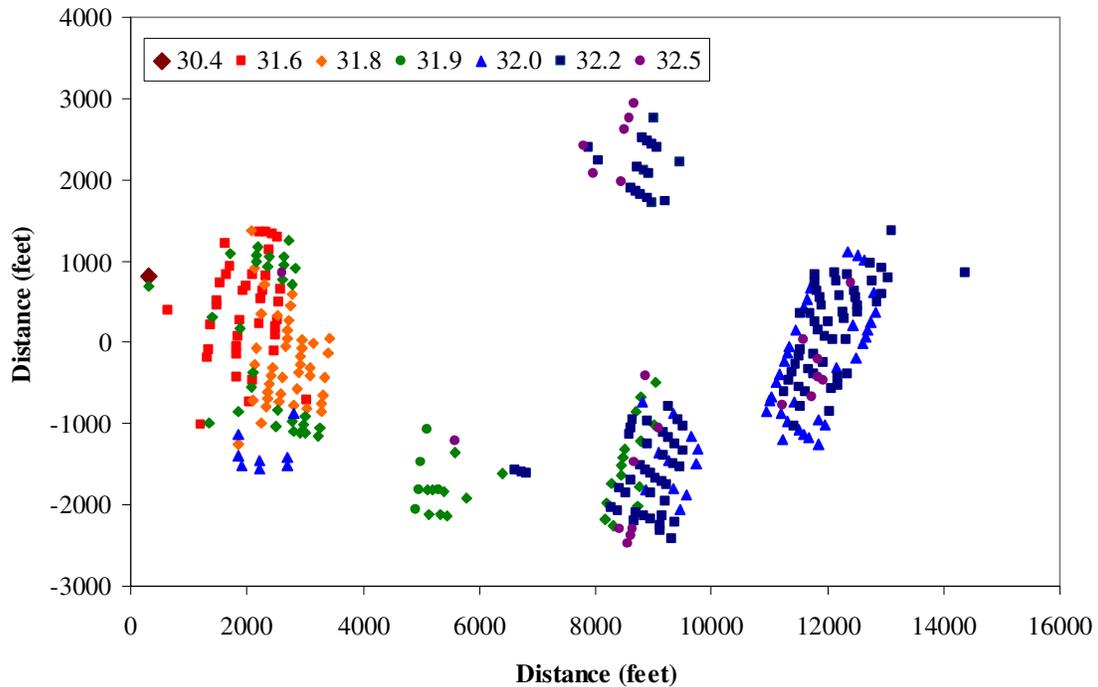


Figure 2.15. Location in transformed coordinates and time interval of observations

Temporal spacing of the observations is of paramount importance in this problem of estimating the first arrival time. Models that are similar at a specific point in time may be very different at other times. Therefore, a large range of possible arrival times are valid if the observations are clustered in time.

This type of data insufficiency is commonplace. The sources of many contaminated sites are often created in a time period in which monitoring for potential contamination is not performed and the potential for adverse health effects are not considered. Observations are made only once problems are suspected, which are often many years after first contaminant releases.

2.6.3.4 Limitations in the Type of Observations

Since the observations available for automatic calibration do not include observed parameter values, the parameter estimation process is performed indirectly with concentration values. The usage of various types of observation data are recommended for successful calibration (Hill, 2006). However, the option to collect additional data does not exist in this case study and alternative approaches to reduce uncertainty should be considered.

2.6.3.5 Limitations in the Number of Observations

The number of observations available is often limited in practice by cost constraints. The minimum number of observations required to reach a unique solution given a system of linear equations is equal to the number of parameters to be estimated. For a non-linear system, the solution space is complex and the number of observations must be greater than the number of estimable parameters to obtain a unique solution. The number of observations required can grow exponentially with increases in the number of parameters to be estimated. If the number of parameters to be estimated are large (i.e., greater than 10), 320 observations can easily be too small in our opinion. Given no alternate data types, a greater number of observations is required.

Chapter 3

Parameter Estimation

Models that simulate physical processes contain unknown parameters that require adjustments so that specific systems can be represented. Many of these parameters cannot be directly measured in the field and thus, need to be inferred from other types of observations. Parameter estimation methods attempt to determine a set of model parameter values so that selected model outputs match field observations. This is usually achieved by minimizing a performance index or an objective function that represents the ability of model results to match field observations; thus objective functions can be viewed as an indication of how “good” parameter estimates are given an observational data set. In general, parameter estimation processes consist of two components: the definition of the objective function and related constraints and the selection of a parameter estimation method. The parameter estimation process selected for the case study investigated here should address issues surrounding the observation data discussed in Chapter 2 and consider the ultimate goal of understanding the uncertainty of first arrival times to be analysed in Chapter 4.

For the given case study, the observations available for use in parameter estimation methods are trichloroethylene (TCE) concentration levels at 320 specific locations and times, as discussed in detail in Chapter 2. Five insufficiencies in the data were noted in Chapter 2: (1) the presence of concentration levels below the detectable limit or non-detects (ND), (2) the presence of an extreme value, (3) the clustering of observations in space and time, (4) the limitations in the type of observations, and (5) the limitations in the number of observations.

A simple substitution method with the concentration level of zero was selected to treat NDs. Substitution methods typically use a value of zero, the detection limit, or half of the detection limit (Singh et al., 2002). Examination of the proportion of NDs in this case study indicate that more of the NDs are likely to represent a value close to zero as opposed to the detection limit (DL) or half the DL; therefore, a simple substitution by the DL or even half the DL can lead to serious bias given the range and distribution of the concentration levels in the data set for this study. Partitioning of the NDs is not performed due to the subjectivity in the partitioning strategy, which is strongly dependent on the delineation of the plume boundary. Also, the uncertainty or the variability in the detectable concentration levels makes the existence of concentration levels of zero for the majority of non-detects a possibility.

The extreme value or outlier is handled by using robust estimators in the objective function definition. The option of simply removing this data point was not exercised since it is located in a region of high concentrations and may not be an outlier. Robust estimators are designed to be less sensitive to deviations in a small subset of the data (Finsterle and Najita, 1998; Wittmeyer and Neuman, 1991).

In our opinion, the problem of data clustering, primarily in the spatial sense, can also be alleviated indirectly by the usage of robust estimators. Robust estimators should reduce the contribution of the residuals in the Leclerq Class area that have relatively larger concentration levels to the parameter estimation process. Another solution to the bias caused by spatial clustering is to use multi-objective optimization that weighs the ability of a parameter set to produce outputs that match observations in each area equally. However, the implementation of both strategies may be redundant and a single objective optimization with robust estimators was used.

The clustering of the data in the temporal sense only conditions the portion of the concentration breakthrough curves that corresponds to the data. Therefore, flexibility exists in the remaining portion of breakthrough curves, and many different parameter sets can produce outputs that match the observations. As a result, the acceptance of multiple parameter sets is unavoidable given the data.

The best way to treat the remaining two data insufficiencies, which are the lack of different types and number of data points, is to collect additional hard data or to use prior knowledge. Resources to acquire additional hard data are not available, and therefore the incorporation of prior knowledge into the parameter estimation process must be considered. Since the creation of soft data based on prior knowledge is highly subjective, prior knowledge is integrated into the parameter estimation process by developing constraints. The availability of prior knowledge is limited in this case study, and only one criterion can be defined based on the information provided in Chapter 2. The criterion constrains the product of the horizontal velocity in the aquifer, v , and the longitudinal dispersivity in the aquifer unit, D_{x_s} , to be less than 250,000 ft²/yr, which is half of the product of the upper limits of the two parameters specified in Chapter 2. This criterion, which is referred to as the *Product Criteria*, is based on expert opinion that limits mechanical dispersion. Many other criteria that should be applied exist but are unknown, and, the treatment of these two insufficiencies using only one criterion is insufficient. Fortunately, the acceptance of multiple parameter sets, which is performed to treat temporal data clusters, can also account for these insufficiencies. In fact, the acceptance of multiple parameter sets can address all five data insufficiencies to some extent. If a large number of parameter sets are accepted, the chance of a parameter set that corrects for these potential problems is greater.

Parameter estimation methods used for groundwater and contaminant transport problems are primarily based on the principles of maximum likelihood and are not designed to find multiple possible solutions. Nonetheless, a review of these methods provides a good basis for selecting an appropriate method to estimate parameters for the given case study. Parameter estimation methods can be categorized as deterministic local optimization methods or global optimization methods. Gradient-based deterministic local optimization methods, such as PEST (Watermark Numerical Computing, 2004) and UCODE (USGS, 2005), are applied most commonly to groundwater flow and contaminant transport problems due to generally lower computation times compared to heuristics methods (Carrera et al., 2005). Mathematically, gradient-based methods are only guaranteed to find local optimums or to find global optimums in smooth convex problems (i.e. well-posed problems) (Gill et al., 1981). The global optimum represents the lowest possible objective function value while a local optimum represents the lowest objective function value in a certain portion of the parameter space. Typically, the goal of optimization is to find the global optimum but gradient-based methods can fail by only finding local optimums. The difficulty in finding the global optimum arises due to ill-posedness caused by non-uniqueness and instability, which are often encountered due to non-linearity and insufficient data (Kirchner, 2006). To overcome these issues, regularization techniques (Tonkin and Doherty, 2005) and pilot points (Moore and Doherty, 2006) are used to smooth out the local optimums by adding prior knowledge. Nonetheless, these techniques still do not guarantee the removal of all the local optimums, and the global optimum may still be difficult to find. The inability of gradient-based methods to guarantee a unique and globally optimal solution has led to the use of heuristic global optimization techniques such as genetic algorithms and simulated annealing (Carrera et al., 2005). The need for a relatively large number of model realizations which grows with the number of uncertain parameters for genetic algorithms and simulated annealing have led to hybrid techniques such as the “Shuffled Complex Evolution” (SCE) (Duan et al., 1992). SCE combines probabilistic and deterministic approaches using the concept of evolution and direct search algorithms, and has shown sufficient success to be accepted as the standard in surface water hydrology. Application of the SCE method to hydrogeology problems has recently been demonstrated by Agyei et al. (2006) using the same concept of evolution but with the efficient gradient-based Levenberg-Marquardt method (Gill et al., 1983). Thus, the SCE method appears to be the most promising approach for identifying a unique and globally optimal solution for difficult problems with complex response surfaces.

The parameter sets generated using methods based on the principles of maximum likelihood do not provide sufficient information on low likelihood outcomes. However, the importance of low likelihood outcomes is equal or possibly greater than maximum likelihood outcomes when assessing uncertainty. Therefore, an examination of multiple possible outcomes, which vary in the degree of likelihood, is necessary. Proponents of common parameter estimation methods based on maximum a posteriori framework have indirectly acknowledged the possibility of finding multiple “good” parameter sets. For example, Moore and Doherty (2005) state that uniqueness is not guaranteed using a gradient-based method with regularization as there are many ways in which the “preferred state” can be defined in the regularization operator and recommend catering calibration based on “the needs of the prediction”. Since the purpose of uncertainty analysis is to capture all “preferred states”, multiple parameter sets can be considered to be “good”, where “good” parameter sets are those that best describe a “preferred state”.

A direct acknowledgement and the consideration of multiple “preferred states” is made by Beven and Binley (1992) in their development of the Generalized Likelihood Uncertainty Estimation (GLUE) method. The GLUE method involves randomly generating a large number of parameter sets from uniform parameter distributions and using only the behavioural parameter sets in uncertainty analysis (Feyen et al., 2001; Beven and Binley, 1992). Behavioural parameter sets, which can be considered to represent “preferred states”, are defined in the GLUE framework as parameter sets with a performance index value below a subjective user-specified behavioural threshold. This threshold is generally determined using prior knowledge and preliminary investigations into possible performance index values. Lower values of the performance index indicate smaller differences between simulations and observations, and therefore are considered behavioural. For some problems, a sufficient number of behavioural parameter sets are difficult to find without arbitrarily increasing the behavioural threshold (Beven, 2006). This difficulty in finding behavioural parameter sets may indicate that the search methodology does not fully capture the parameter space, which grows with increases in the dimensionality of the problem. An increase in the behavioural threshold to identify more parameter sets is undesirable since the selection of behavioural threshold becomes arbitrary. To increase the number of multiple solutions of “good” quality, the Dynamically-Dimensioned Search – Approximation of Uncertainty (DDS-AU) method (Tolson and Shoemaker, 2006) optimizes numerous random initial solutions or parameter sets using a non-gradient-based local search algorithm. If the performance index is a measure of solution quality, each local optimum can be viewed as the “good” solution in its respective region of the parameter space.

The application of any gradient-based search algorithm on multiple initial solutions obtained using the multistart procedure (Duan et al., 1992) may generate results that are similar to DDS-AU results. In fact, there is debate regarding the robustness of the DDS-AU method in finding the global optimum and reductions in computational requirements compared to gradient-based multi-start methods (Yin, 2006). However, the purpose of this thesis is not to compare the DDS-AU method to other methods. The DDS-AU method is selected to estimate parameters because it provides a platform for finding multiple “good” solutions. In addition, the modifications required to accommodate different objective function definitions are more labour-extensive for gradient-based methods. Moreover, the application of the DDS-AU method to groundwater flow or contaminant transport problem has yet to be performed. Nonetheless, the functionalities of DDS-AU are selected based on preliminary comparisons to multiple implementations of PEST to justify its use. In addition, implementations of PEST provide valuable insight since a large amount of information is provided for each implementation.

In addition to the ability of the simulated values to resemble the observed values, the model results may need to satisfy additional requirements. In the given case study, concentration levels greater than the detection limit at specified locations must be simulated within 32 years; this requirement is referred to as the *Location Criteria*. Another requirement that an acceptable solution must meet is the *Product Criteria*, which was described above. Although there may be additional criteria that should be considered, they are not obvious and/or difficult to define; and thus, the effect of applying the two criteria are explored in conjunction with the need for additional criteria in this chapter.

Both criteria are handled by adding penalty functions to the objective function definition and/or using simple rejection during post-processing of the generated parameter sets. The penalty function approach is not guaranteed to find solutions that satisfy the corresponding criteria; therefore, post-processing is required regardless of whether a penalty function is used. Moreover, application of criteria via post-processing can also be useful in examining the effect of applying the criteria. Post-processing of the results using constraints can be viewed as determining its membership to the behavioural or non-behavioural category. Although the concept of behaviour has been primarily used in a binary sense in GLUE with membership dictated by the subjective behavioural threshold as mentioned above, the originators of the concept of behaviour, Hornberger and Spear (1981), did not limit the definition of behaviour by stating that it is “problem-dependent”. Therefore, the “occurrence or non-occurrence of the behaviour” can be defined as whether or not the two criteria are satisfied. To differentiate between these two definitions of “behaviour”, we refer to this new definition as

“acceptance” and the parameter sets that satisfy the two criteria are hereon called acceptable parameter sets. Even though a parameter set can be tested for both its “behaviour” and “acceptance”, the application of a behavioural threshold is abandoned due to its subjectivity and the use of the DDS-AU method. The DDS-AU method is designed to find solutions of “good” quality, and a filtering of unnecessarily high objective function values using a behavioural threshold is somewhat redundant.

In this chapter, multiple parameter sets are generated with the DDS-AU method using various objective function definitions and are tested for acceptance. Various definitions of the objective function or the performance index are used to generate a sufficient number of acceptable parameter sets to be used in the uncertainty analysis conducted in Chapter 4. Finally, the characteristics of parameter spaces are explored to infer the true nature of parameters in terms of their distribution of values and correlation structure. Different parameter spaces are investigated to assess the abilities of the approach to finding multiple acceptable parameter sets adopted in this chapter.

A consideration of the physical processes and the validity of the conceptual model are explored for selected parameter sets using breakthrough concentration profiles at locations of concern established in Chapter 2 and a plan view of the plume. A reference to contaminant arrival times defined by the arrival of a TCE concentration of 0.5 ppb is also made.

3.1 Objective Function Definition

The purpose of parameter estimation is to find the “true” value or the “true” solution given a data set and model containing errors and biases. Parameter estimation methods are generally optimization methods whose sole objective is to minimize or maximize the objective function or performance index value. Therefore, the nature of errors and biases must be captured by the objective function definition. Although the actual error and/or bias of a quantity associated with the model is unknown, it can be inferred by examining the difference between observed and modelled quantities, which is called a residual, r_i . In our case, these quantities are TCE concentration levels and the residual, r_i , is

$$r_i = z_i^* - z_i(\boldsymbol{\alpha}) \quad (11)$$

where z_i^* is the measured concentration at point i , and $z_i(\boldsymbol{\alpha})$ is the simulated concentration given a vector of parameters, $\boldsymbol{\alpha}$. The measured concentration, z_i^* , is composed of three parts (Finsterle and Najita, 1998) as given by

$$z_i^* = z_{i,true} + (e_d + b_d) \quad (12)$$

where $z_{i,true}$ is the true concentration at point i , e_d is the error associated with the data, and b_d is any bias present in the data. Similarly, the simulated concentration, $z_i(\boldsymbol{\alpha})$, given a vector of parameters, $\boldsymbol{\alpha}$, is

$$z_i(\boldsymbol{\alpha}) = z_{i,true} + (e_m + b_m) \quad (13)$$

where e_m is the error associated with the model, and b_m is any bias present in the model. Therefore, the residual represents all of the errors and biases associated with the problem as follows

$$r_i = (e_d + b_d) - (e_m + b_m) \quad (14)$$

Note that the signs associated with the errors and biases are unknown, and the residual can be viewed as a sum of all the errors and biases (Finsterle and Najita, 1998). The contribution of each component of the residual identified in (14) to r_i is unknown. Parameter estimates can end up compensating for model error, and extrapolations made using these estimates must be performed with care.

Nonetheless, the existence of different components assists in inferring the possible nature of the combined effect represented by the residual. For example, if one of the sources generate errors that are not normally distributed or biases that are systematic and their effect is known to be large, the combined effect of errors and biases of both the model and the data is not likely to be normally-distributed.

The objective function used in parameter estimation processes considers the sum of all residuals, and the definition of the objective function controls the relative influence of observations on the parameter estimates. Therefore, the components of objective functions (i.e. estimators and its additions) are introduced in this section.

3.1.1 Estimators

The most commonly-used definition of the objective function in groundwater flow and contaminant transport modelling is the Least Squares estimator, L_2 -estimator, based on the L_2 norm (Carrera et al., 2005). (The word, “estimator”, is used to represent the fact that these expressions are used to “estimate” parameters.) The L_2 -estimator represents the sum of the squared difference between measured and simulated TCE concentrations at specific locations and times as given by

$$L_2\text{-estimator} = \sum_i^{n_{obs}} r_i^2 \quad (15)$$

where n_{obs} represents the number of observations. (Thus, the value of the L_2 -estimator is also referred to as the sum of squares.) The objective function or the performance index of non-Bayesian methods such as the commonly-used nonlinear least squares maximum likelihood methods are “not derived

from probabilistic arguments... but is simply presented as a reasonable basis for estimation” (McLaughlin and Townley, 1996). The popularity of the least squares estimator stems from the fact that it is “equivalent to a Gaussian maximum likelihood estimator with known statistical properties” (McLaughlin and Townley, 1996). Thus, its popularity is based on the assumption that the nature of the combined effect of all errors and biases is normally distributed under the central limit theorem. However, the availability of observations is limited and many sources of error exist, and the cumulative effect of errors and biases can easily disobey normality even with a “perfect” conceptual model. In such situations, the validity of the L_2 -estimator is lost and a L_2 -estimator is as subjective as any other estimator; thus, a reasonable basis for estimation can be provided by many alternate definitions. In addition, the performance of the L_2 -estimator degrades with contamination in the data, and is not robust (Rousseeuw and Leroy, 1987). The L_2 -estimator is highly sensitive to outliers with a breakdown point of $1/n$, which indicates that a single outlier out of n observations can produce large deviations from the general trend of the remaining data. The contribution of this outlier to the total value of the L_2 -estimator is magnified by taking the square of the residual.

The least absolute values estimator based on the L_1 norm (L_1 -estimator) removes this magnification and represents the sum of the absolute residual values as

$$L_1\text{-estimator} = \sum_i^{n_{obs}} |r_i| \quad (3.6)$$

Although the effect of all outliers appears to be reduced, this approach is only robust for outliers in certain directions and the corresponding breakdown point is still $1/n$. Nonetheless, the effect of outliers in some directions is treated and the L_1 -estimator is still considered a robust estimator. In fact, improved results have been observed using the L_1 -estimator over the L_2 -estimator for some groundwater flow and solute transport problems (Xiang et al., 1994).

A drawback of the L_1 -estimator is the fact that the weight assigned to low residual values relative to higher residual values may be too high. In this case, the parameter estimation process attempts to further decrease low residual values at the expense of sacrificing the fit for observations with higher residual values (Boyd and Vandenberghe, 2003). Huber’s M-estimator (Rousseeuw and Leroy, 1987) attempts to resolve this issue by approximating the L_2 -estimator at low residual values while resembling the L_1 -estimator at high residual values. Figure 3.1 shows that Huber’s M-estimator provides a middle ground between the L_1 -estimator and the L_2 -estimator, jointly referred to as L-estimators, by achieving a reduction in the effect caused by large residuals while keeping the effect of

small residuals low. Huber's M-estimator can be represented by more than one shape, which is controlled by the k value in

$$\text{Huber's M-estimator} = \begin{cases} \sum_i^{n_{obs}} r_i^2 & |r_i| \leq k \\ \sum_i^{n_{obs}} k(2|r_i| - k) & |r_i| \geq k \end{cases} \quad (16)$$

As k approaches infinity, Huber's M-estimator behaves identically to the L_2 -estimator; in contrast, as k approaches zero, Huber's M-estimator behaves like the L_1 -estimator. The k values in Figure 3.1 are selected to uniformly represent the region between the two L-estimators.

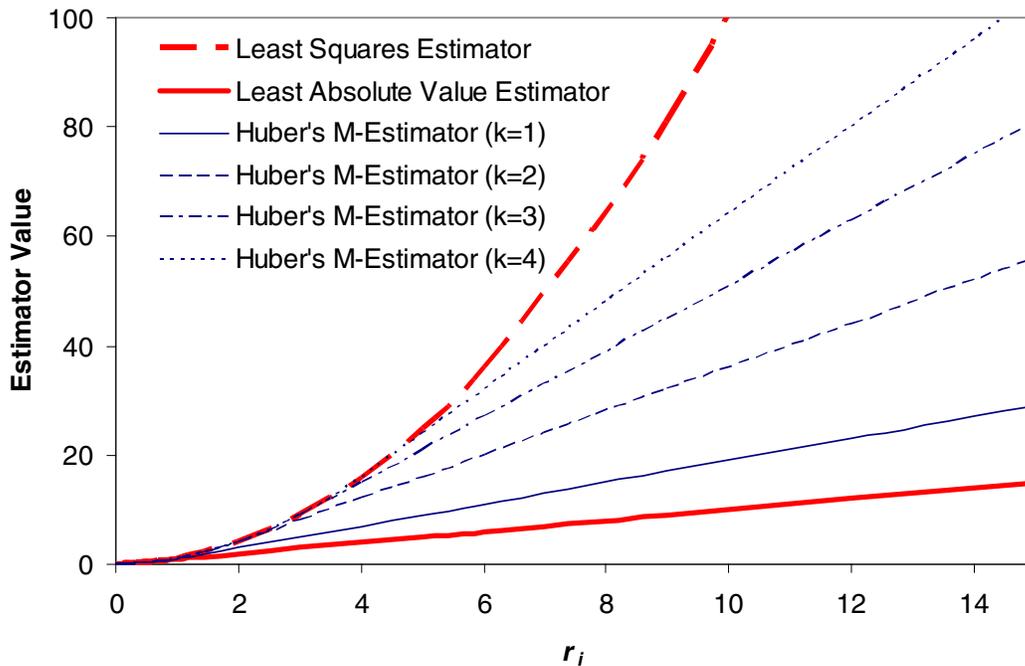


Figure 3.1. Comparison of Huber's M-estimator with L-estimators

In cases where bias caused by outliers is severe, the problem of relatively high weights to low residual values may be overshadowed by the large residual value associated with the outlier(s). Estimators based on a logarithmic distribution such as the Cauchy's M-estimator can achieve a further reduction of the relative weight of high residual values using

$$\text{Cauchy's Estimator} = \sum_i^{n_{obs}} c^2 \cdot \log \left(1 + \left(\frac{r_i}{c} \right)^2 \right) \quad (17)$$

where c is a subjective factor. Figure 3.2 shows that the contribution of each residual to the objective function using Cauchy's M-estimator is similar to the L_1 -estimator for small residuals, while the differences in weights for residuals with changes in its value are minimal for larger residuals. The differences in weights between large residuals are reduced with decreases in the c value. (The selection of c values used in Figure 3.2 is only based on their ability to span the space in the plot between the x-axis and the curve associated with the L_1 -estimator.) The optimal value of c is unknown and thus, a value of 1, which is assumed by Finsterle and Najita, 1998, is adopted. The removal of the power of two applied to the r_i appears to have negligible effect on the estimator for a c value of 1 as shown in Figure 3.3. There is a small difference for low r_i values, which can be significant if there are many low r_i values. This revised Cauchy's M-Estimator is referred to as the LRS1 M-estimator.

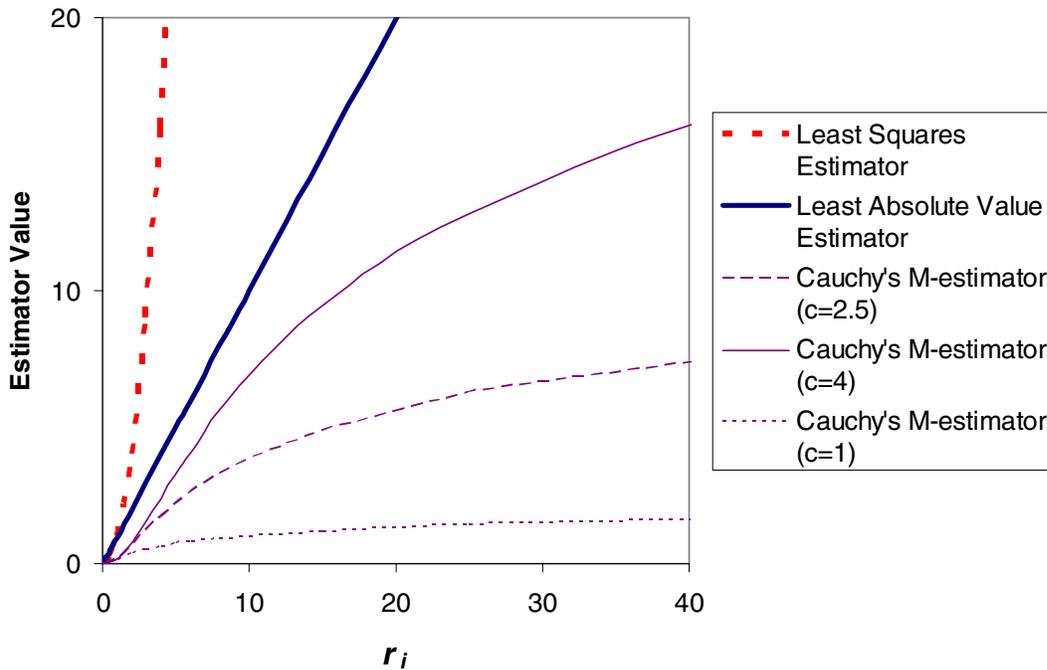


Figure 3.2. Comparison of Cauchy's M-estimator with L-estimators

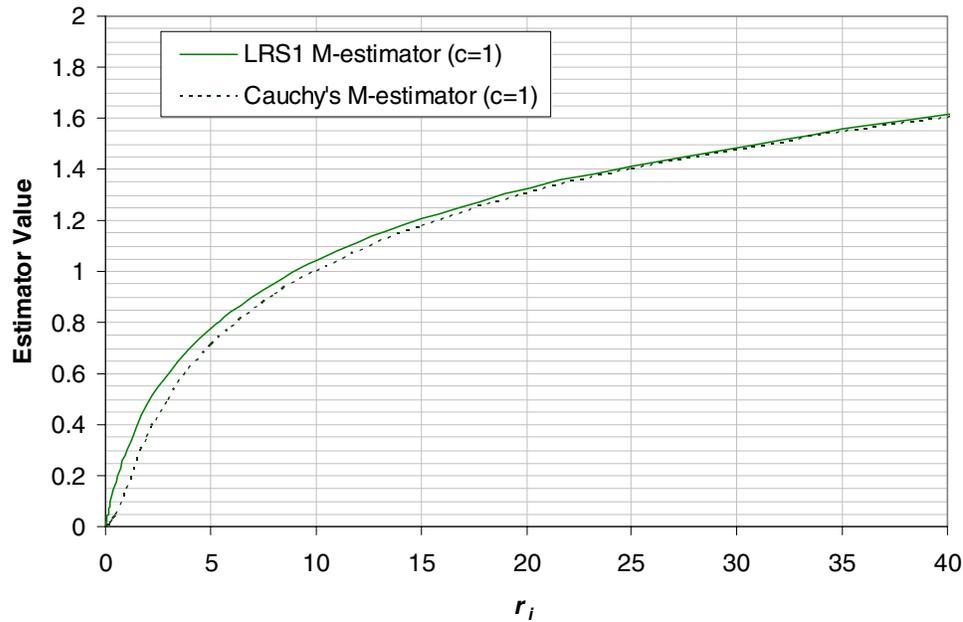


Figure 3.3. Comparison of Cauchy's M-estimator with c value of 1 with revised Cauchy's M-estimator with c value of 1

Huber's M-estimator, Cauchy's M-estimator, and LRS1 M-estimator all belong to a family of robust estimators referred to as M-estimators. M-estimators, which include the L-estimators, express the contribution of a residual with a symmetric, positive-definite function with a unique minimum at zero, $\rho(r_i)$, as given by

$$\text{General M-estimator} = \sum_i \rho(r_i) \quad (18)$$

The breakdown point for M-estimators is $1/n$; but improvements over L-estimators in terms of their ability to reduce the effect of specific types of outliers exist using Huber's, Cauchy's, or LRS1 M-estimators (Rousseeuw and Leroy, 1987).

Since the robustness of an estimator is problem-dependent, the two L-estimators and three M-estimators, which are summarized in Table 3.1, are used in parameter estimation for this case study. Each of these estimators provides a different way to weight the residuals based on the value of the residual.

Table 3.1. Estimators used in parameter estimation process

Description	Type	Abbreviation	$\rho(r_i)$
Least squares or L_2 norm	L-estimator	L_2	r_i^2
Least absolute value or L_1 norm	L-estimator	L_1	r_i
Huber's M-estimator	M-estimator	Huber	r_i^2 $ r_i \leq k$ $k(2 r_i -k)$ $ r_i \geq k$
Logarithm of the sum of the residual and 1	M-estimator	LRS1	$\log(r_i+1)$
Cauchy's estimator	M-estimator	Cauchy	$c^2 \cdot \log(1+(r_i/c)^2)$

Typically, robust estimators are also described using influence functions which represents the slope of the ρ function listed in Table 3.1. Influence functions are not examined here since gradient-based optimization methods are not used. In fact, gradient-based methods are not appropriate for objective function definitions using Cauchy's M-estimator or the LRS1 M-estimator due to descending first derivatives.

3.1.2 Add-ons to Estimators

The problem of relatively large weights applied to low residuals exists for all M-estimators and the L_1 -estimator. A simple approach to circumvent this problem is to create a deadzone, which may be achieved by assigning a weight of zero to low residuals (Boyd and Vandenberghe, 2003) as shown by

$$\text{General Estimator with Deadzone} = \begin{cases} 0 & |r_i| \leq d \\ \sum_i \rho(r_i) & |r_i| > d \end{cases} \quad (19)$$

The application of a deadzone as described by (19) can be applied to any objective function definition. It is important to note that the application of a deadzone leads to the presence of non-unique minimums, which is not problematic if equifinality (Beven, 2006) is assumed.

The two criteria, the *Location Criteria* and the *Product Criteria*, imposed on all parameter sets generated can be included in the objective function via the addition of a penalty function. The limiting criteria is the *Location Criteria* and the corresponding penalty function is defined as follows

$$\text{Penalty Function} = \begin{cases} 0 & c_j \leq c_d \\ S \cdot \sum_j^{n_{loc}} f(c_j, c_d) & c_j \geq c_d \end{cases} \quad (20)$$

where S represents a user-specified, problem-dependent scaling factor, n_{loc} is the number of locations of concern, j is the index representing a location of concern, c_j is the modelled concentration level at j ,

c_d is the desired concentration level at j , and $f(c_j, c_d) = |c_j - c_d|$. (Note the flexibility of using other functions to compare c_j and c_d .) The *Product Criteria* can be incorporated in the same manner as follows

$$\text{Penalty Function} = \begin{cases} 0 & P \leq P_{\max} \\ S \cdot f(P, P_{\max}) & P \geq P_{\max} \end{cases} \quad (21)$$

where P is the product of v and D_v for the parameter set, P_{\max} is 250,000 ft²/yr, and $f(P, P_{\max}) = P - P_{\max}$.

The use of either penalty function requires the identification of an appropriate S . The addition of a penalty function to the objective function changes the response surface to facilitate the search of parameter sets that satisfy the criteria of concern. However, the objective function value can no longer be viewed as only the combined effect of errors and biases.

3.2 Parameter Estimation Methods

Parameter estimation methods based on very different viewpoints for calibrating parameters are adopted in this thesis: the gradient-based local search algorithm, PEST, designed to find one “good” solution and a stochastic search algorithm, DDS-AU, designed to find multiple “good” solutions. A description of each method is followed by the reasoning behind its selection and implementation strategy.

3.2.1 PEST

3.2.1.1 Description

PEST is a gradient-based optimization approach that uses the efficient Levenberg-Marquardt scheme and is thus a local search method. PEST is designed for L_2 -estimators and manipulations must be performed to incorporate other estimators. There are 32 algorithm parameters associated with PEST to provide added flexibility and overcome shortfalls of the optimization strategy. PEST also offers numerous regularization options such as Tikhonov regularization, truncated singular value decomposition, and a combination of the two methods. Other options include logarithmic parameter transformations and observation weighting. Details on all of the functionalities associated with PEST are given by Watermark Numerical Computing (2004).

3.2.1.2 Purpose

The Parameter Estimation (PEST) package appears to be one of the most commonly used calibration tools in groundwater flow and contaminant transport modelling problems due to its presence in modelling packages such as Visual MODFLOW (Waterloo Hydrogeologic, 2006). Since it is a well-accepted method, the use of new alternative methods is only justifiable if its performance is superior or at the least equal to PEST. Therefore, PEST is primarily used to perform a preliminary investigation of the optimization problem so that a basis for evaluating the alternative method (i.e. the DDS method) can be established. The two bases for comparison are: the objective function values since they indicate the quality of the estimate; and the number of function evaluations required for convergence since it indicates the amount of computational resources required.

3.2.2 DDS-AU

3.2.2.1 Description

The Dynamically Dimensioned Search – Approximation of Uncertainty (DDS-AU) algorithm developed by Tolson and Shoemaker (2006) is designed to find multiple good global solutions within the specified parameter ranges by using the objective function value and random initial parameter sets. This algorithm is independent of gradients and sensitivities, and is a heuristic global search algorithm targeting problems with a high number of local optima.

Parameter sets are initialized by selecting the best of 10 parameter sets whose values are randomly sampled from uniform distributions. A user-specified number of trials starting with the initial parameter sets are implemented for each DDS-AU run. For each trial, a probabilistic search strategy that is global in initial iterations but becomes more local in the final iterations is applied. Objective function values are used during optimization but there is no preferred estimator. Optimization for each trial is stopped once the user-specified number of function evaluations has been reached. Although this does not guarantee convergence to precise estimates, applying restrictions on the number of function evaluations is desirable given the computational effort required to run the model and the need to generate many simulations.

There is only one algorithm parameter, the scalar neighbourhood size perturbation parameter, and the default value of 0.2 recommended by Tolson and Shoemaker (2006) sufficiently serves most purposes. The DDS-AU algorithm is also dependent on a random seed value that is assigned for each run.

3.2.2.2 Purpose and Implementation Strategy

The DDS-AU method is used to generate multiple parameter sets of “good” quality and to compare various objective functions. Although the DDS-AU method is designed to fulfill its purpose, it is a novel method and its use must be justifiable. Therefore, a basic comparison with PEST results is carried out. PEST results are also used to set the number of function evaluations per optimization trial in the DDS-AU method.

At least 1000 acceptable parameter sets are assumed to be required to adequately represent the acceptable parameter space. Therefore, more than 1000 DDS trials must be implemented and many function evaluations are required. The increase in required computational effort is met by using the high performance computing (HPC) network, SHARCNET. Multiple serial DDS runs with 100 trials each were submitted instead of a single parallelized run with a high number of trials. This is assumed to create a more thorough search of the solution space since the random seed can be changed for each DDS run. The submission of multiple runs to different SHARCNET processors and compilation of output files are automated using script files. Note that if serial jobs are submitted simultaneously to different processors the number of trials that can be generated is dependent on the number of available nodes for a given amount of time and hence the size and dynamics of the queue.

The effects of various estimators with and without different add-ons on the resulting sets of parameter sets is compared based on their ability to satisfy the two criteria, the *Location Criteria* and the *Product Criteria*, since the case study represents a real problem and “true” solutions are unknown. This ability is measured by the acceptance rate, which is the percentage of trials that satisfy the two criteria out of the total number of trials.

3.3 Parameter Estimation

There are 18 adjustable model parameters; however, only 10 parameters are selected for optimization based on prior knowledge and the degree of uncertainty of the parameters, which were both discussed in Chapter 2. Tables 3.2 and 3.3 list values of the fixed parameters and ranges of the parameters to be estimated respectively.

Table 3.2. Adjustable model parameters not estimated using a parameter estimation method

Description	Abbrev.	Values	Units
Source Length	w_x	150	ft
Source Width	w_y	150	ft
Retardation in overburden	R_t	1.5	n/a
Molecular diffusion times tortuosity	μ_r	0.0	ft ² /yr
Porosity in overburden	n_o	0.2	n/a
Retardation in limestone	R_l	1.5	n/a
Porosity in limestone	n_a	0.05	n/a
Limestone thickness	w_z	100.0	ft

Table 3.3. Adjustable model parameters estimated using a parameter estimation method

Description	Abbrev.	Ranges	Units
Initial concentration	c_0	100.0 – 2000.0	mg/L
Horizontal velocity in limestone	v	100.0 – 1000.0	ft/yr
Longitudinal dispersivity in limestone	D_x	10.0 – 500.0	ft
Transverse dispersivity in limestone	D_y	1.0 – 50.0	ft
Vertical dispersivity in limestone	D_z	0.05 – 5.0	ft
Decay of dissolved TCE in overburden	μ_t	0.00001 – 1.0	1/years
Source reduction rate	λ	0.00001 – 1.0	1/years
Vertical velocity in overburden	v_y	1.0 – 40.0	ft/yr
Dispersivity in overburden	D_v	0.001 – 20.0	ft
Depth to top of limestone	b_t	0.0 – 80.0	ft

3.3.1 PEST Results

3.3.1.1 Preliminary Investigations

A preliminary investigation of the parameter estimation problem was performed using an ad hoc approach with approximately 100 implementations of PEST using 5 different initial parameter sets. Five preliminary observations are made:

1. Any effects caused by fine-tuning of PEST's algorithm parameters are overshadowed by its dependence on the initial parameter set. Also, the need for fine-tuning of a large number of algorithm parameters is undesirable from the perspective of automatic identification of multiple possible parameter sets. Therefore, the algorithm parameters of PEST used in the optimization are the values or midpoint of the range of values recommended in the PEST manual (Watermark Numerical Computing, 2004). These values are listed in Appendix C.

2. Regularization using the *Location Criteria* has minor effects on the objective function value and the parameter values; and such regularization attempts do not convert an initially unacceptable parameter set to an acceptable parameter set. Regularization given a priori knowledge of the relationship between the horizontal velocity in the overburden, v_v , and the longitudinal dispersivity in the aquifer, D_x , is not implemented since the estimated parameter sets obtained during preliminary investigations satisfy the *Product Criteria*. It is important to note that the optimized parameter sets are generally able to satisfy the *Product Criteria* if the initial parameter values satisfy the *Product Criteria*. The opposite also appears to be true. Therefore, regularization using either of these criteria appears to be unable to effectively smooth out local minimums in the response surface and arrive at major optimums.
3. Tests of various weighting strategies, which are developed based on the proximity and number of observations in its neighbourhood, indicate that the effect on the response surface can be significant. Therefore, highly subjective weights should not be applied and equal a priori weighting of observations is more appropriate for the given problem with no supplemental information on the quality of the observations.
4. Logarithmic transformations of selected parameters can aid in finding acceptable parameter set. This can be explained by the fact that the nature of parameter distributions is not likely to obey normality assumptions. However, the strategy to select parameters to be transformed is not obvious.
5. Parameter sets that satisfy both criteria are difficult to find. This is primarily due to the difficulty in reaching a concentration equal or greater than the detection limit at the representative location for Mejdrech Area D, which is one of the four requirements in the *Location Criteria*. This difficulty is expected given the limited number and low magnitude of observed concentrations associated with this area.

These preliminary findings provide insight into the response surface and indicate the importance of initial parameter sets and parameter transformations.

Of the approximately 100 implementations of PEST performed during the preliminary investigations, only three estimated parameter sets satisfy both criteria and can be classified as acceptable. The acceptable parameter set with the lowest objective function value is listed in Table 3.4.

Table 3.4. Estimated parameter set that satisfies both constraints

Parameter	Units	Parameter estimate	Parameter distribution
c_0	mg/L	1000.0	Normal
v	ft/yr	606.1	Normal
D_x	ft	35.3	Lognormal
D_y	ft	38.5	Lognormal
D_z	ft	2.5	Lognormal
μ_t	1/years	3.9E-04	Lognormal
λ	1/years	6.4E-02	Lognormal
v_v	ft/yr	3.6	Lognormal
D_v	ft	1.1	Lognormal
b_t	ft	14.4	Normal

Note: The corresponding sum of square value is 688.4 ppb².

The corresponding deterministic breakthrough concentration profile at the four locations of concern is presented in Figure 3.4. The corresponding first arrival times are 3.8 years, 19.5 years, 24.9 years, and 20.6 years for Leclerq Class Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D respectively. The corresponding plume in plan view at different depths after 33 years is depicted in Figure 3.5. Figure 3.5 indicates that the plume begins to separate into two possibly due to the reduction in the release of TCE at the source with time.

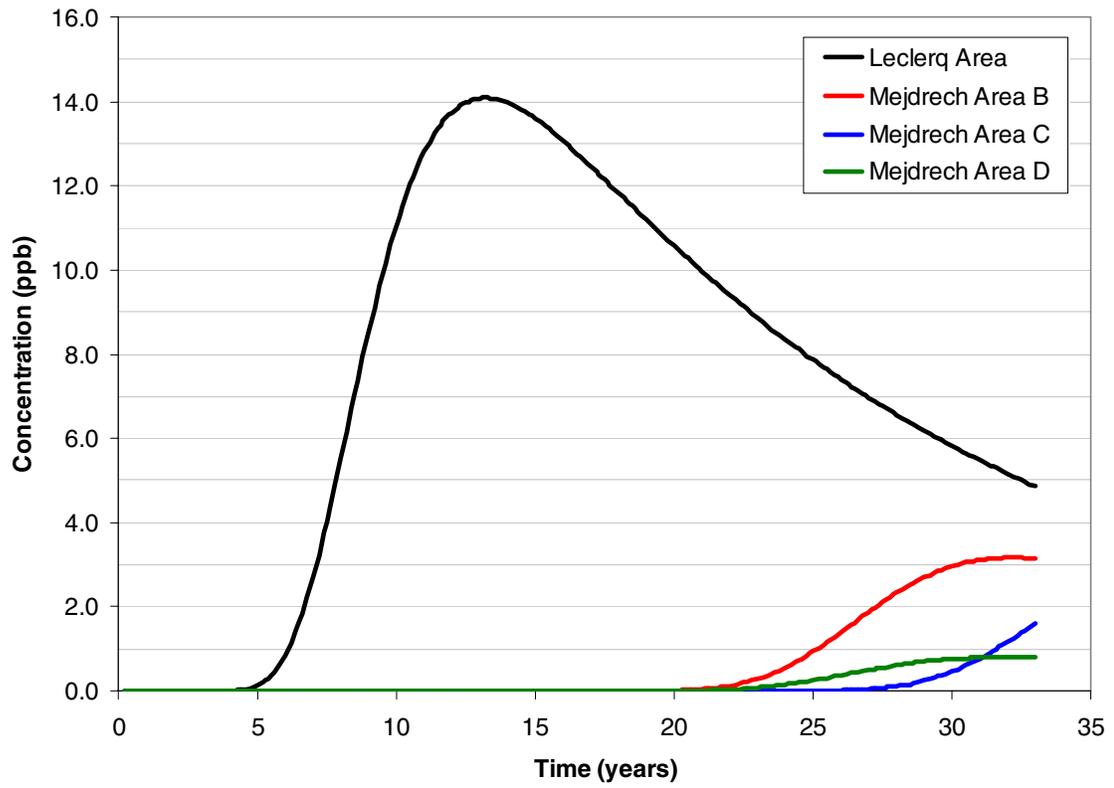


Figure 3.4. Breakthrough concentration profiles at the four locations using the acceptable parameter set presented in Table 3.4

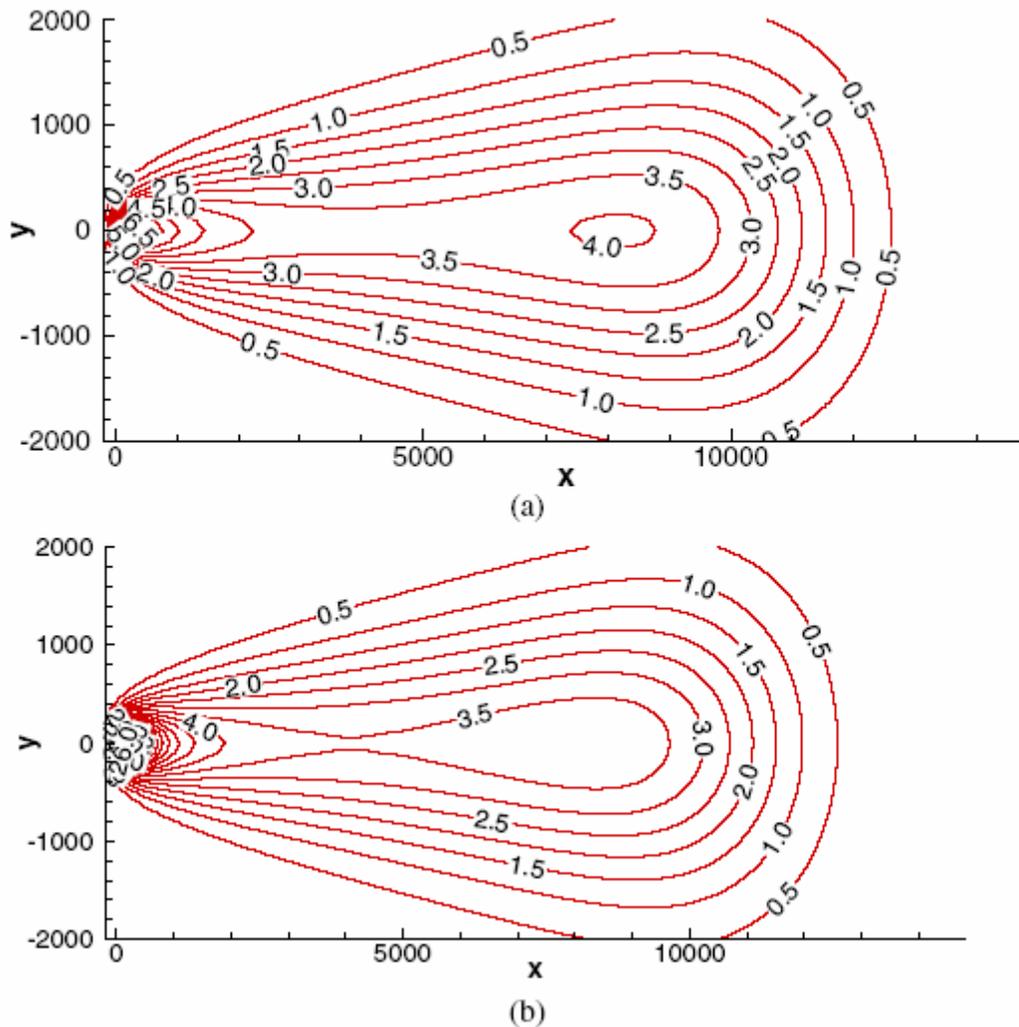


Figure 3.5. Plan view of the TCE plume in ppb after 33 years since the opening of the Metcoil facility at a depth of (a) 50 ft (b) 100 ft from the defined bottom of the limestone using the acceptable parameter set presented in Table 3.4

3.3.1.2 Estimation

Since preliminary investigations show a large dependence on the initial parameter set, 22 different initial parameter sets listed in Table 3.5 are optimized using PEST. Initial parameter sets 1 to 20 are found by scaling random numbers that have been randomly selected from uniform distributions that range from 0 to 1 with parameter ranges specified in Table 3.3. Initial parameter sets 21 and 22 are two of the initial parameter sets used in preliminary investigations. Note that two of the initial random parameter sets, 4 and 9, do not satisfy the *Product Criteria*.

Table 3.5. Random initial parameter values

Parameter Set	c_0	ν	D_x	D_y	D_z	μ_t	λ	ν_v	D_v	b_t
1	1420	599	250	35	1.1	0.02	0.0007	12	7.9	29
2	199	809	50	12	3	0.002	0.0017	22	17	49
3	513	810	67	47	3.6	0.992	0.121	19	18	58
4*§	480	694	442	29	1.3	0.713	0.867	22	2	11
5	1294	155	77	35	3.0	0.046	0.538	21	19	10
6	1061	840	185	16	4.3	0.100	0.948	27	5	32
7	994	475	477	9	0.2	0.164	0.857	3	10	10
8*	1039	186	427	21	1.2	0.710	0.554	32	6	19
9*§	1384	644	479	44	0.1	0.557	0.995	6	4	26
10	700	307	55	1.5	2.1	0.621	0.124	5	9	27
11*	453	137	20	14	2.7	0.376	0.909	1	7	18
12	857	158	165	29	2.1	0.191	0.504	32	19	16
13	1623	293	172	6	4.5	0.081	0.002	12	2	59
14	319	545	115	46	2.6	0.458	0.284	28	11	69
15	1105	286	152	3	0.8	0.481	0.376	23	16	36
16	138	661	214	36	3.8	0.262	0.328	22	13	33
17	229	179	359	43	1.6	0.007	0.609	14	18	70
18*	1725	370	296	5	3.6	0.491	0.554	30	1	76
19*	681	774	124	25	1.9	0.553	0.726	28	1	77
20**	190	292	436	41	0.4	0.562	0.064	38	9	40
21	1000	606	50	20	0.05	0.001	0.001	1.001	5	15
22	1500	1000	50	50	0.05	0.001	0.01	1.6	10	2

* Failed to converge for the cases with normally and log-normally distributed parameters.

** Failed to converge for the cases with log-normally distributed parameters only.

§ Does not satisfy the *Product Criteria*.

Each of the 22 initial parameter sets was used as a starting point for PEST implementations without and with logarithmic transformation of parameters. The results from these optimization efforts are shown in Tables 3.6 and 3.7. Although the acceptable parameter set shown in Table 3.4 found during preliminary investigations applied logarithmic transformations only to selected parameters, the selection was subjectively and somewhat arbitrarily made. As a result, many different combinations of parameters that are and are not transformed must be performed to fully explore the benefits of parameter transformation. However, the ability of parameter sets to satisfy both criteria is consistently and significantly dependent on the values of the initial parameter set, and the effort required to investigate parameter transformation strategies cannot be justified.

Table 3.6. Optimized parameter estimates without logarithmic transformations

Parameter Set	c_0	ν	D_x	D_y	D_z	μ_t	λ	ν_ν	D_ν	b_i	Sum of Squares	No. of FE
1	500	565	55	18	3.8	1.3E-04	8.9E-03	2.1	4.6	2	768	445
2	208	813	11	14	3.5	1.0E-05	6.7E-03	7.3	20.0	73	660	310
3	514	712	30	16	4.1	7.0E-02	4.0E-02	11.9	20.0	51	875	136
4*§	480	694	442	29	1.3	7.1E-01	8.7E-01	22.4	1.6	11	1354	31
5	1293	190	72	22	1.0	8.0E-03	6.7E-01	4.6	8.8	48	722	146
6	1074	589	154	14	5.0	5.6E-04	1.2E-01	4.4	5.8	36	581	393
7	994	104	477	10	0.1	7.1E-02	4.3E-01	5.1	5.3	7	770	88
8*	1039	186	427	21	1.2	7.1E-01	5.5E-01	31.9	5.6	19	1354	31
9*§	1384	644	479	44	0.1	5.6E-01	1.0E+00	6.3	3.7	26	1354	31
10	700	367	73	14	1.3	6.3E-02	5.4E-02	4.5	15.9	20	652	177
11*	453	137	20	14	2.7	3.8E-01	9.1E-01	1.3	7.5	18	1354	31
12	856	112	167	7	1.9	5.1E-02	9.3E-01	27.2	19.0	2	706	105
13	1603	797	102	11	1.7	4.4E-04	3.2E-03	1.1	2.5	16	625	151
14	271	535	117	15	4.3	1.0E-05	9.6E-02	12.9	20.0	80	588	381
15	1108	264	232	5	0.1	7.6E-02	5.4E-01	19.0	18.3	27	1013	108
16	149	856	13	15	2.8	1.8E-04	8.1E-05	7.7	3.9	66	566	558
17	230	180	359	14	0.3	2.4E-03	5.7E-01	16.1	14.8	69	783	102
18*	1725	370	296	5	3.6	4.9E-01	5.5E-01	30.1	1.2	76	1354	31
19*	681	774	124	25	1.9	5.5E-01	7.3E-01	27.7	0.6	77	1354	31
20**	212	352	453	8	0.8	5.3E-02	1.3E-01	40.0	13.0	17	594	104
21	1000	606	50	20	0.1	4.4E-04	1.2E-04	1.1	4.1	13	6195	125
22	1500	1000	56	39	0.1	1.0E-05	3.9E-02	1.0	9.6	18	1255	150

* Failed to converge for the cases with normally and log-normally distributed parameters.

** Failed to converge for the cases with log-normally distributed parameters only.

§ Does not satisfy the *Product Criteria*.

Note: FE=Function Evaluations

All of the estimated parameters sets are unique; thus, convergence to a unique solution from different initial parameter sets may not be possible. Nonetheless, PEST outputs show that some level of optimization is achieved for the majority of the parameter sets since sum of squares values at final iterations are lower in comparison to initial iterations. This supports findings from the preliminary investigations that indicate a response surface with many local minimums. Approximately 30% of the PEST trials reproduced the initial parameter set, regardless of parameter transformations; these parameter sets are marked with one or two asterisk(s) in Tables 3.5, 3.6, and 3.7. The output generated by PEST states that the cause is a gradient of zero. This indicates that the Marquardt Lambda or the parameters governing derivative calculation should be altered. However, further investigations of these parameter sets were not implemented since the effort required cannot be justified given the greater influence of initial parameter values. A change of some of the PEST's

algorithm parameters to all cases is not desirable since it can slow the optimization process by requiring excess computational effort.

Table 3.7. Optimized parameter estimates using logarithmic transformations

Parameter Set	c_0	ν	D_x	D_y	D_z	μ_t	λ	ν_ν	D_ν	b_t	Sum of Squares	No. of FE
1	415	1000	15	17	3.6	4.4E-04	3.3E-05	3.6	0.7	35	566	274
2	124	751	28	15	3.1	1.1E-03	1.2E-03	9.2	8.3	57	562	109
3	1964	599	110	15	3.9	3.4E-02	1.9E-01	8.9	16.0	80	546	169
4*§	480	694	442	29	1.3	7.1E-01	8.7E-01	22.4	1.6	11	1354	31
5	890	223	34	15	4.2	3.1E-02	2.1E-01	4.8	12.7	36	575	163
6	1193	241	62	14	4.4	5.4E-03	7.2E-01	6.8	12.9	71	575	258
7	1090	201	52	14	0.3	1.4E-02	3.5E-01	2.0	9.7	16	577	163
8*	1039	186	427	21	1.2	7.1E-01	5.5E-01	31.9	5.6	19	1354	31
9*§	1384	644	479	44	0.1	5.6E-01	1.0E+00	6.3	3.7	26	1354	31
10	1589	1000	46	15	3.6	1.2E-01	9.0E-02	18.7	20.0	30	577	159
11*	453	137	20	14	2.7	3.8E-01	9.1E-01	1.3	7.5	18	1354	31
12	1752	100	161	8	1.9	7.1E-02	1.0E+00	26.7	19.0	10	738	83
13	2000	1000	72	8	5.0	1.9E-02	4.4E-03	2.4	11.8	80	933	99
14	332	562	127	15	3.8	2.1E-03	1.1E-01	12.6	15.8	80	545	245
15	1166	583	128	15	4.3	2.2E-02	1.7E-01	9.8	16.5	80	546	307
16	257	525	141	15	4.4	9.0E-03	9.3E-02	17.2	20.0	80	547	264
17	289	239	34	14	2.1	4.5E-03	2.7E-01	8.7	18.2	80	595	208
18*	1725	370	296	5	3.6	4.9E-01	5.5E-01	30.1	1.2	76	1354	31
19*	681	774	124	25	1.9	5.5E-01	7.3E-01	27.7	0.6	77	1354	31
20**	190	292	436	41	0.4	5.6E-01	6.4E-02	38.4	8.7	40	1325	59
21	500	732	500	16	0.2	3.5E-04	3.1E-03	1.0	2.4	2	604	156
22	816	468	269	18	0.1	5.0E-04	8.9E-02	1.5	2.6	5	590	200

* Failed to converge for the cases with normally and log-normally distributed parameters.

** Failed to converge for the cases with log-normally distributed parameters only.

§ Does not satisfy the *Product Criteria*.

The relationship between sum of squares values and the number of function evaluations listed in Tables 3.6 and 3.7 are presented in Figure 3.6. No strong relationship is visible with a wide range in the number of function evaluations required for convergence for a given interval of sum of squares values. Similar ranges in sum of squares values for both types of parameter representations are noted when the one extreme sum of squares value, which is obtained without parameter transformations, is disregarded. This high sum of squares value corresponds to the case where the fit of all other observations are sacrificed to fit the extreme measured concentration of 19.5 as shown in Figure 3.7. Typically, the effect of the extreme measurement is reduced as shown in the scatter plot of observed and modelled TCE concentrations for Parameter Set 1 in Figure 3.8. The scatter plots of all the

However, this observation is statistically insignificant since there are only two such cases, and initial parameter sets that did satisfy the *Product Criteria* are also unable to be optimized. Indeed, the major difficulty lies in satisfying the *Location Criteria*; and additional efforts such as changes in objective function definition and a more thorough search of the parameter space is required.

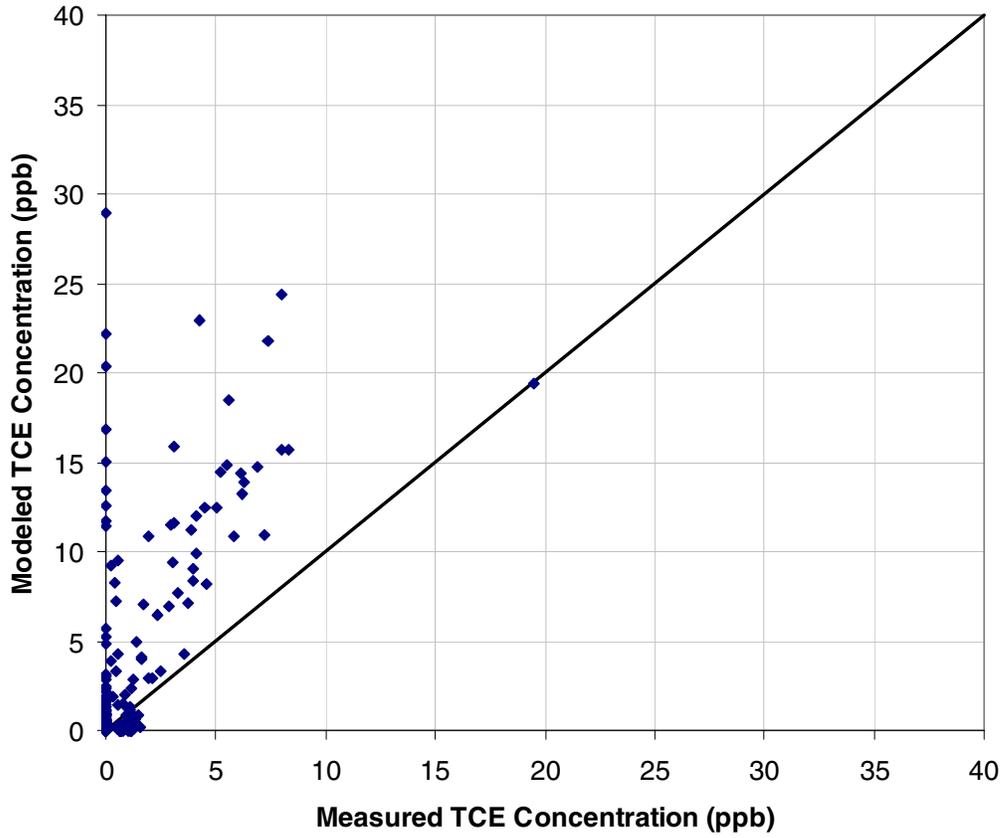


Figure 3.7. Scatter plot of simulated versus measured TCE (ppb) for Parameter Set 21 with no logarithmic transformation

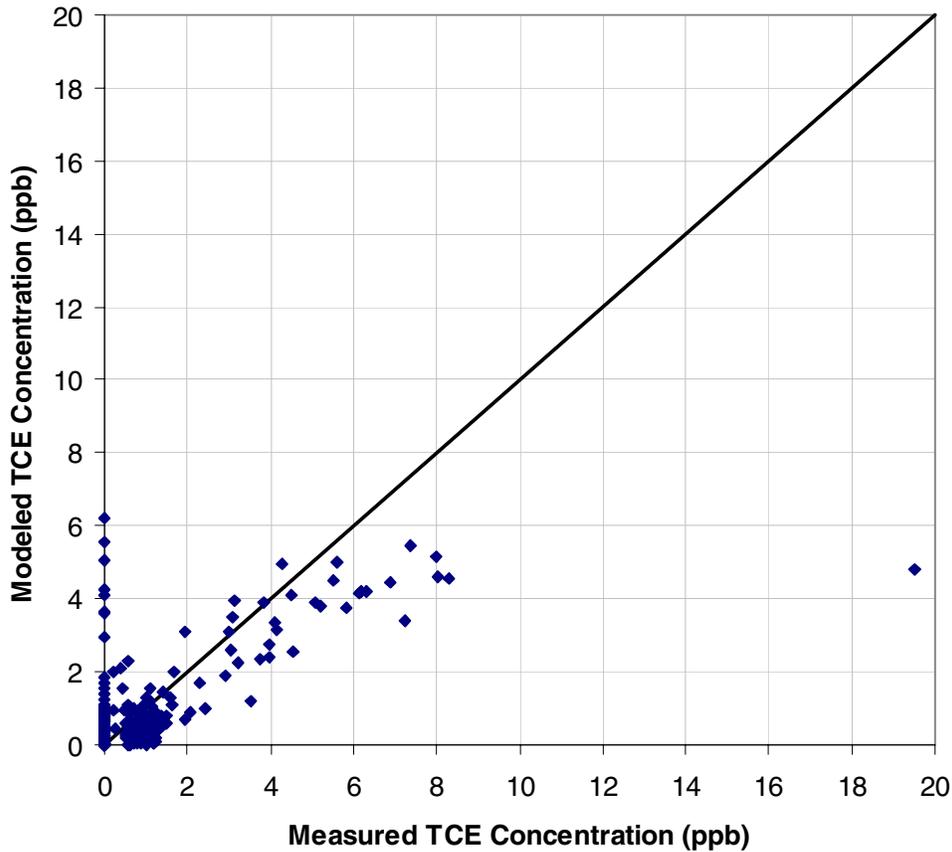


Figure 3.8. Scatter plot of simulated versus measured TCE (ppb) for Parameter Set 1 with no logarithmic transformation

3.3.2 DDS Results

3.3.2.1 Preliminary Investigations: A Comparison to PEST Results

A DDS run with 16 trials is implemented to facilitate comparisons with the PEST results. For each trial, 200 function evaluations are used in DDS to obtain an “optimized” parameter set. The number of function evaluations per trial is based on the average number of function evaluations required by PEST for convergence, which are 217 and 196 for the case without and with logarithmic transformations respectively.

Table 3.8 shows that lower sum of squares values are found with fewer function evaluations using the DDS method over multiple implementations of PEST. Extremely high sum of squares values are

removed; and the number of DDS trials with sum of squares values less than 600 are double the number of PEST trials with sum of squares values less than 600. Even though different initial parameter sets are used and there is a dependence on the seed value, the fact that the first implementation of DDS using the same number of trials as PEST is able to produce slightly lower sum of squares values with fewer function evaluations justifies its use in this thesis. Note that the total number of function evaluations for the PEST results would be higher if the function evaluations for failed PEST implementations are considered.

Table 3.8. Comparison of PEST and DDS implementation

Optimization Method	Number of Trials	Total Number of Function Evaluations	Lowest Sum of Squares Value	Highest Sum of Squares Value	Number of Trials with Sum of Squares Value < 600
PEST	16	3479**	566	6195	4 (25%)
DDS	16	3360	554	794	9 (56%)

*Includes function evaluations performed during the multistart procedure

**Does not include function evaluations that correspond to PEST implementations where no optimization could be performed

It is important to note that these results do not imply that the local greedy search approach is superior to PEST for the given parameter estimation problem since a multistart procedure is not used to initialize PEST results. Better results using PEST may be achieved with the incorporation of a multistart procedure.

3.3.2.2 Estimation

The five objective function definitions described in Section 3.1.1 are used to estimate “optimized” parameter sets with the DDS method, and the range of objective function values prior to the application of criteria are presented in Table 3.9. More than 5000 trials are generated for each definition because similar acceptance rates are achievable using different seed values with 5000 trials. Details on all the trials are presented in Appendix D.

A comparison of the objective function values for the case with no optimization to the cases with optimization presented in Table 3.9 shows that lower objective function values are achieved with optimization. For example, the lower bound for the case with optimization using the L_2 -estimator is 544.2 while the lower bound for the case without optimization produces a sum of squares of 559.8. Similar observations are found for the upper and lower limits of the range of objective function values obtained using LRS1 M-estimator, L_1 -estimator, and L_2 -estimator.

Table 3.9. A comparison of the objective function values prior to the application of criteria

Estimator	Range of Objective Function Values		
Initial DDS Sampling*	559.8 (L_2)	-	6613.1 (L_2)
	204.9 (L_1)	-	473.1 (L_1)
	51.6 (LRS1)	-	109.8 (LRS1)
LRS1 M-estimator	49.7	-	60.6
Cauchy's M-estimator ($c = 1.0$)	24.6	-	37.1
L_1 -estimator	192.8	-	272.3
Huber's M-estimator ($k = 1,2,3,4$)	36.3	-	732.1
L_2 -estimator	544.2	-	938.7

*The initial DDS sample is a result of selecting the “best” parameter set out of 10 random parameter sets.

The improvements are significant when the upper end of the objective function value is considered. This confirms the presumption of redundancy caused by the application of a behavioural threshold when DDS is used to generate parameter sets.

It is interesting to note the large difference between values obtained using the LRS1 M-estimator and Cauchy's M-estimator despite their similarities noted in Section 3.1. The major difference between the two estimators exists for low residual values where the difference can be up to 0.2. The difference noticed can be achieved if there are 126 low residual values.

The disadvantage of using the parameter sets obtained with the DDS method is their lower acceptance rate, which represents the proportion of estimated parameter sets that are able to satisfy the two criteria, as shown in Table 3.10. The limiting constraint is the difficulty in satisfying the *Locations Criteria* at the location of concern for Mejdrech Area D. This confirms the problem of spatial clustering noted in Chapter 2.

The range of objective function values associated with acceptable parameter sets are closer to the lower end of the overall range of objective function values prior to test for acceptance. This indicates the ability of the two physical criteria to act as filter for non-behavioural parameter sets indicating the possible redundancy of the application of a behavioural threshold to acceptable parameter sets. A more detailed investigation of this phenomenon is performed using parameter characteristics.

Table 3.10 shows that the highest acceptance rate of 1.48% is obtained using no optimization. This indicates that the ability of the parameter set to satisfy the limiting constraint, *Location Criteria*, is affected by other causes in addition to the parameter set's ability to represent observations. Although the parameter sets that satisfy criteria have relatively low objective function values, the parameter sets with the lowest objective function values have difficulty satisfying the criteria. This indicates that errors are not fully captured by the objective functions.

Table 3.10. A comparison of the acceptance rates

Estimator	Acceptance rate	Range of Objective Function Values		Range of Objective Function Values of Trials Satisfying Criteria			
Initial DDS Sampling*	1.48%	559.8	-	6613.1	738.9	-	1350.5
		(L2)	-	(L2)	(L2)	-	(L2)
		204.9	-	473.1	250.8	-	473.1 (L1)
		(L1)	-	(L1)	(L1)	-	(L1)
		51.6	-	109.8	56.3	-	109.8
		(LRS1)	-	(LRS1)	(LRS1)	-	(LRS1)
LRS1 M-estimator	1.24%	49.7	-	60.6	55.0	-	58.5
Cauchy's M-estimator ($c = 1.0$)	0.16%	24.6	-	37.1	29.8	-	34.2
L_1 -estimator	0.06%	192.8	-	272.3	235.0	-	264.8
Huber's M-estimator ($k = 1,2,3,4$)	0.00 - 0.06%	36.3	-	732.1	Generated insignificant number of acceptable parameter sets		
L_2 -estimator	0.00%	544.2	-	938.7			

* The initial DDS sample is a result of selecting the "best" parameter set out of 10 random parameter sets.

Of the optimized results, the LRS1 M-estimator generates the greatest number of acceptable parameter sets with an acceptance rate of 1.24%. There is a significant drop in the acceptance rate between the LRS1 M-estimator and Cauchy's M-estimator to 0.16%. Even fewer acceptable parameter sets are found with the remaining estimators. The acceptance rate based on parameter sets obtained with the L_2 -estimator is virtually zero with no acceptable trials being found out of approximately 13,000 trials generated, which is significantly more than the 5000 generated for the other estimators. Since L_2 -estimators are designed to amplify the effects of outliers, the outlier's effect on the ability of the model to satisfy constraints must be significant and robust estimators should be used.

The best performance in terms of the acceptance rate was achieved using the LRS1 M-estimator. The low weight assigned to residuals of high values by the LRS1 M-estimator indicates that its success is based on treating outliers. Although the true nature of errors can be inferred from an estimator's acceptance rate, the level of uncertainty regarding the true nature is still significant; and therefore, multiple parameter sets obtained with the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator are recommended for use in uncertainty analysis. Figures 3.1, 3.2, and 3.3 show that other M-estimators provide a middle ground between the L-estimators and the LRS1 M-estimator, and therefore their consideration was assumed to be redundant for our purposes.

The acceptance rates observed in Table 3.10 results in very few acceptable parameter sets with a reasonable amount of computational effort. At the highest observed acceptance rate in Table 3.10 of

1.5%, approximately 67,000 trials, which correspond to 13,000,000 function evaluations, need to be generated. Therefore, options to increase the acceptance rate such as penalty functions and dead zones in the objective function described in Section 3.1.2 must be adopted. Since the limiting constraint is the *Location Criteria*, a penalty function using (20) with a large scaling factor appears to be the best option. The repeated application of the forward solution used to find the arrival time creates a huge computational burden. As a result, the penalty function chosen is based only on the concentration at 33 years. In general, parameter sets that are able to create model outputs with a concentration of 0.5 ppb at the Leclerq Class Area are easily found. The difficulty lies in modelling a sufficient concentration level in Mejdrech Area D. The corresponding solution may be biased as the concentration at 33 years usually does not represent the peak concentration value as shown in Figures 3.4 and 3.9. For the Leclerq Class Area, 33 years corresponds to the decreasing tail of the breakthrough concentration profile, and the peak occurs at earlier times. However, for the remaining three areas, 33 years corresponds to the increasing leading tail of the breakthrough concentration profile, and the peak is not seen. Since the breakthrough concentration profile is typically at the leading tail for Mejdrech Area D, the simplification in the calculation of the penalty function is justifiable. In fact, great improvements in the acceptance rate are made with penalty functions based on the *Location Criteria*.

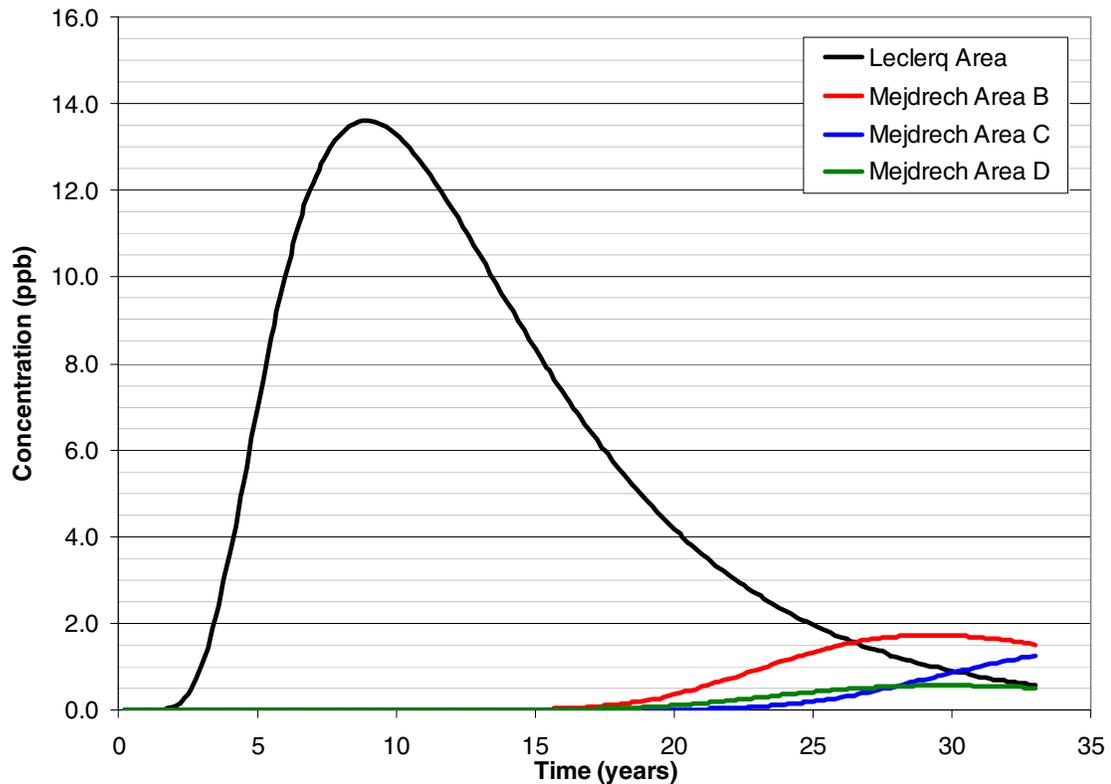


Figure 3.9. Breakthrough concentration after 33 years since the opening of the Metcoil site at the four locations using the acceptable parameter set with the lowest objective function value using the LRS1 M-estimator

A trial-and-error approach is used to determine the optimal scaling factor for the penalty function using the L_2 norm, L_1 norm, and LRS1 M-estimator; the corresponding results are presented in Table 3.11.

The acceptance rates listed in Table 3.11 implies that an optimal scaling factor for all objective function definitions tested is 10,000. A penalty function with scaling factor of 10,000 increases the acceptance rate by approximately 30% to 50% from the rates presented in Table 3.10. Approximately 60% of the trials that satisfy the *Location Criteria* pass the *Product Criteria*, and the need for a penalty applied to increase number of trials that satisfy the *Product Criteria* does not exist.

Table 3.11. Acceptance rate for L_2 -estimator, L_1 -estimator, and the LRS1 M-estimator with respect to changes in the scaling factor for the penalty function

Scaling factors for the penalty function	L_2 -estimator	L_1 -estimator	LRS1 M-estimator
No penalty	0.00%	0.06%	1.24%
100	n/a	n/a	26%
500	0%	20%	43%
1000	4%	29%	42%
5000	25%	39%	49%
10000	34%	42%	54%
50000	21%	24%	29%

Deadzones can also be used as a method to increase the acceptance rate. Deadzone limits of 0.1 ppb and 0.5 ppb are selected based on the detection limit and are tested for several estimators (Table 3.12).

Table 3.12. Acceptance rates for objective function definitions with deadzones

Estimator	Acceptance rate	
	Deadzone Limit = 0.1 ppb	Deadzone Limit = 0.5 ppb
LRS1 M-estimator	1.4%	1.3%
L_1 -estimator	0.2%	0.5%
L_2 -estimator	0.0%	not tested

A comparison of the acceptance rates in Tables 3.10, 3.11, and 3.12 shows that the benefit from applying the deadzones is small relative to benefits achieved with the addition of the penalty.

The earliest arrival times obtained using the various objective functions are all approximately 0.4 years, 8.3 years, 11.7 years, and 10.5 years for Leclercq Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D respectively. The parameter set with slightly earlier arrival times compared to all first arrival times is found using the LRS1 M-estimator. This is most likely due to the fact that this estimator generated the largest number of acceptable trials. The breakthrough concentration curves at the four locations and the corresponding plume in plan view are presented respectively in Figures 3.9 and 3.10 for the case with the lowest objective function value based on the LRS1 M-estimator.

In general, the objective function definition appears to have negligible effects on the first arrival times with similar earliest arrival times found for all four areas. Nonetheless, the effects of the objective function definition on the uncertainty of arrival times are further explored in Chapter 4. The parameter sets recommended for use in uncertainty analysis are presented in Table 3.13.

Table 3.13. Trials to be used in uncertainty analysis

Estimator	Penalty Factor	Number of Acceptable Parameter Sets
LRS1 M-estimator	10,000	1614
L_1 -estimator	10,000	1245
L_2 -estimator	10,000	1019

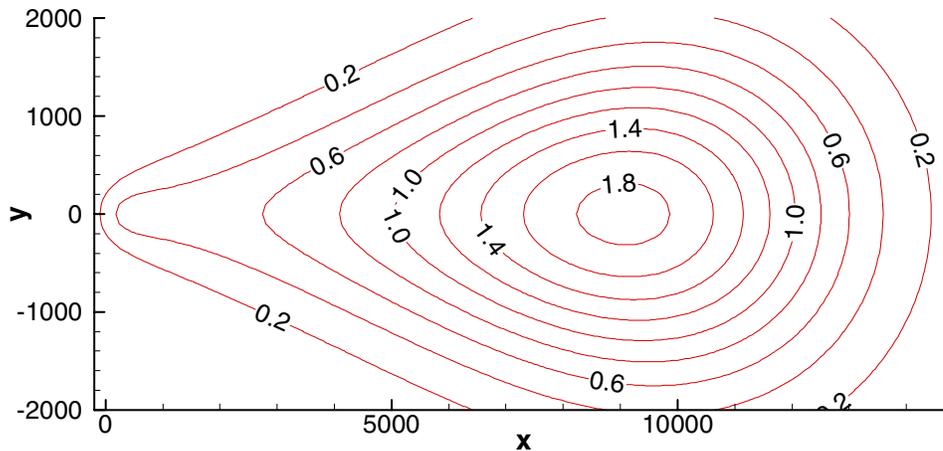


Figure 3.10. Plan view of the plume at a depth of 50 ft after 33 years since the opening of the Metcoil site using the acceptable parameter set with the lowest objective function value using the LRS1 M-estimator

3.4 Parameter Characteristics

Parameter values are constrained to be within a predefined range selected based on subjective interpretations of site-specific information using knowledge of physical processes. These parameter ranges used as inputs in the DDS method have been chosen to be slightly wider than likely to conservatively account for uncertainties. The distributions of parameters in these ranges are assumed to be uniform to avoid making misleading assumptions. The resulting parameter space, which is referred to as the prior parameter space, is expected to evolve towards the “true” parameter space with the implementation of the DDS method and the application of the two criteria.

Valuable insight into the true nature of the parameter distributions and correlations can be gained by examining the characteristics of the parameter spaces. Three parameter spaces are considered: the prior parameter space, the parameter space of DDS results, and the acceptable parameter space. Parameter estimates using PEST results are not considered since convergence of parameter characteristics is not likely to be achieved with 22 parameter sets in this thesis.

3.4.1 Parameter Values

The trends in the distributions must be examined with corresponding objective function values to avoid incorrect conclusions. In general, trends associated with high objective function values must be viewed sceptically. For example, in a bimodal distribution, the mode that corresponds to high objective function values is less likely to represent the true distribution. Therefore, two plots are examined for each parameter: a plot of parameter value versus objective function value (Val-Obj), and a plot of parameter value versus frequency (i.e. a histogram).

Frequency histograms are created by subdividing the range assigned for DDS runs into 10 equal intervals for consistency. Therefore, the bin widths, w , and bins, b_i , are defined as follows

$$w = \frac{u - l}{10} \quad (22)$$

$$b_i = l + \frac{2i - 1}{2} w \quad (23)$$

where i represents the bin number, and u and l represent the upper and lower limit of the range, which are presented in Table 3.3.

A fit to a standard parametric distribution such as the normal or lognormal distribution can smooth and distort real information especially around the tails (USEPA, 1997). Therefore, reference to standard analytic distributions is limited to subjective descriptions of the resulting empirical distributions. Note that trends observed in histograms can vary with bin widths but optimization of the bin width is not necessary for our purposes.

3.4.1.1 Prior Parameter Space

The prior parameter space is characterized by 6000 parameter sets that have been used as the starting point for DDS optimization trials. The number of parameter sets obtained from multiple DDS runs with various seed values is assumed to be sufficient for convergence (i.e. additional parameter sets are assumed to have negligible effects on parameter space characterization). No attempts to minimize the objective function or to consider the two criteria are made for these parameter sets; thus, the values are the result of the multi-starts with parameter sets taken from bounded uniform distributions.

For parameters, c_0 , v , D_{1s} , and v_v , the histograms appear to suggest uniformly distributed parameter values as shown in Figure 3.11. However, indications of non-uniformity exist for these four parameters, which consistently have fewer values in the first interval. Distributions for parameters,

D_y , D_z , and D_v , in Figure 3.11 are multi-modal with two to four modes. This can signify insufficiencies in the number of parameter sets; however, this is unlikely since differences are not noticed with a reduction in the number of parameter sets. The remaining parameters, μ_t , λ , and b_t , exhibit lognormal behaviour in Figure 3.11, which is less pronounced for b_t . The non-uniform nature of the distributions for parameters, μ_t and λ , appears to result due to a bias in the sampling process to values near the lower limit. The non-uniform nature of the prior parameter space must be considered when evaluating subsets of this space.

For most of the parameters, there appears to be no relationship between parameter values and the objective function value, which is defined using the L_2 -estimator in Figure 3.11. However, the Val-Obj plots for parameters, D_y and μ_t , appear to indicate a minimum at approximately 10 ft and 0 yr⁻¹ respectively.

It is also important to note that the majority of parameter sets produce high performance index values as shown in Figure 3.12. The objective function value appears to be bounded by an upper limit around 1354 ppb²; however, there are two parameter sets that significantly exceed this value as shown in Figure 3.13 for the parameter, c_0 . This phenomenon consistently appears for all parameters since one objective function value exists for each parameter set.

3.4.1.2 Optimized Parameter Space

Optimized parameter spaces obtained through the DDS method are found using the estimators selected in Section 3.3, which are the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function. All three optimized parameter spaces are characterized by lower performance indices with no clustering of points at high performance index values, which is noticed in Figure 3.11 for the prior parameter space.

The trends in the Val-Obj (Figure 3.14) and frequency histograms (Figure 3.15) of the optimized parameter space have converged to distributions that no longer resemble those found in Figure 3.11. Two notable exceptions are the parameters, μ_t and λ , for which the trends previously found have been accentuated. In general, the nature of frequency distributions can be qualitatively described as normally-, lognormally-, or uniformly-distributed depending on the parameter. Some distributions are difficult to classify as they can be described using more than one standard distribution. This is further complicated by the fact that the optimized parameter space is clouded with differences in estimator definition. Therefore, only the number of modes present in the empirical distribution of each parameter is presented in Table 3.14 since they are easily distinguishable.

The deviations between estimators shown in Figure 3.15 are more pronounced near the limits. Since these parameter sets may be of greater interest in uncertainty analysis where low likelihood solutions play a significant role in the decision-making process, their presence should be further investigated. Further evaluation is possible using the corresponding Var-Obj plots. The Var-Obj plots for the three sets of parameter sets (Figure 3.14) indicate that the major trends are not affected by the definition of the estimator. (Note the importance of scale in examining the Var-Obj plots.) Two important pieces of information attainable from these plots are: the existence of a global optimum or major optimums, and the existence of regions of high performance index values. Firstly, the response surface appears to be highly irregular with many local minimums regardless of estimator definition. There may be a global optimum or major optimums that are not found by local search methods, which can become trapped in one of the small minimums. In fact, the existence of a global optimum is apparent in the parameters, ν , D_x , μ_t , and λ , where the points in the Val-Obj plot form a concave lower bound. The numbers of major optimums identified in the Val-Obj plots are listed in Table 3.14. The presence of multiple major optimums is assumed if there appear to be more than one concave lower bound. For example, two horizontal curves, which are more visible in the plot based on the LRS1 M-estimator, are formed by clusters of points in the lower half of the Val-Obj plots for b_t ; both curves indicate a slight decrease in performance index values with an increase in b_t and thus, points to a major optimum. Secondly, performance index values are designed to indicate a parameter sets ability to model observations. Therefore, the regions of high performance index values are important in assessing the confidence in the distribution present in histograms. The modes in Figure 3.15 that correspond to these regions of high performance index values are designated as likely to be incorrect in Table 3.14. Note that there are still many parameter sets with low performance index values in the intervals representing possibly incorrect modes; thus, there is a possibility that the problematic modes and the corresponding distribution are correct. Therefore, the presence of high performance index values decreases not eliminates the probability of the corresponding mode.

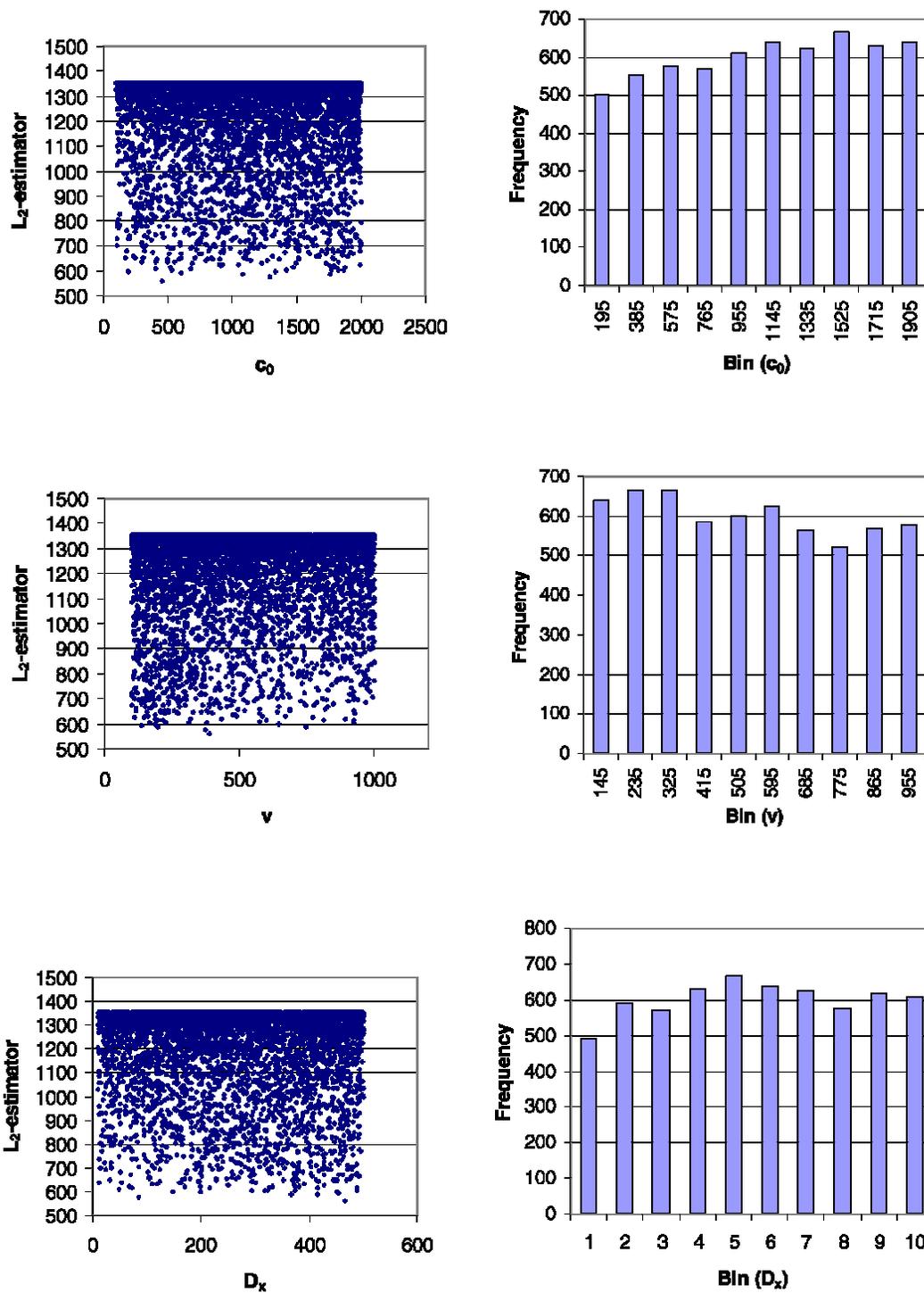


Figure 3.11. Parameter values in terms of frequency histograms and objective function values defined by the L_2 -estimator using parameter sets obtained using the multistart procedure

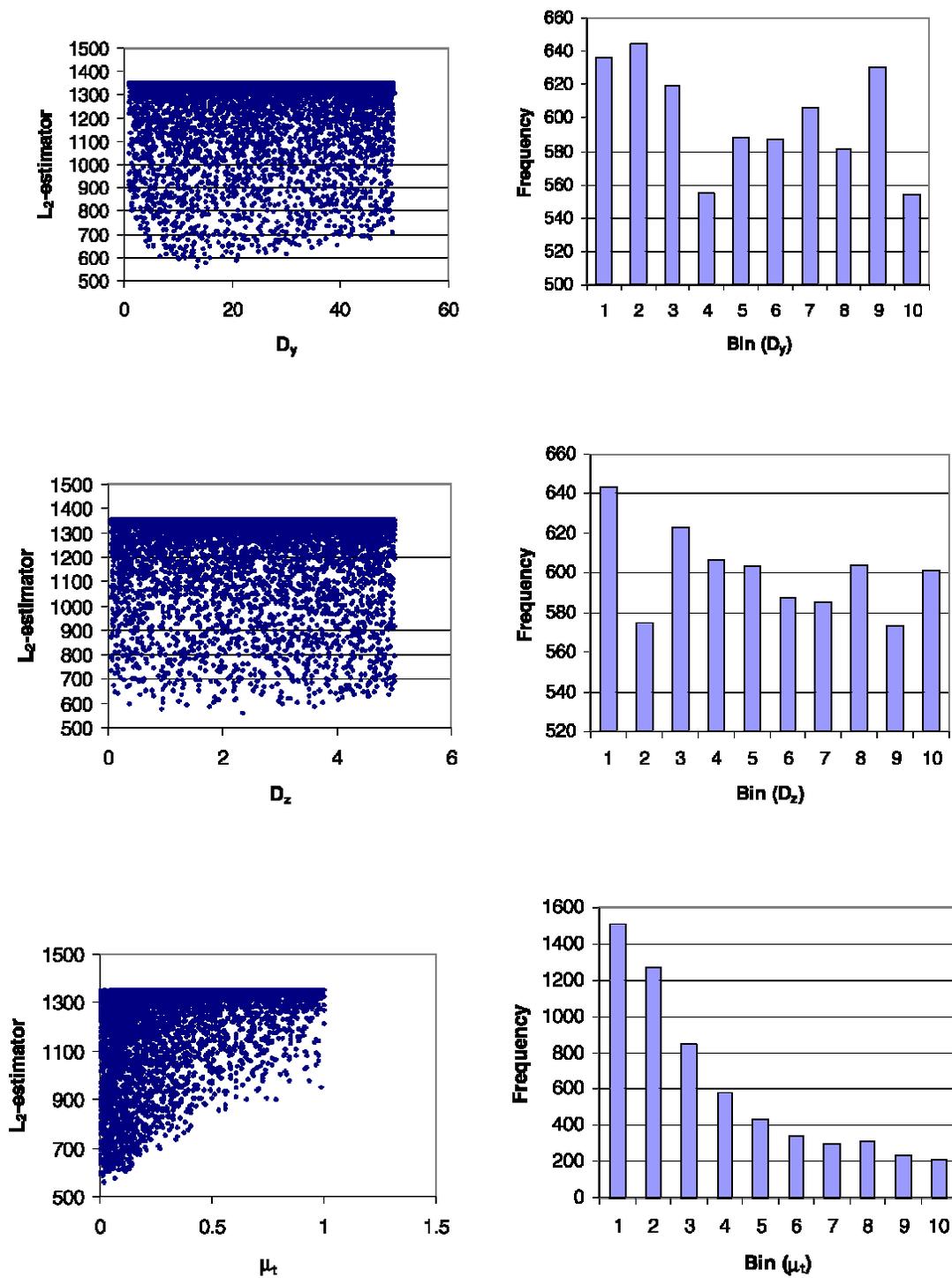


Figure 3.11. Parameter values in terms of frequency histograms and objective function values defined by the L_2 -estimator using parameter sets obtained using the multistart procedure (continued)

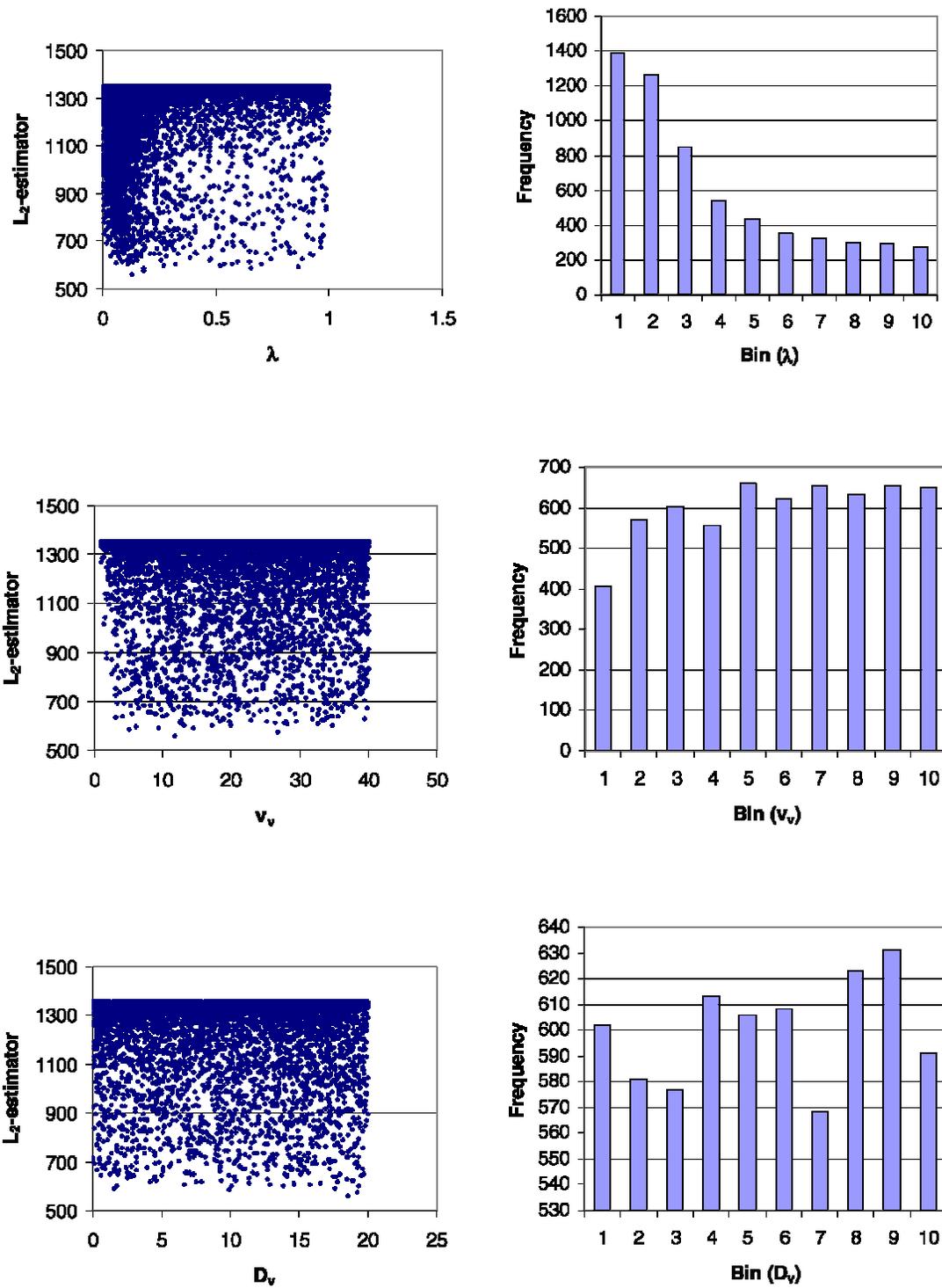


Figure 3.11. Parameter values in terms of frequency histograms and objective function values defined by the L_2 -estimator using parameter sets obtained using the multistart procedure (continued)

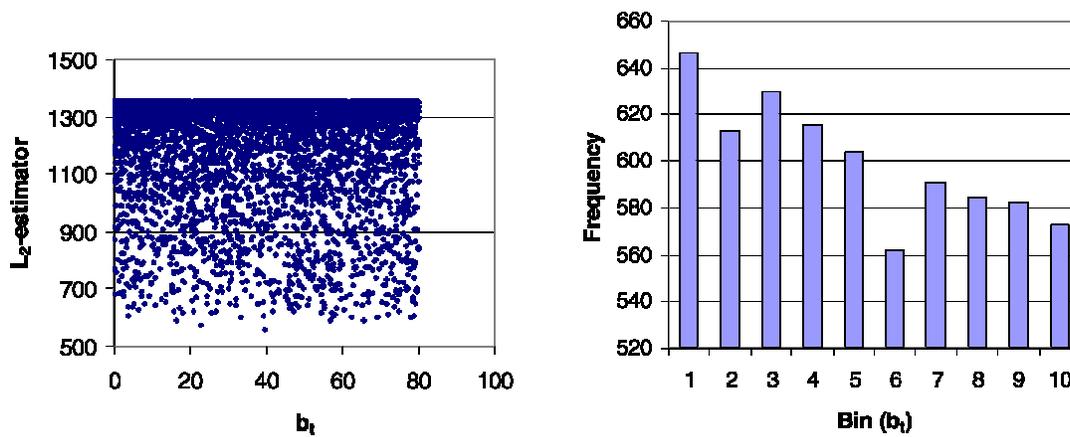


Figure 3.11. Parameter values in terms of frequency histograms and objective function values defined by the L_2 -estimator using parameter sets obtained using the multistart procedure (continued)

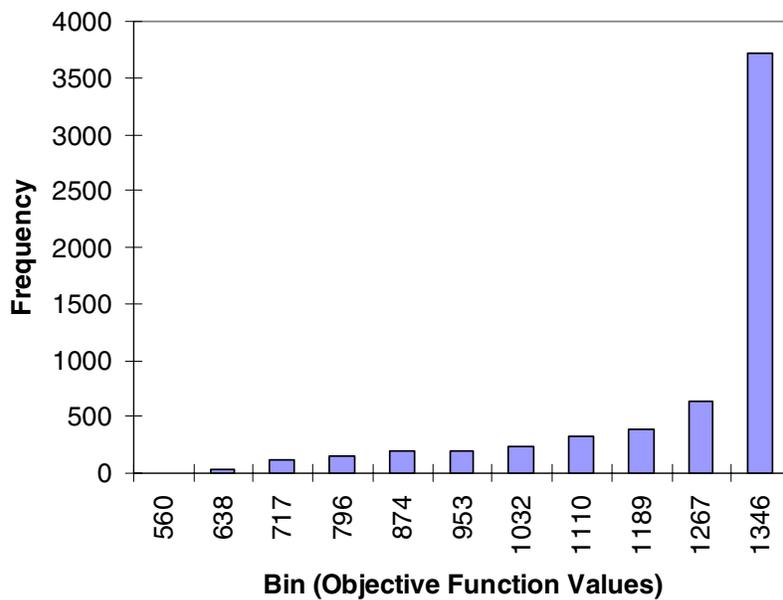


Figure 3.12. Histogram of objective function values defined by the L_2 -estimator for parameter sets obtained using the multistart procedure

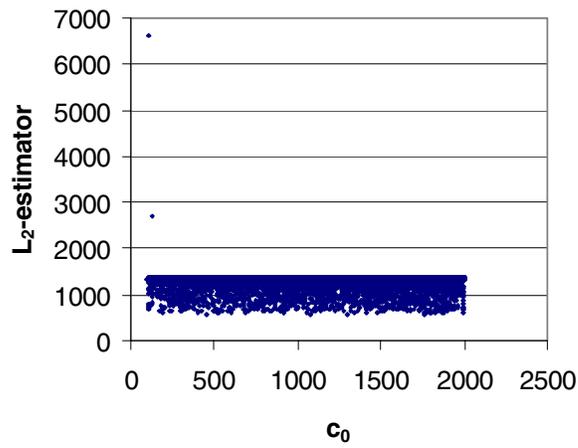


Figure 3.13. Relationship between parameter values and objective function values defined by the L_2 -estimator for parameter sets obtained using the multistart procedure

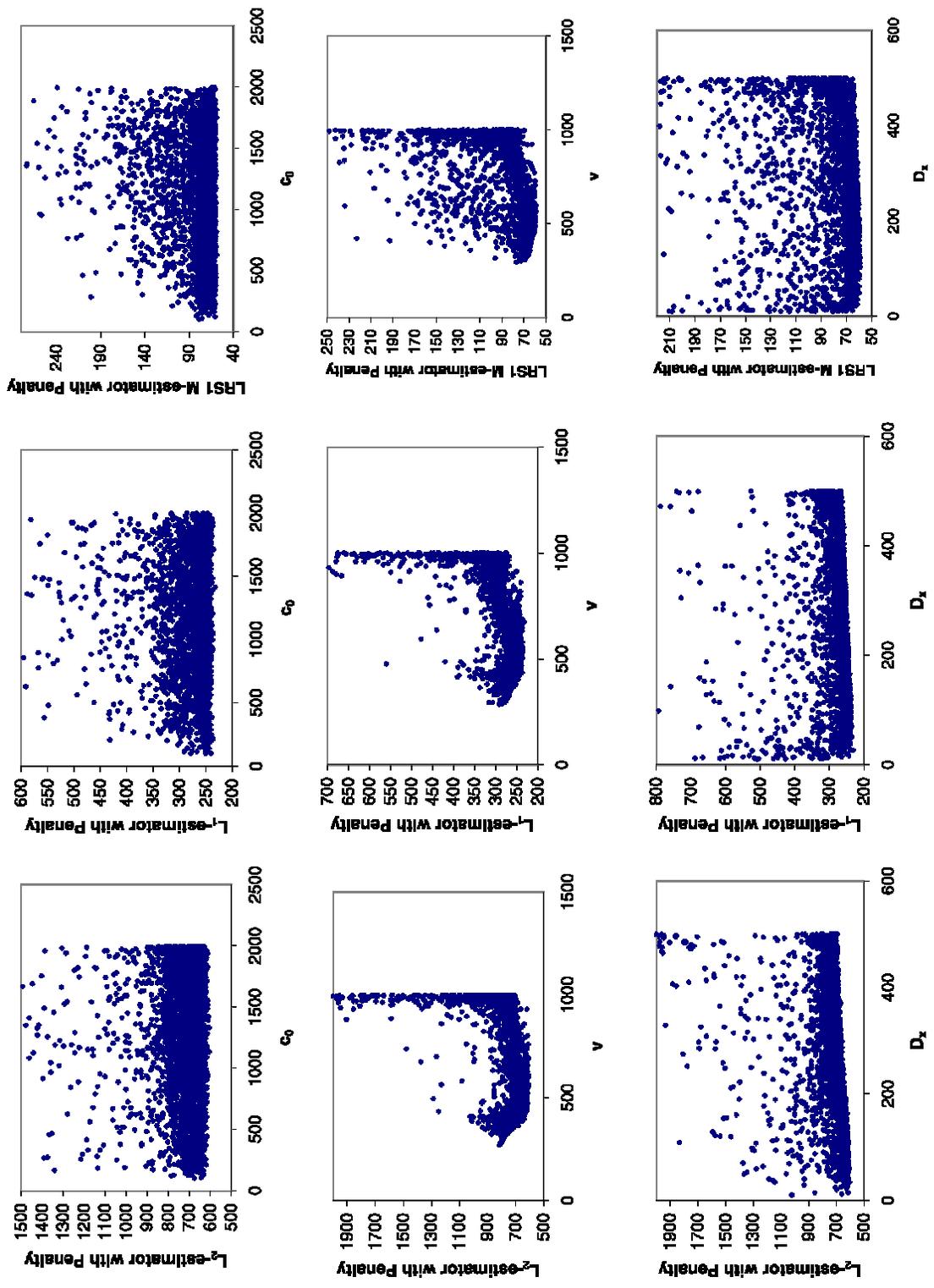


Figure 3.14. Val-Obj plots for parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function

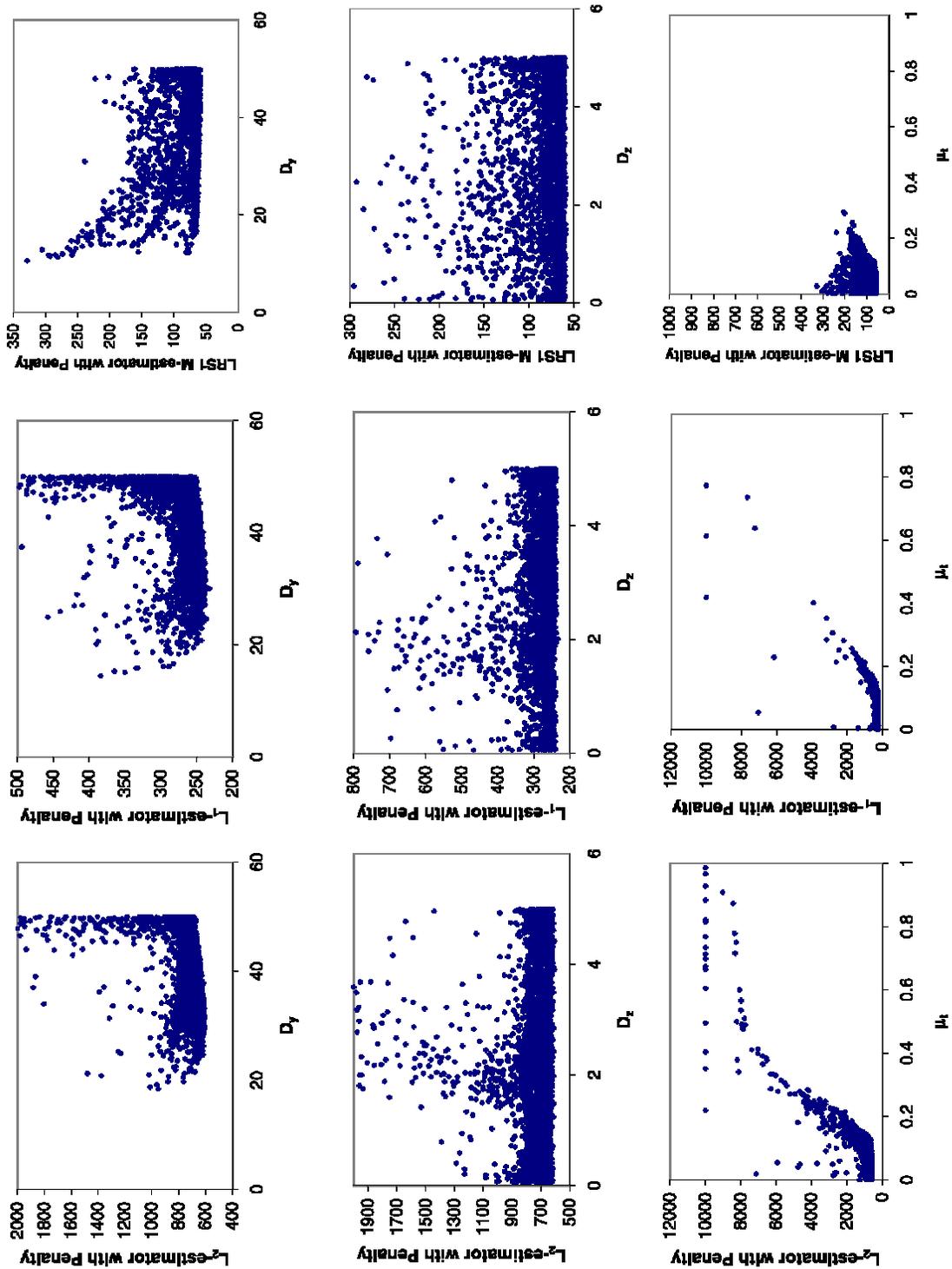


Figure 3.14. Val-Obj plots for parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function (continued)

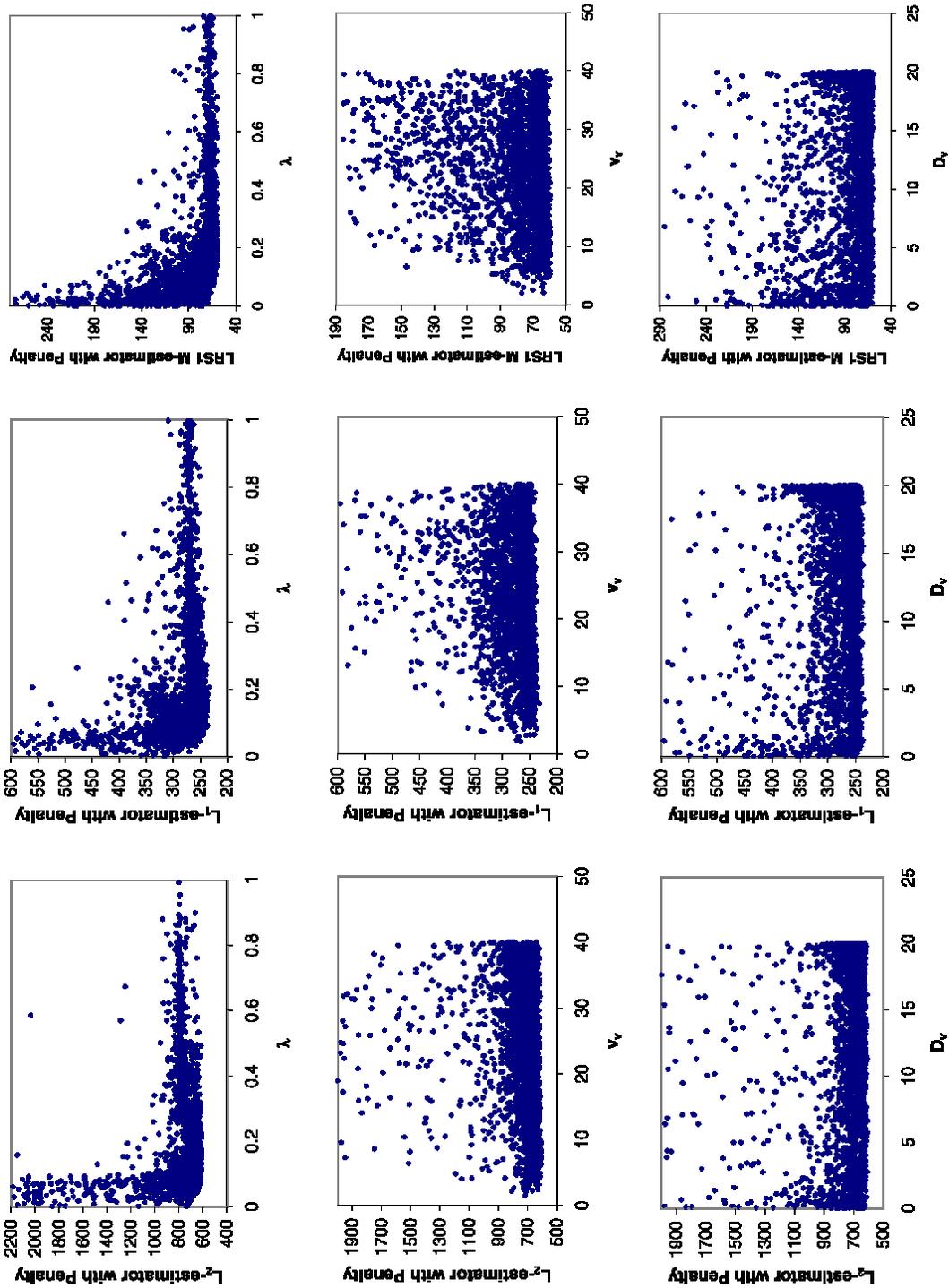


Figure 3.14. Val-Obj plots for parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function (continued)

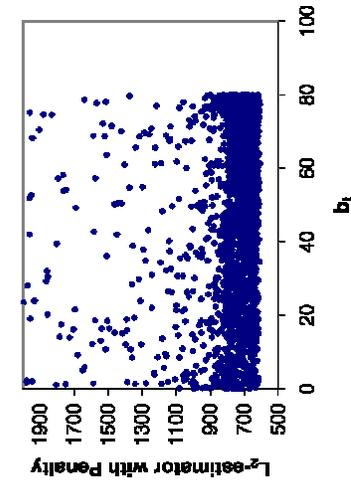
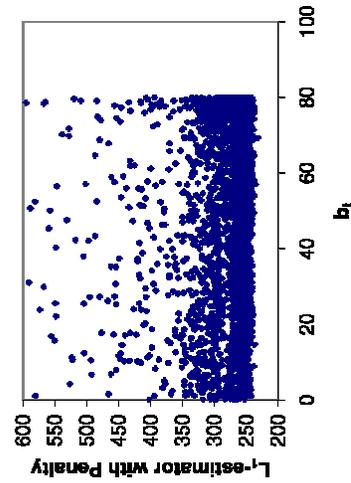
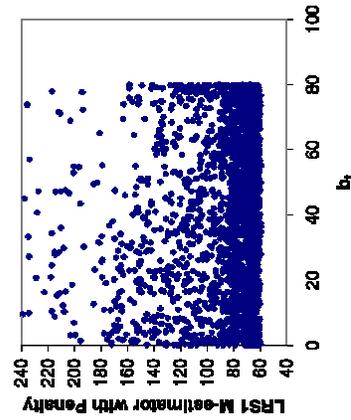


Figure 3.14. Val-Obj plots for parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function (continued)

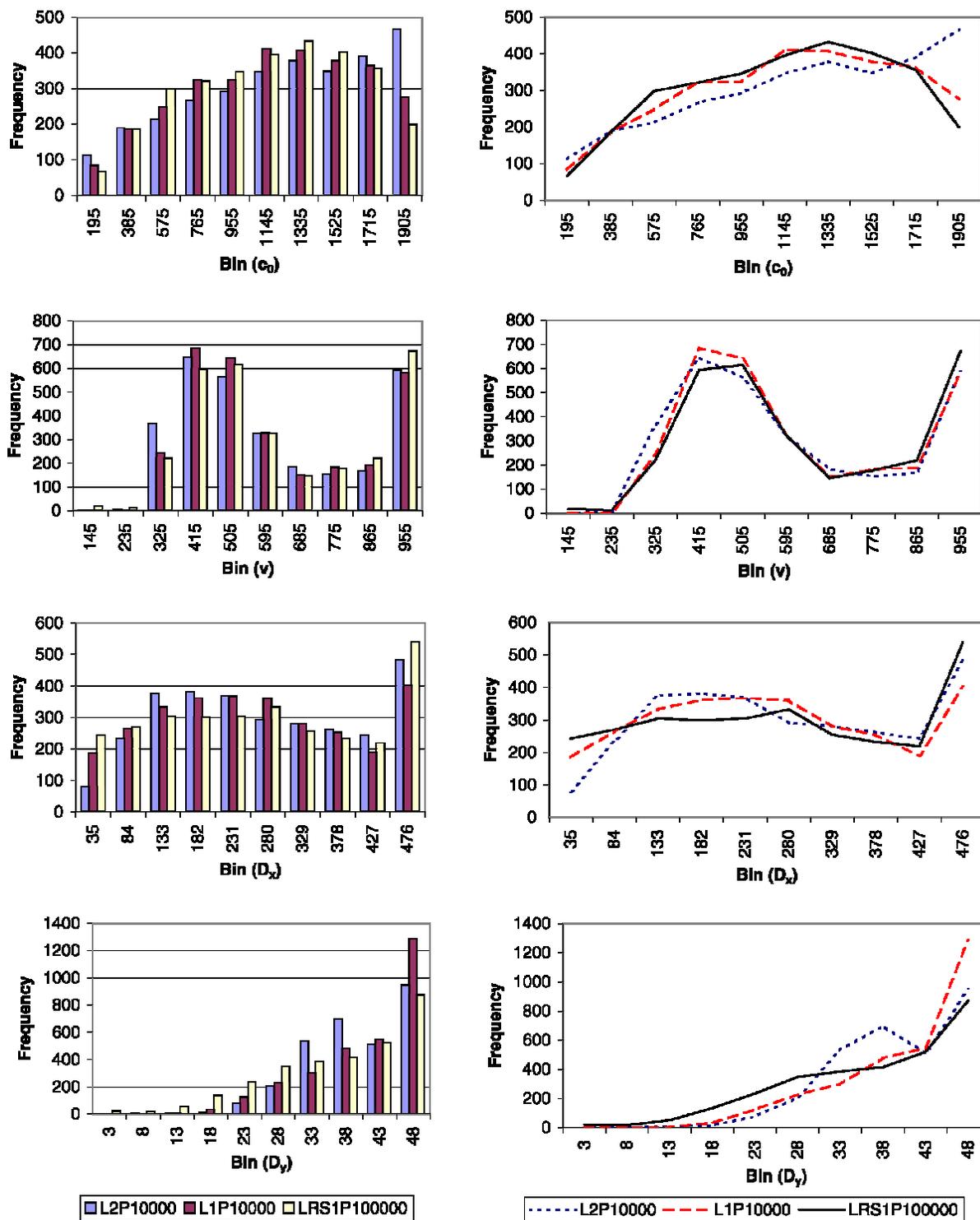


Figure 3.15. Histograms of parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function

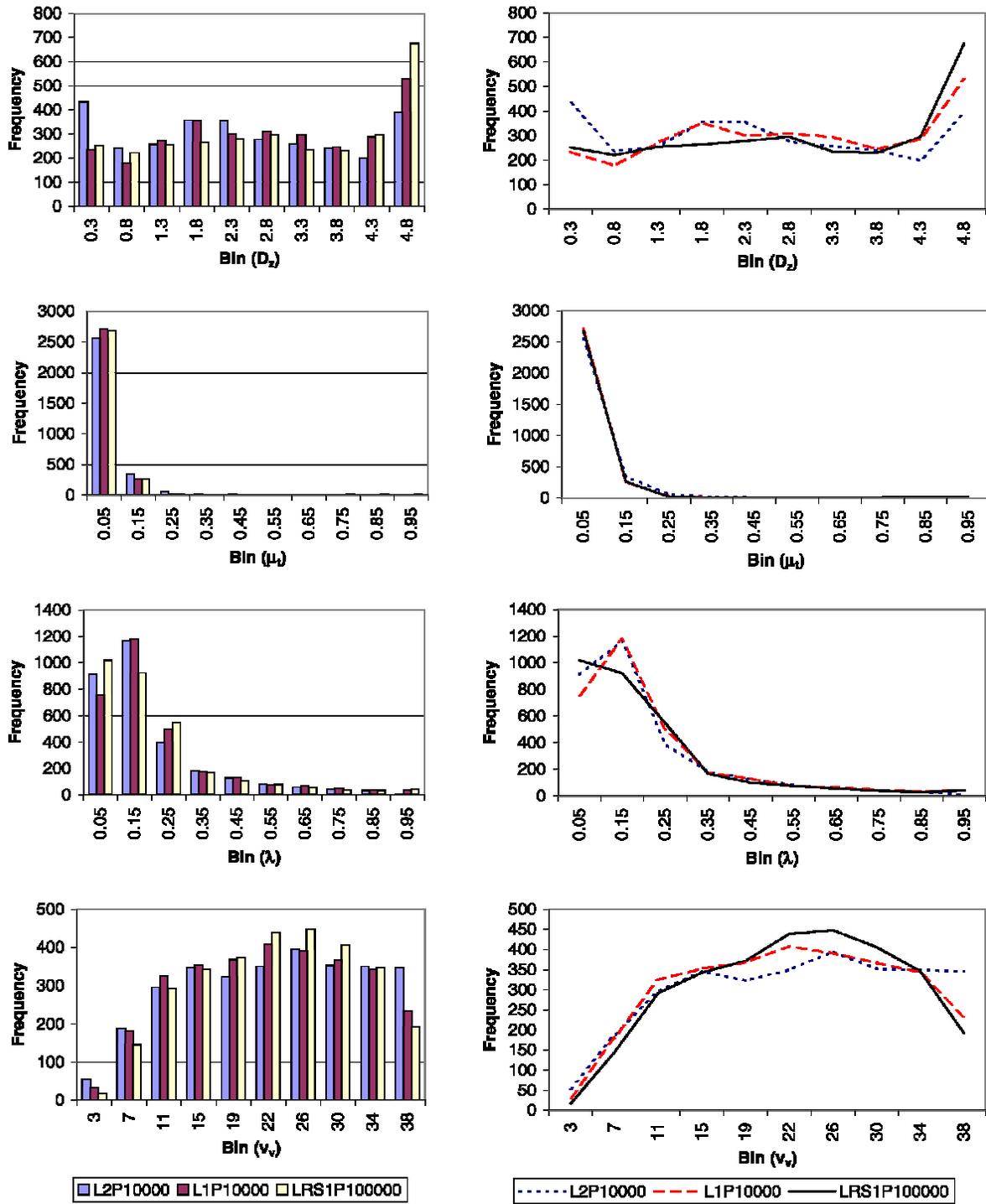


Figure 3.15. Histograms of parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function (continued)

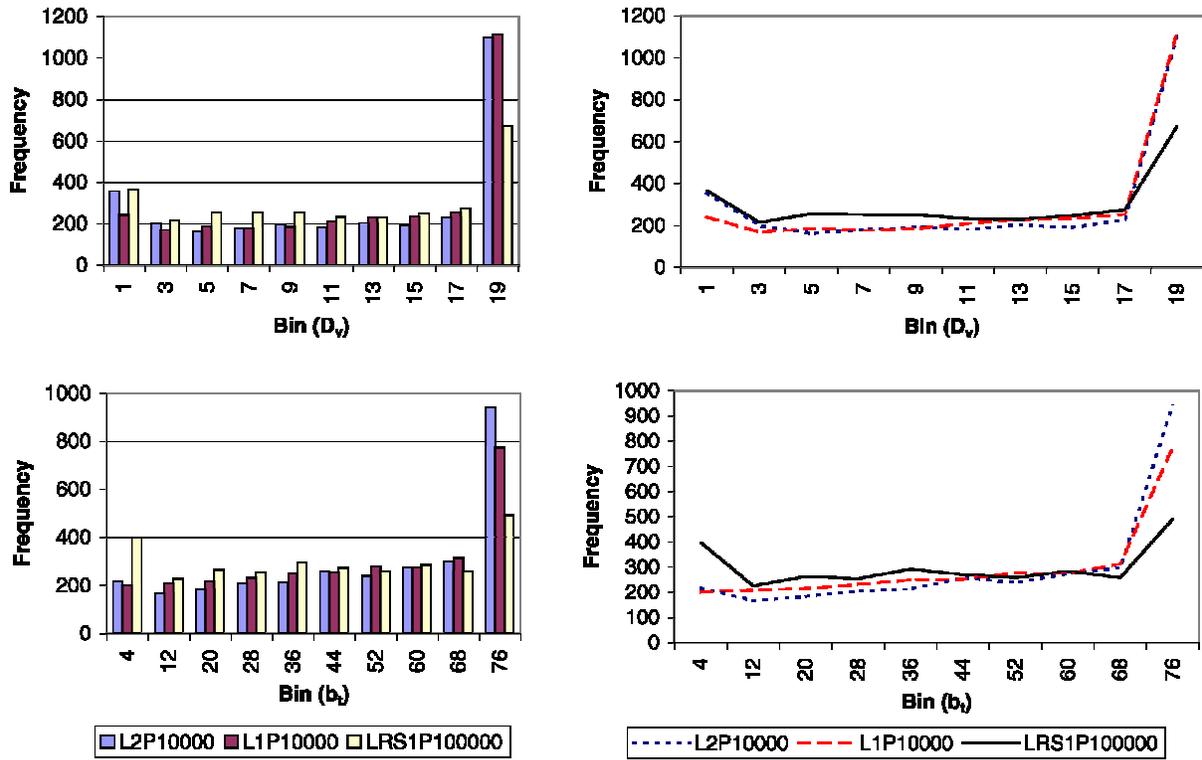


Figure 3.15. Histograms of parameter values obtained using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function (continued)

Table 3.14. Qualitative descriptions of optimized parameter space

Parameter	Modes		Major Optimums		Mode With High Performance Index Values
	Number	Value(s) [*]	Number	Value(s) [§]	
c_0	1	1300	None	n/a	None
v	2	450, 950	1	450 – 500	2 nd Mode
D_x	2	180, 480	1	100 – 200	2 nd Mode
D_y	1 to 2	38 ^{**} , 50	1 or 2	38, 50 ^{**}	None
D_z	1 to 3	0, 2, 5	None	n/a	2 nd Mode
μ_t	1	0.0	1	0.0 – 0.1	None
λ	1	0.15	1	0.15 – 0.2	None
v_y	1 to 2	12, 25	1 or 2	10, 30 ^{**}	None
D_y	1 to 2	1, 20	None	n/a	1 st Mode
b_t	1 to 2	4 ^{**} , 80	1	80	None

^{*} Approximately estimated from histograms

[§] Approximately estimated from Val-Obj plots

^{**} Not present in all three curves

Parameter values at modes typically equal the values at major optimums. The mode that does not coincide with a major optimum represent parameter sets with high performance index values. Therefore, trends in frequency distributions can be refined to create distributions that are more representative of its probability. However, the information available is subjective and trends are difficult to identify for most parameters. Trends found with one estimator are typically amplified or subdued by using other estimators. For example, there appears to be two curves in the Var-Obj plot for D_y , which is most easily seen in the plot based on the LRS1 M-estimator as shown in Figure 3.14. The opposite trend is noticed between the plots for D_z , where the concentration of high performance indices around the center of the plot is no longer present in the plot based on the LRS1 M-estimator.

In general, the trends shown for the parameter sets generated using the L_1 -estimator appears to be a middle ground between the remaining two estimators.

It is important to note that a behavioural threshold is not applied and there are a number of parameter sets with extremely high parameter values as shown in Figure 3.16 for all three estimators. Although Figure 3.16 only shows the parameter, c_0 , the same trend exists for other parameter since there is only one performance index value for each parameter set. The upper limit for performance index value is reached by all three estimators due to the penalty function.

3.4.1.3 Acceptable Parameter Space

Acceptable parameter spaces are subspaces of optimized parameter spaces and have parameter sets that are able to satisfy the two criteria. Given that the effect of various estimators produce similar results, the parameter sets obtained using the L_1 -estimator with a scaling factor of 10,000 for the penalty function are used to illustrate the effect of applying the two criteria.

The major trends in the acceptable parameter space, which is shown in Figure 3.17, are identical to those in the optimized parameter space. The significant difference is the elimination of extremely high performance indices. This confirms that the two criteria can also function as a behavioural threshold.

The nature of the frequency distributions is still difficult to classify, and only the numbers of modes are presented in Table 3.15. Val-Obj plots are examined to identify major optimums and regions with high objective function values; these trends are also presented in Table 3.15.

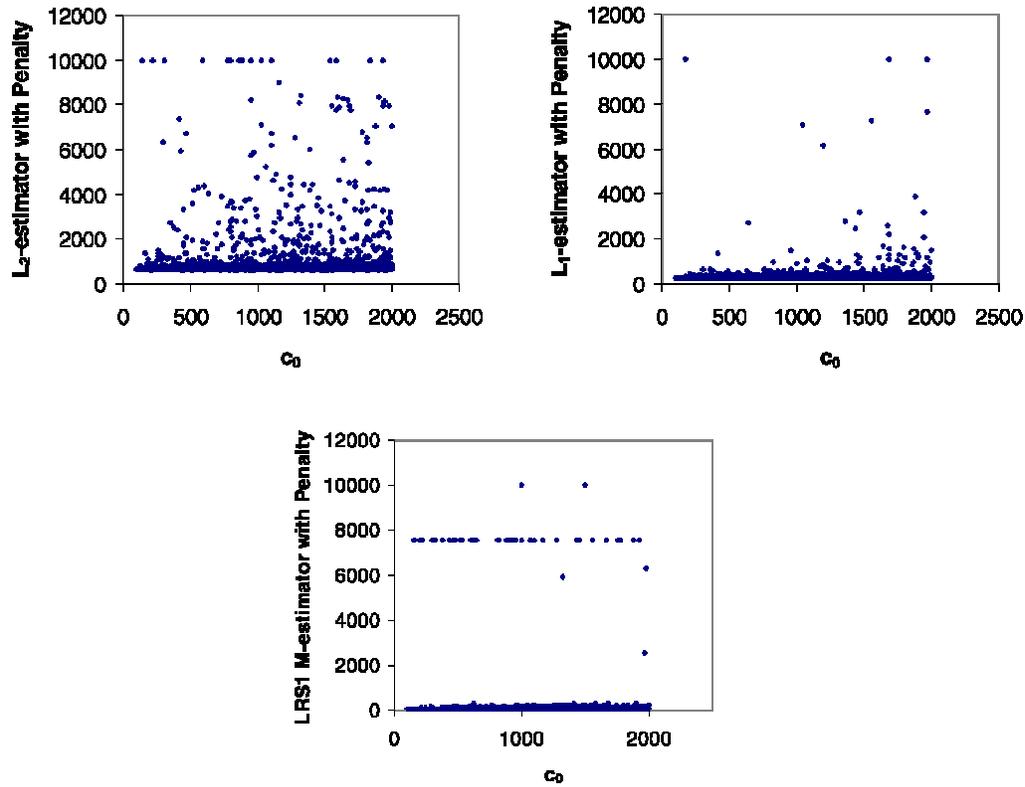


Figure 3.16. Val-Obj plot for entire range of objective function values for the parameter, c_0

Table 3.15. Qualitative descriptions of the acceptable parameter space based on the L_1 -estimator with a scaling factor of 10,000 for the penalty function

Parameter	Modes		Major Optimums		Mode With High Performance Index Values
	Number	Value(s) [*]	Number	Value(s) [§]	
c_0	1	1200	None	n/a	None
v	2	500, 950	1	500	2 nd Mode
D_x	1	180	1	100	None
D_y	1	50	1	38	1 st (and only) Mode
D_z	3	0, 2, 5	None	n/a	2 nd Mode
μ_t	1	0	1	0.0 – 0.1	None
λ	1	0.15	1	0.15	None
v_v	1	25	None	n/a	None
D_v	2	0, 20	None	n/a	None
b_t	1	80	None	n/a	None

Note: The cells highlighted are different from Table 3.14

^{*} Approximately estimated from histograms

[§] Approximately estimated from Val-Obj plots

A comparison of Table 3.14 and Table 3.15 shows that modes or major optimums are eliminated with the application of the two criteria. This is useful for parameters such as D_y , for which the trends led to conflicting indications.

Despite the significant amount of information gained through empirical distributions and Val-Obj plots for the acceptable parameter space, distributions that can reliably indicate probability are difficult to define. However, the improvements noted in the acceptable parameter space indicate that a reliable probability distribution may be achieved with the application of additional physical criteria.

The multiple modes in the histogram may indicate a need to revise the parameter range, especially for parameters such as v , D_v , and b_r . Some modes located near the limits are associated with high performance index values. Since the ranges have been selected to be greater than necessary, a reduction should be considered.

Typically, the relationship between parameter values and performance index values is the combined effect of all parameter values. A notable exception is the decay coefficient in the till unit illustrated in Figure 3.17 where the decay value independently increases the objective function value for decay values greater than approximately 0.12 yr^{-1} . This does not affect the histogram since the peak of the uni-modal distribution is so high.

The parameter characteristics of the L_2 -estimator and the LRS1 M-estimator with a scaling factor for the penalty function of 10,000 included in Appendix E are also examined and compared to Figure 3.17. The parameter sets generated with the L_2 -estimator appear to have very similar characteristics compared to the parameter sets generated using the L_1 -estimator. The LRS1 M-estimator generates a slightly different result. Fewer modes are present with the disappearance of modes identified as possibly incorrect. For example, the higher valued mode in the histogram of horizontal velocities in limestone is eliminated with the LRS1 M-estimator. It also eliminated the second mode and the corresponding distribution for the vertical dispersivity in limestone. In addition, there are very few outlying performance index values associated with the parameter sets generated using the LRS1 M-estimator. In general, the LRS1 M-estimator appears to be the best estimator in terms of achieving believable parameter characteristics as it is able to subdue possibly incorrect modes and distributions in the histogram. Nonetheless, the parameter characteristics of all the parameters are generally similar in terms of major trends. Note that the use of the subjective descriptions should always be accompanied by graphical representations.

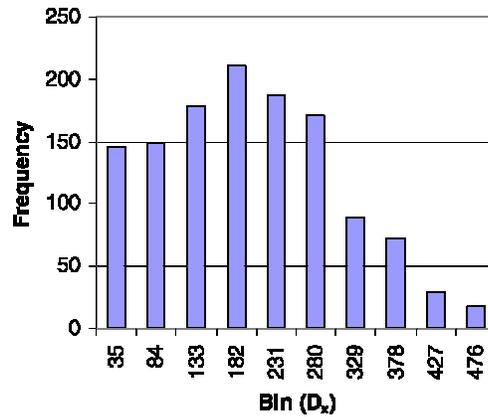
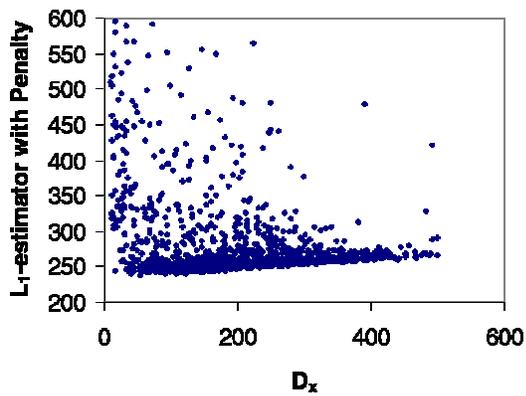
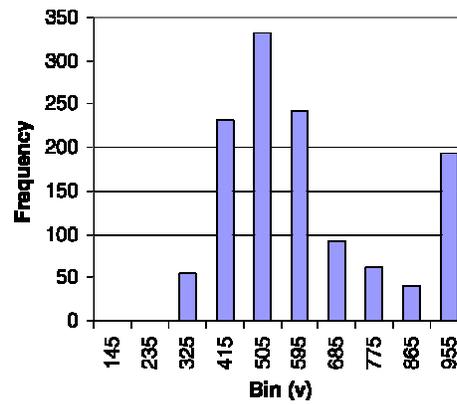
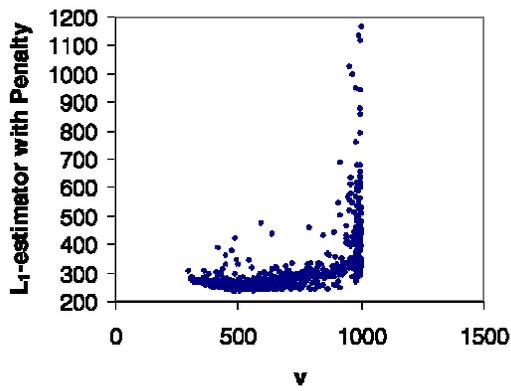
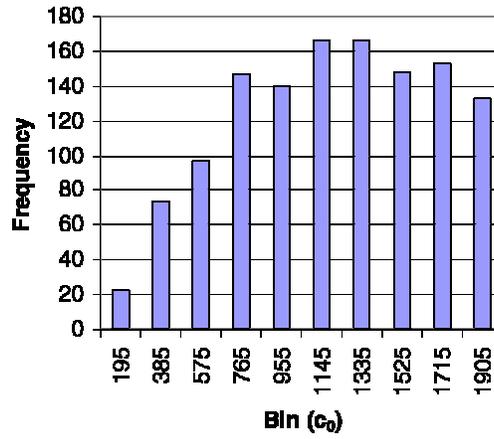
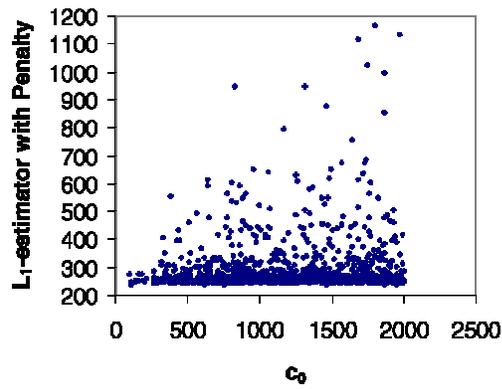


Figure 3.17. Parameter values that were generated using the L_1 -estimator with a penalty of 10,000 in terms of performance index values and frequency for parameter sets that satisfy the physical criteria

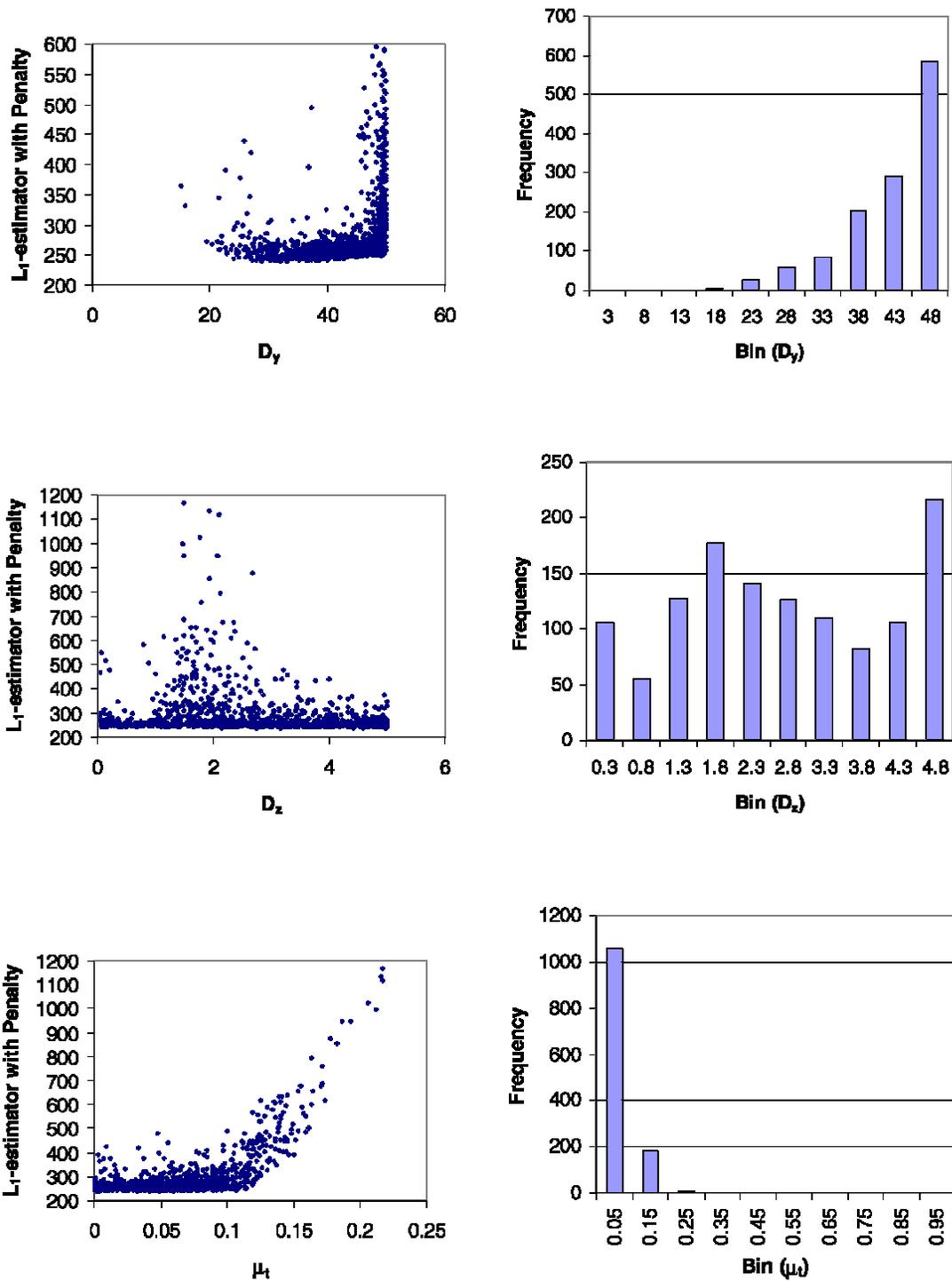


Figure 3.17. Parameter values that were generated using the L_1 -estimator with a penalty of 10,000 in terms of performance index values and frequency for parameter sets that satisfy the physical criteria (continued)

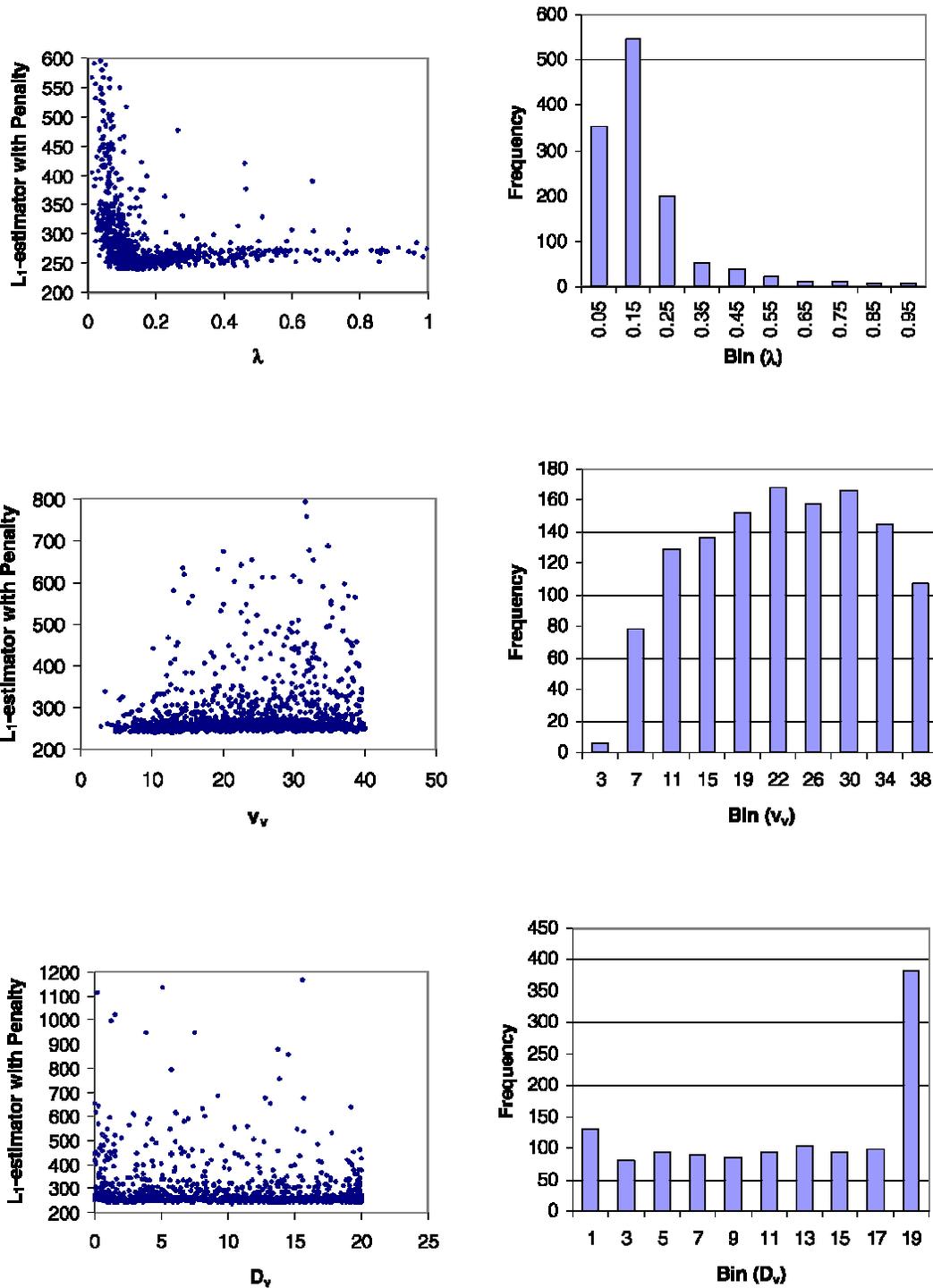


Figure 3.17. Parameter values that were generated using the L_1 -estimator with a penalty of 10,000 in terms of performance index values and frequency for parameter sets that satisfy the physical criteria (continued)

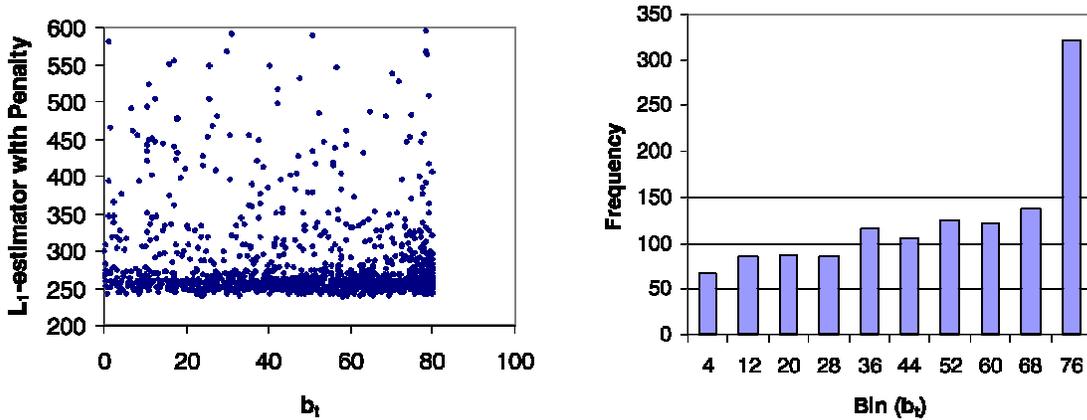


Figure 3.17. Parameter values that were generated using the L_1 -estimator with a penalty of 10,000 in terms of performance index values and frequency for parameter sets that satisfy the physical criteria (continued)

3.4.2 Parameter Correlations

The parameter correlation structure can explain the failure of many parameter estimation processes, which assume independence between parameter sets. It is important to note that problems with high dimensionality such as this case study have a more complex correlation structure and difficulties are likely to be encountered in the parameter estimation processes. Therefore, an understanding of the parameter correlation structure can be used in future estimation and uncertainty analysis efforts. For example, no correlation is likely to be visible for a pair of parameters with uniform distributions, while a strong correlation is likely to be visible for a pair of parameters with non-uniform distributions that have large kurtosis. In general, parameter distributions have been shown to evolve with optimization and refined with the application of the two criteria; as a result, evolution and refinement are expected to also be seen in the correlation structure. Therefore, all three parameter spaces are examined and compared using correlation coefficients.

A common measure of correlation is the Pearson product-moment correlation coefficient, r , which is defined as (Johnson, 2000),

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) \quad (24)$$

where x_i and y_i for $i = 1, 2, \dots, n$ are paired data points, \bar{x} and \bar{y} are the expected values (i.e. means) of the data points (x_i, y_i) , and s_x and s_y are the standard deviations of the data points (x_i, y_i) . An

assumption inherent in this correlation coefficient is normality since means and standard deviations are used to represent the parameter distribution. Since the nature of parameter distributions is uncertain, a measure of correlation that is independent of known parameterized distributions such as Spearman's rank correlation coefficient may be more appropriate. Spearman's rank correlation coefficient, r_s , is determined using (Johnson, 2000),

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (25)$$

where d_i represents the difference between the ranks of the data points (x_i, y_i) .

A basis for interpreting the coefficients that range from -1 to +1 is established using visual inspection of Pearson's coefficient for a linear case. The interpretations, presented in Table 3.16, are also assumed to be applicable for Spearman's coefficient.

Table 3.16. Interpretation of correlation coefficient values

Correlation	Negative	Positive
None	-0.10 to 0	0 to 0.10
Small	-0.29 to -0.10	0.10 to 0.29
Medium	-0.49 to -0.30	0.30 to 0.49
Large	-1.00 to -0.50	0.50 to 1.00

3.4.2.1 Prior Parameter Space

Using the same 6000 parameter sets used in Section 3.4.1.1, correlation tables were produced using (24) and (25) and presented in Tables 3.17 and 3.18. Since non-uniform behaviour is noticed in Section 3.4.1.1 for selected parameters, minor correlations are expected.

The majority of the parameter combinations, which represents 91% of the 45 pairs, are not correlated. One parameter pair is considered to be highly correlated; while three pairs are considered to be slightly correlated. Although the coefficient values are slightly different, the same degree of correlation between parameter pairs is achieved with both coefficients. For example, both Pearson's and Spearman's coefficient for the parameter pair, λ and μ , is less than -0.5. (Note that this was expected since the distributions for these parameters did not exhibit a uniform nature.)

Table 3.17. Pearson product-moment correlation coefficient for random parameter sets

	c_0	v	D_x	D_y	D_z	μ_t	λ	v_v	D_v	b_t
c_0	1									
v	0.03	1								
D_x	0.01	-0.03	1							
D_y	0.02	-0.01	0.00	1						
D_z	0.00	-0.02	0.00	-0.01	1					
μ_t	0.07	0.12*	0.01	-0.01	-0.03	1				
λ	0.05	-0.18*	0.02	-0.03	-0.03	-0.50**	1			
v_v	-0.01	-0.02	0.02	0.01	0.01	0.18*	-0.02	1		
D_v	0.00	0.02	0.00	0.01	0.01	0.01	-0.01	0.00	1	
b_t	0.02	-0.01	-0.01	-0.03	0.03	-0.08	0.06	0.01	0.01	1

*Small correlation between parameters

**Large correlation between parameters

Table 3.18. Spearman's rank correlation coefficient for random parameter sets

	c_0	v	D_x	D_y	D_z	μ_t	λ	v_v	D_v	b_t
c_0	1									
v	0.03	1								
D_x	0.01	-0.03	1							
D_y	0.01	-0.01	0.00	1						
D_z	0.00	-0.02	0.00	-0.01	1					
μ_t	0.08	0.10*	0.00	-0.02	-0.03	1				
λ	0.07	-0.19*	0.02	-0.03	-0.03	-0.56**	1			
v_v	-0.01	-0.02	0.03	0.01	0.01	0.19*	0.00	1		
D_v	0.00	0.02	0.00	0.01	0.01	0.00	-0.01	0.00	1	
b_t	0.02	-0.01	-0.01	-0.03	0.02	-0.08	0.05	0.01	0.01	1

*Small correlation between parameters

**Large correlation between parameters

3.4.2.2 Optimized Parameter Space

Since the difference between the two coefficients are small, only the Pearson product-moment correlation coefficient is used to assess correlations of parameter sets generated using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function. Given that the parameters are not likely to be normally-distributed, the correlations for each

parameter pair are qualitatively described in Table 3.19 using the interpretation presented in Table 3.16.

Table 3.19. Summary of qualitative interpretation of Pearson product-moment correlation coefficients of the parameter sets generated using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function

	c_0	v	D_x	D_y	D_z	μ_t	λ	v_v	D_v	b_t
c_0	1									
v	N-S	1								
D_x	N	S	1							
D_y	N	M-L	M	1						
D_z	N	N	N	N	1					
μ_t	N-S	M-S	N-S	N-S	N-S	1				
λ	N-M	L	N-S	N-M	S	M-S	1			
v_v	N-S	N	N-S	N-S	N	S	S	1		
D_v	N-S	N-M	S	N-S	N	S	N-M	N-S	1	
b_t	N	N-S	N-S	N-S	N	N-S	N-S	N	N	1

Note: “N” = None; “N-S” = None to Small; “S” = Small; “N-M” = None to Medium; “M-S” = Medium to Small; “M” = Medium; “M-L” = Medium to Large; “L” = Large

Only 1 out of the 45 parameter pairs can be considered to be strongly correlated, and 36 out of the 45 pairs are considered to have small to no correlation.

It is important to note that differences arise based on the selection of estimator. Therefore, the results based on all three estimators are combined in Table 3.19 to facilitate comparisons. (The parameter pairs with a difference in correlations are described using hyphenated abbreviations). Twenty of the parameter pairs, which correspond to 44% of the 45 pairs, in Table 3.19 can be considered to have the same degree of correlation. Parameter pairs are assumed to exhibit a similar degree of correlation if the parameter pair can be considered to be in either of two adjacent categorizations. For example, the correlation of a parameter pair with N-S correlation can be considered to be similar since the coefficient values for the small and none category are similar. Table 3.19 indicates that 44% of parameter pairs have similar correlations while 11% of the parameter pairs can be considered to have different degrees of correlation. Therefore, the correlation of the majority of parameter pairs (i.e. 89%) is similar regardless of estimator selection.

3.4.2.3 Acceptable Parameter Space

Despite the consistency between estimators noted in Section 3.4.2.2, the Pearson product-moment correlation coefficients for all three acceptable parameter spaces are combined and qualitatively described in Table 3.20.

Table 3.20. Summary of qualitative interpretation of Pearson product-moment correlation coefficients of the parameter sets generated using the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator with a scaling factor of 10,000 for the penalty function

	c_0	v	D_x	D_y	D_z	μ_t	λ	v_v	D_v	b_t
c_0	1									
v	N	1								
D_x	N	S-L	1							
D_y	N	S-L	M-S	1						
D_z	N-S	N-S	N	N-S	1					
μ_t	M-S	L	M-S	S-L	S	1				
λ	S	L	M	N-M	S	M-L	1			
v_v	S	N	N-S	N-S	N	M-L	N-S	1		
D_v	N	M-S	M-S	M-S	N-S	M-S	M-S	N	1	
b_t	N	N-S	N-S	N-S	N	N-M	N-M	N	N	1

Note: “N” = None; “N-S” = None to Small; “S” = Small; “N-M” = None to Medium; “M-S” = Medium to Small; “S-L” = Small to Large; “M” = Medium; “M-L” = Medium to Large; “L” = Large

Table 3.20 shows that parameter correlations are strengthened by the application of the two criteria. The proportion of parameter pairs that can be considered to be strongly correlated are increased to 4%, and 58% of the pairs are considered to have small to no correlation.

3.5 Summary

The parameter estimation process was catered to the specific insufficiencies noted in Chapter 2 where the model was developed and the observation data was examined. Preliminary investigation of the problem was conducted using a well-accepted parameter estimation method, PEST, and the corresponding findings are used to define characteristics of the parameter estimation process. The selected estimation process includes numerous definitions for the objective function and the consideration of multiple “good” parameter sets. For each objective function definition, three procedures were implemented as a part of the parameter estimation method for the given case study:

an initializing procedure, a search for “good” parameter sets using the DDS-UA method, and a test for acceptance based on physical criteria.

With respect to objective function definitions, the best performance in terms of the acceptance rate was achieved using the LRS1 M-estimator. However, a penalty function based on the limiting criteria, the *Location Criteria*, must be included in the objective function to produce a sufficient number of acceptable parameter sets for use in uncertainty analysis. Since it is difficult to justify the selection of a single objective function definition, the parameter sets obtained using the two L-estimators are also selected based on their characteristics and are recommended for use in conjunction with the parameter sets obtained using the LRS1 M-estimator.

The solution space generated by all of the estimators tested consists of many local minimums, which is addressed by the adoption of equifinality. Optimization of the numerous initial parameter sets appears to have greater effect on converging parameter characteristics compared to the remaining two procedures in the parameter estimation method chosen in this thesis. Optimization is designed to remove parameter sets with high objective function values, which represent a majority in the prior parameter space. Any remaining parameter sets with high objective function in the optimized parameter space are removed with the application of the two criteria. The acceptable parameter space defined by applying the two criteria to the optimized parameter space is also an important step as it further refines parameter characteristics. Note that even further refinements by the addition of other criteria can improve parameter estimates. The combined effect of optimization and the application of the two criteria performs the function of behavioural thresholds without the addition of subjectivity.

Chapter 4

Uncertainty Analysis: First Arrival Estimates

Sources of uncertainty in inverse modelling can be categorized into parameter uncertainty, model uncertainty, and scenario uncertainty (USEPA, 1997). Biases in data can arise from imprecise calibration of devices, inaccurate assumptions, non-representative samples, and misclassification, which all contribute to parameter uncertainty. Model uncertainties can originate from error associated with parameterization, boundary conditions, structure, and resolution. Lastly, scenario uncertainty can be caused by errors in conceptualization and professional judgement. The identification and treatment of all sources of uncertainty is desirable but not always feasible in practice. In addition to uncertainty, there is variability caused by natural random processes such as subsurface heterogeneity. Although there should be methodological differences between the analysis of uncertainty and variability, they are difficult to separate in problems where further data collection is not possible. In this thesis, variability is inherently incorporated into the model development by using the equivalent porous media concept and uncertain averaged model parameters. The different sources of uncertainty associated with the first arrival estimate, which now include variability, are difficult to distinguish and are simultaneously included in the final representation of uncertainty.

All three sources of uncertainty are addressed to differing degrees in Chapters 2 and 3. In Chapter 2, the problem is formulated in such a way that the conceptual model and the corresponding assumptions minimize model and scenario uncertainties, and are reasonable given available data. Nonetheless, there still remain model and scenario uncertainties. These uncertainties are not easily quantifiable using available methods such as the Maximum Likelihood Bayesian Model Averaging method (Neuman, 2003) because of the difficulty associated with justifiably defining other possible models and scenarios. Therefore, the representation of error in the parameter estimation process includes errors and biases caused by both the model and the data as discussed in Chapter 3. Objective functions have been designed to consider model errors such as the total least squares (TLS) method (Sun et al., 2006); however, these methods increase the degree of freedom, which can cause the problem to be more ill-posed and the parameter estimation process to be more difficult. Without a direct consideration of model uncertainties, parameter estimates may compensate for model errors and biases. Fortunately, the consideration of multiple parameter sets as performed in Chapter 3 can indirectly address this problem since the degree to which each of the multiple parameter set represents

uncertainties varies. To treat specific insufficiencies in the observed data, robust statistics in the parameter estimation process are used in Chapter 3. A further refinement in terms of parameter characteristics and the removal of parameter sets with high objective function values are achieved with the application of two physical criteria: *Location Criteria* and *Product Criteria*. The usage of the resulting multiple acceptable parameter sets that include treatments for insufficiencies and satisfy physical criteria is assumed to be sufficient for a comprehensive uncertainty analysis given the available resources.

The uncertainty analysis performed in this thesis focuses on representing input uncertainty and propagating the input uncertainty into output uncertainty. Approaches to perform uncertainty analysis include interval mathematics (Dou et al., 1995), fuzzy set theory (Dou et al., 1995), and probabilistic analysis (USEPA, 1997). Interval mathematics provides a framework for dealing with parameter uncertainties of unknown structure; however, the result may not be indicative of output range since the model output is not a simple monotonic function of the parameters. For example, the use of the upper limit for all parameter values does not guarantee that the upper limit of the solution will be generated given the non-linear nature of the problem and correlations between uncertain parameters. There are alternate approaches that have modified the search method to find the best and worst case scenarios without using interval mathematics (Brooks et al., 1994). Although the best-and-worst case scenarios provide valuable information, the distribution of possible outputs near the parameter limits can have a strong influence on the tails of output distributions, which play an important role in decision-making scenarios. The impreciseness of many inputs to groundwater flow and contaminant transport models have led to formulations of fuzzy groundwater flow models (Dou et al., 1995), which can be useful in decision making scenarios. The most commonly used approach to uncertainty analysis in groundwater flow and contaminant transport modelling is probabilistic analysis, which is designed to generate the nature of output uncertainty. Probabilistic uncertainty analysis can take the form of analytic methods based on stochastic partial differential equations (Dou et al., 1995) or sampling based methods such as Monte Carlo analysis (USEPA, 1997). Analytical methods using linear and non-linear first-order approximations of 95% confidence intervals that can be obtained with gradient-based parameter calibration methods are commonly used and evaluated optimistically (Vecchia and Cooley, 1987; Christensen and Cooley, 1999). However, the reliability of this method is limited to well-posed problems with known error distributions. In contrast, probabilistic analysis using sampling-based methods such as Monte Carlo analysis can provide practical output probability distributions with reasonable intervals even for ill-posed problems. In fact, the use of Monte Carlo

simulations in uncertainty analysis is “most frequently encountered in human health risk assessments” (USEPA, 1997) and used as a bench mark for validating other types of uncertainty estimation methods (Christensen and Cooley, 1999; James and Oldenberg, 1997).

Despite the general acceptance of the results produced by Monte Carlo uncertainty analysis, the results are only probabilistic approximations strongly dependent on input posterior parameter distributions even if numerical stability and convergence of both the central tendency and the tails of probability distributions are achieved using a sufficient number of simulations. Input parameter distributions are typically selected based on the characteristics of the globally optimum parameter set. However, the use of only one “optimum” parameter set is difficult to justify given the nature of real problems (Kuczera and Parent, 1998). Given limited prior knowledge, subjectivity and uncertainty are introduced in the process of describing the nature of parameter distributions, and parameter estimates are sensitive to definitions of the objective function and to the parameter estimation method as discussed in Chapter 3. Recall that the purpose of parameter estimation was not to find the global optimum but to find many local optimums or multiple maximum likelihood parameter estimates. The characteristics of these parameter sets provide valuable information in the form of empirical parameter distributions and correlation structure. Although the empirical parameter distributions can be approximated and used as posterior parameter distributions in the Monte Carlo uncertainty analysis framework, problems can still arise due to the inability of Monte Carlo analysis to account for dependence between parameters, which has been shown to be significant in environmental problems (Hornberger and Spear, 1981). To minimize potential problems caused by the interaction between parameters in Monte Carlo analysis, USEPA (1997) recommends limitations on the number of uncertain parameters through prior knowledge and sensitivity analysis. Additional prior knowledge via data collection is not always possible, while a comprehensive sensitivity analysis can be labour-intensive for non-linear, non-monotonic models and when the parameter values span many orders of magnitude. Converting an uncertain parameter into a constant using either of these approaches can lead to premature and unjustifiable decisions; thus, a continual “review of the basis of fixing certain parameters” is recommended (USEPA, 1997). Furthermore, the accuracy of the tails of the output distribution is questionable given the subjectivity of the input parameter distributions even if numerical stability and convergence of both the central tendency and the tails of probability distributions are achieved. In fact, as discussed in Chapter 3, the quality of the optimal parameter sets, in terms of their conditionality to observations, is low at the tails. This is especially problematic given that we would like to accurately estimate the tails of the output distribution.

Therefore, the concept of examining multiple equally probable parameter sets, also known as equifinality (Beven, 2005), is adopted for use in this uncertainty analysis. Parameter uncertainty analysis approaches based on the concept of equifinality are considered to be Monte Carlo type or sampling-based assessments. The difference between popular Monte Carlo type assessments, such as the traditional Monte Carlo analysis discussed above, the generalized likelihood uncertainty estimation (GLUE) (Beven and Binley, 1992) and the Markov Chain Monte Carlo (MCMC) method (e.g., Kuczera and Parent, 1998), is basically in the sampling approach used. Despite the improved efficiency in exploring the parameter space found with the MCMC method over the GLUE method (Kuczera and Parent, 1998), we observed that the need for improvements at the accuracy of the tails is not directly addressed by any of these methods. The Dynamically Dimensioned Search – Approximation of Uncertainty (DDS-AU) approach introduced by Tolson (2005) uses a parameter optimization scheme that is designed to produce multiple “optimal” solutions as the sampling approach embraces the equifinality concept and samples for multiple "optimal" solutions using multiple independent optimization trials of the DDS algorithm. Relative to GLUE, Tolson (2005) showed an example where GLUE sampling was orders of magnitude less efficient than DDS-AU for identifying high likelihood solutions. Therefore, the DDS-AU approach was selected with the interest of increasing the accuracy at the tails.

The uncertainty of the arrival of a contaminant concentration of 0.5 ppb at the four locations of concern, Leclercq Class Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D, is evaluated in the following analysis, using the steps outlined in Section 4.1. The traditional approach to uncertainty analysis using Monte Carlo sampling is reviewed and its shortcomings are summarised for the contaminant transport problem under consideration in this thesis. The DDS-AU sampling approach is modified as described in Section 4.1.1.2 to accommodate the two physical criteria defined for the given case study. These samples are used to generate output likelihood distributions using various definitions for likelihood functions. Considerations are also made for different objective function definitions, the number of parameter sets, and the factors in likelihood functions. Based on the comparisons of the controlling factors of likelihood distributions, an approach to conservatively represent uncertainty is devised.

4.1 Steps to Uncertainty Analysis

Sampling-based uncertainty analysis can be divided into two major steps: generation of samples, and evaluation and compilation of samples to generate probability or likelihood distributions. In this thesis, a sample is defined as a set of parameter values.

4.1.1 Sampling Approaches

Two viewpoints can be followed in the development of a sampling strategy as discussed in the previous section: the existence of one unique globally optimum parameter set, or the existence of multiple probable parameter sets. Sampling strategies based on both approaches are investigated in this thesis. From the perspective of one unique global optimum, the Monte Carlo sampling approach is used. Multiple probable parameter sets, on the other hand, are generated using a revised DDS-AU method.

4.1.1.1 Monte Carlo or Latin Hypercube Sampling

Despite the need to examine multiple acceptable parameter sets, the Monte Carlo sampling approach based on the existence of a global optimum is implemented due to its popularity in practice.

The commonly-used Monte Carlo sampling strategy assumes independence between parameter sets and randomly samples parameter values from their posterior parameter distributions (USEPA, 1997). A combination of prior knowledge and information gained through the parameter estimation process is used to define the posterior parameter distributions. Examples of its application are widely available in the literature (e.g., Sohrabi et al., 2002).

Efficiency in terms of computational time required for convergence of distributions can be improved using Latin Hypercube sampling, where parameter values are independently divided into equally probable intervals, or Orthogonal sampling, where the parameter space is equally divided. Nonetheless, the resulting sample at convergence is theoretically equal to the converged sample generated using Monte Carlo sampling (McKay et al., 1979).

4.1.1.2 Physically-Based DDS-AU (P-DDS-AU) Sampling

In Chapter 3, the two physical criteria, *Locations Criteria* and *Product Criteria*, are used to determine the acceptability of a parameter set locally optimized using the DDS algorithm. The DDS-AU sampling methodology (Tolson, 2005) outlined in Appendix F is designed only for the application of a behavioural threshold based on performance index values; thus, it is not fully

compatible with the new concept of acceptability introduced in this thesis, and therefore modifications are required.

In the DDS-AU sampling methodology, the number of model evaluations per DDS optimization trial (m) is calculated using the maximum total number of model evaluations (N) and the desired number of behavioural samples to identify (n) as follows

$$m = N/n \quad (26)$$

(26) is only valid if all of the N model evaluations are used to generate behavioural solutions. In essence, this is feasible within the definition of behavioural based on a threshold value and the optimization ability of the DDS algorithm. The DDS algorithm is designed to lower the objective function value with each iteration. A trial that produces a non-behavioural solution can be further optimized using the DDS algorithm with additional model evaluations. If sufficient model evaluations are performed, all DDS trials are assumed to be able to generate behavioural solutions.

Within the new definition referred to as acceptability, the equality in (26) no longer holds and $m < N/n$. The acceptance rate, a , which is calculated in Chapter 3 for various objective function definitions, must be used to find m as follows

$$m = \frac{N \cdot a}{n} \quad (27)$$

where n now represents the desired number of acceptable samples. (27) approaches (26) as a , which can range from 0 to 1, approaches 1. The value of a is dependent on many factors including data, objective function definition, and the three variables, m , n , and N . The value of a for this case study was determined to have a broad range from 0.0 to approximately 0.5 due to uncertainties, errors, and biases originating from various sources. As a result, the a value should be increased during the sampling procedure to a sufficient value by adjusting the objective function definition, m , and/or N .

The DDS-AU sampling methodology is revised to accommodate these differences and is referred to as the physically-based DDS-AU sampling methodology (P-DDS-AU), which is given as:

1. Define the maximum total number of model evaluations for analysis, N . This value is governed by computational resources and time constraints.
2. Define the desired minimum number of acceptable samples, n , that the sampling approach must identify. Review of literature on GLUE applications typically use 1000 to 1500 behavioural samples (Feyen et al., 2001; Christensen, 2003; McMichael et al., 2006) at which convergence is assumed. If no prior knowledge is available, an n value equal to or greater than 1000 should be used.

3. Define the number of model evaluations per DDS optimization trial, m , using DDS performance compared to other parameter estimation methods (e.g., PEST). If convergence is desirable, m values should not be decreased below the assumed number of model evaluations required for convergence. It is important to note that the benefit of DDS can be lost if a sufficient number of model evaluations are not used per DDS optimization trial.
4. Select an objective function definition.
5. Define the total number of DDS trials (t):
 - a) If prior knowledge of acceptance rate is known, $t = n/a$, else
 - b) $t = N/m$.
6. Perform t DDS optimization trials from t random or user-defined initial solutions and save only the final DDS solutions from each optimization trial.
7. If a is unknown, classify each of the n DDS solutions as acceptable or not acceptable using the two criteria: *Locations Criteria* and *Product Criteria*.
8. If a is unknown, calculate $a = n/t$.
9. If the number of acceptable solutions is too low (i.e. $n < 1000$) then choose one of the following options:
 - a) Select a different objective function definition, then go to Step 5,
 - b) Increase N to a value of $\frac{m \cdot n}{a}$, then go to Step 5a,
 - c) Decrease m to a value of $\frac{N \cdot a}{n}$, then go to Step 5b, or
 - d) Increase N to a value less than $\frac{m \cdot n}{a}$ and decrease m to a value greater than $\frac{N \cdot a}{n}$, then go to Step 5b.

4.1.2 Quantitative Representation of Uncertainty

Quantitative representations of uncertainty are assumed to be more useful in decision-making scenarios in comparison to qualitative representations. Quantitative representation of output uncertainty is facilitated by probability or likelihood distributions in both the cumulative and non-cumulative sense.

4.1.2.1 Probability, Likelihood, and Pseudo-Likelihood

In this thesis, the probability or likelihood of the time of arrival of a specific contaminant concentration at the four locations of concern given the observed concentration data is desired. A set of arrival times with respect to location is unique to each parameter set, which represents a contaminant plume history. The parameter sets that do not satisfy the two criteria are deemed not physically possible and are assigned a probability of zero. Therefore, the probability or likelihood of a parameter set, θ , correctly representing the plume is assumed to equal the probability or likelihood of the corresponding vector of contaminant arrival times, τ , and thus, we assume $P(\tau | \theta) = P(\theta)$. Consequently, the model and scenario uncertainty are assumed to be incorporated into parameter uncertainty.

The definition of probability and likelihood, which are generally used interchangeably, depends on the perspective. From the frequentist perspective, probability or likelihood can simply be thought in terms of an occurrence and/or frequency of an event given a population of events. From a Bayesian perspective, probability or likelihood includes conditional and on occasion subjective knowledge, which is incorporated through Bayes' theorem. Probability or likelihood from either perspective are designed to satisfy all three axioms of probability (Johnson, 2000), and are therefore referred to here only as probability. Both interpretations are controversial in different ways and it is difficult to select one perspective. As a result, more than one approach to expressing probability is adopted, which is in line with the inclusive manner of the approaches selected in this thesis.

The frequentist interpretation of probability assumes that each sample is equally probable and the probability assigned to each sample is

$$P_f(\theta) = 1/n \quad (28)$$

where f is used to denote the frequentist interpretation, and n is the number of acceptable parameter sets as defined in the previous section. However, the probability of a parameter set can be considered to be conditional to the observations, c_1, c_2, \dots, c_n , if a maximum likelihood perspective is taken to generate the parameter sets. The Bayes' representation of probability can be implemented by using Bayes' theorem to incorporate this condition as

$$P(\theta | c_1, c_2, \dots, c_n) = \frac{P(\theta) \cdot P(c_1, c_2, \dots, c_n | \theta)}{P(c_1, c_2, \dots, c_n)} \quad (29)$$

The quantification of the conditional probability presented as (29) can be made by inferring the relationship between observations and parameter values from objective functions. Although we can

assume that $P(\theta) = P_f(\theta)$, $P(\theta) \neq P_f(\theta)$ can also be true from a Bayesian perspective. In addition, the correct approach to map the objective function to a probability and the probability of the observations, $P(c_1, c_2, \dots, c_n)$, are unknown.

Therefore, only objective function values are used to provide a measure of the probability of a parameter set. Since these values are not equal to probability, they are referred to as pseudo-likelihoods, L , in this thesis. According to Beven and Binley (1992) only two conditions must be satisfied by the “likelihood” function:

1. “It should be zero for all simulations that are considered to exhibit behaviour dissimilar to the system under study.”
2. “It should increase monotonically as the similarity in behaviour increases.”

For the first condition, the concept of behaviour as interpreted by Beven and Binley (1992) is not used; however, the similar concept, acceptability, is applied in this thesis to the pseudo-likelihood definitions by assigning a value of zero to simulations that do not satisfy the two criteria. In fact, the P-DDS-AU sampling approach simply removes parameter sets that do not satisfy the two criteria from further consideration. For the second condition, the interpretation of behaviour used by Beven and Binley (1992), which is based on the model’s ability to resemble observed condition, is adopted in the pseudo-likelihood definitions used in this thesis.

Feyen et al. (2001) use pseudo-likelihoods based on the L_2 -norm in their implementation of the GLUE method as given by

$$L_{c/d}(\theta_k | c_1, c_2, \dots, c_n) = (P_k)^{-N} \quad (30)$$

with

$$P_k = \frac{\sum_{i=1}^{n_{obs}} (c_{sim,i} - c_{obs,i})^2}{n_{obs}} \quad (31)$$

where c/d denotes the type of likelihood function, N is the shaping factor subjectively chosen by the user, k represents the k^{th} parameter set, n_{obs} is the number observations, $c_{sim,i}$ is the i^{th} simulated concentration at location (x,y,z) at time, t , and $c_{obs,i}$ is the i^{th} observed concentration at location (x,y,z) at time, t . Weights that change the impact of an observation based on its quality can be included in the P_k value but equal weighting is assigned to each observation due to the lack of information regarding the quality of observations. In addition, P_k is redefined in this thesis to be any objective function value divided by n_{obs} since objective function definitions are not limited to the L_2 -norm. Since the parameter sets that did not satisfy the two criteria are not included in the uncertainty

analysis, the effect of the penalty function described in Chapter 3 in the objective function should be small. In fact, our investigations show that the increase in objective function values for acceptable parameter sets due to larger penalty factors is small especially in terms of the lower limit of objective function values generated. Moreover, the objective functions that include a penalty function create likelihoods that are conditional to both the observations and the parameter set's ability to satisfy physical criteria. If the physical criteria are used to determine the acceptance of a parameter set, it makes intuitive sense to also use the criteria to assess the degree of likelihood. Therefore, the existence of the small difference is justifiable and the use of the objective function value even with a penalty in P_k is reasonable.

The likelihood definition parameter, N , is subjectively chosen by the user. If N equals 0, $L_{c/d}$ is 1 and each parameter set is given equal weighting; the pseudo-likelihood is no longer conditional to observations and follows a frequentist perspective of probability. As N is increased, the difference in pseudo-likelihoods between parameter sets is magnified and the pseudo-likelihoods with large P_k values approach zero as shown in Figure 4.1. As N approaches infinity, the parameter sets with the lowest P_k value will have a $L_{c/d}$ value of infinity while the remaining parameter sets will have a value of zero. As a result, the effect of N can be significant in the representation of uncertainty and care must be taken in selecting an appropriate value.

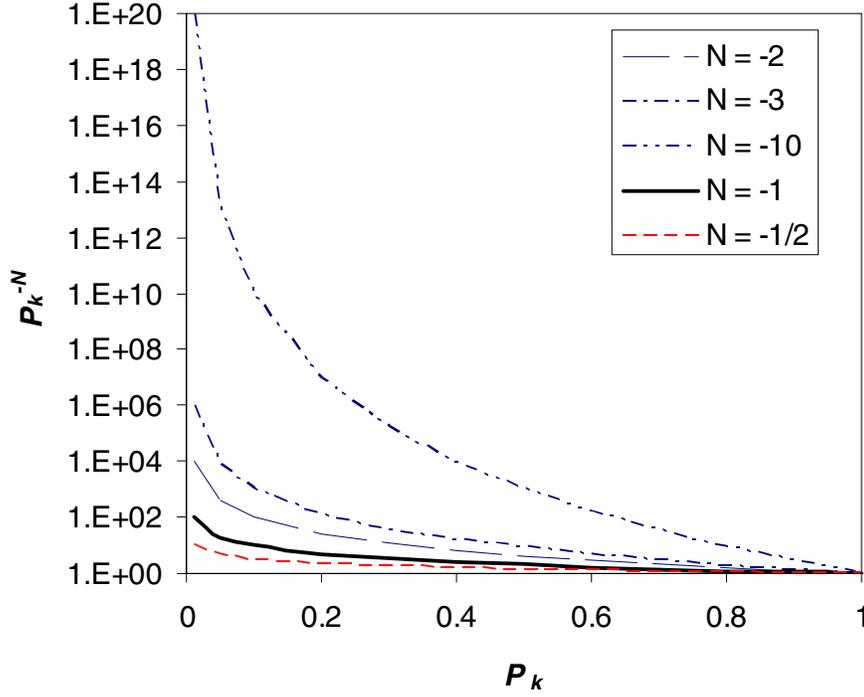


Figure 4.1. Effect of N value on $L_{c/d} = P_k^{-N}$

An alternate form of pseudo-likelihood definition based on Nash and Sutcliffe efficiency criterion (Feyen et al., 2001; Beven and Freer, 2001) is still based on the P_k value but the pseudo-likelihood values are forced to range from 0 to 1 by incorporating the variance of observations, σ_o^2 as given by

$$L_{a/b}(\theta_k | c_1, c_2, \dots, c_n) = (1 - P_k/\sigma_o^2)^N \quad P_k \leq \sigma_o^2 \quad (32)$$

with

$$\sigma_o^2 = \frac{\sum_{i=1}^{n_{obs}} (c_{obs,i} - \overline{c_{obs}})^2}{n_{obs}} \quad (33)$$

where a/b denotes the type of likelihood function. The parameter set is considered to be unlikely when P_k equals or is greater than the variance of the observations, σ_o^2 . In contrast, the parameter set is considered likely if P_k approaches zero, which implies that the residuals approach zero. As shown in Figure 4.2, the value of N in this definition magnifies the relationship by lowering the pseudo-likelihood values for low $(1 - P_k/\sigma_o^2)$ values or high P_k/σ_o^2 values to a greater degree than for higher $(1 - P_k/\sigma_o^2)$ or lower P_k/σ_o^2 values. However, the shape of this relationship is different for N values

greater than 1 and N values less than 1. Therefore, the N values investigated should be 1, a value less than 1, and a value greater than 1.

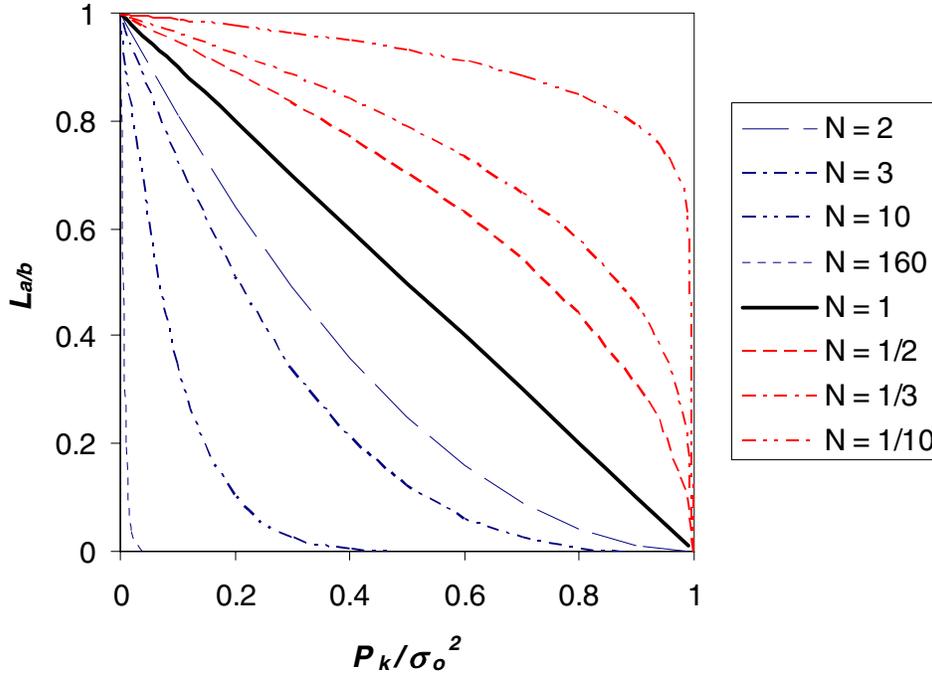


Figure 4.2. Effect of N value on $L_{a/b} = (1 - P_k / \sigma_o^2)^N$

Also, $L_{a/b}$ assigns a value of zero to the parameter set in cases where the inequality, $P_k \leq \sigma_o^2$, is not satisfied. In this thesis, the variance of observations, σ_o^2 , is interpreted as the variance of all observations regardless of time and location. Therefore, the overall structure of the observation data as opposed to uncertainty of the data is being replicated. An alternative is to determine the variance of observations and simulations at specific locations and times. However, this is not feasible given the data provided as repeated observations at a particular location or time are not available. Therefore, the usage of σ_o^2 is questionable; nonetheless, it produces a finite range of pseudo-likelihood values.

The definition, $L_{a/b}$, compares the variance of observation, σ_o^2 , to P_k . It also makes intuitive sense to compare the variance of observations, σ_o^2 , to the variance of residuals, σ_r^2 , which can be calculated from

$$\sigma_r^2 = \frac{\sum_{i=1}^{n_{obs}} (r_i - \bar{r})^2}{n_{obs}} \quad (34)$$

where $r_i = r_i(x,y,z,t) = c_{sim,i} - c_{obs,i}$, and \bar{r} is the average of all r_i values. Although P_k is closely related to σ_r^2 , the resulting values are different. The model efficiency or the coefficient of determination, L_e , (Beven and Binley, 1992) incorporates the variance of residuals, σ_r^2 , into the definition of pseudo-likelihood as

$$L_e(\theta_k | c_1, c_2, \dots, c_n) = (1 - \sigma_r^2 / \sigma_o^2) \quad \sigma_r^2 \leq \sigma_o^2 \quad (35)$$

where e denotes the type of likelihood function. The model efficiency or coefficient of determination, L_e , is similar to the form of $L_{a/b}$ in (32) except that P_k is used instead of σ_r^2 . The incorporation of an N value in a similar manner as $L_{a/b}$ and $L_{c/d}$ into (35) is not done due to the lack of such incorporation in literature.

Similarly, the variance of residuals, σ_r^2 , can replace P_k in $L_{c/d}$ to produce the likelihood function (Beven and Binley, 1992) given by

$$L_b(\theta_k | c_1, c_2, \dots, c_n) = (\sigma_r^2)^{-N} \quad (36)$$

where b denotes the type of likelihood function.

In general, the four definitions for the pseudo-likelihood function, $L_{c/d}$, $L_{a/b}$, L_e , and L_b , summarized in Table 4.1 can be considered to be equally valid.

Some may view the four definitions of pseudo-likelihoods as probabilities or likelihoods in the Bayesian sense if they satisfy all three axioms of probability. The third axiom of probability is easily satisfied since the samples are assumed to be mutually exclusive. Pseudo-likelihood values using the definition, L_e , are designed to be between 0 and 1 and therefore satisfy the first axiom of probability; however, the remaining definitions fail to do so when the value of P_k or σ_r^2 is less than 1. The second axiom of probability can be satisfied if the pseudo-likelihood values are normalized to total 1. However, the set of samples are not likely to fully represent the sample space since the parameters are continuous variables. Even if the three axioms can be satisfied, decision-makers must be aware that an arrival time with a cumulative normalized likelihood value of 10%, for example, does not necessarily imply that there are 1 in 10 chances that the contaminant of concern will arrive at this time. A clear understanding of the definitions and assumptions is crucial when decisions are made based on pseudo-likelihoods or Bayesian probabilities.

Table 4.1. Likelihood definitions to be used to describe the likelihood of a parameter set

Description	Equation	Source
Model efficiency or coefficient of determination based on Nash and Sutcliffe efficiency criterion	$L_e = (1 - \sigma_r^2 / \sigma_o^2); \sigma_r^2 \leq \sigma_o^2$	Beven and Binley (1992)
Based on inverse residual variance with shaping factor, N	$L_b = (\sigma_r^2)^{-N}$	Beven and Binley (1992)
Based on Nash and Sutcliffe efficiency criterion with shaping factor, N	$L_{a/b} = (1 - P_k / \sigma_o^2)^N; P_k \leq \sigma_o^2$	Feyen et al. (2001)
Based on inverse error variance with shaping factor, N	$L_{c/d} = (P_k)^{-N}$	Feyen et al. (2001)

N is a likelihood definition parameter chosen by the user,
 k represents the k^{th} parameter set,
 n_{obs} is the number observations,
 $r_i = r_i(x,y,z,t) = c_{sim,i} - c_{obs,i}$,
 $c_{sim,i}$ is the i^{th} simulated concentration at location (x,y,z) at time, t ,
 $c_{obs,i}$ is the i^{th} observed concentration at location (x,y,z) at time, t ,

$$\sigma_e^2 = \frac{\sum_{i=1}^{n_{obs}} (r_i - \bar{r})^2}{n_{obs}},$$

$$\sigma_o^2 = \frac{\sum_{i=1}^{n_{obs}} (c_{obs,i} - \bar{c}_{obs})^2}{n_{obs}},$$

$$P_k = \frac{\sum_{i=1}^{n_{obs}} (c_{sim,i} - c_{obs,i})^2}{n_{obs}}$$

4.1.2.2 Non-Cumulative and Cumulative Distributions

Given discrete samples of continuous variables, a probability distribution can be approximated in a non-cumulative sense as

$$P(\tau) = \sum_j^{n_\tau} P_j(\tau | \theta_j) \quad (37)$$

where j represents the parameter set that produces the vector of arrival times, τ , and n_τ is the total number of parameter sets that produce the vector of arrival times, τ . (37) takes into consideration that

the same arrival time can be found using different parameter sets. The corresponding cumulative distribution, $P(\tau \leq T)$, can be approximated as

$$P(\tau \leq T) = \sum_t^{n_{\tau \leq T}} P_t(\tau \leq T) \quad (38)$$

where T represents the latest vector of arrival times represented by the cumulative probability. Non-cumulative and cumulative distributions described by (37) and (38) respectively for probability can also be applied to pseudo-likelihoods defined in Section 4.1.2.1 in the same manner. The pseudo-likelihoods should be normalized prior to the creation of distributions to produce a more meaningful representation. Computational details relating to the calculation of cumulative and non-cumulative distributions are presented in Appendix G.

These normalized likelihood distribution of arrival times can provide inferences on confidence limits and the nature of the uncertainty associated with arrival times. It is important to note that the intervals generated by GLUE using the likelihood definitions defined above have been questioned. Studies using synthetic data have been performed to investigate the accuracy of GLUE's 95% intervals to facilitate a comparison with the 95% confidence intervals generated with probabilistic analysis using covariance information. Montanari (2005) used a synthetic hydrological model with known sources of uncertainty and concluded that the interval generated using GLUE is an underestimate. Christensen (2003) introduced a correction to account for observation error, which improves the performance of GLUE intervals for a synthetic problem; nonetheless, he noted that small-scale fluctuations can still cause the GLUE interval to be inaccurate. Therefore, the GLUE method produces a distribution that can be considered to represent the large-scale trends as defined by McLaughlin and Townley (1996). Small-scale fluctuations are important in our quest for low likelihood arrival times. However, there is no method for quantifying the degree of small-scale fluctuations, which are assumed to predominantly originate from measurement errors. Although gradient-based methods are designed to provide an indication of normally-distributed small-scale fluctuations, the high dimensionality and nonlinearity causes these methods to fail. Therefore, the small-scale fluctuations should be estimated using prior knowledge on measurement errors and considered in the final decision.

4.2 Uncertainty Analysis Using Monte Carlo Sampling

Monte Carlo sampling requires information regarding the nature of parameter uncertainty. This nature can be approximated by standard analytic distributions such as the normal distribution or the

lognormal distribution, or estimated empirically as done in Chapter 3. The use of standard analytic distributions is popular in practice and easily implemented using available algorithms. The fact that slightly earlier convergence is achieved using logarithmic transformations of uncertain parameters in the optimization algorithm, PEST, may be an indication of log-normality in parameter distribution. Although the parameter characteristics found using DDS results do not appear to unanimously follow a lognormal distribution, the evidence supporting the use of Gaussian distributions to represent parameters is weaker. Therefore, the lognormal distribution was selected to represent all posterior parameter distributions. The parameters (the mean and the lognormal standard deviation) that are used to approximate posterior parameter distributions are estimated using two different parameter estimation methods: PEST and DDS. The corresponding output distributions were created using the approach described in Section 4.1.2 for the four locations of concern: Leclerq Class Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D. Recall that the output represents the arrival time of a contaminant concentration of 0.5 ppb.

Since the Monte Carlo sampling approach only considers the posterior parameter distributions and does not incorporate the two criteria, the parameter sets are not guaranteed to be acceptable. Even if acceptable parameter sets are found, there is no evidence suggesting that the parameter set with the lowest objective function value represents the global minimum unless the full extent of the parameter space is thoroughly searched. A thorough search of a 10-dimensional problem is not practically feasible, as 10^{10} function evaluations are required to examine 10 different values for each parameter. In fact, a reduction of dimensionality may be necessary for a meaningful uncertainty analysis given a practical amount of computational resources. In addition, many common random number algorithms such as Monte Carlo sampling are dependent on the user-specified seed value if convergence via a sufficient number of samples has not occurred. Therefore, a more efficient approach using the Latin Hypercube method is also implemented to check for convergence.

4.2.1 Monte Carlo Sampling with PEST Results

The parameter set determined using PEST that satisfies both criteria and produce relatively low sum of squares values is presented in Table 4.2. It is important to note that there is no evidence suggesting that this parameter set represents the global optimum as defined by the L_2 -norm since lower sum of squares values were obtained using the L_2 -norm with the DDS method.

Table 4.2. Result of Parameter Estimation using PEST

Parameter	Units	Estimated Value	95% Lower Confidence Limit	95% Upper Confidence Limit
c_0	mg/L	1000	-6602	8602
v	ft/yr	606	-620	1832
D_x	ft	35	0.049	25590
D_y	ft	39	30*	49*
D_z	ft	2.5	0.0043	1467
μ_t	1/years	0.0004	2.1E-09	72
λ	1/years	0.06	0.0054*	1*
v_v	ft/yr	3.6	0.0024*	5487
D_v	ft	1.1	0.0006	1919
b_t	ft	14.4	-152	181

* Parameter value lies within desired range specified for DDS runs.

The 95% confidence limits shown in Table 4.2 correspond to unrealistic parameter values and fail to lie within the desired parameter ranges. All of the approximately 160 PEST implementations in Chapter 3 produce similar results indicating the need for additional information and alternate methods. In fact, the values are orders of magnitude different from the desired ranges. These results point towards possible violations of underlying assumptions in PEST regarding the data, which is assumed to have normally distributed errors and no biases, and the model, which is assumed to be approximately linear. As discussed in Chapter 3, the error associated with the data is likely to contain biases and to follow a non-Gaussian distribution. In addition, the model developed in Chapter 2 is represented using a highly non-linear analytical solution. Furthermore, the high dimensionality of the problem can also lead to unrealistic results. As a result, the confidence limits cannot be used to indicate lognormal standard deviations of parameters.

Instead, prior knowledge regarding parameter values obtained using expert opinion presented in Chapter 2 is used to estimate lognormal standard deviations. The parameters describing posterior distributions used in Monte Carlo sampling are presented in Table 4.3.

One thousand Monte Carlo samples were generated using the posterior distributions in Table 4.3. A seed of 200 was arbitrarily chosen to be used for all four locations for consistency. If convergence was achieved, the seed value is not important. However, stability was not thoroughly investigated; and 1000 Monte Carlo simulations were assumed to be sufficient for convergence based on previous studies and typically-available computational resources. Nonetheless, the Monte Carlo simulations produce lognormal standard deviations that follow the input lognormal standard deviations very closely, as shown by the outputs included in Appendix #. Also, the majority of the sum of squares

values, which range from 554 to 6725 with a mode of 760, fall in the lower reasonable intervals as shown in Figure 4.3. Nevertheless, further study into convergence and the effect of the seed value is recommended if decisions are to be based on these Monte Carlo samples, and these results are used to provide general insight.

Table 4.3. Posterior parameter distribution to be used in Monte Carlo sampling

Parameter	Units	Mean [§]	Lognormal Standard Deviation [*]
c_0	mg/L	1000	0.1
v	ft/yr	606	0.025
D_x	ft	35	0.2
D_y	ft	39	0.2
D_z	ft	2.5	0.2
μ_t	1/years	0.0004	0.05
λ	1/years	0.06	0.01
v_v	ft/yr	3.6	0.1
D_v	ft	1.1	0.1
b_t	ft	14.4	0.2

^{*}Based on Prior Knowledge

[§]Based on PEST implementations

The breakthrough concentrations from the Monte Carlo samples are represented by their mean and standard deviation in Figure 4.4. The deterministic breakthrough curve is also shown in Figure 4.4. In general, the mean breakthrough curve corresponds to higher concentration values at early times and confirms the importance of a stochastic uncertainty analysis for the problem of estimating first arrival times. The large standard deviation values relative to the mean and deterministic values indicate that there is a large degree of uncertainty associated with the results. However, it is important to note that the mean and standard deviation profiles are dependent on uncertain posterior parameter distributions and conclusions based on these curves are questionable.

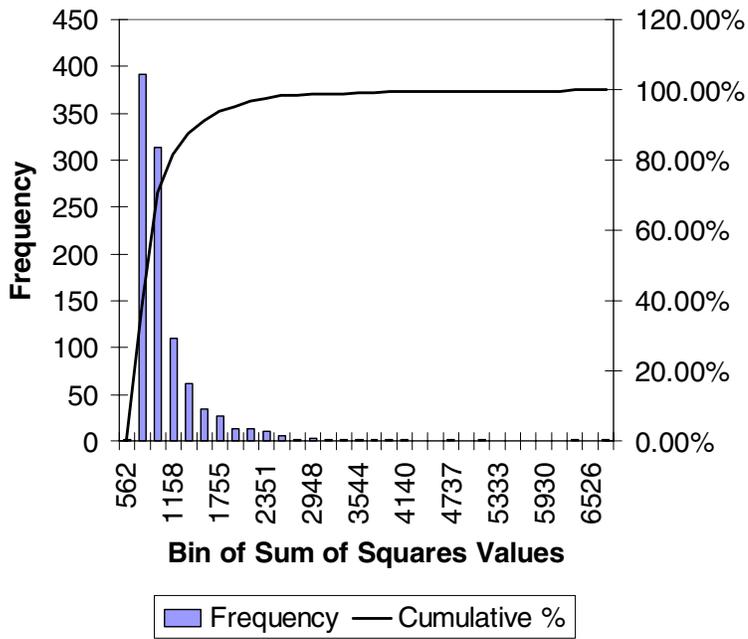


Figure 4.3. Histogram of sum of squares value for Monte Carlo samples generated using posterior distributions defined by PEST results and prior knowledge

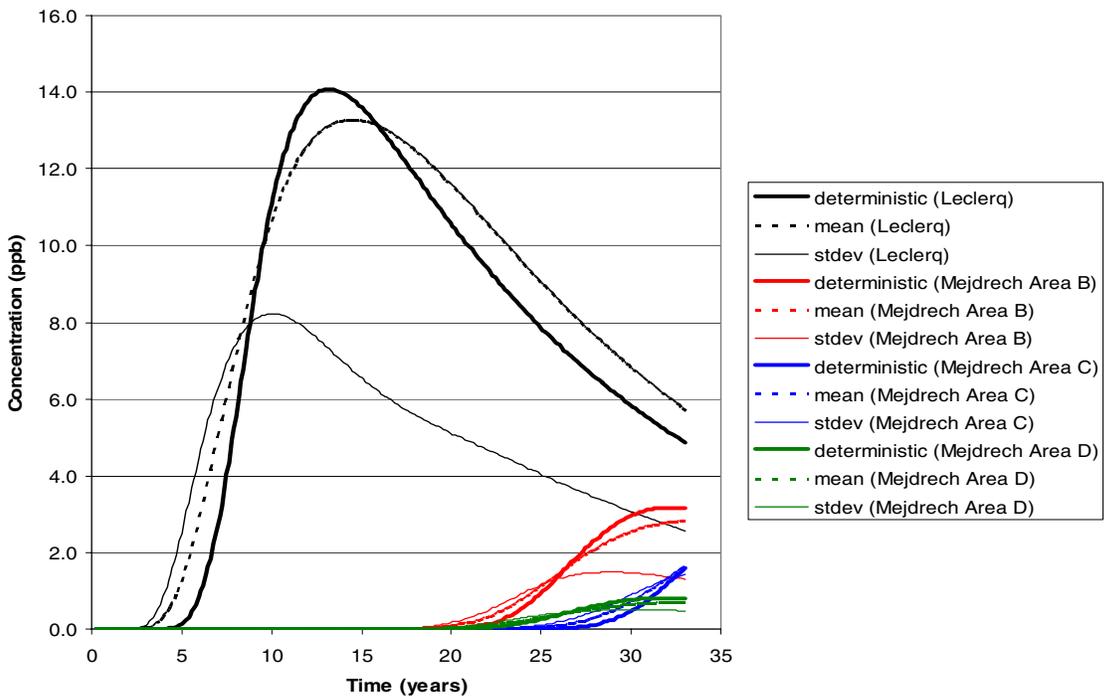


Figure 4.4. Deterministic, mean, and standard deviation of TCE concentrations over time at the four locations of concern using Monte Carlo samples based on PEST results

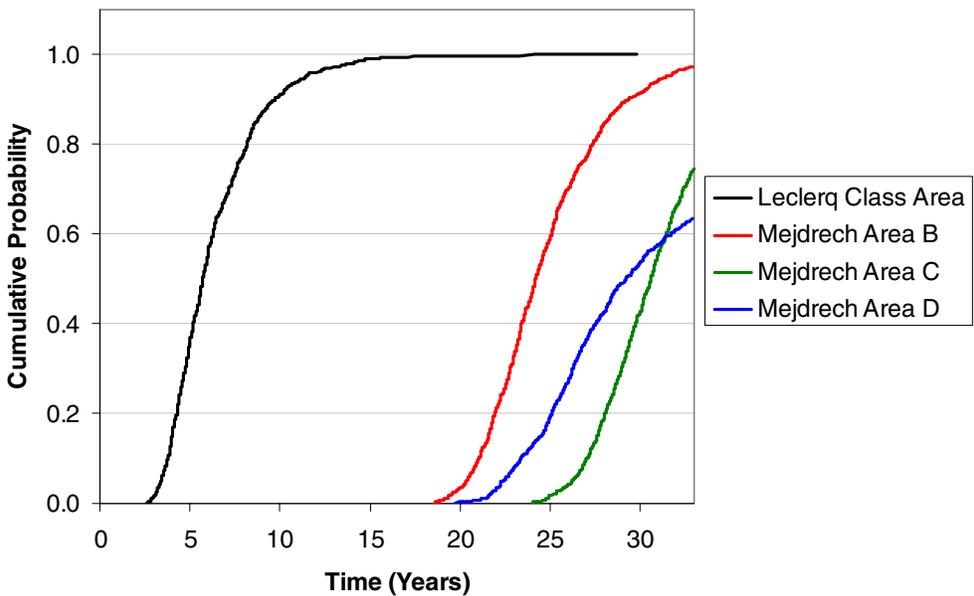


Figure 4.5. Cumulative probability distribution using equally weighted Monte Carlo samples based on PEST results

Cumulative output distributions assuming equally probable parameter sets (i.e. $N = 0$ for $L_{c/d}$, $L_{a/b}$, or L_b) are shown in Figure 4.5. These cumulative distributions represent the probability of the arrival time based on the frequency of parameter sets that generate the arrival time given no conditions, and thus, can be expressed as $P_j(\theta)$. Figure 4.5 shows that the slope of the curve becomes flat for probabilities < 0.05 , and therefore a greater range of possible arrival times exist for low probabilities indicating a larger degree of uncertainty at the tails. Since the portion of the tail region of the probability distribution with a relatively flat slope is shorter for the Leclerq Class area in Figure 4.5, the uncertainty in first arrival times is smallest for Leclerq Class Area in comparison to the uncertainties associated with the remaining three locations.

A cumulative probability of 1.0 is not reached within the 33 years for Mejdrech Areas B, C, and D. The difference between the maximum cumulative probability value at 33 years and 1.0 represents the probability that a concentration of 0.5 ppb will not have reached the given location in 33 years. Therefore, the probability that a concentration of 0.5 ppb is reached within the 33 years is 100%, 97%, 74%, and 63% for the Leclerq Class Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D respectively. The *Location Criteria* represents the intersection of all four events, where an event is a concentration of 0.5 ppb or greater being reached at the given location. If the events are assumed to be independent, the probability of the intersection of the four events is given by

$$\begin{aligned} P(L \cap B \cap C \cap D) &= P(L) \cdot P(B) \cdot P(C) \cdot P(D) \\ &= 1.00 \cdot 0.97 \cdot 0.74 \cdot 0.63 = 0.45 \end{aligned} \quad (39)$$

where L is the event that a concentration of 0.5 ppb is reached within 33 years in the Leclerq Class Area, B is the event that a concentration of 0.5 ppb is reached within 33 years in Mejdrech Area B, C is the event that a concentration of 0.5 ppb is reached within 33 years in Mejdrech Area C, and D is the event that a concentration of 0.5 ppb is reached within 33 years in Mejdrech Area D. However the assumption of independence is invalid since the same physical processes govern all the events. One expression for the probability of the intersection of the four events assuming dependence is given by

$$P(L \cap B \cap C \cap D) = P(L) \cdot P(B|L) \cdot P(C|L \cap B) \cdot P(D|L \cap B \cap C) \quad (40)$$

Based on knowledge of physical processes, it makes intuitive sense that $P(B|L) > P(B)$, $P(C|L \cap B) > P(C)$, and $P(D|L \cap B \cap C) > P(D)$. If these inequalities are true, the probability of the intersection of the four events is greater assuming dependence and the independent case represents the lower limit of the probability of the intersection of the four events.

Although a minimum of 45% of the Monte Carlo parameter sets are expected to satisfy the *Location Criteria*, all of the 1000 parameter sets sampled using the Monte Carlo method given a lognormal distribution failed to satisfy the two physical criteria even though the deterministic breakthrough curve passes both criteria. Additional 700 random parameter sets were generated with the Latin Hypercube sampling method to test if the samples (i.e., parameter sets) were not representative. These parameter sets also failed to satisfy the two criteria. A reduction of the standard deviations appears to be an attractive approach since the deterministic results are able to satisfy the two criteria. However, a 25% reduction in lognormal standard deviations still did not produce acceptable sets. A further reduction in lognormal standard deviations is not justifiable given the purpose of characterising uncertainty. The solution may be to use additional parameter sets or alternate sampling approaches. One of the underlying assumptions is that all of the parameter distributions follow a lognormal distribution, which is not supported by the findings in Chapter 3. Therefore, it may be prudent to either change the parameter distributions to alternate distributions. However, the interaction between parameters is likely to be the most significant factor given the number of uncertain parameters in this case study. As a result, the Monte Carlo sampling approach may not be appropriate. This is further investigated in the next section.

It is important to mention that pseudo-likelihoods with $N \neq 0$ are not considered since the results of this sampling approach may not be applicable for decision-making purposes and the extra effort is not justifiable.

4.2.2 Monte Carlo Sampling with DDS Results

The characteristics of the empirical parameter distributions determined in Chapter 3 appear to be consistent regardless of the objective function definition. If independence between parameters is assumed, the empirical distributions can be used as the posterior parameter distribution in Monte Carlo sampling. Empirical distributions are assumed to resemble the histogram of parameter values presented in Chapter 3. Visual inspection was used to approximately match a lognormal distribution to the empirical distributions based on the parameter sets generated using the L_1 -estimator with a penalty factor of 10,000. An example is shown for the horizontal velocity in the limestone in Figure 4.6, where a good match is made to the modal distribution associate with lower objective function values. Poor matches result for parameters that appear to be uniformly distributed since lognormal distributions are used. The mean and lognormal standard deviations that correspond to the approximated lognormal distribution are presented in Table 4.4 for each parameter.

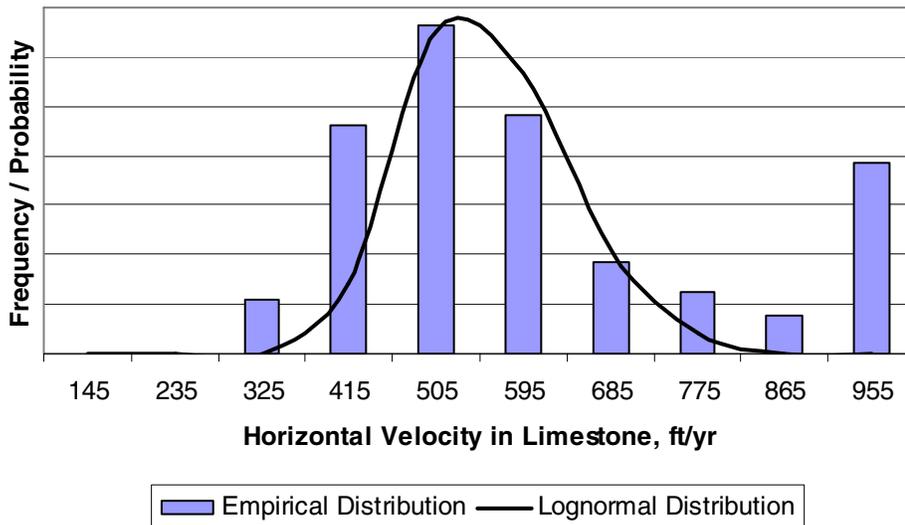


Figure 4.6. Comparison of empirical and approximated lognormal distribution for the horizontal velocity in limestone

Table 4.4. Mean and standard deviation of parameter values generated using the DDS method

Parameter	Units	Mean	Lognormal Standard Deviation
c_0	mg/L	1000	0.20
v	ft/yr	550	0.15
D_x	ft	255	0.23
D_y	ft	30	0.10
D_z	ft	2.5	0.45
μ_t	1/years	0.08	0.45
λ	1/years	0.2	0.25
v_v	ft/yr	18	0.25
D_v	ft	5	0.60
b_t	ft	25	0.50

A comparison of the lognormal standard deviations and means of the parameters corresponding to Monte Carlo simulations to the input lognormal standard deviations and means indicate that the Monte Carlo simulations adequately represents the input parameter uncertainty. The sum of squares value corresponding to the Monte Carlo samples range from 565 to 145,870 with a mode of 1500, which is significantly greater than noted in Figure 4.3. The unrealistically high sum of squares value indicates problems in the posterior parameter distributions or the Monte Carlo sampling strategy.

The resulting breakthrough curves are presented in Figure 4.7 while the corresponding cumulative probability curves are presented in Figure 4.8. The trends shown in Figure 4.7 are similar to those in

Figure 4.4 except that the profiles are shifted to the left indicating earlier arrival times. The shape of the cumulative probability curves is different with steeper slopes than in Figure 4.5 due to the generally larger standard deviations.

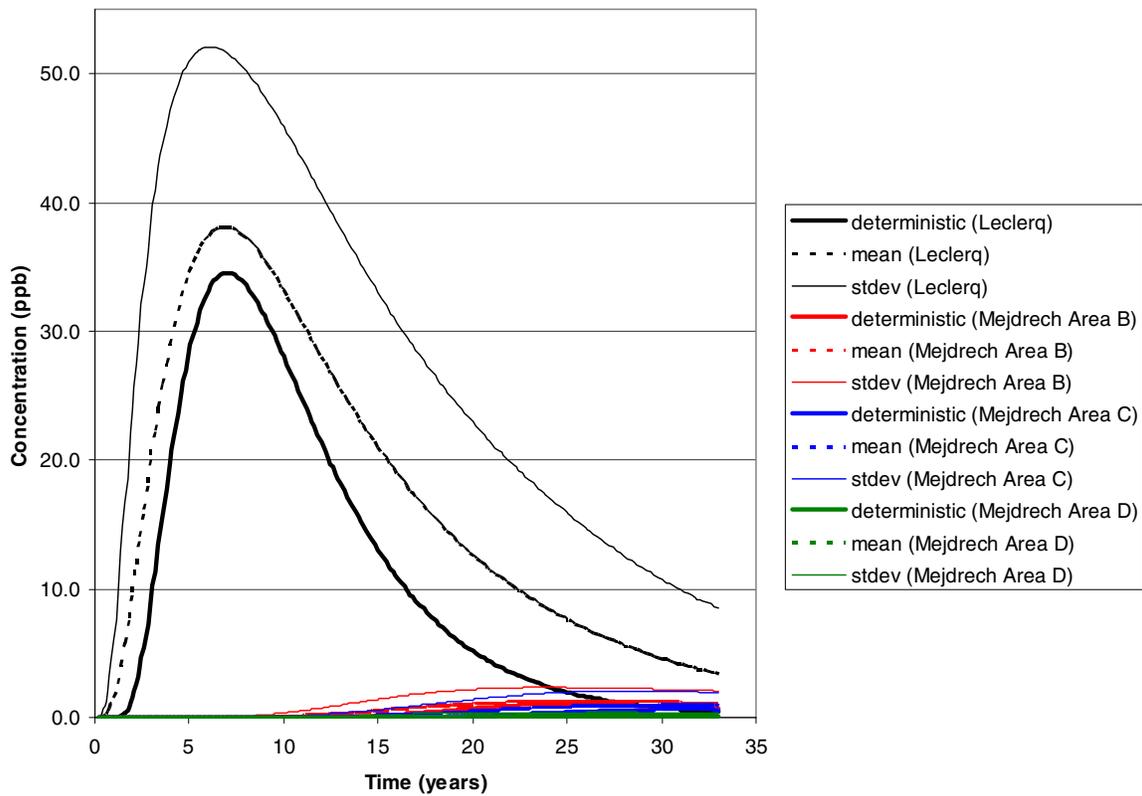


Figure 4.7. Deterministic, mean, and standard deviation of TCE concentrations over time at the four locations of concern using Monte Carlo samples based on DDS results

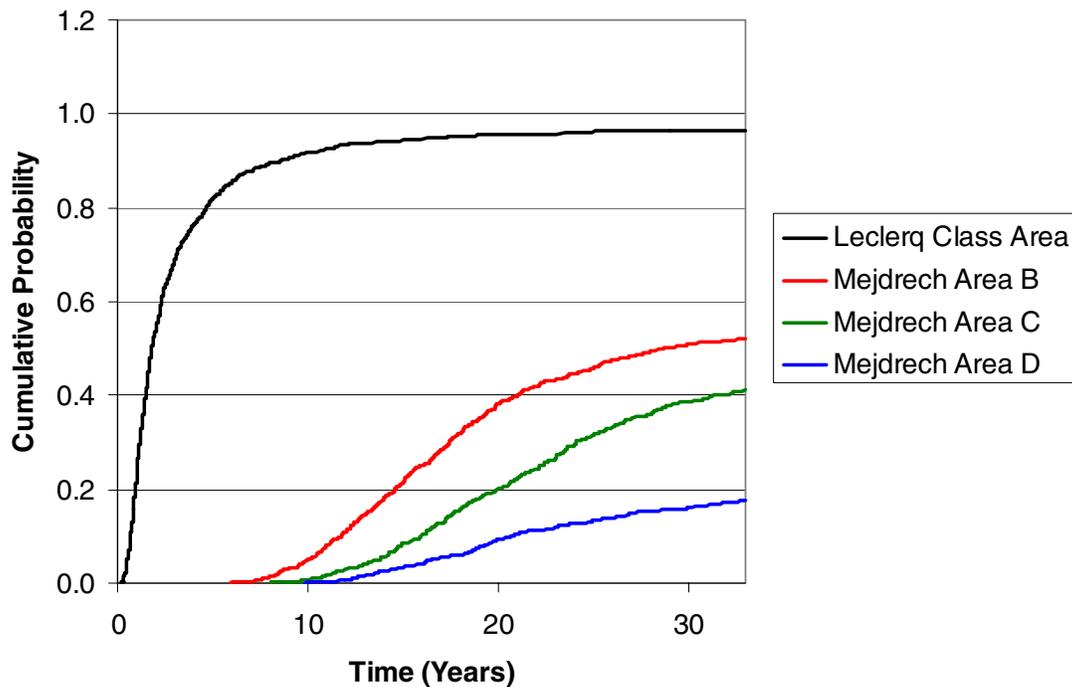


Figure 4.8. Cumulative probability distribution using equally weighted Monte Carlo samples based on DDS results

The probability curves in Figure 4.8 appear to reach a stable value as the curve flattens out at probabilities < 1.0 . This indicates that the probability of arrival of a concentration of 0.5 ppb may never reach 1.0 for some locations or a concentration of 0.5 ppb will never reach the specified location. The probability that a concentration of 0.5 ppb is reached within the 33 years is lower in Figure 4.8 than in Figure 4.5; and the values are 96%, 52%, 41%, and 17% for the Leclerq Area, Mejdrech Area B, Mejdrech Area C, and Mejdrech Area D. The lower limit of probability is 3.5%, which corresponds to the probability of the intersection of the four events is 3.5% if the events are assumed to be independent.

Again, none of these parameter sets satisfies the two criteria, and no parameter sets are accepted. Since the probability of satisfying the *Locations Criteria* is smaller than for Monte Carlo samples obtained using PEST results discussed in Section 4.2.1, an acceptable parameter set is not likely to be found even with a large number (>1000) of Latin Hypercube simulations.

4.2.3 Shortcoming of the Traditional Monte Carlo Analysis

Despite the insight regarding the uncertainty associated with this problem using the traditional Monte Carlo analysis approach, there are a number of shortcomings that invalidates its use in decision-making scenarios. The major shortcoming is the inability of the traditional Monte Carlo sampling strategy to find acceptable parameter sets. The most likely explanations are incorrect assumptions: independence of parameters, and log-normality. The means of the parameter distributions are selected to correspond to parameter sets that satisfy both criteria and thus, are acceptable. It is logical to expect that the reduction in standard deviations can lead to more acceptable parameter sets; however, reduced standard deviations showed no improvements and therefore parameter dependence appears to be the most significant factor.

4.3 Uncertainty Analysis Using Physically-Based Dynamically-Dimensioned Search Approximation of Uncertainty (P-DDS-AU) Sampling

The sampling procedure, P-DDS-AU, described Section 4.1.1.2 was used to generate a sufficient number of acceptable parameter sets to be used in this uncertainty analysis. In addition, parameter dependence is inherently accounted for by the sampling approach which looks at the combined set and not only the individual parameter values. All four definitions of pseudo-likelihood functions (Table 4.1), which are simply referred to as likelihood, are employed; and both cumulative and non-cumulative distributions are examined. For likelihood definitions, L_e and $L_{a/b}$, the variance in observations, σ_o^2 , is found to be 3.4 ppb² by applying (33) to the 320 observations described in Chapter 2.

The number of acceptable parameter sets, n , used to generate each likelihood distribution was made to be equal or greater than 1000 based on performances of the GLUE method in literature (Feyen et al., 2001). Option (a) in Step 9 of P-DDS-AU sampling methodology as outlined in Section 4.1.1.2 is exercised to find the sample sets to ensure convergence while limiting computational effort. To test the effect of objective function definition, more than 1000 acceptable parameter sets were found using the three objective function definitions selected for uncertainty analysis in Chapter 3; the corresponding n values are summarized in Table 4.5.

Table 4.5. Number of acceptable parameter sets used in developing the likelihood function

Objective Function Definition	Number of behavioural parameter sets, n
L ₁ -estimator with a penalty of 10,000	1245
L ₂ -estimator with a penalty of 10,000	1019
LRS1-estimator with a penalty of 10,000	1614

The m value, which is the number of function evaluations per DDS optimization trial, was set at 200 based on preliminary investigations using the optimization algorithm, PEST, as discussed in Chapter 3. The total number of DDS trials, t , and the total number of function evaluations, N , vary according to the value of m .

4.3.1 Factors Affecting Likelihood Distributions

4.3.1.1 Effect of Likelihood Function Definitions on Likelihood Distributions of Arrival Times

The effect of likelihood definition on the likelihood of arrival times is dependent on the objective function definition and the data. Differences in the likelihood distribution with parameter sets generated using the LRS1 M-estimator, L₁-estimator, and the L₂-estimator is noted by comparing Figures 4.9, 4.10, and 4.11. Although a further investigation into the effects of objective function definitions is considered in Section 4.3.1.4, the effect of likelihood function varies with objective function definition and must be considered concurrently. Similar trends are noted for the LRS1 M-estimator and the L₁-estimator with the lowest likelihoods generated using L_e and the highest likelihoods generated using the $L_{a/b}$. Opposite trends are noted for the L₂-estimator but with L_b representing the highest likelihoods. This makes intuitive sense as there is a squared term in the L₂-estimator that is not present in the other two estimators.

Equal weighting of parameter sets by setting N -values to zero creates a profile with the earliest arrival times for any given level of likelihood. This profile represents the upper limit of the likelihood distribution as the other likelihood definitions will approach this profile as the N -value approaches zero. Similar findings are noted for other locations.

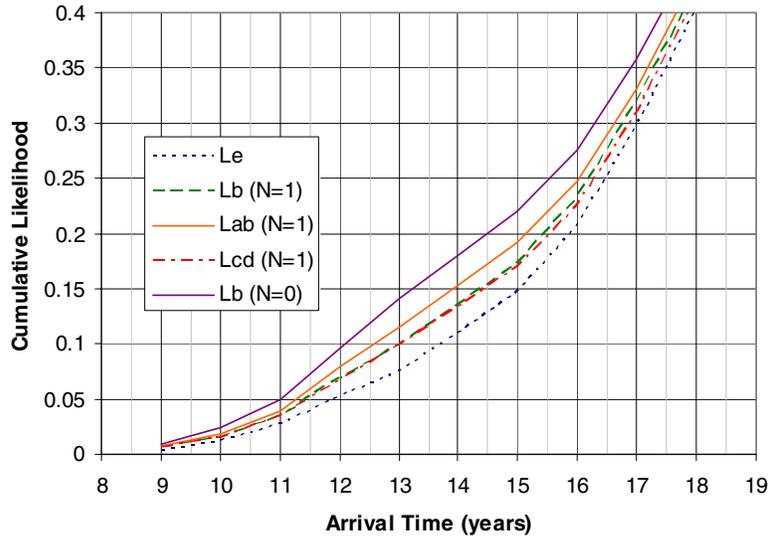


Figure 4.9. The tail-end of the cumulative normalized likelihood distribution plot for Mejdrech Area B using the L_1 -estimator with a penalty factor of 10,000

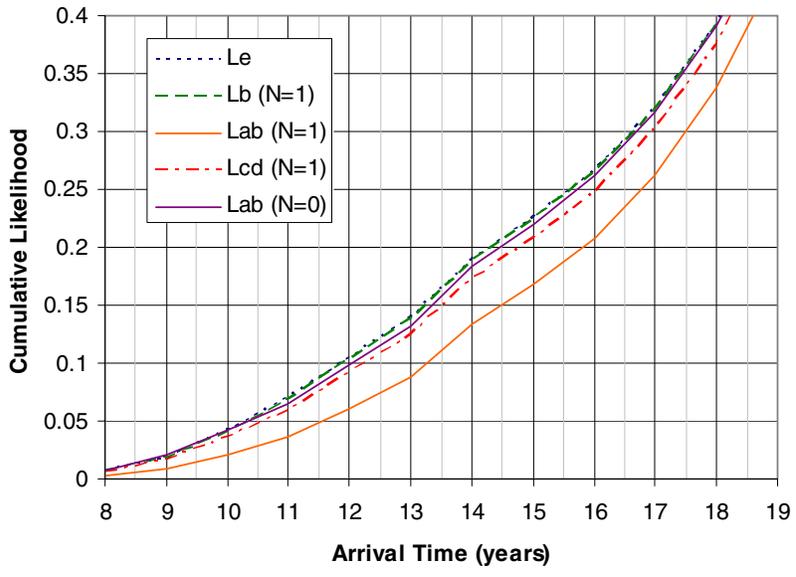


Figure 4.10. The tail-end of the cumulative normalized likelihood distribution plot for Mejdrech Area B using the L_2 -estimator with a penalty factor of 10,000

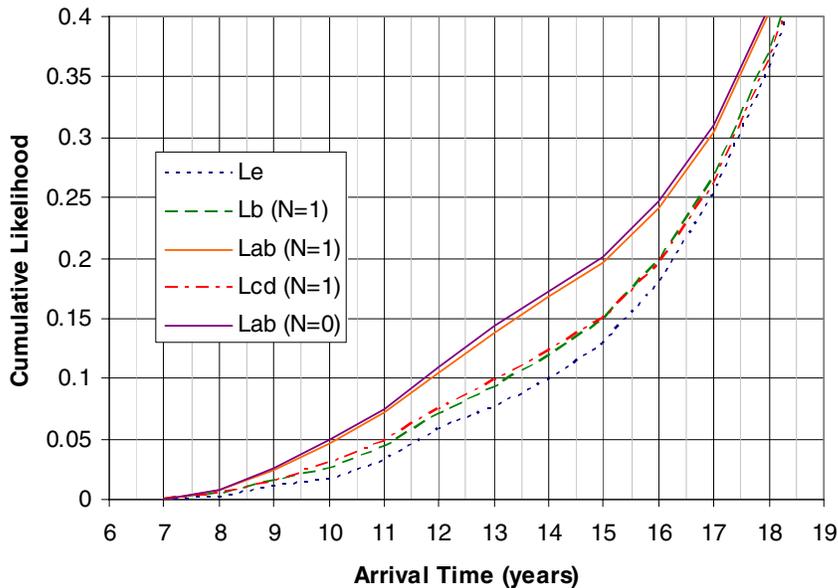


Figure 4.11. The tail-end of the cumulative normalized likelihood distribution plot for Mejdrech Area B using the LRS1-estimator with a penalty factor of 10,000

Figure 4.12 shows the cumulative likelihood distribution envelopes for all the locations for the parameter sets generated using the L_1 -estimator with a penalty of 10,000 and an N -value of 1. It is important to note that these envelopes are only presented to show the effects of likelihood distributions on the final uncertainty envelope to be developed. In Figure 4.12, the difference in cumulative normalized likelihood values becomes negligible as the arrival time and the cumulative likelihood value increases. This indicates that the likelihood function definitions diverge in the tail region, and the likelihood of early arrival times has a greater range of possible values. The greatest range is noted for Mejdrech Area D as evidenced by the larger envelope leading to almost a two year difference in arrival time at the 10% level. In contrast, insensitivity to the likelihood definition is noted for the Leclerq area with a maximum difference of only 0.05 years. A possible cause is the smaller time step leading to fewer errors generated by rounding down but it is difficult to accept that the difference is greater than the time step. Therefore, the main cause of the difference is attributed to the amount of data available at each location. This is supported by the fact that the largest number of data is available for the Leclerq Area while the fewest number of data is available for Mejdrech Area D.

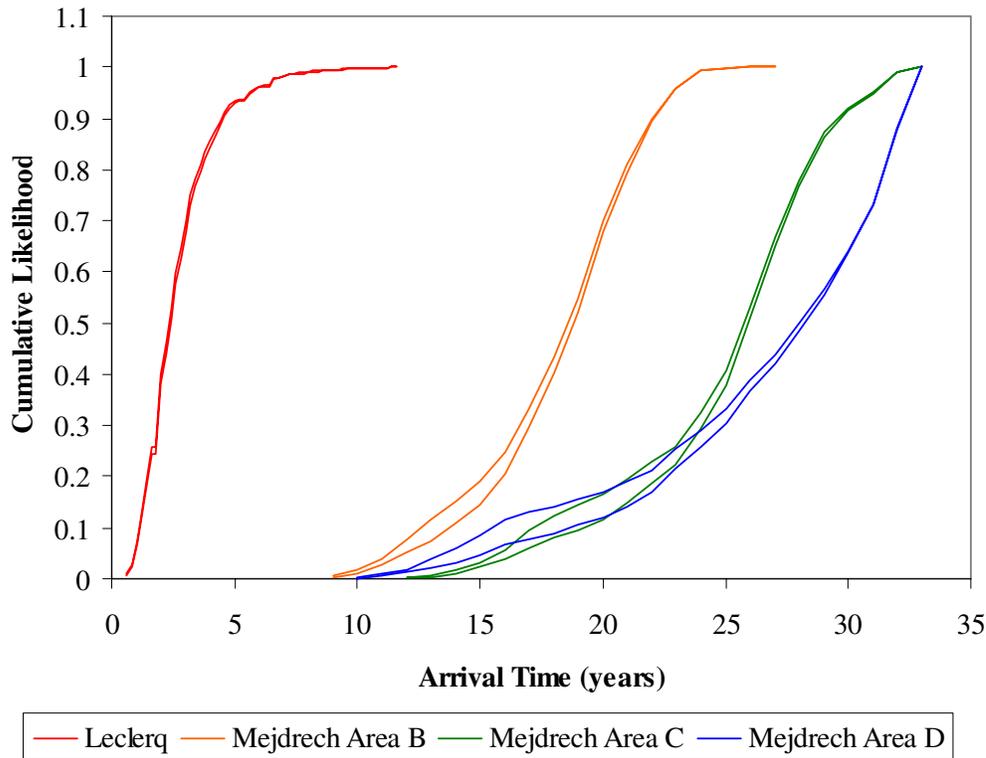


Figure 4.12. Cumulative normalized likelihood for the parameter sets generated using the L_1 -estimator with a penalty factor of 10,000 calculated with L_e and $L_{a/b}$ with $N=1$

It is important to note that the number of parameter sets used in uncertainty analysis is reduced in the definition of L_e and $L_{a/b}$ if the P_k or σ_e^2 values are less than σ_o^2 . Fortunately, Figures 4.9, 4.10, and 4.11 show that effects are small and insignificant for all objective function definitions.

In summary, a single likelihood function definition may lead to non-conservative output distributions and more than one definition should be considered. Analysis on the effect of the likelihood definition should be evaluated on a case-by-case basis. If decision-making only requires the first arrival time independent of the likelihood value, any likelihood function definition can be used since the likelihood functions only change the relationship between arrival times.

4.3.1.2 Effect of N -value in Likelihood Function on Likelihood Distributions of Arrival Times

The likelihood function definitions, L_b , $L_{a/b}$, and $L_{c/d}$, have a subjective likelihood definition parameter, N , for which values of 1, 2, and $n_{obs}/2$ have been suggested (Beven and Binley, 1992; Feyen et al., 2001). Therefore, N values of 0, 1, 5, 20, and 160 were examined. The condition where

N equals zero represents the case where each parameter set is given equal weighting. In contrast, only a single solution is considered likely when N approaches infinity.

The effect of likelihood function definitions discovered in Section 4.3.1.1 are amplified by increases in the N -value. Figure 4.13 shows the reduction in the slope of the cumulative likelihood curve and the corresponding likelihood value as N is increased. This trend is characteristic of all locations and likelihood definitions using the observation data with non-detect values set to zero.

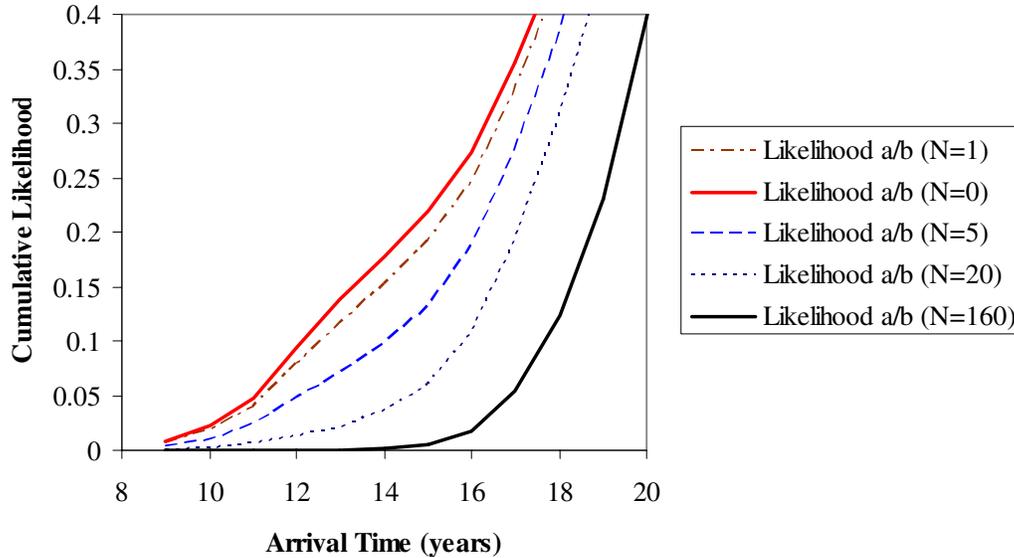


Figure 4.13. Cumulative normalized likelihood plot for Mejdrech Area B using $L_{a/b}$ for various values of N using parameter sets generated using the L_1 -estimator with a penalty factor of 10,000

Examination of Figure 4.13 and the desire to be conservative may lead to the selection of equal weights for parameter sets (i.e. N -value of zero). However, the normalized likelihood distributions created using an N -value of zero leads to an early peak when using the observation data with NDs set to zero as shown in Figure 4.14. The disappearance of this early peak indicates that the corresponding arrival times are less conditional to observation data. The nature of the distribution appears to follow a lognormal distribution in most cases. However, the lack of a right tail in most of the distributions makes it difficult to confirm this. The disappearance of the early peak when $N=10$ is noted starting from an N -value of 5 for L_b and $L_{c/d}$ and an N -value of 10 for $L_{a/b}$. Similar trends in the non-cumulative distribution are also present at other locations. However, preliminary studies with N -values using different data sets indicate that these N -values are case-specific. Nonetheless, similar

trends, where an N -value of zero produces the most conservative results, are observed for the L_2 -estimator, the L_1 -estimator, and the LRS1 M-estimator.

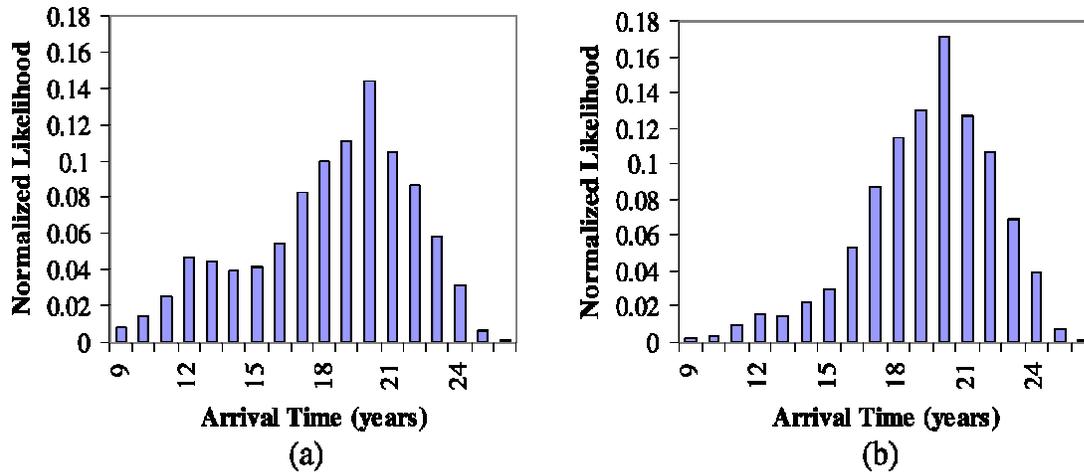


Figure 4.14. Normalized likelihood distribution for Mejdrech Area B using $L_{a/b}$ using (a) $N = 0$ and (b) $N = 10$

Due to its subjectivity, the usage of N -values greater than zero is difficult to justify unless a distribution created using an N -value greater than zero matches the distribution based on prior knowledge. The safest approach is to use an N -value of zero and address the problem of multiple modes in the normalized likelihood distribution using additional physical criteria. However, it is important to note that an N -value of zero may not always produce the most conservative results for a different problem and data set.

4.3.1.3 Effect of the Number of Acceptable Parameter Sets on Likelihood Distribution of Arrival Times

To test the effect of n on the likelihood distribution, the number of parameter sets for each of the three acceptable parameter sets generated using different objective function definitions were randomly reduced to 1000. It is important to note that the inequalities in the likelihood definitions, L_e and $L_{a/b}$, can cause the effect of some additional parameter sets to be omitted.

The profiles in the region of low likelihoods appear to be slightly more dependent on the number of behavioural parameter sets used in the analysis as shown in Figure 4.15. However, an increase in the number of parameter sets from 1000 to 1614 has negligible effect relative to differences observed using various definitions for objective functions. Therefore, convergence can be assumed for n values, which represent the desired number of behavioural samples, greater than 1000. The number

of acceptable parameter sets to be used in the uncertainty analysis should be determined by the convergence of parameter characteristics. It is important to note that this implies the effect of the random seed value does not need to be considered further for this case study.

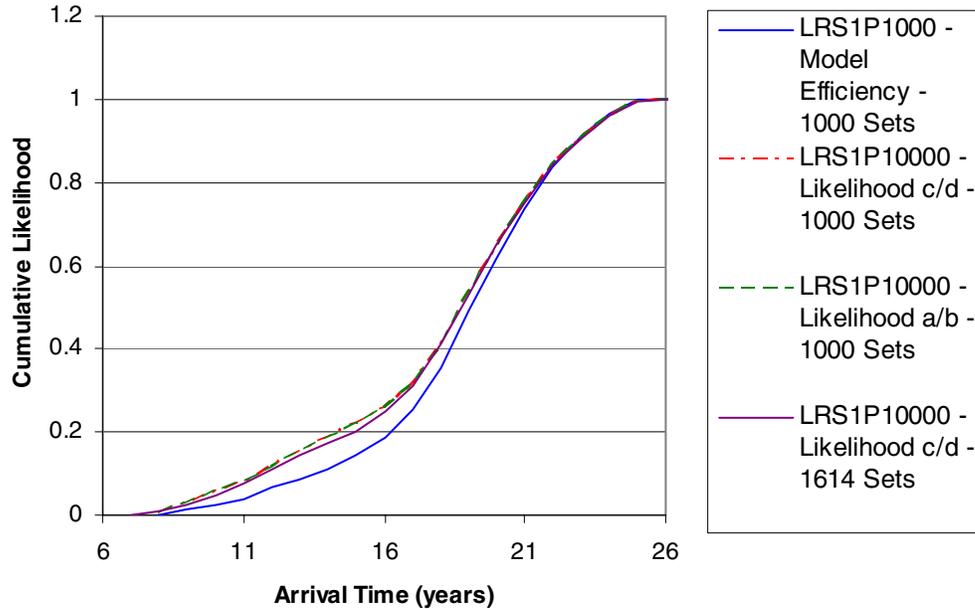


Figure 4.15. Cumulative normalized likelihood plot at Mejdrech Area B using various likelihood definitions with an N -value of 0, where applicable, and 1000 and 1614 parameter sets generated using the LRS1 M-estimator with a penalty factor of 10,000

4.3.1.4 Effect of the Objective Function Definition on Likelihood Distributions of Arrival Times

The general effect of objective function definition was noted in Section 4.3.1.1, and a more rigorous examination of its effect discussed here. 1000 parameter sets obtained using each of the three objective functions are used for consistency even though convergence is assumed to have occurred. The comparison is not biased to any particular objective function definition since all of the objective function definitions used have the same scaling factor in the penalty function.

Three objective functions defined by the L_1 -estimator, L_2 -estimator, and the LRS1 M-estimator are compared in Figure 4.16. The likelihood curves generated using “optimized” parameter sets all follows an S-shape, with a large range of arrival times associated to lower cumulative likelihood

functions. The relationship between the curves in Figure 4.16 is different for high and low likelihoods.

The relationship between the curves generated with different objective function definition becomes magnified with increases in the N -value as noted in a previous section. However, it is clear that this effect is overshadowed by the selection of N -value.

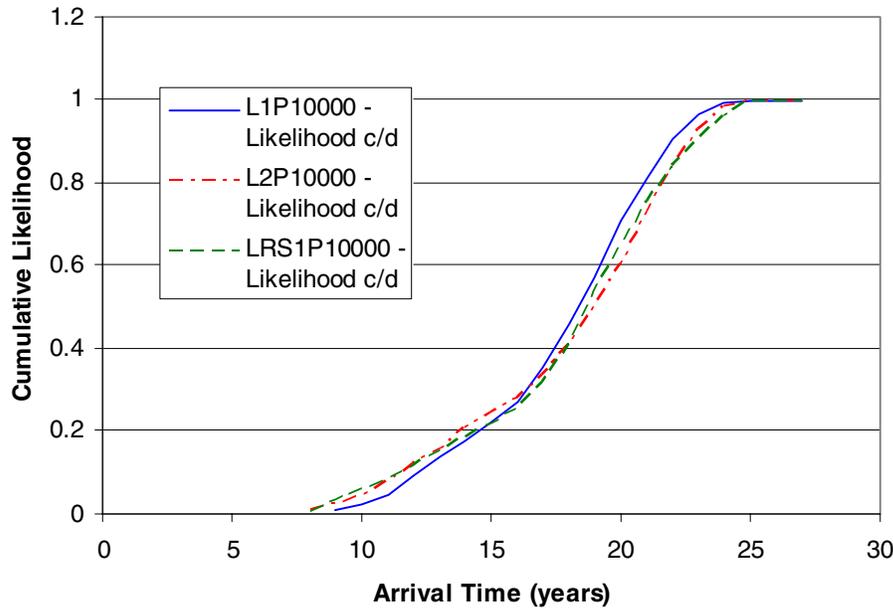


Figure 4.16. Cumulative normalized likelihood plot at Mejdrech Area B using $L_{c/d}$ with an N -value of 0 and 1000 parameter sets generated using various objective function definitions with a penalty factor of 10,000

4.3.2 Steps to a Conservative Uncertainty Envelope

It is practically infeasible to examine each of the likelihood distributions and it is difficult to justify a selection of one distribution over another. Therefore, uncertainty envelopes that represent the bounds of cumulative likelihood distributions are created in order to maximize information available to decision-makers while minimizing assumptions. The effects noted in Sections 4.3.1.1 to 4.3.1.4 are compiled to develop a generalized approach to arriving at a conservative cumulative normalized likelihood distribution. It is important to note that the order of the steps must be followed to ensure that the final cumulative normalized likelihood distribution is indeed conservative. The four steps in their respective order are:

1. Ensure that a sufficient number of acceptable parameter sets are used.
2. Select the likelihood function definition corresponding to the most conservative cumulative likelihood curve.
3. Select the N -value corresponding to the most conservative cumulative likelihood curve.
4. Select the objective function definition corresponding to the most conservative cumulative likelihood curve.

More than one definition or value may be found for Steps 2 to 4 since the relationship can change with cumulative likelihood values. If only one definition or value is selected, the definition or value that produces the most conservative results for low cumulative likelihood values should be chosen. In such case, the plot should only show the portion corresponding to low cumulative likelihood values.

The purpose of this study is to provide the best indication of uncertainty with the available resources. The selection of a single representative curve for each location is not representative of the uncertainty investigation performed. The sufficiency in the number of parameters is questionable and the likelihood function definition is somewhat subjective. Therefore, an uncertainty envelope taking into consideration all of the cumulative likelihood curves selected is created for each location of concern as shown in Figure 4.17. In general, the uncertainty at the tails is greater as indicated by the larger widths. For a refinement of the envelopes, the application of additional physical criteria is recommended.

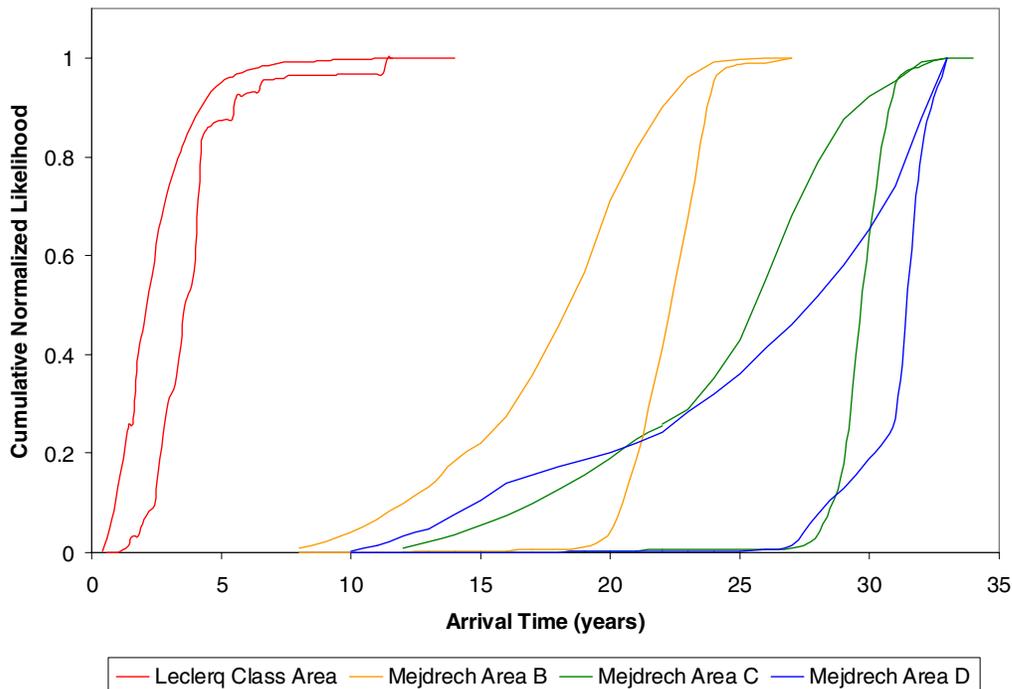


Figure 4.17. Uncertainty envelopes for the four locations of concern

4.4 Summary

Monte Carlo samples are inadequate for uncertainty analysis due to the difficulties associated with defining the input uncertainty and finding acceptable parameter sets. P-DDS-AU sampling strategy produces a sufficient number of acceptable parameter sets to conduct a meaningful uncertainty analysis.

The most conservative results are achieved by assigning equal likelihoods to parameter sets in this case study. However, the approach to arriving at conservative results varies with problems since the choice of likelihood function and objective function definitions and their properties are problem-dependent.

All four effects must be considered with respect to its order in the creation of uncertainty envelopes. In general, there appears to be greater uncertainty at the tails of the distribution. For a refinement of the envelopes, the application of additional physical criteria is recommended.

Chapter 5

Conclusions and Recommendations

The principle purpose of uncertainty analysis is to facilitate decision-making. Therefore, the confidence on the conclusions drawn from uncertainty analysis must be acceptable. To achieve this, “good scientific practices of clarity, consistency, transparency, reproducibility, and the use of sound methods” (USEPA, 1997) along with justifiable assumptions must be considered in the development of the approach to characterizing uncertainty. These considerations can be incorporated into many aspects of uncertainty analysis to maximize information available to decision-makers. To demonstrate this, the uncertainty of arrival times for a real contaminant transport problem which involves TCE contamination due to releases from the Lockformer Company Facility in Lisle, Illinois is used.

In this chapter, the conclusions and recommendations are presented with respect to the two major components of the approach to characterizing uncertainty: parameter estimation and uncertainty analysis. A comprehensive uncertainty analysis is not possible without an understanding of the parameter estimation process and the corresponding parameter spaces. In fact, the significant insight gained during the parameter estimation process can lead to the selection of appropriate uncertainty analysis methods and provide explanation for the failure of other methods.

5.1 On Parameter Estimation

The parameter estimation process for the case study investigated here was selected based on insufficiencies in the model and observational data due to errors, biases, and limitations. A consideration of its purpose, which is to aid in characterising uncertainty, was also made in the process by including many possible variations in attempts to minimize assumptions. Preliminary investigation of the problem was conducted using a well-accepted parameter estimation method, PEST, and the corresponding findings were used to define characteristics of the parameter estimation process. Numerous objective function definitions, which include the well-known L_2 -estimator, robust estimators (L_1 -estimators and M-estimators), penalty functions, and deadzones, were incorporated in the parameter estimation process to treat specific insufficiencies. The concept of equifinality was adopted and multiple maximum likelihood estimates were accepted if pre-defined physical criteria are met. For each objective function definition, three procedures were implemented as a part of the

parameter estimation method for the given case study: a multistart procedure, a heuristic local greedy search using the Dynamically-Dimensioned Search (DDS), and a test for acceptance based on predefined physical criteria. The two physical criteria applied in this thesis are: the *Location Criteria* and the *Product Criteria*.

The best performance in terms of the ability of parameter sets to satisfy the physical criteria, which is quantified by the acceptance rate, was achieved using a Cauchy's M-estimator that was modified for this study and designated as the LRS1 M-estimator. The low weight assigned to residuals of high values by the LRS1 M-estimator indicates that its success is based on treating outliers. Although the true nature of errors can be inferred from an estimator's acceptance rate, the level of uncertainty regarding the true nature is still significant; and therefore, multiple parameter sets obtained with the LRS1 M-estimator, the L_1 -estimator, and the L_2 -estimator are recommended for use in uncertainty analysis. Other M-estimators provide a middle ground between the L-estimators and the LRS1 M-estimator. Therefore, the inclusion of the two L-estimators is made to minimize assumptions regarding the error distribution.

For all estimators, penalty function based on the *Location Criteria* had to be incorporated into the objective function definitions to generate a sufficient number of acceptable parameter sets. The limiting functionality of the *Location Criteria* is a result of the difficulty of data insufficiencies that are not fully addressed by treatments such as robust estimators. A penalty function based on the *Product Criteria* was clearly not a limiting factor in generating acceptable parameter sets; therefore, it was not applied in the search procedure when finding multiple acceptable parameter sets to be used in uncertainty analysis. Similarly, deadzones proved to produce negligible benefits and are not used to generate multiple acceptable parameter sets to be used in uncertainty analysis.

The characteristics for the multiple parameter sets to be used in uncertainty analysis were examined in terms of frequency histograms and plots of parameter value versus objective function value to infer the nature of the likelihood distributions of parameters. The correlation structure was estimated using Pearson's product-moment correlation coefficient. The parameters are generally distributed uniformly or appear to follow a random nature with few correlations in the parameter space that results after the implementation of the multistart procedure. The execution of the search procedure results in parameter distributions that appear to follow lognormal, normal, or uniform distributions and in the introduction of many correlations. The application of the physical criteria refines the parameter characteristics in the parameter space resulting from the search procedure by reducing anomalies. The combined effect of optimization and the application of the physical criteria performs

the function of behavioural thresholds by removing parameter sets with high objective function values.

In general, parameter estimation was performed in an inclusive manner that minimizes assumptions. This is not to say that the parameter estimation process used in this thesis can be approached as a black-box. In fact, preliminary investigations should always precede parameter estimation and should be considered a part of the parameter estimation process. However, given the uncertainties surrounding the analysis of real problems, preliminary investigations are not likely to justify the abandonment of possible solutions and solution methods. In fact, ambiguities in inferring probability distributions based on parameter characteristics remain. With the advent of computing power, it is increasingly difficult to justify a decision not to explore additional options and perform a more thorough search through the solution space if resources are available.

One possible option to further refine parameter characteristics is the application of additional physical criteria, which can be defined using prior knowledge and characteristics observed in the multiple parameter sets. To minimize effort and maximize benefits, additional physical criteria should be designed to target specific anomalies in the parameter characteristics.

Although a need to use other objective function definitions are not clearly evident through error analysis, indirect benefits may be achieved in terms of refinements of parameter characteristics. Alternative families of robust estimators include the R-estimators (Rousseeuw and Leroy, 2003), which are based on the ranks, and the S-estimators (Rousseeuw and Leroy, 2003), which are generalizations of the least median of squares estimators. These estimators may be a good approach to reduce the effect of some problems in the data set and should be investigated in future studies, especially if M- and L-estimators are known to be ineffective as treatments to contaminated data.

It is important to note that the objective function definitions investigated in this chapter are designed to find the large-scale trend of the data or one of the “preferred states”. Small-scale fluctuations as defined by McLaughlin and Townley (1996), which cannot be ignored in our quest of defining uncertainty, are generally assumed to be independent-and-identically distributed random variables, and methods to define its nature are not typically implemented. In fact, for this case study, a definition of the nature of small-scale fluctuations is difficult given limitations in the data. Therefore, a strategy to estimate the significance and magnitude of small-scale fluctuations should be researched.

5.2 On Uncertainty Analysis

The steps to uncertainty analysis of arrival times are separated into two separate tasks in this thesis: sampling of parameter sets and quantitative representation of likelihood. Two different sampling methodologies are applied: the Monte Carlo sampling methodology, which randomly and independently samples from user-defined distributions, and the physically-based DDS-UA (P-DDS-UA) sampling methodology, which is developed based on the multiple parameter sets acquired during the parameter estimation process. The P-DDS-UA sampling methodology was designed to account for parameter correlations and does not require assumptions regarding parameter distributions. These features are advantageous for the given case study where a significant number of parameter correlations have been observed and where parameter distributions cannot be suitably represented by parametric distributions. For both sampling methodologies, the quantitative representation of likelihood is facilitated by likelihood distributions in both the cumulative and non-cumulative sense by assigning likelihood values to each sample or parameter sets. Four pseudo-likelihood function definitions are used to represent likelihoods: two based on the Nash and Sutcliffe efficiency criterion, L_e and L_{ab} , one based on inverse residual variance, L_b , and one based on inverse error variance, L_{cd} . All of these definitions are controlled by a shaping factor, while two (L_{ab} and L_{cd}) are influenced by objective function definitions. All variations are considered in the development of cumulative likelihood distribution envelopes, which are designed to maximize the amount of information available to decision-makers. These envelopes are referred to as uncertainty envelopes.

For uncertainty analysis of this case study, Monte Carlo samples are found to be inadequate due to its inability to find parameter sets that meet the predefined physical criteria. Successful results are achieved using the P-DDS-UA sampling methodology, and its results are used to generate uncertainty envelopes. Shaping factors have a strong effect on the likelihood distribution that overshadows the effects of both likelihood and objective function definitions. The effect of likelihood function and objective function definitions on the uncertainty envelopes is problem-dependent. Therefore, no general assumptions can be made regarding the abilities of likelihood function definitions and objective function definitions to produce conservative results. In fact, preliminary investigations that consider the characteristics of the acceptable parameter space, the objective function definition, and the likelihood function definition are important to the creation of uncertainty envelopes. Nonetheless, the relative importance of each factor appears to be consistent and conservative uncertainty envelopes can be obtained if the recommended order in which each factor is considered is followed. In general, greater uncertainty appears to be present at the tails of the distribution for all locations. This is

evident by the greater slope in the cumulative likelihood distributions and the greater width of the uncertainty envelope at low likelihood values.

Since decision-making is often governed by the information at the tails of likelihood distributions, a refinement of uncertainty envelopes is desirable. Although refinements are feasible by changing the shaping factor based on its non-cumulative distribution, this approach is subjective, and the application of additional physical criteria is recommended.

Decision-makers must be aware of the assumptions upon which the entire analysis is based on. Although the entire process is designed to be conservative and inclusive, certain decisions must be made and their effects should be considered. For example, the objective function and likelihood function definitions used to generate the uncertainty envelopes are not exhaustive. In addition, the layers of assumptions to ensure conservative results may have led to an overly conservative estimate. However, it is difficult to justify the removal of any measure taken to be conservative, and any refinement of the envelopes should be performed with the application of additional physical criteria. Also, the accuracy of the likelihood function obtained by any method is still dependent on how thoroughly the solution space is searched and the convergence issue remains. Even though “good looking bell shaped histograms” are observed, the results are spurious if the “existence of the limiting distribution” is questionable (Kuczera and Parent, 1998). Nonetheless, it is practically infeasible to consider every single variation and apply all possible methods given the plethora of possible likelihood and objective function definitions. The selection of likelihood and objective function definitions and their properties are made based on the needs of the problem; therefore, preliminary investigations should always be conducted to provide a basis for selecting appropriate methods and definitions.

It is important to note the importance of parameter estimation in uncertainty analysis. In fact, uncertainty analysis is not justifiable without a valid understanding of the parameter solution space.

Finally, it is imperative to remember that the communication of assumptions and definitions used in both parameter estimation and uncertainty analysis is crucial in decision-making scenarios.

Appendix A

Analytical Solution to the One-Dimensional Advection-Dispersion Equation

The analytic solution of dissolved TCE transport through the overburden is obtained using the C13 Solution by van Genuchten and Alves (1982). The resulting solution for migration of a dissolved species through the till is

$$c(z,t) = \begin{cases} c_0 e^{-\lambda t} A(z,t) & 0 \leq t \leq t_r \\ c_0 e^{-\lambda t} [A(z,t) - A(z,t-t_r)] & t_r < t \leq \infty \end{cases}$$

with

$$A(z,t) = \left\{ \frac{1}{2} \exp\left[\frac{(v_v - w)z}{2D_v}\right] \operatorname{erfc}\left[\frac{R_t z - wt}{2(D_v R_t t)^{1/2}}\right] + \frac{1}{2} \exp\left[\frac{(v_v + w)z}{2D_v}\right] \operatorname{erfc}\left[\frac{R_t z + wt}{2(D_v R_t t)^{1/2}}\right] \right\}$$

where

$$w = v_v \left[1 + \frac{4D_v}{v_v^2} (u_t - \lambda R_t) \right]^{1/2}$$

in which c_0 is the initial dissolved TCE concentration in the source zone [M/L³]; λ accounts for the reduction of the dissolved TCE source concentration [T⁻¹]; t represents time since the first release of contaminant [T]; t_r represent the time at which the source was removed [T]; v_v is the linear velocity in the overburden [L/T]; D_v is the longitudinal dispersion coefficient [L²/T]; and R_t is the retardation factor for the dissolved TCE in the overburden [-].

Appendix B

Green's Function Used in the Solution for the Three-Dimensional Advection-Dispersion Equation

The Green's function, $h(x, x', y, y', z, z', t, t')$, used in the solution for the three-dimensional advection-dispersion equation is

$$h(x, x', y, y', z, z', t, t') = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x - [x' + v_x(t-t')/R_r])^2}{2\sigma_x^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(y-y')^2}{2\sigma_y^2}} \cdot \left\{ \sum_{n=-1}^1 \left[\frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{(z-z'+2nz_0)^2}{2\sigma_z^2}} + \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{(z+z'+2nz_0)^2}{2\sigma_z^2}} \right] \right\} e^{-\mu_r(t-t')}$$

Appendix C

Algorithm Parameters for Pest

The following is the PEST control file. For a description of the algorithm parameters, refer to the PEST manual (Watermark Numerical Computing, 2004).

```
pcf
* control data
restart estimation
10 320 1 0 1
1 1 single point 1 0 0
5.0 2.0 0.3 0.01 10
5.0 5.0 1.0e-3
.1
30 0.005 4 3 .01 3
1 1 1
* parameter groups
pargrp1 relative 0.01 0.00001 switch 2.0 parabolic
* parameter data
c0 none factor 1420. 100. 2000. pargrp1 1.0 0.0 1
v none factor 599. 100. 1000. pargrp1 1.0 0.0 1
dispx none factor 250. 10. 500. pargrp1 1.0 0.0 1
dispy none factor 35. 1. 50. pargrp1 1.0 0.0 1
dispz none factor 1.1 0.05 5. pargrp1 1.0 0.0 1
rlam none factor .02 0.00001 1. pargrp1 1.0 0.0 1
gam none factor .0007 0.00001 1. pargrp1 1.0 0.0 1
vv none factor 12. 1.0001 40. pargrp1 1.0 0.0 1
dispv none factor 7.9 0.001 20. pargrp1 1.0 0.0 1
xv none factor 29. 0. 80. pargrp1 1.0 0.0 1
* observation groups
obsgrp1
* observation data
conc1 0.0000 1.0 obsgrp1
conc2 0.0000 1.0 obsgrp1
conc3 0.0000 1.0 obsgrp1
conc4 7.9900 1.0 obsgrp1
conc5 0.0000 1.0 obsgrp1
conc6 4.2800 1.0 obsgrp1
conc7 0.0000 1.0 obsgrp1
conc8 0.0000 1.0 obsgrp1
conc9 0.0000 1.0 obsgrp1
conc10 0.2400 1.0 obsgrp1
conc11 0.4400 1.0 obsgrp1
conc12 0.0000 1.0 obsgrp1
conc13 0.0000 1.0 obsgrp1
conc14 0.0000 1.0 obsgrp1
conc15 0.0000 1.0 obsgrp1
conc16 0.0000 1.0 obsgrp1
conc17 0.5700 1.0 obsgrp1
conc18 0.0000 1.0 obsgrp1
conc19 0.0000 1.0 obsgrp1
conc20 7.3600 1.0 obsgrp1
conc21 3.1300 1.0 obsgrp1
conc22 0.2400 1.0 obsgrp1
conc23 0.0000 1.0 obsgrp1
conc24 0.0000 1.0 obsgrp1
conc25 0.0000 1.0 obsgrp1
conc26 0.0000 1.0 obsgrp1
conc27 0.0000 1.0 obsgrp1
conc28 0.0000 1.0 obsgrp1
conc29 0.0000 1.0 obsgrp1
conc30 0.0000 1.0 obsgrp1
conc31 19.5000 1.0 obsgrp1
conc32 0.0000 1.0 obsgrp1
conc33 0.0000 1.0 obsgrp1
conc34 0.0000 1.0 obsgrp1
conc35 0.0000 1.0 obsgrp1
conc36 2.3000 1.0 obsgrp1
conc37 0.0000 1.0 obsgrp1
```

conc38	0.5700	1.0	obsgrp1
conc39	0.0000	1.0	obsgrp1
conc40	2.9900	1.0	obsgrp1
conc41	0.0000	1.0	obsgrp1
conc42	5.5900	1.0	obsgrp1
conc43	0.2800	1.0	obsgrp1
conc44	5.2200	1.0	obsgrp1
conc45	0.0000	1.0	obsgrp1
conc46	0.0000	1.0	obsgrp1
conc47	0.4000	1.0	obsgrp1
conc48	0.0000	1.0	obsgrp1
conc49	0.0000	1.0	obsgrp1
conc50	0.0000	1.0	obsgrp1
conc51	0.0000	1.0	obsgrp1
conc52	0.0000	1.0	obsgrp1
conc53	0.0000	1.0	obsgrp1
conc54	1.9300	1.0	obsgrp1
conc55	0.0000	1.0	obsgrp1
conc56	0.0000	1.0	obsgrp1
conc57	0.0000	1.0	obsgrp1
conc58	0.0000	1.0	obsgrp1
conc59	0.0000	1.0	obsgrp1
conc60	1.6200	1.0	obsgrp1
conc61	1.6900	1.0	obsgrp1
conc62	1.9500	1.0	obsgrp1
conc63	0.0000	1.0	obsgrp1
conc64	6.1400	1.0	obsgrp1
conc65	8.0200	1.0	obsgrp1
conc66	8.3000	1.0	obsgrp1
conc67	0.0000	1.0	obsgrp1
conc68	0.0000	1.0	obsgrp1
conc69	0.0000	1.0	obsgrp1
conc70	0.0000	1.0	obsgrp1
conc71	4.1000	1.0	obsgrp1
conc72	2.9100	1.0	obsgrp1
conc73	3.0500	1.0	obsgrp1
conc74	0.0000	1.0	obsgrp1
conc75	0.0000	1.0	obsgrp1
conc76	0.0000	1.0	obsgrp1
conc77	0.0000	1.0	obsgrp1
conc78	0.5010	1.0	obsgrp1
conc79	1.4300	1.0	obsgrp1
conc80	3.2300	1.0	obsgrp1
conc81	3.0900	1.0	obsgrp1
conc82	6.3300	1.0	obsgrp1
conc83	6.8900	1.0	obsgrp1
conc84	5.5200	1.0	obsgrp1
conc85	0.0000	1.0	obsgrp1
conc86	1.2300	1.0	obsgrp1
conc87	0.0000	1.0	obsgrp1
conc88	3.5400	1.0	obsgrp1
conc89	3.9700	1.0	obsgrp1
conc90	0.0000	1.0	obsgrp1
conc91	0.0000	1.0	obsgrp1
conc92	6.1900	1.0	obsgrp1
conc93	0.7770	1.0	obsgrp1
conc94	1.1300	1.0	obsgrp1
conc95	5.0800	1.0	obsgrp1
conc96	0.0000	1.0	obsgrp1
conc97	0.0000	1.0	obsgrp1
conc98	7.2200	1.0	obsgrp1
conc99	3.9700	1.0	obsgrp1
conc100	0.0000	1.0	obsgrp1
conc101	0.0000	1.0	obsgrp1
conc102	4.5000	1.0	obsgrp1
conc103	4.1300	1.0	obsgrp1
conc104	4.5600	1.0	obsgrp1
conc105	2.4500	1.0	obsgrp1
conc106	0.0000	1.0	obsgrp1
conc107	0.5630	1.0	obsgrp1
conc108	0.0000	1.0	obsgrp1
conc109	0.0000	1.0	obsgrp1
conc110	0.0000	1.0	obsgrp1
conc111	3.8500	1.0	obsgrp1
conc112	5.8100	1.0	obsgrp1
conc113	3.7400	1.0	obsgrp1

conc114	1.6000	1.0	obsgrp1
conc115	2.0600	1.0	obsgrp1
conc116	0.8300	1.0	obsgrp1
conc117	0.0000	1.0	obsgrp1
conc118	0.0000	1.0	obsgrp1
conc119	0.0000	1.0	obsgrp1
conc120	0.0000	1.0	obsgrp1
conc121	0.0000	1.0	obsgrp1
conc122	0.0000	1.0	obsgrp1
conc123	0.0000	1.0	obsgrp1
conc124	0.0000	1.0	obsgrp1
conc125	0.0000	1.0	obsgrp1
conc126	0.0000	1.0	obsgrp1
conc127	0.0000	1.0	obsgrp1
conc128	0.0000	1.0	obsgrp1
conc129	0.0000	1.0	obsgrp1
conc130	0.0000	1.0	obsgrp1
conc131	0.0000	1.0	obsgrp1
conc132	0.0000	1.0	obsgrp1
conc133	0.0000	1.0	obsgrp1
conc134	0.0000	1.0	obsgrp1
conc135	0.0000	1.0	obsgrp1
conc136	0.0000	1.0	obsgrp1
conc137	0.0000	1.0	obsgrp1
conc138	0.0000	1.0	obsgrp1
conc139	0.0000	1.0	obsgrp1
conc140	0.0000	1.0	obsgrp1
conc141	0.5850	1.0	obsgrp1
conc142	0.0000	1.0	obsgrp1
conc143	0.0000	1.0	obsgrp1
conc144	0.6180	1.0	obsgrp1
conc145	0.0000	1.0	obsgrp1
conc146	0.0000	1.0	obsgrp1
conc147	0.0000	1.0	obsgrp1
conc148	0.0000	1.0	obsgrp1
conc149	0.8680	1.0	obsgrp1
conc150	0.0000	1.0	obsgrp1
conc151	0.6000	1.0	obsgrp1
conc152	0.8050	1.0	obsgrp1
conc153	0.0000	1.0	obsgrp1
conc154	0.0000	1.0	obsgrp1
conc155	1.1900	1.0	obsgrp1
conc156	1.2200	1.0	obsgrp1
conc157	1.2000	1.0	obsgrp1
conc158	1.2100	1.0	obsgrp1
conc159	1.0000	1.0	obsgrp1
conc160	0.0000	1.0	obsgrp1
conc161	1.0100	1.0	obsgrp1
conc162	1.2000	1.0	obsgrp1
conc163	1.1000	1.0	obsgrp1
conc164	1.4800	1.0	obsgrp1
conc165	1.2300	1.0	obsgrp1
conc166	1.0200	1.0	obsgrp1
conc167	1.0100	1.0	obsgrp1
conc168	0.5210	1.0	obsgrp1
conc169	0.0000	1.0	obsgrp1
conc170	0.0000	1.0	obsgrp1
conc171	0.0000	1.0	obsgrp1
conc172	0.0000	1.0	obsgrp1
conc173	0.0000	1.0	obsgrp1
conc174	1.1200	1.0	obsgrp1
conc175	0.0000	1.0	obsgrp1
conc176	0.0000	1.0	obsgrp1
conc177	1.4100	1.0	obsgrp1
conc178	1.3300	1.0	obsgrp1
conc179	0.5480	1.0	obsgrp1
conc180	0.0000	1.0	obsgrp1
conc181	0.0000	1.0	obsgrp1
conc182	0.0000	1.0	obsgrp1
conc183	0.0000	1.0	obsgrp1
conc184	1.3200	1.0	obsgrp1
conc185	1.0100	1.0	obsgrp1
conc186	1.2300	1.0	obsgrp1
conc187	0.0000	1.0	obsgrp1
conc188	0.9450	1.0	obsgrp1
conc189	1.0000	1.0	obsgrp1

conc190	0.0000	1.0	obsgrp1
conc191	0.7610	1.0	obsgrp1
conc192	0.0000	1.0	obsgrp1
conc193	0.0000	1.0	obsgrp1
conc194	0.0000	1.0	obsgrp1
conc195	1.2000	1.0	obsgrp1
conc196	0.0000	1.0	obsgrp1
conc197	1.3300	1.0	obsgrp1
conc198	0.0000	1.0	obsgrp1
conc199	0.5400	1.0	obsgrp1
conc200	0.0000	1.0	obsgrp1
conc201	0.0000	1.0	obsgrp1
conc202	1.0300	1.0	obsgrp1
conc203	1.3500	1.0	obsgrp1
conc204	1.3000	1.0	obsgrp1
conc205	0.9850	1.0	obsgrp1
conc206	0.0000	1.0	obsgrp1
conc207	0.9190	1.0	obsgrp1
conc208	0.0000	1.0	obsgrp1
conc209	1.2700	1.0	obsgrp1
conc210	1.3200	1.0	obsgrp1
conc211	0.5600	1.0	obsgrp1
conc212	0.8680	1.0	obsgrp1
conc213	0.0000	1.0	obsgrp1
conc214	0.7520	1.0	obsgrp1
conc215	0.7050	1.0	obsgrp1
conc216	0.6530	1.0	obsgrp1
conc217	0.0000	1.0	obsgrp1
conc218	0.0000	1.0	obsgrp1
conc219	0.9900	1.0	obsgrp1
conc220	0.0000	1.0	obsgrp1
conc221	0.9370	1.0	obsgrp1
conc222	0.0000	1.0	obsgrp1
conc223	0.0000	1.0	obsgrp1
conc224	0.8940	1.0	obsgrp1
conc225	0.0000	1.0	obsgrp1
conc226	0.7200	1.0	obsgrp1
conc227	0.0000	1.0	obsgrp1
conc228	0.7000	1.0	obsgrp1
conc229	0.0000	1.0	obsgrp1
conc230	0.0000	1.0	obsgrp1
conc231	0.6570	1.0	obsgrp1
conc232	0.8970	1.0	obsgrp1
conc233	0.9470	1.0	obsgrp1
conc234	1.1600	1.0	obsgrp1
conc235	0.0000	1.0	obsgrp1
conc236	0.9320	1.0	obsgrp1
conc237	0.0000	1.0	obsgrp1
conc238	0.0000	1.0	obsgrp1
conc239	1.0400	1.0	obsgrp1
conc240	0.0000	1.0	obsgrp1
conc241	0.0000	1.0	obsgrp1
conc242	0.6310	1.0	obsgrp1
conc243	0.0000	1.0	obsgrp1
conc244	0.0000	1.0	obsgrp1
conc245	1.5200	1.0	obsgrp1
conc246	0.9680	1.0	obsgrp1
conc247	0.7200	1.0	obsgrp1
conc248	1.1200	1.0	obsgrp1
conc249	1.1500	1.0	obsgrp1
conc250	0.7020	1.0	obsgrp1
conc251	0.5420	1.0	obsgrp1
conc252	0.0000	1.0	obsgrp1
conc253	0.0000	1.0	obsgrp1
conc254	0.0000	1.0	obsgrp1
conc255	0.0000	1.0	obsgrp1
conc256	1.0900	1.0	obsgrp1
conc257	0.9980	1.0	obsgrp1
conc258	0.6340	1.0	obsgrp1
conc259	0.0000	1.0	obsgrp1
conc260	0.9870	1.0	obsgrp1
conc261	0.0000	1.0	obsgrp1
conc262	1.1400	1.0	obsgrp1
conc263	1.0300	1.0	obsgrp1
conc264	0.5710	1.0	obsgrp1
conc265	0.6870	1.0	obsgrp1

```
conc266 0.0000 1.0 obsgrp1
conc267 0.0000 1.0 obsgrp1
conc268 0.9270 1.0 obsgrp1
conc269 0.8620 1.0 obsgrp1
conc270 0.0000 1.0 obsgrp1
conc271 1.0800 1.0 obsgrp1
conc272 0.0000 1.0 obsgrp1
conc273 0.0000 1.0 obsgrp1
conc274 0.7900 1.0 obsgrp1
conc275 0.5880 1.0 obsgrp1
conc276 0.5740 1.0 obsgrp1
conc277 0.0000 1.0 obsgrp1
conc278 0.0000 1.0 obsgrp1
conc279 0.0000 1.0 obsgrp1
conc280 1.0700 1.0 obsgrp1
conc281 0.0000 1.0 obsgrp1
conc282 0.0000 1.0 obsgrp1
conc283 0.8960 1.0 obsgrp1
conc284 0.5310 1.0 obsgrp1
conc285 0.9020 1.0 obsgrp1
conc286 0.7220 1.0 obsgrp1
conc287 0.0000 1.0 obsgrp1
conc288 0.0000 1.0 obsgrp1
conc289 0.6250 1.0 obsgrp1
conc290 0.7070 1.0 obsgrp1
conc291 1.1600 1.0 obsgrp1
conc292 0.8310 1.0 obsgrp1
conc293 0.8500 1.0 obsgrp1
conc294 0.5860 1.0 obsgrp1
conc295 0.5300 1.0 obsgrp1
conc296 0.7560 1.0 obsgrp1
conc297 0.0000 1.0 obsgrp1
conc298 1.2300 1.0 obsgrp1
conc299 0.7570 1.0 obsgrp1
conc300 0.9800 1.0 obsgrp1
conc301 0.0000 1.0 obsgrp1
conc302 0.5910 1.0 obsgrp1
conc303 0.0000 1.0 obsgrp1
conc304 0.9280 1.0 obsgrp1
conc305 0.8200 1.0 obsgrp1
conc306 0.0000 1.0 obsgrp1
conc307 0.0000 1.0 obsgrp1
conc308 1.0000 1.0 obsgrp1
conc309 1.1400 1.0 obsgrp1
conc310 0.7780 1.0 obsgrp1
conc311 0.7110 1.0 obsgrp1
conc312 0.6420 1.0 obsgrp1
conc313 0.8080 1.0 obsgrp1
conc314 0.0000 1.0 obsgrp1
conc315 1.0600 1.0 obsgrp1
conc316 0.9810 1.0 obsgrp1
conc317 0.6760 1.0 obsgrp1
conc318 0.8230 1.0 obsgrp1
conc319 0.0000 1.0 obsgrp1
conc320 1.1700 1.0 obsgrp1
* model command line
ModForPEST
* model input/output
plume_param.tpl plume_param.dat
metcoil_b.ins CFIT2.OUT
* prior information
```

Appendix D

DDS Results

Table D.1. Investigation of Effect of Various Estimators and Deadzones

Objective Function Definition							Location Criteria				
Estimator	Add-Ons	Trials	Function Evaluations	Range of Objective Function Values After DDS		No. Pass	% Pass of Total	Range of Objective Function Values After Applying Location Criteria			
No Optimization (L ₁ -estimator)	n/a	6000	1,200,000	204.908	- 473.14	115	1.9%	250.779	-	473.14	
No Optimization (LRS1 M-estimator)	n/a	6000	1,200,000	51.5945	- 109.763	115	1.9%	56.3101	-	109.763	
L ₂ -estimator	None	13268	2,653,600	544.201	- 938.72	0	0.0%				
L ₂ -estimator	None	3000	600,000	545.413	- 919.82	0	0.0%				
L ₂ -estimator	None	9952	1,990,400	544.201	- 938.72	0	0.0%				
L ₂ -estimator	None	100	20,000	553.163	- 849.814	0	0.0%				
L ₂ -estimator	None	100	20,000	545.192	- 807.807	0	0.0%				
L ₂ -estimator	None	100	20,000	550.26	- 888.788	0	0.0%				
L ₂ -estimator	None	16	3,200	553.528	- 573.426	0	0.0%				
L ₂ -estimator	Deadzone (d=0.1)	5000	1,000,000	543.747	- 993.194	0	0.0%				
L ₁ -estimator	None	10936	2,187,200	192.789	- 272.283	22	0.201%	234.959	-	270.883	
L ₁ -estimator	None	1000	200,000	193.526	- 271.82	2	0.2%	252.988	-	270.89	
L ₁ -estimator	None	9936	1,987,200	192.789	- 272.283	20	0.2%	234.959	-	270.883	
L ₁ -estimator	Deadzone (d=0.5)	5000	1,000,000	160.466	- 257.782	40	0.8%	193.323	-	244.864	
L ₁ -estimator	Deadzone (d=0.1)	5000	1,000,000	189.837	- 273.475	10	0.200%	219.69	-	289.863	
L ₁ -estimator	Deadzone (d=0.5); Penalty (S=10,000)	3000	600,000	190.058	- 9999	1935	64.500%	193.484	-	3455	
LRS1 M-estimator	no dead	4030	806,000	49.7006	- 60.6159	57	1.4%	55.0338	-	58.54	
LRS1 M-estimator	Deadzone (d=0.1)	5000	1,000,000	49.7834	- 60.6159	89	1.8%		-		
LRS1 M-estimator	Deadzone (d=0.5)	3000	600,000	49.8005	- 60.6159	46	1.5%	55.1313	-	58.6804	
Cauchy's M-estimator (c=1)	None	5000	1,000,000	24.5819	- 37.1213						
Cauchy's M-estimator (c=1/2)	None	5000	1,000,000			0	0.00%				
Fair	None	4985	997,000	193.857	- 296.011	7	0.1%	212.284	-	290.144	
LRS2 M-estimator	None	5000	1,000,000	17.3383	- 26.8076	4	0.08%	24.1037	-	26.8076	
LRS2 M-estimator	Deadzone (d=0.1)	7000	1,400,000	16.2124	- 27.1624	9	0.13%	21.7009	-	27.1624	
LRS2 M-estimator	Deadzone (d=0.1)	2000	400,000	16.238	- 27.0856	4	0.2%	25.3546	-	27.0856	
LRS2 M-estimator	Deadzone (d=0.1)	5000	1,000,000	16.2124	- 27.1624	5	0.1%	21.7009	-	27.1624	
Huber's M-estimator (k=0.1)	None	4911	982,200	36.2582	- 51.6412	9	0.2%	42.2652	-	51.4822	
Huber's M-estimator (k=1.0)	None	5000	1,000,000	220.424	- 370.751	2	0.04%	342.952	-	370.37	
Huber's M-estimator (k=3.0)	None	5000	1,000,000	378.089	- 686.591	0	0.0%				
Huber's M-estimator (k=4.0)	None	5000	1,000,000	415.688	- 732.149	0	0.0%				
PEST (L ₂ -estimator)	No Transformation	15	3479	566.01	- 6195.1	0	0.0%				
PEST (L ₂ -estimator)	Lognormal Transformation	16	3128	545.44	- 1353.9	0	0.0%				
PEST (L ₂ -estimator)	Selected Lognormal Transformation	16	3128	562.1	- 1353.5	0	0.0%				

Table D.1. Investigation of Effect of Various Estimators and Deadzones (continued)

Objective Function Definition		Product Criteria					First Arrival Times			
Estimator	Add-Ons	No. Pass	% Pass of Total	% Pass of Passed Location Criteria	Range of Objective Function Values After Applying The Two Criteria		Leclercq Class Area	Majdrech Area B	Majdrech Area C	Majdrech Area D
No Optimization (L ₁ -estimator)	n/a	89	1.48%	77%	250.779	- 473.14	0.6	7.6	10.6	9.8
No Optimization (LRS1 M-estimator)	n/a	89	1.48%	77%	56.3101	- 109.763	0.6	7.6	10.6	9.8
L ₂ -estimator	None									
L ₂ -estimator	None									
L ₂ -estimator	None									
L ₂ -estimator	None									
L ₂ -estimator	None									
L ₂ -estimator	None									
L ₂ -estimator	None									
L ₂ -estimator	Deadzone (d=0.1)									
L ₁ -estimator	None	7	0.064%	32%	234.959	- 264.777	1.8	12.6	16.8	14
L ₁ -estimator	None	1	0.10%	50%	252.988	- 252.988	2.4	19	26.6	23.6
L ₁ -estimator	None	6	0.080%	30%	234.959	- 264.777	1.8	12.6	16.8	14
L ₁ -estimator	Deadzone (d=0.5)	26	0.52%	65%	193.323	- 244.864	0.6	8.8	11.8	12.8
L ₁ -estimator	Deadzone (d=0.1)	9	0.18%	90%	219.89	- 269.883	1.0	9.8	13.4	11.2
L ₁ -estimator	Deadzone (d=0.5); Penalty (S=10,000)	1317	43.9%	68%	193.464	- 1045.26	0.4	7.8	11	9.4
LRS1 M-estimator	no dead	50	1.2%	88%	55.0338	- 58.54	0.4	7.8	10.6	12.4
LRS1 M-estimator	Deadzone (d=0.1)	72	1.4%	81%						
LRS1 M-estimator	Deadzone (d=0.5)	40	1.3%	67%	55.1313	- 58.8804	0.6	9.2	12.2	12.4
Cauchy's M-estimator (c=1)	None	8	0.16%		29.76	- 34.22	0.8	10.8	14.4	11.4
Cauchy's M-estimator (c=1/2)	None									
Fair	None	3	0.06%	43%	212.264	- 290.144	1	8.8	11.6	15.4
LRS2 M-estimator	None	1	0.02%	25%	24.1037	- 24.1037	1.0	8.4	11.2	12.2
LRS2 M-estimator	Deadzone (d=0.1)	2	0.03%	22%	21.7009	- 23.1356	2.0	12.6	16.6	13.8
LRS2 M-estimator	Deadzone (d=0.1)	0			n/a					
LRS2 M-estimator	Deadzone (d=0.1)	2	0.04%	40%	21.7009	- 23.1356	2	12.6	16.6	13.8
Huber's M-estimator (k=0.1)	None	3	0.1%	33%	42.2652	- 47.5855				
Huber's M-estimator (k=1.0)	None	1	0.0%	50%	342.952	- 342.952	2.4	13.2	17.2	13.4
Huber's M-estimator (k=3.0)	None									
Huber's M-estimator (k=4.0)	None									
PEST (L ₂ -estimator)	No Transformation	0	0.00%							
PEST (L ₂ -estimator)	Lognormal Transformation	0	0.00%							
PEST (L ₂ -estimator)	Selected Lognormal Transformation	0								

Table D.2. Investigation of Effect of Scaling Factor in Penalty Function

Estimator	Scaling Factor for Penalty Function	Location Criteria						
		Trials	Function Evaluations	Range of Objective Function Values After DDS		No. Pass	% Pass of Total	Range of Objective Function Values After Applying Location Criteria
L_2 -estimator	no penalty	13268	2,653,600	544.201 - 938.72		0	0.0%	n/a
L_2 -estimator	500	1000	200,000	592.536 - 1258.76		1	0.1%	708.258 - 708.258
L_2 -estimator	1,000	1000	200,000	592.141 - 1668.97		116	11.6%	624.789 - 798.786
L_2 -estimator	5,000	2974	594,800	604.122 - 4653.23		1209	40.7%	612.518 - 1531.93
L_2 -estimator	10,000	3000	600,000	606.005 - 9999		1502	50.1%	609.338 - 2764.15
L_2 -estimator	50,000	3022				852	28.2%	
L_2 -estimator	50,000	1000	200,000	610.147 - 9999		304	30.4%	610.147 - 9999
L_2 -estimator	50,000	2022	404,400	614.333 - 9999		548	27.1%	614.333 - 9999
L_1 -estimator	no penalty	10936	2,187,200	192.789 - 272.283		22	0.2%	234.959 - 270.883
L_1 -estimator	500	2929				647	22.1%	
L_1 -estimator	500	2929	585,800	226.881 - 639.476		647	22.1%	238.145 - 398.191
L_1 -estimator	1000	3000		227 - 1,013		1022	34.1%	
L_1 -estimator	1000	1000	200,000	231.477 - 1012.56		347	34.7%	240.997 - 459.934
L_1 -estimator	1000	2000	400,000	227.088 - 997.435		675	33.8%	239.76 - 400.183
L_1 -estimator	5,000	2973	594,800	233.231 - 3993.94		1600	53.8%	237.57 - 1036.3
L_1 -estimator	10,000	3000				1825	60.8%	
L_1 -estimator	10,000	3000	600,000	232.472 - 9999		1825	60.8%	238.495 - 1185.38
L_1 -estimator	50,000	2351	470,200	235.076 - 9999		846	36.0%	235.076 - 9999
LRS1 M-estimator	no penalty	4030	806000	49.7006 - 60.6159		57	1.41%	55.0338 - 58.54
LRS1 M-estimator	100	2987				805	27.0%	
LRS1 M-estimator	100	2987	597,400	58.2404 - 136.168		805	27.0%	58.4087 - 82.1161
LRS1 M-estimator	500	2999				1459	48.6%	
LRS1 M-estimator	500	2999	599,800	58.864 - 436.149		1459	48.6%	58.9358 - 183.379
LRS1 M-estimator	1000	7497				4042	53.9%	
LRS1 M-estimator	1000	2590	518,000	58.9899 - 811.034		1382	53.4%	58.9899 - 162.862
LRS1 M-estimator	1000	4907	981,400	59.0513 - 811.298		2660	54.2%	59.0513 - 243.774
LRS1 M-estimator	5,000	3000	600,000	59.0967 - 3811.05		2171	72.4%	59.0967 - 263.86
LRS1 M-estimator	10,000	3000	600,000	59.2425 - 9999		2390	79.7%	59.2425 - 305.184
LRS1 M-estimator	50,000	2998	599,600	59.0894 - 9999		1228	41.0%	59.0894 - 9999

Table D.2. Investigation of Effect of Scaling Factor in Penalty Function (continued)

Estimator	Scaling Factor for Penalty Function	Product Criteria				First Arrival Times			
		No. Pass	% Pass of Total	% Pass of Location Criteria	Range of Objective Function Values After Applying The Two Criteria	Leclercq Class Area	Mejdrech Area B	Mejdrech Area C	Mejdrech Area D
L ₂ -estimator	no penalty	n/a			n/a				
L ₂ -estimator	500	0	0.0%	0%	n/a				
L ₂ -estimator	1,000	42	4.2%	36%	624.789 - 740.408	0.6	10	14.8	13.4
L ₂ -estimator	5,000	756	25.4%	63%	612.518 - 1155.43	0.4	8.2	12	10.8
L ₂ -estimator	10,000	1019	34.0%	68%	609.338 - 2764.15	0.4	8.4	12	10.6
L ₂ -estimator	50,000	647	21.4%						
L ₂ -estimator	50,000	231	23.1%	76%	610.147 - 9999	0.6	8.8	12	11.4
L ₂ -estimator	50,000	416	20.6%	76%	614.333 - 9999	0.6	9	12.4	10.4
L ₁ -estimator	no penalty	7	0.06%	32%	234.959 - 264.777	1.8	12.6	16.8	14
L ₁ -estimator	500	574	19.6%						
L ₁ -estimator	500	574	19.6%	89%	238.145 - 398.191	0.4	8	11.6	11
L ₁ -estimator	1000	881	29.4%	86%					
L ₁ -estimator	1000	305	30.5%	88%	240.997 - 331.412	0.4	8.4	12	10.8
L ₁ -estimator	1000	576	28.8%	85%	239.76 - 400.183	0.4	8	11.8	10.4
L ₁ -estimator	5,000	1173	39.5%	73%	237.57 - 667.843	0.4	8.2	11.4	9.8
L ₁ -estimator	10,000	1245	41.5%						
L ₁ -estimator	10,000	1245	41.5%	68%	238.495 - 1169.15	0.6	9	12.2	10.2
L ₁ -estimator	50,000	553	23.5%	65%	235.076 - 9999	0.4	8.4	11.6	10.8
LRS1 M-estimator	no penalty	50	1.24%	88%	55.0338 - 58.54	0.4	7.8	10.6	12.4
LRS1 M-estimator	100	782	26.2%						
LRS1 M-estimator	100	782	26.2%	97%	58.4087 - 82.1161	0.4	9.6	13.2	12
LRS1 M-estimator	500	1291	43.0%						
LRS1 M-estimator	500	1291	43.0%	88%	58.9358 - 183.379	0.4	8.2	11.6	10.6
LRS1 M-estimator	1000	3153	42.1%						
LRS1 M-estimator	1000	1074	41.5%	78%	58.9899 - 162.431	0.6	8.2	11.6	10.4
LRS1 M-estimator	1000	2079	42.4%	78%	59.0513 - 187.552	0.4	8.4	12	9.8
LRS1 M-estimator	5,000	1482	49.4%	68%	59.0967 - 262.568	0.4	7.8	11	9.6
LRS1 M-estimator	10,000	1614	53.8%	68%	59.2425 - 305.184	0.4	7.4	10.4	9.2
LRS1 M-estimator	50,000	876	29.2%	71%	59.0894 - 9999	0.4	7.8	10.6	10.6

References

- Agyei, E. and K. Hatfield (2006). "Enhancing gradient-based parameter estimation with an evolutionary approach." *Journal Of Hydrology* 316(1-4): 266-280.
- Ball, R. O. (2003). Expert Report: Contamination of the Ground Water in Lisle, Illinois and Neighboring Suburbs, ENVIRON International Corporation.
- Beven, K. (2006). "A manifesto for the equifinality thesis." *Journal Of Hydrology* 320(1-2): 18-36.
- Beven, K. and A. Binley (1992). "The Future Of Distributed Models - Model Calibration And Uncertainty Prediction." *Hydrological Processes* 6(3): 279-298.
- Beven, K. and J. Freer (2001). "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology." *Journal Of Hydrology* 249(1-4): 11-29.
- Boyd, S. and L. Vandenberghe (2003). *Convex Optimization*, Cambridge University Press.
- Brooks, R. J., D. N. Lerner, et al. (1994). "Determining The Range Of Predictions Of A Groundwater Model Which Arises From Alternative Calibrations." *Water Resources Research* 30(11): 2993-3000.
- Carrera, J., A. Alcolea, et al. (2005). "Inverse problem in hydrogeology." *Hydrogeology Journal* 13(1): 206-222.
- Clayton Group Services, Inc. (2002). *Comprehensive VOC Investigation Report: The Lockformer Company, Lisle, Illinois. Clayton Project No. 15-65263.01.008*
- Christensen, S. (2004). "A synthetic groundwater modelling study of the accuracy of GLUE uncertainty intervals." *Nordic Hydrology* 35(1): 45-59.
- Christensen, S. and R. L. Cooley (1999). "Evaluation of prediction intervals for expressing uncertainties in groundwater flow model predictions." *Water Resources Research* 35(9): 2627-2639.
- Dou, C. H., W. Woldt, et al. (1995). "Steady-State Groundwater-Flow Simulation With Imprecise Parameters." *Water Resources Research* 31(11): 2709-2719.
- Duan, Q. Y., S. Sorooshian, et al. (1992). "Effective And Efficient Global Optimization For Conceptual Rainfall-Runoff Models." *Water Resources Research* 28(4): 1015-1031.

- Feyen, L., K. J. Beven, et al. (2001). "Stochastic capture zone delineation within the generalized likelihood uncertainty estimation methodology: Conditioning on head observations." *Water Resources Research* 37(3): 625-638.
- Finsterle, S. and J. Najita (1998). "Robust estimation of hydrogeologic model parameters." *Water Resources Research* 34(11): 2939-2947.
- Gill, P. E., W. Murray, et al. (1983). *Practical Optimization*, Academic Press Inc.
- Green, E. D., P. L. Patton, et al. (2006). "Future claimant trust and "channeling injunctions" to resolve mass tort environmental liability in bankruptcy: The Met-Coil Model." *Emory Bankruptcy Developments Journal* 22.
- Groundwater Services Inc. (GSI) (2003). *Expert Opinion of John A. Conner, P.E., Regarding Soil And Groundwater Conditions Associated with the Lockformer Facility, Lisle, Illinois*
- Hill, M. C. (2006). "The Practical Use of Simplicity in Developing Ground Water Models." *Ground Water* 44(6): 775-781.
- Hornberger, G. M. and R. C. Spear (1981). "An Approach To The Preliminary-Analysis Of Environmental Systems." *Journal Of Environmental Management* 12(1): 7-18.
- James, A. L. and C. M. Oldenburg (1997). "Linear and Monte Carlo uncertainty analysis for subsurface contaminant transport simulation." *Water Resources Research* 33(11): 2495-2508.
- Johnson, R. A. (2000). *Miller & Freund's Probability and Statistics for Engineers: 5th Edition*. Upper Saddle River, New Jersey, Prentice Hall.
- Kirchner, J. W. (2006). "Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology." *Water Resources Research* 42(3).
- Kuczera, G. and E. Parent (1998). "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm." *Journal Of Hydrology* 211(1-4): 69-85.
- McKay, M. D., R. J. Beckman, and W. J. Conover (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.
- McLaughlin, D. and L. R. Townley (1996). "A reassessment of the groundwater inverse problem." *Water Resources Research* 32(5): 1131-1161.

- McMichael, C. E., A. S. Hope, et al. (2006). "Distributed hydrological modelling in California semi-arid shrublands: MIKE SHE model calibration and uncertainty estimation." *Journal Of Hydrology* 317(3-4): 307-324.
- Montanari, A. (2005). "Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations." *Water Resources Research* 41(8).
- Moore, C. and J. Doherty (2005). "Role of the calibration process in reducing model predictive error." *Water Resources Research* 41(5).
- Moran, J. M., J. S. Zogorski, and P. J. Squillace (2006). "Chlorinated Solvents in Groundwater of the United States." *Environ. Sci. & Technol.* Published on Web 12/01/2006.
- Neuman, S. P. (2003). "Maximum likelihood Bayesian averaging of uncertain model predictions." *Stochastic Environmental Research And Risk Assessment* 17(5): 291-305.
- Rousseeuw, P. J. and A. M. Leroy (2003). *Robust Regression and Outlier Detection*. Hoboken, New Jersey, John Wiley & Sons, In.
- Singh, A. and J. Nocerino (2002). "Robust estimation of mean and variance using environmental data sets with below detection limit observations." *Chemometrics And Intelligent Laboratory Systems* 60(1-2): 69-86.
- Sohrabi, T. M., A. Shirmohammadi, et al. (2002). "Uncertainty in nonpoint source pollution models and associated risks." *Environmental Forensics* 3(2): 179-189.
- Suarez, M. P. and H. S. Rifai (1999). "Biodegradation Rates for Fuel Hydrocarbons and Chlorinated Solvents in Groundwater." *Bioremediation Journal* 3(4): 337-362.
- Sun, A. Y. P., Scott L.; Wittmeyer, Gordon W. (2006). "A constrained robust least squares approach for contaminant release history identification." *Water Resources Research* 42(4).
- Sykes, J.F. (2004). *An Assessment of Soil Contamination and TCE Migration in Groundwater from the Lockformer Company Facility, Lisle, Illinois*.
- Tolson, B. A. (2005). *Automatic Calibration, Management and Uncertainty Analysis: Phosphorus Transport in the Cannonsville Watershed, Cornell University*. Ph. D.: 227.

- Tolson, B. A. and C. A. Shoemaker "The Dynamically Dimensioned Search (DDS) Algorithm for Computationally Efficient Watershed Model Calibration." *Water Resources Research* 43(1).
- Tonkin, M. J. and J. Doherty (2005). "A hybrid regularized inversion methodology for highly parameterized environmental models." *Water Resources Research* 41(10).
- U.S. Environmental Protection Agency (1989). *Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part A)*, Office of Emergency and Remedial Response. EPA/540/1-89/002.
- U.S. Environmental Protection Agency (1991). *Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part B, Development of Risk-Based Preliminary Remediation Goals)*, Office of Emergency and Remedial Response. EPA/540/R-92/003.
- U.S. Environmental Protection Agency (1997). *Guiding Principles for Monte Carlo Analysis, Risk Assessment Forum*. EPA/630/R-97/001.
- U.S. Environmental Protection Agency (2000). "Characterization of Data Variability and Uncertainty: Health Effect Assessments in the Integrated Risk Information System (IRIS)." EPA/635/R-00/005F.
- U.S. Environmental Protection Agency (2001). *Risk Assessment Guidance for Superfund, Volume 3: Process for Conducting Probabilistic Risk Assessment (Part A)*, Office of Emergency and Remedial Response. EPA 540-R-02-002.
- U.S. Environmental Protection Agency (2002). *Calculating Upper Confidence Limits for Exposure Point Concentrations at Hazardous Waste Sites*, Office of Emergency and Remedial Response. OSWER Directive No. 9285.6-10.
- U.S. Environmental Protection Agency (2007). *List of Contaminants & their MCLs*
<http://www.epa.gov/safewater/contaminants/index.html>
- U.S. Environmental Protection Agency (2006). *Ground Water & Drinking Water, Consumer Factsheet on: TRICHLOROETHYLENE*
http://www.epa.gov/safewater/contaminants/dw_contamfs/trichlor.html (Accessed: November 25, 2006)

- U.S. Geological Survey (2006). UCODE_2005 and Six Other Computer Codes for Universal Sensitivity Analysis, Calibration, and Uncertainty Evaluation; <http://water.usgs.gov/software/ucode.html> (Accessed: November 25, 2006)
- Vecchia, A. V. and R. L. Cooley (1987). "Simultaneous Confidence And Prediction Intervals For Nonlinear-Regression Models With Application To A Groundwater-Flow Model." *Water Resources Research* 23(7): 1237-1250.
- Federal-Provincial-Territorial Committee on Drinking Water (2004). *Trichloroethylene in Drinking Water*.
- Watermark Numerical Computing (2004). *PEST - Model-Independent Parameter Estimation User Manual: 5th Edition*.
- Wittmeyer, G. and S. P. Neuman (1991). "Monte-Carlo Experiments With Robust Estimation Of Aquifer Model Parameters." *Advances In Water Resources* 14(5): 252-272.
- Woller, D. M., E. W. Sanderson, et al. (1986). *Public Ground-Water Supplies in DuPage County, Illinois, Department of Energy and Natural Resources. ISWS/BUL-60(32)/86*.
- Xiang, Y., J. F. Sykes, and N.R. Thomson (1992). "A Composite L_1 Parameter Estimator For Model-Fitting In Groundwater-Flow And Solute Transport Simulation." *Water Resources Research* 29(6): 1661-1673.
- Yin, Y. (2006). Personal Communication.