



3022

INTEGRATION OF DATA FUSION MODELING (DFM) WITH VAM3DF CONTAMINANT TRANSPORT CODE

Prepared for:

**Fluor Fernald
7400 Willey Road
Hamilton, OH 45013-9402**

April 2000

**HydroGeoLogic, Inc.
1155 Herndon Parkway
Suite 900
Herndon, Virginia 20170
(703) 478-5186**

000001

**INTEGRATION OF DATA FUSION MODELING (DFM) WITH
VAM3DF CONTAMINANT TRANSPORT CODE**

Prepared for:

**Fluor Fernald
7400 Willey Road
Hamilton, OH 45013-9402**

Prepared by:

**HydroGeoLogic, Inc.
Herndon, VA 20170**

April 2000

TABLE OF CONTENTS

	Page
1.0 INTRODUCTION	1-1
1.1 BACKGROUND	1-1
1.2 PRODUCT DEFINITION AND CAPABILITIES	1-1
1.2.1 Operational Environment	1-1
1.2.2 Capabilities	1-2
2.0 DATA FUSION METHOD (DFM) THEORETICAL BACKGROUND	2-1
2.1 GENERAL THEORY	2-2
2.1.1 Models	2-2
2.1.1.1 Physical Model	2-2
2.1.1.2 Statistical Model	2-3
2.1.1.3 Measurement Model	2-5
2.1.2 Penalty Function	2-5
2.1.3 Gauss-Newton Method	2-6
2.1.4 Extended Kalman Filter Updates	2-9
2.1.5 Tests of Model Robustness	2-10
2.1.6 Uncertainties	2-11
3.0 IMPLEMENTATION	3-1
3.1 SOFTWARE STRUCTURE	3-1
3.2 OVERVIEW OF SOFTWARE APPLICATION	3-1
3.2.1 Nonlinear Batch Estimation	3-1
3.2.2 Estimation Updates Using Extended Kalman Filter	3-3
3.2.3 Testing of Model's Robustness	3-3
3.2.4 Parameter Estimate Error Covariance	3-4
3.2.5 Prediction and Prediction Error Covariance Simulation	3-4
4.0 DFM/VAM3DF USER'S GUIDE	4-1
4.1 DFM/VAM3DF INPUT	4-1
4.1.1 Mandatory Input Files	4-1
4.1.2 Optional Input Files	4-1
4.1.4 DFM/VAM3DF Input Details	4-2
4.2 DFM/VAM3DF EXECUTION	4-5
4.3 DFM/VAM3DF OUTPUT	4-6
5.0 VERIFICATION: SMALL-SCALE TESTING	5-1
5.1 INTRODUCTION	5-1
5.2 TESTING RANDOM FIELD GENERATION	5-1
5.2 TEST PROBLEM DESCRIPTION	5-2
5.2.1 Discretization	5-2
5.2.2 Initial Condition	5-2
5.2.3 Boundary Conditions	5-2
5.2.4 Transport Parameters	5-2
5.2.5 Selection of Field and Concentration Measurement Locations	5-2
5.2.6 Description of Test Scenarios	5-3

000003

TABLE OF CONTENTS

	Page
5.3	SIMULATION PROCEDURES 5-3
5.4	TEST METHODS 5-4
5.4.1	Visual Inspection 5-4
5.4.2	Chi-Squared Test 5-4
5.5	SIMULATION RESULTS AND DISCUSSION 5-5
5.5.1	Estimation of Initial Concentration 5-5
5.5.2	Estimation of Log_{10} (Effective Porosity) 5-6
5.5.3	Estimation of $\text{Log } K_d$ 5-6
5.5.4	Estimation of Polynomial Trend in Log Effective Porosity 5-6
5.5.5	Estimation of Polynomial Trend in $\text{Log } K_d$ 5-7
5.5.6	Estimation of Polynomial Trend in Longitudinal Dispersivity 5-7
5.5.7	Estimation of the $\text{Log}_{10} K_d$ with Freundlich Isotherm 5-7
5.5.8	Estimation of Polynomial Trend in Desorption Rate Coefficient 5-7
5.6	SUMMARY OF RESULTS 5-8
6.0	SITE-SPECIFIC APPLICATION EXAMPLE: FERNALD SITE 6-1
6.1	INTRODUCTION 6-1
6.2	MODEL DESCRIPTION 6-1
6.2.1	Discretization 6-1
6.2.2	Initial and Boundary Conditions 6-1
6.3	SENSITIVITY ANALYSIS 6-2
6.3.1	Sensitivity Description 6-2
6.3.2	Summary 6-2
6.4	DFM/VAM3DF APPLICATION 6-3
6.4.1	Measurements 6-3
6.4.2	Statistics Model 6-3
6.4.3	Estimation of the 1993 Initial Conditions 6-4
6.4.4	Estimation of the Distribution Coefficient 6-4
6.4.5	Computational Requirements 6-5
6.4.6	DFM/VAM3DF Application Summary 6-5
7.0	SUMMARY, CONCLUSIONS AND RECOMMENDATIONS 7-1
7.1	SUMMARY AND CONCLUSIONS 7-1
7.2	RECOMMENDATIONS 7-1
8.0	REFERENCES 8-1

LIST OF FIGURES

	Page
Figure 3.1	Software Structure. 3-5
Figure 3.2	Gauss-Newton Iterative Procedure. 3-6
Figure 3.3	Post Gauss-Newton Computational Steps. 3-7
Figure 5.1	Simulated and True Variograms for the Generated Iostropic Markov Random Field 5-19
Figure 5.2	Simulated and True Variograms for the Anistropic Markov Random Field Generated. 5-20
Figure 5.3	Numerical Discretization of Computational Domain 5-21
Figure 5.4	Estimation of Initial Plume: True and Estimated Initial Plumes 5-22
Figure 5.5	Estimation of Initial Plume: Standard Deviations Associated with Log Initial Plume Estimates 5-23
Figure 5.6	Estimation of Initial Concentrations: True and Simulated Breakthrough Curves 5-24
Figure 5.7	Estimation of Spatial Random Porosity with Underlying Correlation Length = 4m: True and Estimated Log Porosity Fields 5-25
Figure 5.8	Estimation of Spatial Random Porosity with Underlying Correlation Length = 4m: Standard Deviations Associated with Estimated Log Porosity Fields 5-26
Figure 5.9	Estimation of Spatial Random Porosity with Underlying Correlation Length = 4m: True and Simulated Breakthrough Curves 5-27
Figure 5.10	Estimation of Spatial Random Porosity with Underlying Correlation Length = 8m: True and Estimated Log Porosity Fields 5-28
Figure 5.11	Estimation of Spatial Random Porosity with Underlying Correlation Length = 8m: Standard Deviations Associated with Estimated Log Porosity Fields 5-29
Figure 5.12	Estimation of Spatial Random Porosity with Underlying Correlation Length = 8m: True and Simulated Breakthrough Curves 5-30
Figure 5.13	Estimation of Spatial Random Kd with Underlying Correlation Length = 4m: True and Estimated Log Kd Fields 5-31
Figure 5.14	Estimation of Spatial Random Kd with Underlying Correlation Length = 4m: Standard Deviations Associated with Estimated Log Kd Fields 5-32
Figure 5.15	Estimation of Spatial Random Kd with Underlying Correlation Length = 4m: True and Simulated Breakthrough Curves 5-33
Figure 5.16	Estimation of Spatial Random Kd with Underlying Correlation Length = 8m: True and Estimated Log Kd Fields 5-34
Figure 5.17	Estimation of Spatial Random Kd with Underlying Correlation Length = 8m: Standard Deviations Associated with Estimated Log Kd Fields 5-35
Figure 5.18	Estimation of Spatial Random Kd with Underlying Correlation Length = 8m: True and Simulated Breakthrough Curves 5-36
Figure 5.19	Estimation of Linear Trend in Log Porosity: True and Simulated Log Porosity 5-37
Figure 5.20	Estimation of Linear Trend in Log Porosity: True and Simulated Breakthrough Curves 5-38

LIST OF FIGURES

		<u>Page</u>
Figure 5.21	Estimation of Log Kd with a Linear Trend Based on Log Concentration Measurement Error of 0.1: True and Estimated Log Kd Fields	5-39
Figure 5.22	Estimation of Log Kd with a Linear Trend Based on Log Concentration Measurement Error of 0.1: True and Simulated Breakthrough Curves	5-40
Figure 5.23	Estimation of Log Kd with a Linear Trend Based on Log Concentration Measurement Error of 0.01: True and Estimated Log Kd Fields	5-41
Figure 5.24	Estimation of Log Kd with a Linear Trend Based on Log Concentration Measurement Error of 0.01: True and Simulated Breakthrough Curves	5-42
Figure 5.25	Estimation of Longitudinal Dispersivity with a Linear Trend: True and Estimated Dispersivity a_L Fields	5-43
Figure 5.26	Estimation of Longitudinal Dispersivity with a Linear Trend: True and Simulated Breakthrough Curves	5-44
Figure 5.27	Estimation of Random Freundlich Isotherm Kd with Underlying Correlation Length=4m: True and Estimated Log Kd Fields	5-45
Figure 5.28	Estimation of Random Freundlich Isotherm Kd with Underlying Correlation Length=4m: Standard Deviations Associated with Log Kd Estimates	5-46
Figure 5.29	Estimation of Random Freundlich Isotherm Kd with Underlying Correlation Length=4m: True and Simulated Breakthrough Curves	5-47
Figure 5.30	Estimation of Desorption Rate Coefficient with a Linear Trend: True and Estimated Dispersivity a_d Fields	5-48
Figure 5.31	Estimation of Desorption Rate Coefficient with a Linear Trend: True and Simulated Breakthrough Curves	5-49
Figure 6.1	Regional and Local Model Domains.	6-13
Figure 6.2	Average Uranium Concentrations for 2000 wells in 1993 Used for Initial Conditions	6-14
Figure 6.3	Average Uranium Concentrations for 3000 wells in 1993 Used for Initial Conditions	6-15
Figure 6.4	Average Uranium Concentrations for 4000 wells in 1993 Used for Initial Conditions	6-16
Figure 6.5	Baseline Conditions for Time = 1 day.	6-17
Figure 6.6	Baseline Conditions for Time = 1641 days.	6-18
Figure 6.7	Sensitivity Simulation 1 at Time = 1 day.	6-19
Figure 6.8	Sensitivity Simulation 1 at Time = 1641 days.	6-20
Figure 6.9	Sensitivity Simulation 2 at Time = 1 day.	6-21
Figure 6.10	Sensitivity Simulation 2 at Time = 1641 days.	6-22
Figure 6.11	Sensitivity Simulation 3 at Time = 1 day.	6-23
Figure 6.12	Sensitivity Simulation 3 at Time = 1641 days.	6-24
Figure 6.13	Sensitivity Simulation 4 at Time = 1 day.	6-25
Figure 6.14	Sensitivity Simulation 4 at Time = 1641 days.	6-26
Figure 6.15	Sensitivity Simulation 5 at Time = 1 day.	6-27

LIST OF FIGURES

	Page
Figure 6.16	Sensitivity Simulation 5 at Time = 1641 days. 6-28
Figure 6.17	Sensitivity Simulation 6 at Time = 1 day. 6-29
Figure 6.18	Sensitivity Simulation 6 at Time = 1641 days. 6-30
Figure 6.19	Sensitivity Simulation 7 at Time = 1 day. 6-31
Figure 6.20	Sensitivity Simulation 7 at Time = 1641 days. 6-32
Figure 6.21	Sensitivity Simulation 8 at Time = 1 day. 6-33
Figure 6.22	Sensitivity Simulation 8 at Time = 1641 days. 6-34
Figure 6.23	Sensitivity Simulation 9 at Time = 1 day. 6-35
Figure 6.24	Sensitivity Simulation 9 at Time = 1641 days. 6-36
Figure 6.25	Sensitivity Simulation 10 at Time = 1 day. 6-37
Figure 6.26	Sensitivity Simulation 10 at Time = 1641 days. 6-38
Figure 6.27	Sensitivity Simulation 11 at Time = 1 day. 6-39
Figure 6.28	Sensitivity Simulation 11 at Time = 1641 days. 6-40
Figure 6.29	Sensitivity Simulation 12 at Time = 1 day. 6-41
Figure 6.30	Sensitivity Simulation 12 at Time = 1641 days. 6-42
Figure 6.31	Sensitivity Simulation changing Vertical Discretization at Time = 1 day. 6-43
Figure 6.32	Sensitivity Simulation changing Vertical Discretization at Time = 1641 days. 6-44
Figure 6.33	Example of a Breakthrough Curve at Extraction Wells. 6-45
Figure 6.34	2000 Series Monitoring Wells. 6-46
Figure 6.35	3000 Series Monitoring Wells. 6-47
Figure 6.36	Extraction Wells. 6-48
Figure 6.37	Geoprobe Locations. 6-49
Figure 6.38	Slice of the VAM3DF Grid. 6-50
Figure 6.39	Initial Conditions Estimation #1 - Model Layer 5. 6-51
Figure 6.40	Initial Conditions Estimation #1 - Model Layer 8. 6-52
Figure 6.41	Initial Conditions Estimation #2 - Model Layer 5. 6-53
Figure 6.42	Initial Conditions Estimation #2 - Model Layer 8. 6-54
Figure 6.43	IC Estimation #2 Monitoring Well Residuals (1st control point). 6-55
Figure 6.44	IC Estimation #2 Monitoring Well Residuals (2nd control point). 6-56
Figure 6.45	IC Estimation #2 Extraction Well Residuals. 6-57
Figure 6.46	IC Estimation #2 Geoprobe Residuals. 6-58
Figure 6.47	Breakthrough Summary for Monitoring Well 2545. 6-59
Figure 6.48	Breakthrough Summary for Monitoring Well 3550. 6-60
Figure 6.49	Breakthrough Summary for Extraction Well 3924. 6-61
Figure 6.50	Breakthrough Summary for Extraction Well 31567. 6-62
Figure 6.51	Concentration Profile Summary at Geoprobe 12373E. 6-63
Figure 6.52	Concentration Profile Summary at Geoprobe 12442. 6-64
Figure 6.53	K_d Estimation #1 - Model Layer 5. 6-65
Figure 6.54	K_d Estimation #1 - Model Layer 8. 6-66
Figure 6.55	K_d Estimation #1 Monitoring Well Residuals (1st control point). 6-67
Figure 6.56	K_d Estimation #1 Monitoring Well Residuals (2nd control point). 6-68
Figure 6.57	K_d Estimation #1 Extraction Well Residuals. 6-69
Figure 6.58	K_d Estimation #1 Geoprobe Residuals. 6-70
Figure 6.59	K_d Estimation #2 - Model Layer 5. 6-71

000007

LIST OF FIGURES

	Page
Figure 6.60	K_d Estimation #2 - Model Layer 8. 6-72
Figure 6.61	K_d Estimation #2 Monitoring Well Residuals (1st control point). 6-73
Figure 6.62	K_d Estimation #2 Monitoring Well Residuals (2nd control point). 6-74
Figure 6.63	K_d Estimation #2 Extraction Well Residuals. 6-75
Figure 6.64	K_d Estimation #2 Geoprobe Residuals. 6-76

LIST OF TABLES

	Page
Table 5.1	Description of Test Problems 5-9
Table 5.2	Estimation of initial plume: summary of results 5-11
Table 5.3	Estimation of spatially random porosity field: summary of results (τ = correlation length) 5-12
Table 5.4	Estimation of spatially random K_d field: summary of results (τ = correlation length) 5-13
Table 5.5	Estimation of linear trend in log porosity: summary of results 5-14
Table 5.6	Estimation of linear trend in log K_d : summary of results 5-15
Table 5.7	Estimation of linear trend in longitudinal dispersivity α_L : summary of results 5-16
Table 5.8	Estimation of spatially random Freundlich K_d : summary of results 5-17
Table 6.1	Parameter Values 6-6
Table 6.2	Sensitivity Simulations 6-7
Table 6.3	Number of Measurements 6-8
Table 6.4	Residual Summary 6-9
Table 6.5	Estimated Trend Coefficients 6-10
Table 6.6	Sum-of-Squares Summary 6-11
Table 6.7	CPU Usage 6-12

000009

LIST OF ACRONYMS

ASCII	American Standard Code for Information Interchange
CDF	Cumulative distribution function
DFM	Data fusion modeling
EKFU	Extended Kalman filter update
FEMP	Fernald Environmental Management Corporation
GMA	Great Miami Aquifer
GSLIB	Geostatistical Software Library
IC	Initial concentrations (or initial conditions)
MAP	Maximun a posteriori
MRF	Markov random field
RM	Representer method
RMS	Root mean squared
SOS	Sum of squares
TVD	Total Variation Diminishing Solution Method
VAM3DF	Variably Saturated Analysis Model in Three-dimensions for the Data Fusion system

1.0 INTRODUCTION

1.1 BACKGROUND

HydroGeoLogic developed and verified a robust numerical flow and transport modeling code for Fernald Environmental Management Project (FEMP) as Phase I of work for the development of aquifer restoration management support tools. The work of Phase I involved customizing a HydroGeoLogic code to provide specific simulation capabilities for the simulation of groundwater flow and mass transport in the Great Miami Aquifer (GMA) at the Fernald Environmental Management Project (FEMP). Specific customization included: kinetic contaminant mass transfer between liquid and solid phases, a multi-dispersivity model, and the Total Variation Diminishing (TVD) solution to accommodate high groundwater velocity in the vicinity of the extraction/injection wells. The HydroGeoLogic code that was customized was the Variably Saturated Analysis Model in 3-Dimensions for the Data Fusion System (VAM3DF) (HydroGeoLogic, 1995). The goal was to provide a transport simulator that can handle Fernald Environmental Management Project (FEMP) specific conditions and is efficient enough for Data Fusion Modeling (DFM).

In Phase II, the integration of DFM with the VAM3DF contaminant transport code was accomplished for the calibration of VAM3DF transport parameters and the performance of transport prediction. It is anticipated that Phase III will provide tools to optimize monitoring and remediation and that Phase IV will provide technology transfer.

Phase II is the linchpin of support tool development because it is in this phase that all the relevant site information comes together. Phase II uses the VAM3DF code developed in Phase I and provides the basis for model predictions and quantified uncertainties that will be used in the tools provided by Phase III to optimize monitoring and remediation.

The DFM/VAM3DF system combines information from solute concentration time histories, the VAM3DF transport model, geostatistical variability, and prior knowledge of constant parameters. The system computes parameter estimates and model predictions and quantifies uncertainty in terms of estimate and prediction error covariances.

The DFM/VAM3DF system was developed using customized FORTRAN software. Testing and performance benchmarking has been performed to demonstrate functionality and to show that the software can be efficiently applied to contaminant transport modeling at the FEMP.

This document presents the software capabilities, input/output specification and test results.

1.2 PRODUCT DEFINITION AND CAPABILITIES

1.2.1 Operational Environment

It is anticipated that, once accepted, the integrated DFM/VAM3DF system will be part of a suite of Aquifer Restoration Management Support Tools. DFM/VAM3DF will be used to help monitor compliance and to provide the basis for optimization of remediation. The DFM/VAM3DF system

000011

will estimate transport parameters and provide model predictions for the fate and transport of chemicals in groundwater. Uncertainties will be quantified in terms of estimate and prediction error covariances. Based on experience in the groundwater community, quantification of uncertainty for parameters with substantial spatial variation requires geostatistical parameterization. Spatial variation is modeled as a random field variation about a trend where the trend is a function of constant parameter states. Parameters having less spatial variation can be modeled with a blocked parameterization such as a trend that is a function of constant parameter states.

The DFM/VAM3DF system is an integration and customization of FORTRAN programs. A graphical interface and input/output file structure is used. This will provide the starting point for Phase III development of an efficient user interface and post-processor for remediation optimization.

1.2.2 Capabilities

The DFM/VAM3DF software provides the capability to process monitoring data, to estimate transport parameters and predict future chemical concentrations in groundwater with quantified uncertainties. Information is combined from VAM3DF flow and transport models, geostatistical variability, prior statistical knowledge of constant parameters, and the monitoring data. A capability is provided to remove data outliers with statistical data editing. Further, calculation of data fit error uncertainties provides the capability to test model robustness and to tune statistical parameters to achieve model robustness.

Prior knowledge of geostatistical model parameters such as correlation distances and standard deviations (provided by a geostatistical software package or technical judgment) must be supplied to DFM/VAM3DF. Then the correlation distances and standard deviations can be manually adjusted in the DFM/VAM3DF system based on data and model fit errors. Prior knowledge of constant parameter standard deviations may be provided by technical judgment. Alternatively, the constant parameter may be viewed as fixed but unknown and estimated directly from the data when there are sufficient data.

The measured data include time histories of aqueous uranium concentration obtained from monitoring and extraction wells, and from geoprobe data. Measurement errors are assumed to be scale errors and multiplicative. They are transformed to additive by using the log of the measurement. The measurement error is modeled as a white noise.

Prior knowledge of the measurement noise standard deviation is obtained from technical judgment and separate analysis of the data. The measurement noise comes from instrumentation and recording errors and from small scale fluctuations in concentrations. The measurement noise standard deviation can be manually tuned in DFM/VAM3DF based on data fit errors. Further, automated data editing is provided based on excessive data fit errors to help prevent estimation errors due to data outliers.

The customized VAM3DF transport model from Phase I is used where it is assumed that flow is known. Since flow calibration uncertainties could be important, spatial variation in effective porosity is used to help account for flow uncertainties.

DFM/VAM3DF is currently designed to provide estimates for the following parameters:

- **Kinetic mass transfer parameters.** A Freundlich isotherm model is used with a distribution coefficient, K_d , and a desorption rate constant, α_d . Since spatial variation in K_d is important, a geostatistical model can be used with $\log_{10} K_d$ modeled as a random field with a polynomial trend. α_d can be modeled as a polynomial trend or a random field with a polynomial trend.
- **Effective Porosity.** Since an effective porosity is used to help account for flow uncertainties, a geostatistical model can be used with the \log_{10} (effective porosity) modeled as a random field with a polynomial trend that is a function of constant parameters.
- **Initial uranium solute concentration.** Since there is no time when initial concentrations (IC's) are known perfectly, a geostatistical model can be used for the initial solute concentration. The \log_{10} IC is modeled as a random field. The initial condition for uranium adsorbed mass fraction is assumed to be in equilibrium so it is a function of the IC and isotherm parameters. Initial chemisorbed mass fraction can be set to zero since the data are not likely to be available.
- **Dispersivities.** Dispersivities are modeled as polynomial trends that are functions of constant parameters.

All constant parameters in the above can be viewed as fixed but unknown parameters which is equivalent to viewing them as structural parameters in kriging. If insufficient data are available for their estimation, then they can be viewed as having Bayesian prior for estimation purposes. The Bayesian prior can come from technical judgment or separate data analysis.

A dual-grid approach is used with separate grids for the physical model and for the random field parameters. This allows the physical model grid to be selected based on the needs of accurate physical modeling and the parameter grid to be selected based on the level of spatial variability in the parameters. The nodes of the physical model grid must be inside the parameter grid, since parameter values are interpolated from the parameter grid onto the model grid. In other words, the physical model grid is used to provide numerical solutions for the transport equation, using the transport parameter values mapped from the parameters grid.

DFM/VAM3DF uses Gauss-Newton nonlinear iteration with backtracking for convergence. The Gauss-Newton step is obtained using the direct method, LSQR method, or representer method. Automated data editing is provided based on excessive data fit errors. Separate use of VAM3DF (with parameter estimates as input) is required for model predictions of uranium concentration and clean up time. The system can be initialized from prior estimates or as a restart from a previous run. During run time, output will be provided on nonlinear backtracking, values of the Gauss-Newton convergence criterion, and data editing so the user can intervene if necessary.

000013

DFM/VAM3DF can provide nonlinear batch estimation for full nonlinear accuracy. Alternatively, the extended Kalman filter provides quick monitoring updates with some degradation in nonlinear accuracy. The extended Kalman filter performs nonlinear estimation by doing Gauss-Newton relinearization only for the current measurements. The extended Kalman filter is initialized using nonlinear batch estimation.

DFM/VAM3DF can directly compute the uncertainty for selected parameters by performing covariance simulation. Model predictions are obtained by running VAM3DF with parameter estimates as inputs. Then uncertainties in model predictions are computed using a Monte Carlo method for covariance simulation.

2.0 DATA FUSION METHOD (DFM) THEORETICAL BACKGROUND

Estimation and prediction is based on combining information from physical models, statistical models, and measurement models. Physical models for flow and transport provide relationships between dependent states (such as contaminant concentrations) and parameter states described below.

Statistical models include models for spatial variation and constants. Spatial variation in parameter states such as kinetic mass transfer parameters can be modeled using a geostatistical Markov random field. Bayesian prior knowledge of constant parameter states such as trend parameters can be modeled using prior estimates and prior uncertainty. Since constant parameters have only one constant realization, the validity of prior knowledge of constant parameters cannot be tested. An alternative in DFM is to treat constant parameters as structural parameters with no prior knowledge except that they are fixed but unknown. The structural parameter treatment is consistent with the way constant parameters are treated in ordinary and universal kriging. Parameter states can include quantities such as kinetic mass transfer parameters, dispersivities, effective porosity, and initial conditions.

Calibration targets and measurements can include contaminant concentration time histories and direct parameter measurements. It is sometimes useful to transform the measurements to the log of the measurements so scale errors that are multiplicative become additive. Measurement error models are additive and reflect the measurement dynamics that can include measurement noise.

As explained by McLaughlin and Townley (1996), measurement errors are not just made up of instrumentation and recording errors but also include small scale fluctuations in the quantities being measured. A good approach for measurement error modeling is to do an audit of error contributors to at least identify the major contributors. Then the bias and trend parameters can be estimated using DFM state estimation and the noise standard deviation can be tested and tuned using DFM structural parameter estimation.

A dual-grid formulation is used throughout the DFM approach. The parameter grid is specified consistent with the need to represent spatial variability and the physical model grid is specified consistent with the need to represent physical relationships. Part of the modeling is the specification of interpolation from the parameter grid to the physical model grid consistent with modeling parameter assumptions. The dual-grid formulation provides the flexibility to honor the physical relationships and parameter variations while maintaining computational efficiency.

The DFM method uses Bayesian state estimation and Maximum Likelihood structural parameter estimation methods. Bayesian estimation is accomplished using Markov process and data equation methods from the DFM methods in Porter, Gibbs, Jones, Huyakorn, Hamm, and Flach (2000) combined with the representer method from physical oceanography in Bennett (1992). The computationally intensive part of the approach involves simulation that runs forward and backward in time where the number of simulations depends on modeling conditions. The key to application of this technology is the marriage of estimation and numerical modeling methods.

000015

The state estimation part of DFM solves the least squares problem examined by McLaughlin and Townley (1996) where DFM uses the somewhat more general model advocated by Bennett (1992). The paper by McLaughlin and Townley (1996) provides an excellent review of the least squares problem formulation and practical solution issues, and is a good companion to read with this report.

In practice, useful analysis can be done using technical judgment to provide prior structural parameters. Some of the structural parameters such as uncertainties for constant parameters can not be tested because only one realization of the constant parameter is observed. Other structural parameters such as the standard deviations of measurement noises and the standard deviations and correlation parameters (correlation distances) of spatial processes can be tested. Then analysis of residual model and data fit errors can be used to adjust prior structural parameters. This is the way many Bayesian estimation methods are successfully used. However, when there are sufficient data and general structural analysis methods are available, then structural analysis can improve the objectivity and accuracy of predictive modeling. Structural parameters that cannot be tested such as uncertainties for constant parameters can be eliminated by making the constant parameter itself a structural parameter. The remaining structural parameters can be tested and tuned.

The distinction between structural (bias, trend, standard deviation, and correlation distance) parameters and parameters that are the values of a spatially varying random field is at the heart of geostatistical approaches such as kriging (Kitinadis, 1997). Since DFM can estimate both parameter states and structural parameters, it can be viewed as a generalization of kriging to incorporate nonlinear physical models.

2.1 GENERAL THEORY

The following sections begin with a description of the model, including a model for spatial parameter variation using Markov processes. The next few sections provide key equations for state estimation methods, including the measurement equations, the penalty function formulation, the Gauss-Newton method, extended Kalman filter updates, and uncertainties.

2.1.1 Models

Since the models are ultimately used for computer implementations, they are discretized at the outset. The following sections describe physical, statistical, and measurement models.

2.1.1.1 Physical Model

The discretized transport equations are implemented in VAM3DF over time $l=1, 2, \dots, L$ for the dependent contaminant state vector c^l on the transport grid and for the independent parameter states θ on the parameter grid as:

$$A^l(c^l, \theta) c^l = B^{l-1}(\theta) c^{l-1} + D^{l-1}(\theta) c^l + g^{l-1}(c^{l-1}, \theta) + f^l(\theta), l = 1, 2, \dots, L \quad (2.1)$$

000010

Where A^l and B^{l-1} contain storage, advection-dispersion, and radioactive decay terms, D^{l-1} is the cross-component dispersion term, g^{l-1} is the kinetic mass transfer term, and f^l is the mass flux term.

In Equation (2.1), parameters are represented on the transport grid by linearly interpolating the parameter states from the parameter grid. The transport equations stacked over time are treated as hard constraints where the contaminant states c^l over time $l=1,2,\dots,L$ are dependent states and the independent parameter states are θ . In Equation (2.1), c^o is a direct function of independent parameter states θ for initial concentration (IC) parameters described in Section 1.2.2. Consequently, c^o is not treated as a dependent state.

Parameter states in θ can include kinetic mass transfer parameters, effective porosity, the initial concentration (IC) for uranium (or concentrations of other chemicals), and dispersivities. Spatial variation in the parameters can be modeled using geostatistical random field parameters and a polynomial trend model that is functions of constant parameters as described in Section 1.2.2. The random field model is specified by 3-dimensional correlation distances and a standard deviation for the process.

Bayesian prior knowledge about constant parameters is provided as prior estimates and prior estimate error covariances. The prior on the constant parameters is a subjective Bayesian prior that is useful to incorporate technical judgment, particularly when data are limited.

2.1.1.2 Statistical Model

Spatial variations of parameter states are modeled using geostatistical Markov process. Prior knowledge of constant parameter states is modeled using prior estimates and uncertainties.

A wide class of 2-dimensional and 3-dimensional spatial random fields can be described or approximated by the following first order process f

$$f - \nabla \cdot (C \nabla f) = w \quad (2.2)$$

where the matrix C is a function of the correlation distances and w is a zero mean white noise input with noise density q . The first order model is assumed here, but it readily generalizes to higher orders (see Yucel and Shumway (1996) for ideas on the use of higher order models in hydrology and geology). Equation (2.2) can be used to model spatial variation in parameters such as the kinetic mass transfer parameters.

A finite difference discretization of Equation (2.2) over a 3-dimensional curvilinear grid gives the discrete Markov random field

$$f_{i,j,k} = \sum_S a_{l,m,n} f_{i+l,j+m,k+n} + w_{i,j,k} \quad (2.3)$$

000017

where the a coefficients are defined over a region of support S that is the adjacent nodes and the integrated noise $w_{i,j,k}$ is a zero mean white noise process with variance $q_{i,j,k}$. It is assumed that the correlation distance is at least twice the grid spacing in each direction. Similar relationships exist on a 2-dimensional grid.

Large scale spatial variation is modeled by the Markov model of Equation (2.3) using only local equations. This property is one of the reasons that DFM achieves computational efficiency for geostatistical estimation.

The summation on the right hand side of Equation (2.3) is an interpolation for $f_{i,j,k}$ given f on the adjacent nodes and $w_{i,j,k}$ is the interpolation error. As will be seen later, state estimation selects states to minimize these interpolation errors in combination with other errors.

In order to keep relationships simple, the variances of the interpolation errors are normalized to the same value in Equation (2.3). This will be important later for estimation of the variance of the interpolation errors.

State estimation can begin with reasonable values for the standard deviations and correlation distances of parameters that have spatial variation. This provides the noise density and correlation distances needed by Equation (2.2) and the discretized model of Equation (2.3) which is then normalized. If there are sufficient data, then the structural parameters consisting of the variance of the interpolation error and the correlation distances can be adjusted.

The statistical model is the sum of the Markov random field Equation (2.3), and a polynomial trend. The coefficients of the polynomial trend are constant parameters θ_c . Prior knowledge about constant parameters θ_c is provided as a prior estimate $\hat{\theta}_c$ and prior estimate error covariance $P_{\hat{\theta}_c}$ that can be expressed as

$$\begin{aligned} \theta_c &= \hat{\theta}_c + \tilde{\theta}_c \\ E[\tilde{\theta}_c \tilde{\theta}_c^T] &= P_{\tilde{\theta}_c} \end{aligned} \tag{2.4}$$

where $\tilde{\theta}_c$ is the constant parameter estimate error. The prior covariance will be diagonal. Only one realization of constant parameters is observed so it is not possible to adjust the prior estimate error covariance. The prior on the constant parameters is a subjective Bayesian prior that is useful to incorporate technical judgment, particularly when data are limited. An alternative is to view the constant parameters as structural parameters that are fixed but unknown. This can be accomplished without changing the computer code by setting the prior estimate error variances to very large values to use the "uninformative prior" approach (see Porter (1979) to understand this idea in another application.). For groundwater this should be a well conditioned calculation since there will usually be only a small number of constant parameters.

The statistical model for the parameters in Equations (2.3) and (2.4) can be expressed as the following linear system:

$$L\theta = s + v_s \quad (2.5)$$

where f and θ_c parameters are stacked into the vector θ , L includes the coefficients and other relationships between parameters in (2.3) and (2.4), s includes the prior constant parameter estimates, and v_s is the marker interpolation error and constant parameter prior estimate error with covariance Q_s . The terms of v_s corresponding with random field variation are normalized to have unity variance for ease of testing model fit errors later.

2.1.1.3 Measurement Model

Time histories of solute concentration measurements are available from extraction wells, monitoring wells, and geoprobe data. The possibility is also left open for data that can be interpreted as direct measurements of parameters. Concentration measurements are modeled by linear interpolation from the transport grid and timeline to the measurement location and time. Initial condition (IC) concentration measurements are modeled by linear interpolation from the parameter grid since the IC's are parameters. Other direct measurements of parameters are also modeled by interpolation from the parameter grid. For processing, the actual measurements of concentration are taken to be the \log_{10} of the concentrations due to the possible spread in the concentration data. The measurement errors are assumed to be additive. The measurements are normalized by dividing by the standard deviation of the measurement noise. This produces measurement equations for concentrations and other direct parameter measurements of the form:

$$\begin{aligned} z_1 &= h_1(c^1, \dots, c^L) + v_{z1} \\ z_2 &= h_2(\theta) + v_{z2} \end{aligned} \quad (2.6)$$

where z_1 is a historical concentration measurement, h_1 is a linear function of the simulated concentrations, z_2 is a direct parameter measurement, h_2 is linear interpolation from the parameter grid, the concentration measurement noise v_{z1} has covariance R_1 , and the direct measurement noise v_{z2} has covariance R_2 . Since the measurements are normalized, R_1 and R_2 are actually identity matrices, but the more general notation is used so the theory is clear.

2.1.2 Penalty Function

State estimation is performed by selecting estimates that minimize model and data fit errors. This is accomplished by defining a penalty function that penalizes excessive model and data fit errors. Then state estimates are selected that minimize the penalty function.

The penalty function P is defined to be the sum-of-squares of the interpolation errors in spatial variation, prior estimate errors in constant parameters, and data fit errors. All of the errors are normalized by the expected statistical variation in the error to give P as

$$\begin{aligned}
P = & (s - L\theta)^T Q_s^{-1} (s - L\theta) + \\
& (z_1 - h_1(c^1(\theta), \dots, c^L(\theta)))^T R_1^{-1} (z_1 - h_1(c^1(\theta), \dots, c^L(\theta))) + \\
& (z_2 - h_2(\theta))^T R_2^{-1} (z_2 - h_2(\theta))
\end{aligned} \tag{2.7}$$

The first term penalizes interpolation errors in spatial variation and prior estimate errors in constant parameters. The second and third terms penalize data fit errors.

If the only statistical assumptions are that the first and second moments of statistical quantities are known, then the penalty function approach provides Bayesian least-squares estimates. Assuming estimate errors are linear perturbations within the expected statistical variation, the estimate error covariance for the Bayesian least-squares estimate can be computed using equations given later in this section. If the additional assumption is made that the statistical quantities are Gaussian, then the estimates are *Maximum A Posteriori* (MAP). By definition, MAP estimates produce the values of the states that have the highest probability of being true given the measured data. If statistical assumptions are dropped altogether, then the penalty function approach for estimates still makes sense. The covariances in the penalty function are reinterpreted as weighting coefficients in smoothness conditions on the estimates.

The penalty function is minimized by solving equivalent nonlinear regression problem with measurement equations:

$$\begin{aligned}
L\theta &= s + v_s \\
h_1(c^1(\theta), \dots, c^L(\theta)) &= z_1 - v_{z_1} \\
h_2(\theta) &= z_2 - v_{z_2}
\end{aligned} \tag{2.8}$$

The solution of the regression problem is obtained by minimizing a least squares cost function that is the same as the Bayesian penalty function. Further, perturbation analysis shows that the estimate error covariance is the same for the Bayesian estimation and regression problems.

2.1.3 Gauss-Newton Method

The regression problem is solved using the Gauss-Newton method where the Gauss-Newton step for iteration r is the solution to the linearized equations:

$$\begin{aligned}
L\delta\theta &= \delta s + v_s \\
\begin{bmatrix} J_1 \\ J_2 \end{bmatrix} \delta\theta &= \begin{bmatrix} \delta z_1 \\ \delta z_2 \end{bmatrix} - \begin{bmatrix} v_{z1} \\ v_{z2} \end{bmatrix}
\end{aligned} \tag{2.9}$$

where $\delta\theta$, δs , δz_1 , δz_2 , are perturbation quantities referenced to state estimates from the previous iteration $\hat{\theta}_{r-1}$ as:

$$\begin{aligned} \delta\theta &= \theta - \hat{\theta}_{r-1} \\ \delta s &= s - L\hat{\theta}_{r-1} \\ \delta z_1 &= z_1 - h_1(c^1(\hat{\theta}_{r-1}), \dots, c^L(\hat{\theta}_{r-1})) \\ \delta z_2 &= z_2 - h_2(\hat{\theta}_{r-1}) \end{aligned}$$

and where J_1 and J_2 are measurement sensitivity matrices for the measurement equations in Equation (2.6).

The partial derivatives in Equation (2.9) are evaluated at the parameter estimates from the previous iteration and the dependent states are computed by simulation of Equation (2.1) using the parameter estimates.

The adjoint equation given below is solved backwards in time for each measurement j to produce the j^{th} row of J_1 . Using (2.1) for the physical model and Equation (2.6) for the measurement model, the adjoint equation is given by:

$$\begin{aligned} (A^{l^T} + S_A^{l^T} + D^{l-1^T})\lambda_j^l &= (B^{l^T} + \frac{\partial g^l}{\partial c^l})\lambda_j^{l+1} + \frac{\partial h_y}{\partial c^l}, \quad l = L, L-1, \dots, 1 \\ \lambda_j^{L+1} &= 0 \\ k^{\text{th}} \text{ column of } S_A^l &= \frac{\partial A^l}{\partial c_k} c^l \end{aligned} \tag{2.10}$$

where the measurement model for the j^{th} measurement is denoted by h_y and where λ_j^l is the adjoint solution for the j^{th} measurement at time step l . Using Equation (2.1) again for the physical model, the element of J_1 for the j^{th} measurement in z_1 and i^{th} parameter becomes:

$$\begin{aligned} J_{1ji} &= - \sum_{l=1}^L \lambda_j^{l^T} R_i^l \\ R_i^l &= \frac{\partial A^l}{\partial \theta_i} c^l - \frac{\partial B^{l-1}}{\partial \theta_i} c^{l-1} - \frac{\partial D^{l-1}}{\partial \theta_i} c^l - \frac{\partial g^{l-1}}{\partial \theta_i} - \frac{\partial f^l}{\partial \theta_i}, \quad l = 2, \dots, L \\ R_i^1 &= \frac{\partial A^1}{\partial \theta_i} c^1 - \frac{\partial B^0}{\partial \theta_i} c^0 - B^0 \frac{\partial c^0}{\partial \theta_i} - \frac{\partial D^0}{\partial \theta_i} c^1 - \frac{\partial g^0}{\partial \theta_i} - \frac{\partial g^0}{\partial c^0} \frac{\partial c^0}{\partial \theta_i} - \frac{\partial f^1}{\partial \theta_i} \end{aligned} \tag{2.11}$$

Using Equation (2.6) with the measurement model for the j^{th} measurement in z_2 denoted by h_{2j} , the element of J_2 for the j^{th} measurement in z_2 and i^{th} parameter becomes:

$$J_{2ji} = \frac{\partial h_{2j}}{\partial \theta_i} \quad (2.12)$$

Equations (2.11) and (2.12) completely determine J_1 and J_2 . The computations are dominated by the backward simulations of Equation (2.10) for each measurement and the associated computations of Equation (2.11). Notice that the backward simulation of Equation (2.10) is zero until the time of the measurement, so the simulation essentially runs backward from the time of the measurement to the initial time.

The Gauss-Newton method enjoys rapid quadratic convergence locally. However, global convergence can be erratic. Consequently, modification is required to achieve reliable convergence. Backtracking is used because it is simple and effective. To implement backtracking, additional evaluations of the penalty function P of Equation (2.7) are required for each iteration as explained in Press (1992). Each evaluation of the penalty function requires simulation of Equation (2.1) to compute the dependent states $c^l(\theta)$. If P is not converging properly, then the step length is reduced.

The state estimates can be constrained in the nonlinear iterations to be physically realistic such as rate constants being non-negative. For backtracking, this can be done in an *ad hoc* manner by further reducing the step length until the constraints are satisfied.

As convergence nears completion and becomes quadratic, a convergence criterion is needed. A natural approach for Bayesian estimation is to stop iterating when the numerical optimization errors become small compared with the expected statistical estimation errors. Since convergence should be quadratic near completion, the Gauss-Newton step provides a measure of the numerical error. Consequently, the following is a good test for convergence based on the normalized convergence criterion C where:

$$C = \sqrt{\frac{1}{p} \delta \hat{\theta}^T P_{\delta \hat{\theta}}^{-1} \delta \hat{\theta}} = \sqrt{\frac{1}{p} \delta \hat{\theta}^T \left(L^T Q_s^{-1} \delta s + J^T \begin{bmatrix} R_1^{-1} & 0 \\ 0 & R_2^{-1} \end{bmatrix} \begin{bmatrix} \delta z_1 \\ \delta z_2 \end{bmatrix} \right)} \leq 0.1, \quad (2.13)$$

$$J = \begin{bmatrix} J_1 \\ J_2 \end{bmatrix}$$

where $P_{\delta \hat{\theta}}$ is the computed covariance of the state estimate errors $\delta \hat{\theta}$, and p is the dimension of θ . Satisfaction of Equation (2.13) says the numerical optimization errors are approximately 10%

of the statistical estimation errors. The second equality in Equation (2.13) avoids the need to explicitly compute the estimate error covariance.

The Gauss-Newton iteration can be initialized by the prior estimates for the parameter states. Alternatively, the iteration can be restarted from the end result of previous iterations. If different conceptual models are being hypothesized, it may be better to restart from the converged results of a somewhat different model than to go all the way back to the prior.

The Gauss-Newton step is obtained by solving (2.9) using the direct method, LSQR method or representer method. The direct method (Lawson and Hanson, 1974) uses Householder transformations on the least squares system Equation (2.9). The resulting upper triangular linear system is solved by backward substitution. The LSQR method (Paige and Saunders, 1982) provides an iterative solution to the least squares system Equation (2.9).

The representer solution of Bennett (1992) is given by

$$\delta \hat{\theta} = \delta \hat{\theta}_0 + Y \left(JY + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} (\delta z - J\delta \hat{\theta}_0) \tag{2.14}$$

$$\delta \hat{\theta}_0 = L^{-1} \delta s$$

where

$$\begin{aligned} LY &= Q_s \Lambda \\ L^T \Lambda &= J^T \end{aligned} \tag{2.15}$$

The inversion of the second equation in Equation (2.14) is performed efficiently using a standard iterative conjugate gradient type of algorithm. The second equation in Equation (2.15) is solved for Λ , and then the first equation is solved for Y . Y and Λ are intermediate matrices. The solutions are sought a column at a time where the number of columns is the same as the total number of scalar measurements. Each solution is performed efficiently using a standard conjugate gradient linear solver. As each column of Y is formed, the multiplication of Y by J in the first equation of Equation (2.14) is performed, and the column of Y is discarded to minimize storage. Y does not need to be saved for the other multiplication by Y in Equation (2.14) because there is a more efficient way to do this. The application of the inverse in Equation (2.14) is achieved using Cholesky methods. Then the results of applying the inverse are post multiplied to the equations in Equation (2.15). Solving the post multiplied equations produces the right-hand side of the first equation in (2.14) by solving two linear systems.

2.1.4 Extended Kalman Filter Updates

Monitoring updates can be performed efficiently using extended Kalman filtering ideas from Jazwinski (1970). The idea is to linearize new measurements in time around the best estimate at the current time. Old measurements from previous times are re-referenced to new estimates using the previous linearization. The extended Kalman filter assumes that the current measurements are the most important for estimating what is desired.

Assume that measurements at previous times have been processed to produce a current estimate. Now assume new measurements become available at the current time. The terms in the measurement model in the second equation in Equation (2.9) that are relative to the current time are linearized around the current estimate as before. However, the terms that are relative to previous times are simply re-referenced to the current estimate using the previous linearization. The assumption is that the previous linearization is still good at new estimates. Also, the terms in the statistical model represented by the first equation in Equation (2.9) are re-referenced to the current estimate since the statistical model is linear. This means that the left hand side of Equation (2.9) that is relative to previous times and provides statistical models remains unchanged, but the right hand side δs , δz terms do change.

Now perform a nonlinear iteration about the new measurements as before. Notice that the backward simulations of Equation (2.10) and the associated computations of Equations (2.11) and (2.12) are only performed for the new measurements for each iteration. Also, notice that the columns of Y in Equation (2.15) for previous measurements remain unchanged. If the new measurements are appended at the end of the old measurements, then the Cholesky factors of the matrix being inverted in Equation (2.14) can be computed by starting with the Cholesky factors from the previous measurements.

2.1.5 Tests of Model Robustness

A powerful test for identifying data and model problems is based on the minimized penalty function that is the sum-of-squares (SOS) of the normalized model fit and data fit errors

$$\begin{aligned} \text{SOS} = & (s - L\hat{\theta})^T Q_s^{-1} (s - L\hat{\theta}) + \\ & (z_1 - h_1(c^1(\hat{\theta}), \dots, c^L(\hat{\theta})))^T R_1^{-1} (z_1 - h_1(c^1(\hat{\theta}), \dots, c^L(\hat{\theta}))) + \\ & (z_2 - h_2(\hat{\theta}))^T R_2^{-1} (z_2 - h_2(\hat{\theta})) \end{aligned} \quad (2.16)$$

The assumption is made that the estimation errors are linear perturbations within the expected statistical variation. SOS should have an expected value equal to the number of scalar measurements m when all constant parameters have Bayesian prior or where is the number of constant parameters that are treated as fixed but unknown parameters (see Porter (1979) for an explanation of the distribution of residuals for uninformative prior and structural parameters). A value of SOS that is very different from its expected value fails the test.

If the data fit errors are Gaussian, then the test can be refined. A good approach is to test the data fit errors to see if they are Gaussian. Even non-Gaussian model and data noise can produce nearly Gaussian data fit errors if there are sufficient local linear operations on the data to produce the estimates. If the data fit errors pass the Gaussian test, then SOS should be distributed as a chi-squared random variable with degrees of freedom k equal to m when all constant parameters have Bayesian prior or $m - P_c$ where P_c is the number of constant parameters that are treated as fixed but unknown parameters. Then the variance of SOS should be $2k$. For k larger than 20, the chi-squared distribution is approximately Gaussian so a good test is to reject model validity if SOS is more than five standard deviations from its mean according to

$$SOS \geq k + 5\sqrt{2k} \text{ or } \leq k - 5\sqrt{2k} \rightarrow \text{Reject model validity} \quad (2.17)$$

If problems are identified using SOS, then a test that helps to isolate data and model problems is to compare the data fit error for each measurement to the standard deviation of the data fit error. The covariance of the data fit error for the representer method is:

$$E \begin{bmatrix} z_1 - h_1(c^1(\hat{\theta}), \dots, c^L(\hat{\theta})) \\ z_2 - h_2(\hat{\theta}) \end{bmatrix} \begin{bmatrix} z_1 - h_1(c^1(\hat{\theta}), \dots, c^L(\hat{\theta})) \\ z_2 - h_2(\hat{\theta}) \end{bmatrix}^T = \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} (JY + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix})^{-1} \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \quad (2.18)$$

But the Cholesky factors of the quantity being inverted have already been computed for the representer solution of Equation (2.14). Since R is diagonal, the covariance of the data fit error can be computed efficiently. The square roots of the diagonal elements of Equation (2.18) provide the standard deviations of the data fit errors. Places where the magnitudes of the data fit errors are much larger than the standard deviations can help indicate the source of problems.

The uncertainty of data fit errors can also be used as the basis of data editing. The diagonals of Equation (2.18) can be shown to be less than the variances of the measurement errors. A conservative data editing approach is to normalize the magnitudes of the residual data fit errors by the square roots of the diagonals of R and edit the data for which the normalized residual error is excessive.

The model fit errors can also help to isolate the sources of problems. The standard deviations from the diagonal elements of Q_s are upper bounds for the standard deviations for the model fit errors where the model fit errors are:

$$s - L\hat{\theta} \quad (2.19)$$

Places where the model fit errors are much larger than the standard deviation from the diagonal of Q_s can help to indicate the source of problems.

2.1.6 Uncertainties

The parameter estimate error covariance for the representer method is given by:

$$E(\theta - \hat{\theta})(\theta - \hat{\theta})^T = P_0 - Y \left(JY + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} Y^T \quad (2.20)$$

Selected diagonal and off diagonal elements of the covariance can be computed. As the representer matrix Y is being computed for the representer solution, elements needed to compute selected covariances are saved. Since the Cholesky factors of $JY+R$ are already computed, the selected elements of the right hand term in Equation (2.20) can be efficiently computed. P_0 is the prior covariance of the parameters. The selected elements of P_0 are available from the prior spatial autocorrelation functions and from the prior constant estimate error covariance.

Monte-Carlo methods are used to compute prediction error uncertainties for the representer method. The following provides Monte Carlo equations for the parameter estimate errors:

$$\theta - \hat{\theta} = L^{-1}v_s - Y \left(JY + \begin{bmatrix} R_1 & 0 \\ 0 & R_2 \end{bmatrix} \right)^{-1} \left(\begin{bmatrix} v_{z1} \\ v_{z2} \end{bmatrix} + JL^{-1}v_s \right) \quad (2.21)$$

The noise terms v_s, v_{z1}, v_{z2} can be produced with random number generators with covariances Q_s, R_1, R_2 , respectively. The Cholesky factors of the quantity being inverted are available from the representer solution.

The parameter estimate errors from the Monte Carlo runs are added to the parameter estimates to produce conditioned realizations of the parameters. Then these parameter realizations are used to simulate the transport model predictions. Then differences between the mean prediction and the prediction realization produces prediction errors. Statistics computed from prediction errors such as standard deviations quantify prediction uncertainty.

3.0 IMPLEMENTATION

3.1 SOFTWARE STRUCTURE

The software flow chart in Figure 3.1 summarizes the DFM/VAM3DF implementation. Figure 3.2 shows the Gauss-Newton iterative procedure. For each iteration, a VAM3DF transport simulation is performed, the least squares system is constructed, and the least squares system is solved. Construction of the least squares system requires an adjoint transport simulation (backwards in time) for each concentration measurement. Since the adjoint transport simulation requires more computations than the VAM3DF transport simulation, the number of concentration measurements will have a significant impact on software run-times. For example, if the VAM3DF transport simulation uses one time unit of CPU then the adjoint transport simulation will use approximately one time unit of CPU for each concentration measurement. If the DFM/VAM3DF run has 100 concentration measurements, then each iteration will use approximately 101 time units of CPU for the VAM3DF transport simulation and the 100 adjoint transport simulations. A time unit is dependent on source factors including: problem size, number of time steps, and computational speed of the computers. Figure 3.3 summarizes the post Gauss-Newton computational procedure. The covariance and Monte Carlo computations depend on the type of least squares solver used. The representer method allows computation of the data fit error covariances, parameter estimate error covariances, and Monte Carlo realizations. The direct method computes the parameter estimate error covariance matrix. The LSQR method estimates the parameter estimate error variances.

3.2 OVERVIEW OF SOFTWARE APPLICATION

Modeling input is needed for the physical models, statistical models, and measured data, and control input is needed to select functions, for Gauss-Newton iterations, and for the least squares solvers. Output is provided during run time of batch estimation and monitoring updates to check on convergence. At the completion of estimation and update iterations, output can be provided to test robustness to help identify and isolate places where the models can be improved. Selected parameter estimate error covariances and prediction error covariances can be output based on covariance simulation. Covariance simulation can be performed for data that has been processed to quantify uncertainty. Covariance simulation can also be performed for possible future data before it is acquired to help make decisions to optimize monitoring.

The following functions are performed by DFM/VAM3DF:

- Nonlinear batch estimation
- Monitoring updates using extended Kalman filter
- Testing robustness
- Parameter estimate error covariances
- Prediction and prediction error covariance simulation

3.2.1 Nonlinear Batch Estimation

Estimation begins with a conceptual understanding of what is desired to be estimated and what models and data are needed to perform the estimation. Then model and data information are gathered for inputs and control inputs are selected.

A physical model grid is selected that will allow the physical relationships between quantities to be estimated, parameters, and data to be adequately represented without excessive simulation computations. A parameter grid is selected that allows adequate representation of spatial variability without an excessive number of parameters. The parameter grid dimension should be more than a few times the longest random field correlation distance and the shortest correlation distance should be at least twice the grid spacing. The physical model grid must be contained inside the parameter grid.

Known physical model inputs are provided and parameters are specified to be estimated. Parameters can be modeled as a random field plus a polynomial trend for heterogeneous variation or just as a trend for smooth variation. For the 3-dimensional random fields, three correlation distances (x, y, and z directions) and the standard deviations are specified. These parameters are modified when testing model robustness. The trend models are functions of constant parameters. If there are adequate data, then the constant parameters can be viewed as fixed but unknown. This is accomplished by setting the prior standard deviation on the parameter to a large number. It is only necessary to set this number two orders of magnitude larger than the largest value of the parameter that is considered possible in order to avoid any conditioning problems from too large a number. If there are not adequate data to estimate all the fixed but unknown parameters, then a Bayesian prior estimate and estimate error covariance can be specified based on technical judgment.

Measurement model and measured data inputs are provided for time histories of solute concentration measurements from extraction wells, monitoring wells, and geoprobe data. The possibility is also left open for input of data that can be interpreted as direct measurements of parameters to further constrain the estimation. For processing, the actual measurements of concentration are taken to be the \log_{10} of the concentrations due to the possibly large spread in the concentration data. The measurement errors are assumed to be additive. The measurements are normalized by dividing by the standard deviation of the measurement noise. This is helpful later for testing data fit errors.

The concentration measurement locations, times, measurement noise standard deviations, and data editing threshold are specified. The measurement noise includes instrumentation and recording errors and includes small-scale fluctuations in the concentrations from modeling errors. The noise standard deviation is initially specified using technical judgment and separate analysis of the data. Then it can be modified during testing of model robustness.

The measurements of initial concentrations do not contain small-scale fluctuation errors from the transport model since the IC's are parameters and not computed by the transport model. Consequently, it may be appropriate to use a smaller standard deviation for noise on the IC measurements. Also, small-scale fluctuation errors may be location dependent. For example,

fluctuations due to the yearly wet and dry cycle might be more pronounced near Paddy's Run and the storm sewer outfall ditch so a larger noise standard deviation might be appropriate. The prior standard deviations can be selected and modified differently for measurements in different regions.

Control input is specified for the nonlinear iteration such as the threshold value of the convergence criterion, maximum number of iterations and least squares solver. Standard control input must be provided for the VAM3DF simulation of dependent contaminant states which are needed at the beginning of each iteration and for each backtracking step.

During run time, output is provided to check on convergence. The following are shown in the output:

- Backtracking Steps
- Convergence criterion
- Value of the penalty function
- Edited data

Normal convergence usually begins with some backtracking. Then it settles into quadratic convergence until the convergence criterion is satisfied. If convergence is unsuccessful there may be problems with the data or the model and the run time output may suggest where the problem is.

At run completion, the penalty function is broken down into components from data fit errors, random field interpolation errors, and constant parameter prior estimate errors. The data fit errors are output and summarized and the parameter estimates are output.

3.2.2 Estimation Updates Using Extended Kalman Filter

The input and interpretation of the output for estimation updates is similar to batch estimation. The updates are initialized with a batch run for a large enough data set to sufficiently reduce nonlinear errors. The subsequent updating is faster than a full nonlinear batch run, but there is some loss in nonlinear accuracy since only the new measurements are relinearized during nonlinear iterations. It is assumed that past estimation has produced adequate linearization of past measurements so the initial batch run must obtain a good linearization for the first set of data.

For updates, the physical and statistical models have already been set up and past data have already been processed. The new measurement models and measured data need to be input. The same kind of control inputs as for batch estimation are required. The same kind of output subject to the same kind of interpretation is provided.

3.2.3 Testing of Model's Robustness

A test that can be performed to help isolate data and model problems is to compare the data fit error for each measurement to the standard deviation of the data fit error. Locations and times where the magnitudes of the data fit errors are much larger than the standard deviations can help indicate the source of problems.

The statistical model fit errors can also be used to test for the sources of problems. The model fit errors are the interpolation errors for the random field models and the prior constant parameter estimate errors. The model fit errors can be computed and compared with an upper bound on the model fit error standard deviation. Locations where the model fit errors are much larger than the standard deviation help to isolate the source of problems. For example, the location of large interpolation errors for a random field indicates the need for larger spatial variation in the model or possibly a spatial discontinuity.

3.2.4 Parameter Estimate Error Covariance

Selected diagonal and off diagonal elements of the parameter estimate error covariance matrix can be computed. The input is similar to what is required for one iteration of nonlinear batch estimation or a monitoring update. The best estimate from previous estimation or prior information is input for the initial estimate. Also, the diagonal and off-diagonal elements that are to be computed are specified.

3.2.5 Prediction and Prediction Error Covariance Simulation

Model predictions and prediction error covariances can be computed by Monte Carlo analysis based on VAM3DF simulation. Standard inputs for the VAM3DF simulation for the prediction conditions are required. DFM/VAM3DF computes Monte Carlo realizations of parameter estimates. The realizations are input to VAM3DF transport model to produce Monte Carlo realizations of prediction. Prediction error statistics such as error covariances can be computed from the prediction realizations.

Software Structure

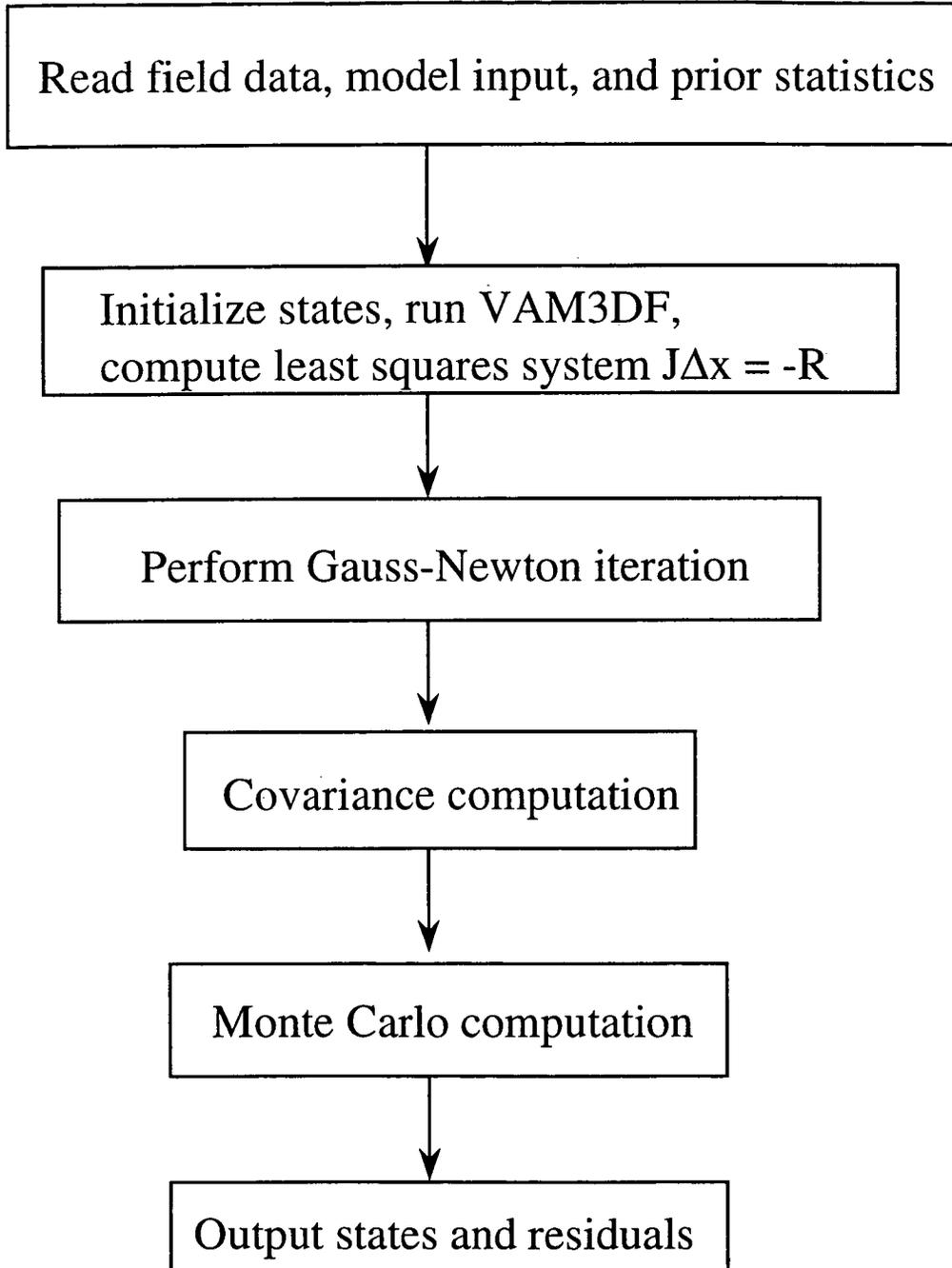


Figure 3.1 Software Structure.

Gauss-Newton Iterative Procedure

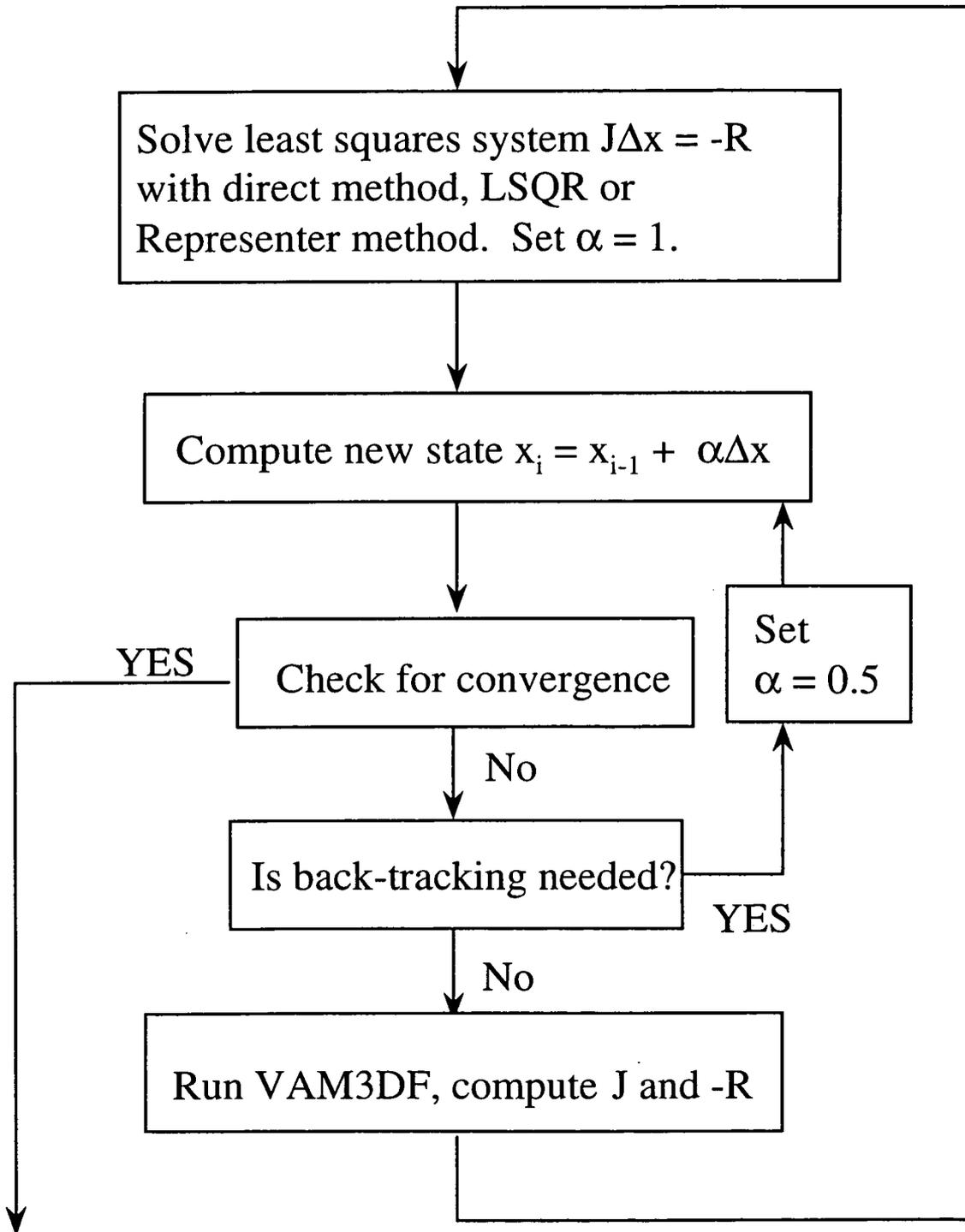


Figure 3.2 Gauss-Newton Iterative Procedure.

Post Gauss-Newton Computational Steps

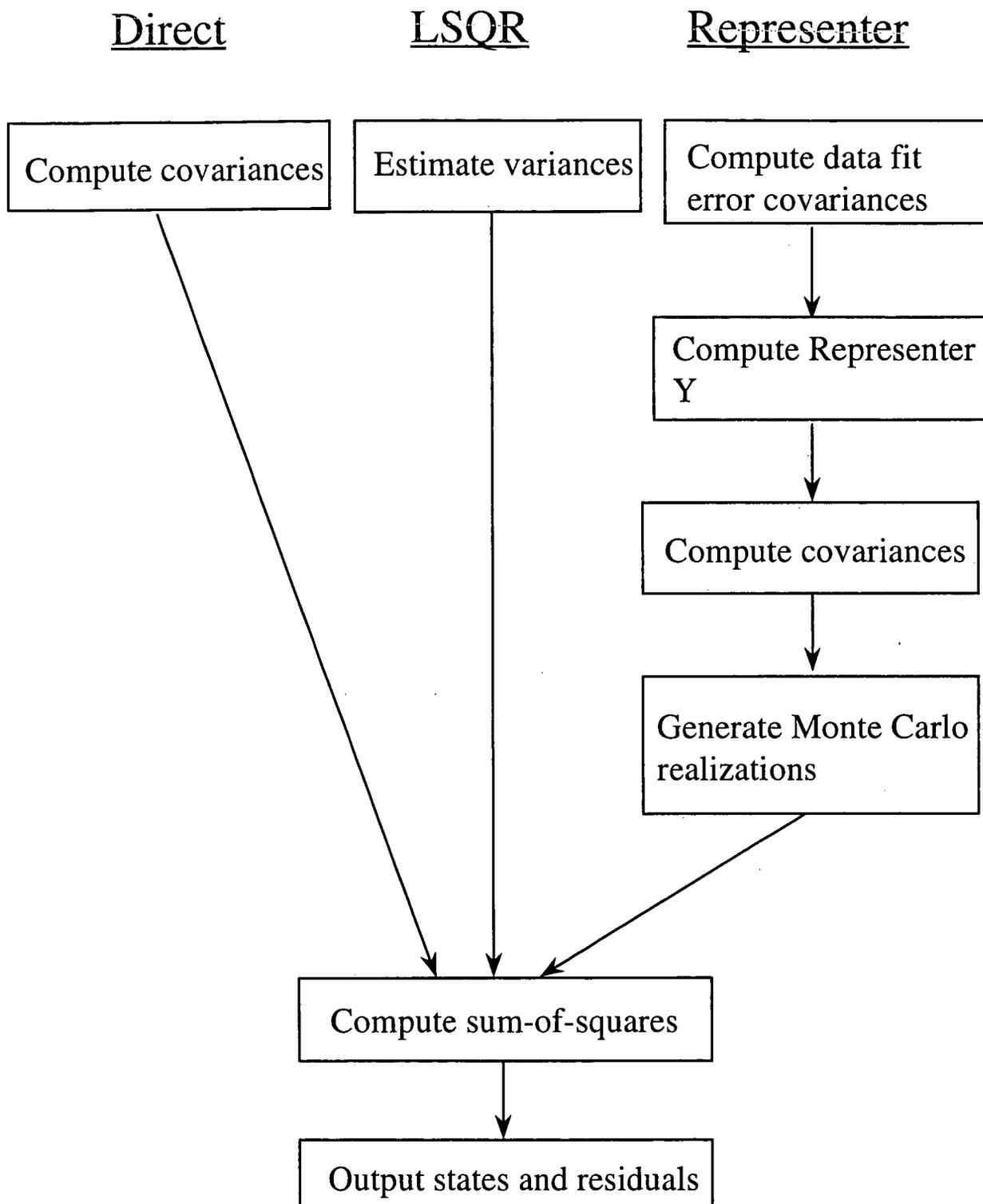


Figure 3.3 Post Gauss-Newton Computational Steps.

000033

4.0 DFM/VAM3DF USER'S GUIDE

The DFM/VAM3DF software requires the user to construct an ASCII input file for DFM/VAM3DF and an ASCII input file for VAM3DF. The software has a windows interface for starting, monitoring, and stopping the DFM/VAM3DF run. The software writes an ASCII output file for DFM/VAM3DF and an ASCII output file for VAM3DF. In addition, an ASCII output file is written for post-processing purposes. The software is Windows 95/NT compatible.

4.1 DFM/VAM3DF INPUT

4.1.1 Mandatory Input Files

The software requires the following three ASCII input files in one working directory:

Dfmvam.in uses a keyword format to specify control input, parameter grid input, and field data input from direct measurements of parameters and concentration data input from extraction wells, monitoring wells, and Geoprobe data. An example input file with a detailed input explanation is given in Appendix A.

Run.in lists the input/output file names used by VAM3DF. Below is an example input file.

```
vam3df.in    -- (or UserDefined - see below)
vam3df.out
vam3df.vel
vam3df.nod
vam3df.ele
vam3df.flx
```

UserDefined lists the standard VAM3DF input with control input, physical model grid, boundary conditions and initial conditions. This user-defined input file name must appear on the first line of **run.in**.

The VAM3DF simulation may require additional input for the Darcy velocities (with the file name listed in **run.in**) See details below.

4.1.2 Optional Input Files

The software allows for the following two optional input files in the working directory:

UserDefined is a binary file containing restart information from a previous DFM/VAM3DF simulation with the same parameters estimated and same concentration data. This file is used if the RESTART option is specified in **Dfmvam.in**.

Dfmvam.sav is a binary file containing bookkeeping information from a previous DFM/VAM3DF simulation with the same parameter grid and physical model grid. Reading the bookkeeping information from **Dfmvam.sav** reduces the start-up

000034

computations. NOTE: there is not a flag for this option, the software will use any **Dfmvam.sav** in the working directory.

4.1.4 DFM/VAM3DF Input Details

The technical details needed to specify the input file **Dfmvam.in** are discussed in this section. A keyword format is used to specify control input, parameter grid input, and field data input. Detailed input explanation is given in Appendix A.

Parameter Grid

Keywords: XGRID, YGRID, ZGRID, ZGRID3D, SITE2MODEL

The parameter grid for the random field must include the VAM3DF numerical grid. The parameter grid is based on the level of spatial variability in the parameters. In each coordinate direction, the correlation length should be at least twice the grid spacing. The keywords XGRID and YGRID define a non-uniform (x,y) grid with (0,0) at the lower-left corner. The keyword ZGRID can be used to define a grid with flat layers with zero elevation on the bottom layer. Another option is to use the keyword ZGRID3D to define a grid with non-uniform layers. The keyword SITE2MODEL defines the coordinate transformation from the VAM3DF grid to the parameter grid.

Parameter Estimate Control

Keywords: NRFNTRD, PARAMS

The keyword NRFNTRD defines the number parameters estimated by random field with a polynomial trend and the number of parameters estimated by a polynomial trend. The keyword PARAMS specifies which parameters are to be estimated (i.e., distribution coefficient (K_d), effective porosity, initial concentrations, α_d , dispersivities). DFM/VAM3DF estimates K_d , effective porosity, initial concentrations in the \log_{10} domain.

Measurements

Keywords: FIELDDATA, CONCDATA, SCALECONC, NSCREENPTS

DFM/VAM3DF fits direct measurements of the parameter being estimated and historical concentration measurements. The keyword FIELDDATA defines direct measurements, while the keyword CONCDATA defines historical concentration measurements. If the parameter being estimated is in the \log_{10} domain the direct measurement must be in the \log_{10} domain. The concentration measurements must be given in the \log_{10} domain. The keyword SCALECONC allows for the concentration measurements to be converted to the units in the VAM3DF model (i.e., $\mu\text{g/L}$ to nano lb/ft³). When the initial concentrations are being estimated, the keyword SCALECONC will also scale the direct measurement data.

The historical concentration measurements are input with a specified top of screen elevation and bottom of screen elevation. When the top and bottom screen elevations are given as the same elevation, the concentration measurement is treated as a point measurement (i.e. Geoprobe data). Otherwise, the keyword NSCREENPTS is used to specify the number of points along the well screen used to approximate the average well concentration.

All measurements are input with an error or noise standard deviation. The historical concentration error standard deviations are in the log₁₀ domain. The error standard deviations are obtained from technical judgment and separate analysis of the data.

Parameter Geostatistics

Keywords: HYDROUNITSi, VARIOGRAMi, TRENDi

A geostatistical model for the three-dimensional spatial variation may be used for each hydrostratigraphic unit (grouped by layers of the parameter grid) and for each parameter estimated. The keyword HYDROUNITSi specifies the layers in each hydrostratigraphic unit for the ith parameter estimated. Correlated spatial variability is modeled by the sum of a spatial polynomial trend and a Markov Random Field (MRF). The random variation about the polynomial trend is modeled as an anisotropic first-order autoregression with an exponential autocorrelation function. For an isotropic model we have

$$E \left[u(x_1, y_1, z_1) - \bar{u}(x_1, y_1, z_1) \right] \left[u(x_2, y_2, z_2) - \bar{u}(x_2, y_2, z_2) \right] \approx \sigma^2 \exp \left[- \frac{\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}}{\tau} \right] \tag{4.1}$$

Where u is the parameter estimate, \bar{u} is the polynomial trend, σ is the user defined standard deviation of the MRF, and τ is the user defined correlation length. The keyword VARIOGRAMi specifies the standard deviation and correlation lengths for the ith parameter estimated. The keyword TRENDi specifies the initial polynomial trend coefficients and their prior error standard deviations for the ith parameter estimated. The maximum allowed polynomial trend in DFM/VAM3DF is a partial 4-th order polynomial (15 coefficients), but in most cases a simple linear trend ($C_0 + C_1x + C_2y + C_3z$) is used. The polynomial trend is a function of normalized coordinates (x,y,z) given by:

$$x = (x - x_0) / \Delta x, \quad y = (y - y_0) / \Delta y, \quad z = (z - z_0) / \Delta z, \tag{4.2}$$

where (x,y,z) are VAM3DF grid coordinates, (x₀,y₀,z₀) is the center of the VAM3DF grid, and (Δx,Δy,Δz) are reference distances.

Measurement Editing

Keyword: EDITTHRES

The Gauss-Newton iteration includes an automated data editing option. Data editing occurs when a measurement residual is several times larger than the prior error standard deviation and DFM/VAM3DF removes the measurement from the cost function. The data editing test is given by:

$$\left| \frac{\bar{z}_K}{w_K} \right| > \alpha \cdot RMS_{i-1} \quad (4.3)$$

where \bar{z}_k is the K^{th} measurement residual, w_k is the specified measurement noise standard deviation of the K^{th} measurement, and $RMS_{i-1} = \sqrt{\frac{1}{n} \sum_{k=1}^n \left(\frac{\bar{z}_k}{w_k} \right)^2}$ for all n measurement residuals at Gauss-Newton iteration $i-1$.

If the absolute value of the normalized residual exceeds a user specified threshold (α) times the residual root-mean-squared from the previous Gauss-Newton iteration, then the measurement is edited. The keyword EDITTHRES defines the data editing threshold (α). Setting the threshold to a large value (i.e., 100) will disable the data editing option.

Least Squares Solvers

Keyword: SOLVER

The keyword SOLVER defines the least squares solution method. The direct solver uses Householder transformations followed by backward substitution to solve the least squares system and compute the estimate error covariance matrix (Lawson and Hanson, 1974). The LSQR solver (Paige and Saunders, 1982) provides an efficient solution of the least squares system and approximates the estimate error variances. The representer method (Bennett, 1992) computes selected estimate error covariances and parameter Monte Carlo realizations.

Gauss-Newton Iteration Control

Keywords: NITER, CONVTEST

The keyword NITER defines the maximum number of Gauss-Newton iterations. The keyword CONVTEST defines the convergence criterion.

Covariance Output

Keyword: COVARIANCE

The keyword COVARIANCE is used to specify the estimate error covariance output.

Monte Carlo Output

Keyword: MONTECARLO

The keyword MONTECARLO specifies the number of output realizations and the negative integer seed for random number generation.

Histogram Plots

Keyword: HISTOGRAM

The keyword HISTOGRAM specifies the number of histogram plots, the number of bars on each plot, and the title, parameter, and location for each histogram plot. This option must be used with the MONTECARLO option.

Monte Carlo Runs

Keyword: MONTECRUN

The keyword MONTECRUN is used to run VAM3DF for each parameter realization, which was computed by a previous run using the MONTECARLO option. When this option is used, the Gauss-Newton iteration is not performed. For a given output time, this option produces the concentration variance at each node of the VAM3DF grid, the maximum concentration histogram, and the maximum concentration cumulative distribution function.

Restart Option

Keyword: RESTART

During each Gauss-Newton iteration, DFM/VAM3DF writes a binary restart file containing current estimates and Jacobian terms for the historical concentration measurements. The keyword RESTART informs DFM/VAM3DF to read a restart file from a previous run with the same states and the same number of historical concentration measurements. If the restart option is not used, then the prior trend is used to initialize parameter estimates and the full Jacobian is constructed for the first iteration.

Extended Kalman Filter Update

Keywords: EKFUNDATE, FIELDDATA, CONCDATAE

When new data is acquired, parameter estimates can be updated without repeating the batch estimation. Both direct measurements and historical concentration measurements can be used to in the update. The keyword EKFUPDATE is used to specify the extended Kalman filter update. The extended Kalman filter update requires using the LSQR solver or representer method and is normally used with the restart option. The direct measurement updates are put in the FIELDATA section, while the keyword CONCDATAE defines historical concentration measurement updates.

Chemsorption for Linear K_d

Keyword: CHEM4LINRKD

This option is used when estimating K_d without kinetic mass transfer. The keyword CHEM4LINRKD specifies α_c , the chemisorption mass transfer rate for contaminant from adsorbed state to a bonded chemisorbed state. Chemisorption is modeled without kinetics by defining the first order decay rate:

$$\lambda = \frac{\alpha_c}{\left(1 + \frac{\phi S_w}{\rho_b k_d}\right)}$$

4.2 DFM/VAM3DF EXECUTION

The Windows interface has the following options:

Starting DFM/VAM3DF

From Windows Explorer, double-click on the dfmvam executable icon. Then from the Dfmvam Window, click on File/Start Dfmvam.

Pausing DFM/VAM3DF during execution

To pause execution and free-up CPU usage for other applications, click on State/Pause. To resume execution click on State/Resume.

Stopping DFM/VAM3DF during execution

From the Dfmvam Window, click on File/Stop Dfmvam. Since software checks for the 'Stop Dfmvam' command after each major computation unit, there may be a delay before the execution is stopped. For a quicker termination, click on File/User Exit.

Stopping DFM/VAM3DF after execution

From the Dfmvam Window, click on File/User Exit.

The Windows interface contains the following two subwindows during executions:

- 1) DFM/VAM3DF run summary window shows the convergence test and cost function for each Gauss-Newton iteration. After the nonlinear iterations, the window shows the trend coefficients estimates, trend coefficients error standard deviations, field data average error and error root-mean-squared, and concentration data average error and error root-mean-squared. The window also shows the spatial variability sum-of-squares, trend polynomial sum-of-squares, field data sum-of-squares, and concentration data sum-of-squares.
- 2) Graphic 1 window shows the progress of the VAM3DF transport simulations and the adjoint backward simulations.

The status bar at the bottom of the Window interface shows the current task

4.3 DFM/VAM3DF OUTPUT

The software writes the following three ASCII files in the working directory:

Dfmvam.out is a run summary file containing the convergence test and cost function for each Gauss-Newton iteration. After the nonlinear iterations, the file contains the trend coefficients estimates, trend coefficients error standard deviations, field data average error and error root-mean-squared, and concentration data average error and error root-mean-squared. The file also contains the spatial variability sum-of-squares, trend polynomial sum-of-squares, field data sum-of-squares, and concentration data sum-of-squares.

UserDefined lists the standard VAM3DF output. This user defined output file name must appear in **run.in**.

Dfmvam.tec is a Tecplot input file containing the parameter estimates on the parameter grid, field data estimates and residuals, and concentration data estimates and residuals (all in physical model coordinates).

Dfmvam.rst is a binary file containing restart information.

Dfmvam.sav is a binary file containing bookkeeping information.

Dfmvammc.tec is a Tecplot input file containing the parameter Monte Carlo realizations (all in physical model coordinates).

Dfmvammc.bin is a binary output file containing the random field parameter monte Carlo realizations.

Dfmvammc.max is an ASCII output file containing the minimum and maximum parameter value ad location.

5.0 VERIFICATION: SMALL-SCALE TESTING

5.1 INTRODUCTION

Small-scale testing was conducted to verify the functionalities of the DFM/VAM3DF code using synthetic datasets. The small-scale tests included as many of the features of the full-scale system as possible so that the results are representative of the full-scale system. Synthetic data were generated with the VAM3DF model. A dual-grid approach was used with different grids for the physical and statistical models. Two-dimensional and three-dimensional small-scale system cases were considered where direct measurements and concentration measurements were taken at specified locations.

The objectives of these tests were to verify the capability of the DFM/VAM3DF model in estimating random spatial variability and pure trend, to examine the effect of number of direct parameter measurements and concentration measurements, on the estimates, and to illustrate the impact that measurement errors and correlation lengths have on the prediction of concentration distributions.

5.2 TESTING RANDOM FIELD GENERATION

Correlated spatial variability was modeled by DFM/VAM3DF using variables that were the sum of a spatial polynomial trend and a first-order Markov Random Field (MRF). The testing of the generation of spatially correlated random field by the DFM/VAM3DF model was first conducted. In this test, the focus was on the generation of first-order MRFs, since the generation of the polynomial trend is trivial. The objective was to test whether the generated MRFs preserve their underlying statistical structure or pre-specified correlation functions or variograms.

In the DFM/VAM3DF code, the spatial variability model has a correlation function which is exponential in three dimensions. The exponential correlation function or variogram is characterized by the variance (σ^2), and correlation lengths (τ_x , τ_y , and τ_z).

To perform the testing, two types of MRFs were generated. The first MRF was an isotropic field in which the correlation lengths in all three dimensions were equal. The second MRF was anisotropic with all three correlation lengths unequal. The statistical grid was chosen to be 32m \times 32m \times 32m with a grid block size of be 2m \times 2m \times 2m. For the isotropic case, the underlying statistical structure has a variance of $\sigma^2 = 0.05^2$ and correlation length of $\tau_x = \tau_y = \tau_z = 4$ m. For the anisotropic case, the underlying statistical structure has a variance of $\sigma^2 = 0.05^2$ and correlation length of $\tau_x = 4$ m, $\tau_y = 6$ m, $\tau_z = 8$ m. Once the MRFs were generated, the variogram model in GSLIB (Standard Center for Reservoir Forecasting, 1997) was employed to estimate the variograms from the MRFs generated. The calculated variograms were compared to the theoretical ones in Figures 5.1 and 5.2 for both the isotropic and anisotropic cases, respectively. The simulated variograms agree favorably with the corresponding theoretical results.

000042

5.2 TEST PROBLEM DESCRIPTION

5.2.1 Discretization

All the test runs were based on the transport of a conservative solute in a pseudo three-dimensional domain of $32\text{m} \times 16\text{m} \times 1\text{m}$. Flow was assumed to be in the x-direction. Synthetic data were used, since the true variables and their states are always known. The generated data provided a basis for testing the performance of the DFM/VAM3DF model.

The computational domain was discretized into $32 \times 16 \times 1$ elements, with an element size of $1\text{m} \times 1\text{m} \times 1\text{m}$. Figure 5.3 shows the discretization of the computational domain. In the DFM/VAM3DF model, the parameters estimated were defined on a parameter grid, composed of $16 \times 8 \times 1$ zones of size $2\text{m} \times 2\text{m} \times 1\text{m}$. The simulation of the pseudo three-dimensional transport using DFM/VAM3DF was made possible by assuming that the soil and transport properties and all initial and boundary conditions did not vary in the z-direction.

5.2.2 Initial Concentrations

Initial concentrations was assumed to be zero everywhere at start of simulation, unless otherwise stated.

5.2.3 Boundary Conditions

The boundary conditions were as follows. No mass fluxes were specified on the boundary faces at $z = 0$, $z = 1\text{m}$, $y = 0$, and $y = 16\text{m}$. No diffusive fluxes were specified across the boundary face at $x = 32\text{m}$. Contaminant mass was introduced at the six nodes located in the center of the $x = 0$ plane, which were assigned the Dirichlet boundary condition of $C = 1$ at the middle nodes ($y = 8\text{m}$) and $C = 0.5$ at the off-middle nodes ($y = 7\text{m}$ and 9m). The Dirichlet boundary conditions lasted for the first five days of the simulation after which the prescribed concentrations dropped to zero.

5.2.4 Transport Parameters

The default transport parameters were: Darcy velocity = 0.5m/day ; effective porosity = 0.3 ; distribution coefficient (K_d) = $0.5 \text{ cm}^3/\text{g}$ with linear isotherm; bulk density = 1.85 g/cm^3 ; longitudinal dispersivity (α_L or α_{LH}) = 1m ; transverse dispersivity (α_T or α_{TH}) = 0.2m ; and decay rate = 0 1/day . Adsorption/desorption was assumed linear unless stated otherwise.

One or more of the above default parameters were overwritten by the corresponding synthetically generated fields.

Time step size of 1 day was used. In all cases, a simulation time of 120 days was assumed.

5.2.5 Selection of Field and Concentration Measurement Locations

Field variables refer to variables that are unknown except at scattered measurement locations and have to be estimated using the DFM/VAM3DF model. Examples are the initial concentration,

effective porosity, partitioning coefficient (K_d), dispersivity, etc. Concentration measurements refer to concentration values measured over time.

For all test cases, various combinations of field and concentration measurements were chosen for the purpose of sensitivity analysis. Actual measurements were made at one or more potential measurement locations. Potential field measurement locations were assumed to coincide with that for concentration measurements except when the field variable was initial concentration.

Twelve potential field measurement locations (see Figure 5.4) were located at a 3m interval in the x-direction, and 2m apart in the y-direction centered at $\{x=6.5m, y=8m\}$. The locations were $\{x=2m, y=6m\}, \{x=5m, y=6m\}, \{x=8m, y=6m\}, \{x=11m, y=6m\}, \{x=2m, y=8m\}, \{x=5m, y=8m\}, \{x=8m, y=8m\}, \{x=11m, y=8m\}, \{x=2m, y=10m\}, \{x=5m, y=10m\}, \{x=8m, y=10m\}$, and $\{x=11m, y=10m\}$. Nine potential concentration measurement locations (see Figure 5.4) were spread 8m apart in the x-direction, and 4m apart in the y-direction centered at $\{x=16m, y=8m\}$ of the domain. The locations were: $\{x=8m, y=4m\}, \{x=8m, y=8m\}, \{x=8m, y=12m\}, \{x=16m, y=4m\}, \{x=16m, y=8m\}, \{x=16m, y=12m\}, \{x=24m, y=4m\}, \{x=24m, y=8m\},$ and $\{x=24m, y=12m\}$.

5.2.6 Description of Test Scenarios

Table 5.1 presents a description of the scenarios for seven sets of test problems. In each case, estimation of only one variable is considered given specified combination of field and concentration measurements.

5.3 SIMULATION PROCEDURES

The following steps were followed for each test problem:

1. Generate a "true" field on the VAM3DF numerical grid.
 - If the variable to be estimated is spatially random, a 3-dimensional random field is generated with the underlying statistical parameters such as the trend coefficients, standard deviation and correlation lengths. The generation is done on a 3-dimensional parameter grid of $32m \times 16m \times 16m$ with an element size of 2m in all three dimensions. Only one slice (at $z=8m$) of the 3-dimensional field generated is selected and used for the subsequent simulation and testing purpose. The slice represents a 2-dimensional random field. A linear interpolation is performed to map this 2-dimensional field onto that of the VAM3DF numerical (x,y) grid, so that the nodal properties at the top ($z=1m$) and bottom ($z=0m$) planes of the numerical grid are the same.
 - If the parameter to be estimated is the initial concentrations, the generation of the "true" field is achieved by running a transport simulation given prescribed initial concentrations, boundary conditions, and transport parameters. The solution at a desired time is considered the "true" initial concentrations.

- If the parameter to be estimated is a linear trend, the generation of the trend is simply based on the prescribed linear trend.
2. Sample the “true” field at selected locations inside the domain to obtain the measurement data for the field. Add noise (or measurement error) to the sampled field data values. The noise term at each measurement location was Gaussian, identical, independently distributed (i.i.d), with zero mean and a prescribed error standard deviation.
 3. Assume the “true” field (without noise) is completely known, run the transport problem using VAM3DF and obtain the concentration solutions at all possible time levels given appropriate initial and boundary conditions. Sample these solutions at selected time levels and locations to obtain the concentration measurement data. Add noise (or measurement error) to the sampled concentration data. Again the noise term at each measurement location was assumed to be Gaussian, i.i.d, with zero mean and a prescribed standard deviation.
 4. The field and concentration measurements with noises are then considered as if they were observed on the XY plane midway ($z=0.5m$) between the top and bottom planes.
 5. Prepare the main input data file for DFM/VAM3DF by incorporating the field and concentration measurements. Run DFM/VAM3DF with convergence criterion of 0.2, and the LSQR solver option.
 6. For prediction purpose, run VAM3DF with the nodal values of the unknown parameters replaced with the corresponding estimates.

5.4 TEST METHODS

5.4.1 Visual Inspection

One way of examining the performance of DFM/VAM3DF is to compare the estimated parameters with the true field. This can be accomplished by plotting the contours of the true and estimated parameters. Another way of examining the performance of DFM/VAM3DF is to compare the true and simulated breakthrough curves at selected spatial locations. The true breakthrough curves were obtained by running VAM3DF with the “true” field. The simulated breakthrough curves were obtained in a similar manner except that the nodal values of the parameters are replaced by the corresponding DFM/VAM3DF estimates. Four observation wells were assumed to be present, which are located at the cross-section, $x = 16m$. The closer the simulated breakthrough curves are to the true ones, the better the parameter estimates will be.

5.4.2 Chi-Squared Test

A useful test for identifying problems is based on the total sum-of-squares (SOS) of the normalized model fit and data fit errors. The total SOS should be distributed as a Chi-squared random variable with degree of freedom (D.F.) k equal to m (the number of scalar measurements) when all constant parameters have Bayesian prior, or $m - p_c$ where p_c is the number of constant

parameters that are treated as fixed but unknown parameters. For $k > 20$, the Chi-squared distribution is approximately Gaussian with mean k and variance $2k$.

SOS should lie within five standard deviations from the mean, or:

$$k - 5\sqrt{2k} \leq SOS \leq k + 5\sqrt{2k} \tag{5.1}$$

Should SOS be smaller than k , Equation (5.1) may be written as:

$$n = \frac{k - SOS}{5\sqrt{2k}} \leq 1 \tag{5.2}$$

or, should SOS be greater than k , Equation (5.1) may be written as:

$$n = \frac{SOS - k}{5\sqrt{2k}} \leq 1 \tag{5.3}$$

where n is the multiple of five standard deviations. The above condition is based on an assumption that components of the SOS are normally-distributed random variates with unit variance. This assumption may not be valid for some situations. The application of the above limits should be considered a qualitative indicator of model's robustness. In the event that n is greater than unity, caution should be exercised.

5.5 SIMULATION RESULTS AND DISCUSSION

Following the simulation procedures outlined earlier, the DFM/VAM3DF model was employed to solve each of the test problems mentioned in Table 5.1. The following provides a summary and discussion of the simulation results.

5.5.1 Estimation of Initial Concentration

The estimation of the spatial variability of the initial concentration was conducted based on 12 potential field measurements and 36 potential concentration measurements at 9 potential observation locations. (Four concentration measurements at days 10, 30, 60, and 90 at each observation location). A prior constant trend of -5 and a small prior standard deviation was specified, implying that the trend parameter was essentially known.

Table 5.2 presents the result summary for this problem. Figures 5.4 and 5.5 shows the initial plume estimates and the associated parameter error standard deviations, respectively, for three combinations of the field and concentration measurements. In the figures, circles represent concentration measurement locations and the triangles denote field or initial condition measurement locations. Figure 5.6 presents a comparison of simulated breakthrough curves with the true breakthrough curves. Figure 5.4 shows that Estimates 12IC 4C and 12IC 36C match the true log IC field better than Estimate 11C 36C. Figure 5.6 shows that 12 IC, 4C and 12IC 36C

000046

breakthrough curves are closer to the true breakthrough curves than the 1IC 36C breakthrough curves. The results suggest that direct measurements of IC have a larger impact on the estimate than concentration measurements, which is consistent with the measurement error given in Table 5.1. Figure 5.5 shows that the parameter uncertainty is smallest at the direct measurement locations.

5.5.2 Estimation of Log Effective Porosity

The estimation of the spatial variability of log (effective porosity) was conducted based on 9 potential effective porosity (or field) measurements and 45 potential concentration measurements at 9 potential observation locations. (Five concentration measurements at days 10, 20, 30, 60 and 90 at each observation location). The 9 field and concentration observation locations coincide with one another. Sensitivity of the log (effective porosity) estimates to changes in measurement errors, correlation length, and number of measurements was examined. A prior constant trend of log (effective porosity) = -0.6 with a standard deviation of 0.2 was specified, implying that the trend parameter is basically unknown and has to be estimated as well. Measurement errors were set equal to 0.01 for log (effective porosity) and 0.1 for log concentrations.

Table 5.3 presents the result summary. Notice that the concentration sum of squares is smaller when all 9 porosity measurements are used. Figures 5.7-5.8 show the log porosity estimates and the associated parameter error standard deviations, respectively. Figure 5.9 depicts the true and simulated breakthrough curves at the observation locations. Figure 5.7 shows that Estimates 9 porosity 5C and 9 porosity 45C match the true field at the 9 porosity measurement locations. Also, the two estimates with 45 concentration measurements have low log (effective porosity) at $x = 4m$, $y = 8m$. Figure 5.8 shows that parameter uncertainty is smallest at measurement locations. Figure 5.4 shows that all of the breakthrough curves based on the estimated log (effective porosity) fields matched the true breakthrough curves very well. This suggests that the concentration distribution are not sensitive small-scale variations in porosity. Figures 5.7-5.9 correspond to the case where the underlying correlation length for log porosity equals 4m (four times the numerical grid size). Figures 5.10-5.12 are the same except that the underlying correlation length for log (effective porosity) was increased to 8m. Notice that the true field in Figure 5.10 is smoother than the true field in Figure 5.7. The estimates, error standard deviations, and breakthrough curves for the 8m correlation length case are similar to the 4m correlation length case.

5.5.3 Estimation of Log K_d

The estimation of the spatial variability of the K_d variable was conducted based on 9 potential K_d (or field) measurements and 45 potential concentration measurements at 9 potential observation locations. The 9 field and concentration observation locations coincide with one another. Five concentration measurements representing snapshots at days 10, 20, 30, 60 and 90 at each observation location were assumed available. Sensitivity of the K_d estimates to changes in measurement errors, underlying correlation length and amount of measurements was examined. A constant trend value of -3 with a standard deviation of 0.2 was specified. Measurement errors were set equal to 0.01 for log K_d and 0.1 for log concentrations. Sensitivity of results with respect to the underlying correlation length was investigated.

Table 5.4 summarizes the results from DFM/VAM3DF. Figures 5.13-5.14 show the $\log K_d$ estimates and the associated standard errors, respectively. Figure 5.15 depicts the true and simulated breakthrough curves at the observation locations. In Figure 5.13 all estimates have the same general trend as the true field. Figure 5.15 shows that the 1 K_d 45C and 9 K_d 45C breakthrough curves match the true breakthrough curves better than the 9 K_d 5C breakthrough curves. Figures 5.13-5.15 correspond to the case where the underlying correlation length equals 4m or four times the numerical grid size. Similar results with the underlying condition length increased to 8m are presented in Figures 5.16-5.18. The estimates, error standard deviations, and breakthrough curves for the 8m correlation length case are similar to the 4m correlation length case.

5.5.4 Estimation of Polynomial Trend in Log Effective Porosity

The estimation of pure polynomial trend in porosity field was conducted by first generating a "true" linear trend in the log porosity field (which varies from $\log(0.5)$ at $x=0m$ to $\log(0.2)$ at $x=32m$). The Dirichlet boundary conditions were altered so that the prescribed concentration values prevailed throughout the simulation period. The trend consisted of an unknown constant plus an unknown x-direction trend coefficient. Measurement errors were set to 0.01 for log porosity and 0.1 for log concentrations.

Table 5.5 summarizes the results from DFM/VAM3DF runs. Figure 5.19 shows the log porosity estimates. Figure 5.20 illustrates the true and simulated breakthrough curves at the observation locations. Note that, even though the 1 porosity 15C estimates do not match the true linear trend in figure 5.19, the resulting breakthrough curves are close to the true breakthrough curves in Figure 5.20.

5.5.5 Estimation of Polynomial Trend in Log K_d

The estimation of pure polynomial trend in K_d field was conducted by first generating a "true" linear trend in the $\log K_d$ field (which varies from $\log(0.1)$ at $x=0m$ to $\log(10)$ at $x=32m$). The Dirichlet boundary conditions were altered so that the prescribed concentration values prevailed throughout the simulation period. The trend consisted of an unknown constant plus an unknown x-direction trend coefficient. Measurement errors were set equal to 0.01 for $\log K_d$. Sensitivity of estimation analysis with respect to concentration measurement errors was conducted by setting log concentration measurement errors to 0.1 and 0.01.

Table 5.6 summarizes the results from DFM/VAM3DF runs. Figure 5.21 shows the $\log K_d$ estimates. Figure 5.22 illustrates the true and simulated breakthrough curves at the observation locations. Figure 5.21 shows that, of the three cases, the 1 K_d 15C estimates do not match the true linear trend, and Figure 5.22 shows that the resulting breakthrough curves do not match true breakthrough curves. Figures 5.23-5.24 present similar results except that the concentration measurement errors were reduced by one order of magnitude to 0.01. A lower concentration error improved the 1 K_d 15C results.

5.5.6 Estimation of Polynomial Trend in Longitudinal Dispersivity

The estimation of pure polynomial trend in longitudinal dispersivity field was conducted by first generating a "true" linear trend in the α_L field (which varied from 0.5 at $x=0\text{m}$ to 1.5 at $x=32\text{m}$). The trend consisted of an unknown constant plus a unknown x-direction trend coefficient.

Table 5.7 presents a summary of the results. Figure 5.25 show the dispersivity estimates. Figure 5.26 illustrates the true and simulated breakthrough curves at the observation locations. Eventhough the $1\alpha_L$ 15C estimates do not match the true linear trend in Figure 5.25, the resulting breakthrough curves are close to the true breakthrough curves in Figure 5.26.

5.5.7 Estimation of the $\text{Log}_{10} K_d$ with Freundlich Isotherm

The estimation of Freundlich isotherm parameter K_d was conducted assuming that n was equal to a constant value of 1.0. The purpose was to test the capability of DFM/VAM3DF in estimating kinetic related parameters. Five concentration measurements representing snapshots at days 5, 10, 20, 30 and 60 at each concentration measurement location were assumed available. A constant unknown trend value of -0.3 with a standard deviation of 0.2 was specified.

Table 5.8 summarizes the DFM/VAM3DF runs. Figures 5.27-5.28 show the $\log K_d$ estimates and the associated standard errors, respectively. Figure 5.29 depicts the true and simulated breakthrough curves at the observation locations. Notice that all three $\log K_d$ estimates resulted in higher concentrations than the true breakthrough curves. This may be caused by low K_d at $x = 4\text{m}$, $y=8\text{m}$ compared to the true K_d field.

5.5.8 Estimation of Polynomial Trend in Desorption Rate Coefficient

The estimation of pure polynomial trend in desorption rate coefficient α_d was conducted by first generating a "true" linear trend in the α_d field (which varies from 10^{-4} at $x=0\text{ m}$ to 10^{-3} at $x=32\text{ m}$). The trend consists of an unknown constant and an unknown trend coefficient in the x-direction.

Table 5.9 presents a result summary. Figure 5.30 shows the desorption rate estimates. Figure 5.31 illustrates the true and simulated breakthrough curves at the four observation locations. All three α_d field estimates resulted in breakthrough curves that match the true breakthrough without matching the true α_d field.

5.5.9 Estimation of $\log K_d$ using Extended Kalman Filters Update (EKFU)

The estimation of the spatial variability of $\log K_d$ was conducted using the EKFU. This test problem is identical to the problem in Section 5.5.3. First, $\log K_d$ was estimated using 9 K_d measurements and 36 concentration measurements at locations at 10, 20, 30 and 60 days. These results were used to initialize the EKFU estimation, which used 9 concentration measurements at 90 days.

Table 5.10 summarizes the results. The 9 K_d 45C results from Section 5.5.3 are shown for comparison. Figures 5.32 - 5.33 show the $\log K_d$ estimates and the associated standard errors, respectively. Notice that 9 K_d 35C and EKFU estimation gave similar results. Figure 5.34 depicts the true and simulated breakthrough curves at the observation locations.

5.6 SUMMARY OF RESULTS

The results presented in the previous section demonstrate the software functionality with respect to the determination of spatial variability and regional trends of transport parameters. Parameter identifiability are the behavior of the robustness indicator, n , are parameter dependent. For most of the test problems, fewer direct measurements leads to larger concentration residuals. Some of the parameters may not be identifiable from concentration measurements alone.

Table 5.1
Description of Test Problems

No.	Parameters Estimated	Estimated Polynomial Trend	Test Scenarios
1	Spatially variable initial concentration with $\mu = \log(10^{-5})$ $\sigma = 2.0$ $\tau_x = 5m, \tau_y = \tau_z = 4m$	Constant trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.01$ for log initial concentration (IC), and $\sigma=0.1$ for log concentration data • Concentration measurements made at days 10, 30, 60, and 90. • Number of measurements: <ul style="list-style-type: none"> - 1 IC and 36 concentrations; - 12 IC and 4 concentrations; - 12 IC and 36 concentrations
2	Spatially variable porosity with $\mu = \log(0.3)$ $\sigma = 0.065$ $\tau_x = \tau_y = \tau_z = 4m$ and 8m	Constant trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.01$ for log porosity, and $\sigma=0.1$ for log concentrations • Concentration measurements made at days 10, 20, 30, 60 and 90. • Measurement combinations: <ul style="list-style-type: none"> - 1 porosity and 45 concentrations; - 9 porosity and 4 concentrations; - 9 porosity and 45 concentrations.
3	Spatially variable K_d with $\mu = \log(0.5)$ $\sigma = 0.05$ $\tau_x = \tau_y = \tau_z = 4m$ and 8m	Constant trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.01$ for log K_d, and $\sigma=0.1$ for log concentrations • Concentration measurements made at days 10, 20, 30, 60 and 90. • Measurement combinations: <ul style="list-style-type: none"> - 1 K_d and 45 concentrations; - 9 K_d and 4 concentrations; - 9 K_d and 45 concentrations.
4	(Linear) Polynomial trend in log porosity True log porosity varies from $\log(0.5)$ at $x=0$ to $\log(0.2)$ at $x=32m$	Linear trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.01$ for log porosity, and $\sigma=0.1$ for log concentrations • Concentration measurements made at days 10, 20, 30, 60 and 90. • Measurement combinations: <ul style="list-style-type: none"> - 1 porosity and 15 concentrations; - 3 porosity and 5 concentrations; - 3 porosity and 15 concentrations
5	(Linear) Polynomial trend in log K_d True log K varies from $\log(0.1)$ at $x=0m$ to $\log(10)$ at $x=32m$	Linear trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.01$ for log K_d, and $\sigma=0.1$ or 0.01 for log concentrations. • Concentration measurements made at days 10, 20, 30, 60 and 90. • Measurement combinations: <ul style="list-style-type: none"> - 1 K_d and 15 concentrations; - 3 K_d and 5 concentrations; - 3 K_d and 15 concentrations.

000051

Table 5.1 (continued)
Description of Test Problems

No.	Parameters Estimated	Estimated Polynomial Trend	Test Scenarios
6	(linear) polynomial trend in longitudinal dispersivity α_L True α_L varies from 0.5m at $x = 0\text{m}$ to 1.5m at $x = 32\text{m}$	Linear trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.0001$ for α_L, and $\sigma=0.1$ for log concentrations. • Concentration measurements made at days 10, 20, 30, 60 and 90. • Measurement combinations: <ul style="list-style-type: none"> - 1 α_L and 15 concentrations; - 3 α_L and 5 concentrations; - 3 α_L and 15 concentrations.
7	Spatial variable (Freundlich with $n=1$) K_d $\mu = -0.3$ $\sigma = 0.6$ $\tau_x = \tau_y = \tau_z = 4\text{m}$	Constant trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.01$ for log K_d, and $\sigma=0.1$ for log concentrations. • Concentration measurements made at days 5, 10, 20, 30, and 60. • Measurement combinations: <ul style="list-style-type: none"> - 1 K_d and 45 concentrations; - 3 K_d and 5 concentrations; - 3 K_d and 45 concentrations.
8	(Linear) Polynomial trend in desorption rate coefficient α_d True α_d varies from 10^{-4} at $x = 0\text{m}$ to 10^{-3} at $x = 32\text{m}$	Unknown linear trend parameters	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.00001$ for α_d, and $\sigma=0.01$ for log concentrations. • Concentration measurements made at days 10, 20, 30, 60 and 90. • Measurement combinations: <ul style="list-style-type: none"> - 1 α_d's and 15 concentrations; - 3 α_d's and 5 concentrations; - 3 α_d's and 15 concentrations.
9	Spatially variable K_d with $\mu = \log(0.5)$ $\sigma = 0.05$ $\tau_x = \tau_y = \tau_z = 8\text{m}$	Constant trend	<ul style="list-style-type: none"> • Measurement errors: $\sigma=0.01$ for log K_d, and $\sigma=0.1$ for log concentrations. • Concentration measurements made at days 10, 20, 30, 60 and 90. • Measurement combinations: <ul style="list-style-type: none"> - 9 α_d's and 36 concentrations; - 3 α_d's and 36 concentrations and EKFU 9 concentrations; - 3 α_d's and 15 concentrations.

000052

Table 5.2
Estimation of initial plume: summary of results

Number of Measurements*	Degree of Freedom, k	Sum of Squares				n
		Field	Spatial Variability	Concentration	Total	
1 IC 36 C	37	2.46x10 ⁻³	38.0	56.3	94.4	1.32
12 IC 4 C	16	9.82x10 ⁻³	97.9	5.8	103.8	3.10
12 IC 36 C	48	1.02x10 ⁻²	100.0	29.2	129.4	1.66

* Notes:

1 IC 36 C implies that the measured data include 1 initial concentration value and 36 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

Table 5.3
Estimation of spatially random porosity field: summary of results
 (τ = correlation length)

τ (m)	Number of Measurements*	Degree of Freedom, k	Sum of Squares				n
			Field	Spatial Variability	Concentration	Total	
4	1 Por 45 C	45	1.1	23.5	49.4	74.3	0.62
	9 Por 5 C	13	3.0	37.0	3.2	43.6	1.2
	9 Por 45 C	53	3.5	52.9	41.	98.3	0.88
8	1 Por 45 C	45	2.7	27.5	50.8	81.1	0.76
	9 Por 5 C	13	7.0	45.2	3.1	55.7	1.67
	9 Por 45 C	53	7.9	65.1	39.8	113.1	1.17

* Notes:

1 Por 45 C implies that the measured data include 1 porosity value and 45 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

Table 5.4
Estimation of spatially random K_d field: summary of results
 (τ = correlation length)

τ (m)	Number of Measurements*	Degree of Freedom, k	Sum of Squares				n
			Field	Spatial Variability	Concentration	Total	
4	1 Kd 45 C	45	0.0	14.4	288.3	303.2	5.44
	9 Kd 5 C	13	0.0	19.3	22.5	42.5	1.16
	9 Kd 45 C	53	0.0	32.7	171.	203.9	2.93
8	1 Kd 45 C	45	0.0	15.8	274.9	291.0	5.18
	9 Kd 5 C	13	0.1	24.3	30.5	55.0	1.65
	9 Kd 45 C	53	0.0	32.3	157.8	190.3	2.66

* Notes:

1 Kd 45 C implies that the measured data include 1 Kd value and 45 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

Table 5.5
Estimation of linear trend in log porosity: summary of results

Number of Measurements*	Degree of Freedom, k	Coefficient 1 Estimate	Coefficient 2 Estimate	Sum of Squares			n
				Field	Concentration	Total	
1 Por 15 C	14	-0.494	-0.053	0.0	25.1	25.1	0.42
3 Por 5 C	6	-0.503	-0.189	1.1	7.5	10.6	0.27
3 Por 15 C	16	-0.503	-0.190	1.1	19.6	22.7	0.24

* Notes:

1 Por 15 C implies that the measured data include 1 porosity value and 15 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

Table 5.6
Estimation of linear trend in log K_d : summary of results

Number of Measurements*	Degree of Freedom, k	Coefficient 1 Estimate	Coefficient 2 Estimate	Sum of Squares			n
				Field	Concentration	Total	
$\sigma = 0.1$ for log concentration							
1 Kd 15 C	14	-0.019	-0.213	6.1	2066.5	2075.0	77.9
3 Kd 5 C	6	-.0025	1.013	1.2	6.2	28.3	1.29
3 Kd 15 C	16	-.0027	1.0129	1.2	12.5	34.6	0.66
$\sigma = 0.01$ for log concentration							
1Kd 15C	14	-0.004	0.989	1.1	53.4	74.3	2.3
3Kd 5C	6	0.0004	1.014	1.5	6.5	28.9	1.3
3Kd 15C	16	-0.001	1.009	1.6	14.8	37.1	0.75

* Notes:

1 Kd 15 C implies that the measured data include 1 Kd value and 15 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

Table 5.7
Estimation of linear trend in longitudinal dispersivity α_L : summary of results

Number of Measurements*	Degree of Freedom, k	Coefficient 1 Estimate	Coefficient 2 Estimate	Sum of Squares			n
				Field	Concentration	Total	
1 α_L 15 C	14	1.015	0.0142	0.0	44.9	45.1	1.18
3 α_L 5 C	6	0.995	0.505	5.5	4.6	34.8	1.66
3 α_L 15 C	16	0.999	0.505	6.0	12.1	42.8	0.94

* Notes:

1 α_L 15 C implies that the measured data include 1 α_L value and 15 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

000055

Table 5.8
Estimation of spatially random Freundlich K_d : summary of results

Number of Measurements*	Degree of Freedom, k	Sum of Squares				n
		Field	Spatial Variability	Concentration	Total	
1 K_d 45 C	45	0.0	32.9	114.6	147.5	2.16
9 K_d 5 C	13	0.0	26.7	21.0	47.8	1.36
9 K_d 45 C	53	0.1	64.9	100.5	165.5	2.19

* Notes:

1 K_d 45 C implies that the measured data include 1 K_d value and 45 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

Table 5.9
Estimation of linear trend in desorption rate coefficient α_d : summary of results

Number of Measurements*	Degree of Freedom, k	$5\sqrt{2k}$	Coefficient 1 Estimate	Coefficient 2 Estimate	Sum of Squares			n
					Field	Concentration	Total	
1 α_d 15 C	14	10.6	0.000554	0.000013	0.20	28.7	29.0	0.56
3 α_d 5 C	6	17.3	0.000552	0.000158	432.1	5.1	655.9	37.5
3 α_d 15 C	16	28.3	0.000552	0.000158	431.5	22.4	672.7	23.22

* Notes:

1 α_d 15 C implies that the measured data include 1 α_d value and 15 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

Table 5.10
Estimation of Spatially Random Kd Field with Extended Kalman File Update

Number of Measurements*	Degree of Freedom, k	Sum of Squares				n
		Field	Spatial Variability	Concentration	Total	
9 Kd 36 C	44	0.0	32.6	147.9	180.6	2.91
9 Kd 36 C and EKFU 9 C	8	0.0	32.2	163.9	196.2	9.41
9 Kd 45 C	53	0.0	32.3	157.8	190.3	2.66

* Notes:

9 Kd 45 C implies that the measured data include 9 Kd value and 45 concentration values, etc.

$$n = \frac{\text{Total} - k}{5\sqrt{2k}}$$

Sum of squares associated with the trend polynomial are included in totals but not listed in the table.

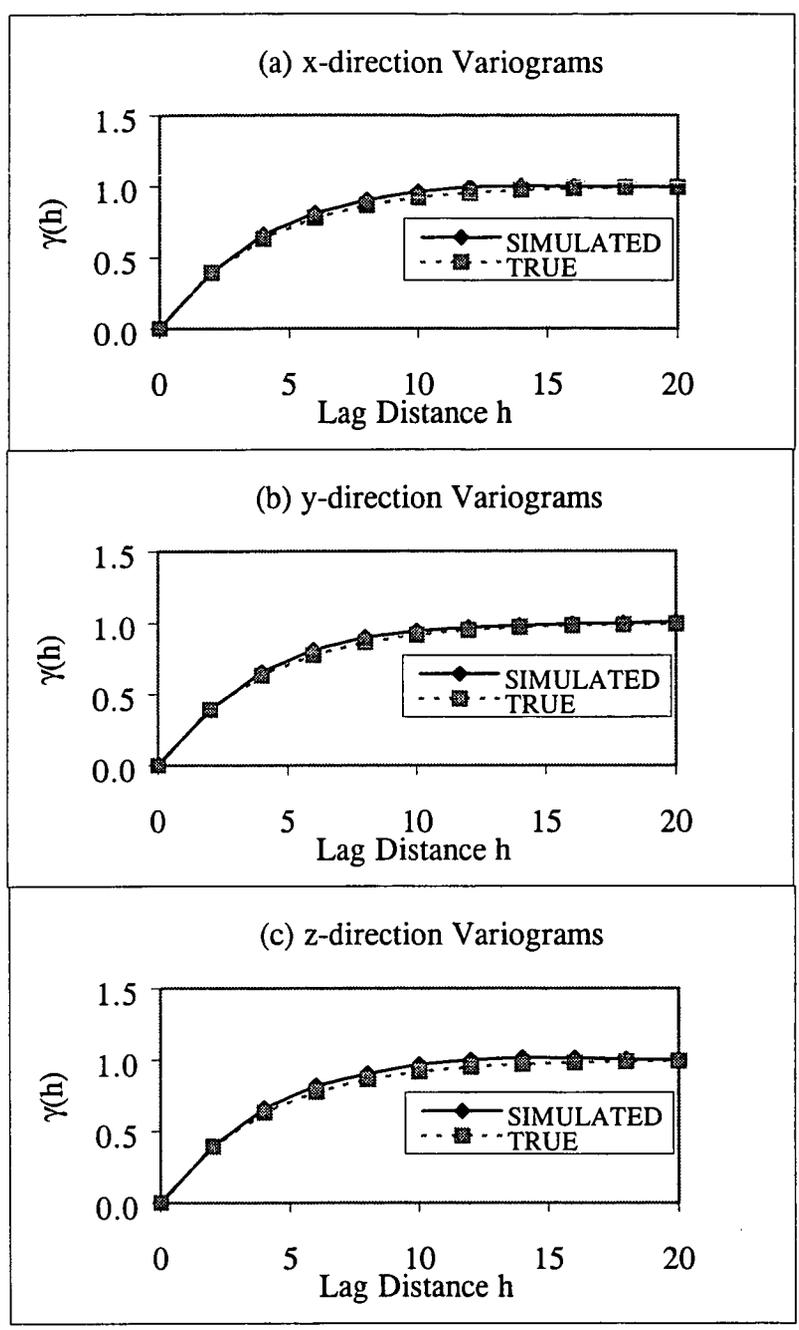


Figure 5.1 Simulated and True Variograms for the Generated Iostropic Markov Random Field.

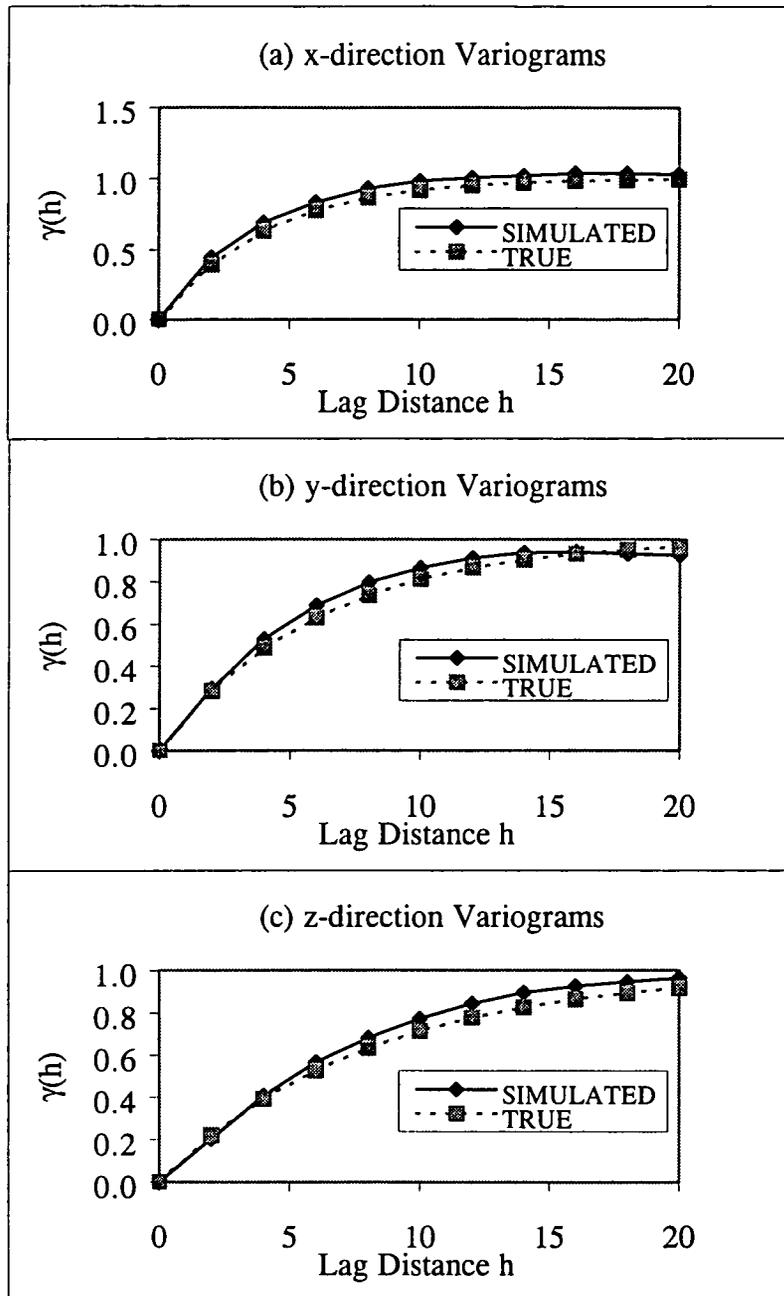
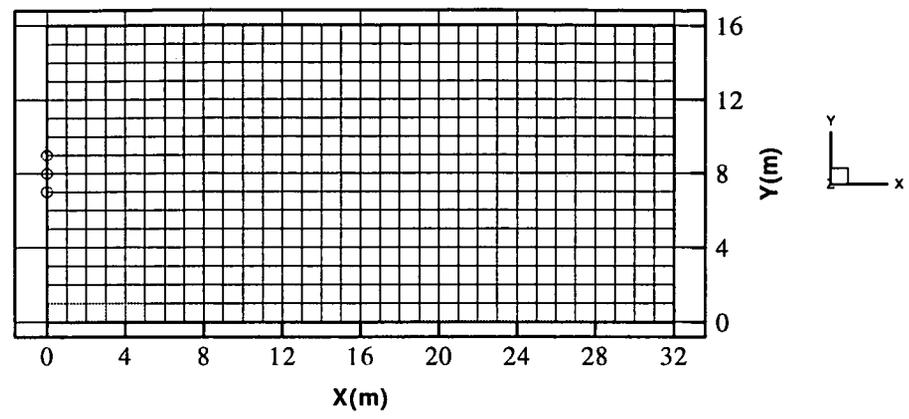


Figure 5.2 Simulated and True Variograms for the Anistropic Markov Random Field Generated.

(a) XY Plane View



(b) YZ Plane View (at X=0)

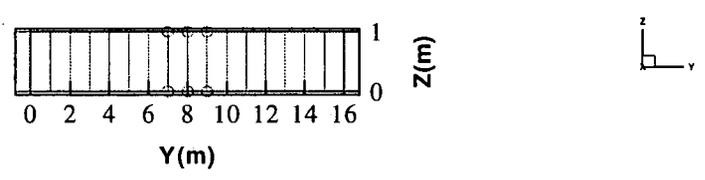


Figure 5.3 Numerical Discretization of Computational Domain (Circles indicate nodes with Dirichlet boundary conditions).

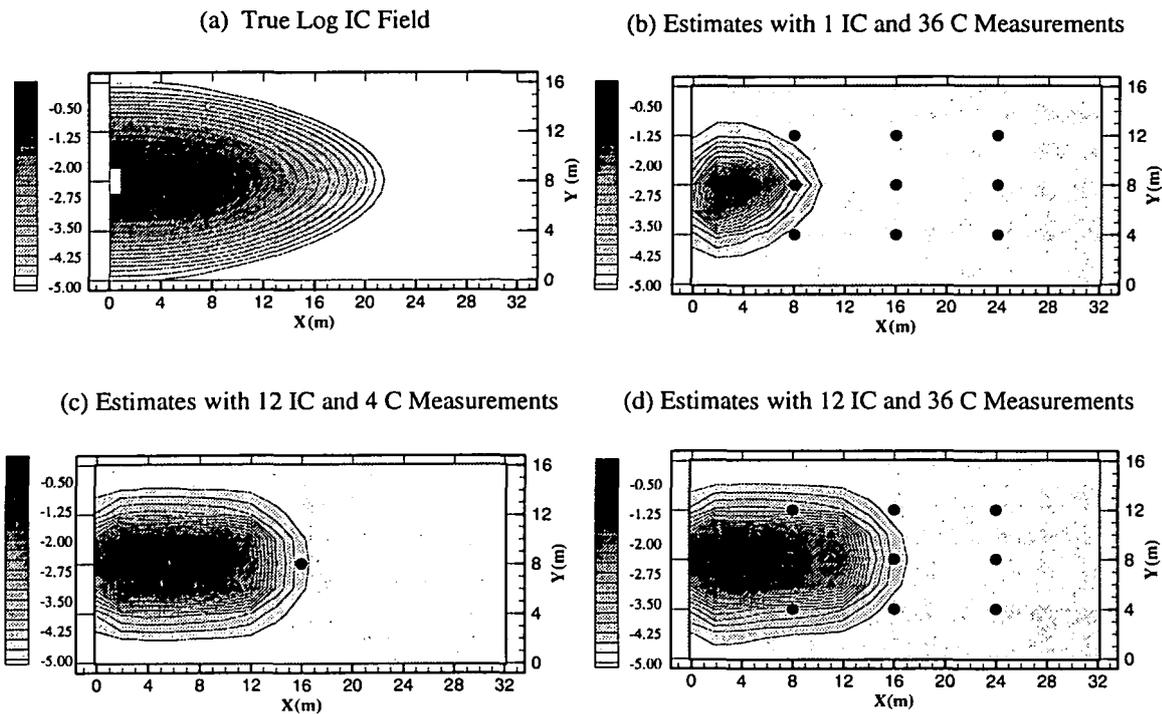
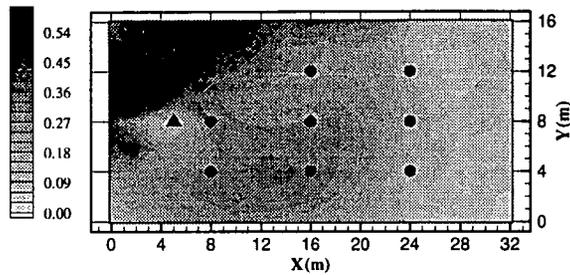


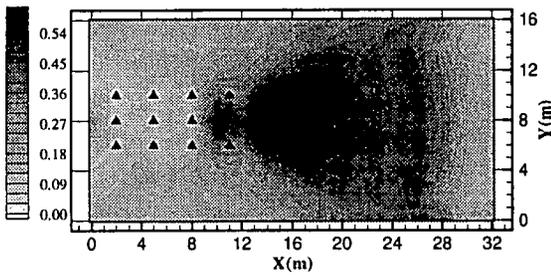
Figure 5.4 Estimation of Initial Plume: True and Estimated Initial Plumes (Circles represent concentration measurement locations, and triangles initial concentration measurement locations. Four concentration measurements were made at each concentration measurement location. 1 IC 36 C implies 1 Initial Concentration and 36 (9 locations \times 4 values/each location) Concentration values, etc.)

000061

(a) 1 IC 36 C Measurements



(b) 12 IC and 4 C Measurements



(c) 12 IC and 36 C Measurements

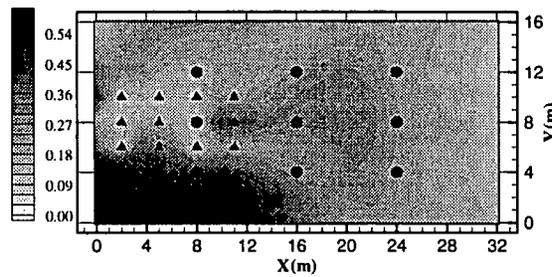


Figure 5.5 Estimation of Initial Plume: Standard Deviations Associated with Log Initial Plume Estimates (Circles represent concentration measurement locations, and triangles initial concentration measurement locations. Four concentration measurements were made at each concentration measurement location. 1 IC 36 C implies 1 Initial Concentration and 36 (9 locations \times 4 values/each location) Concentration values, etc.).

000062

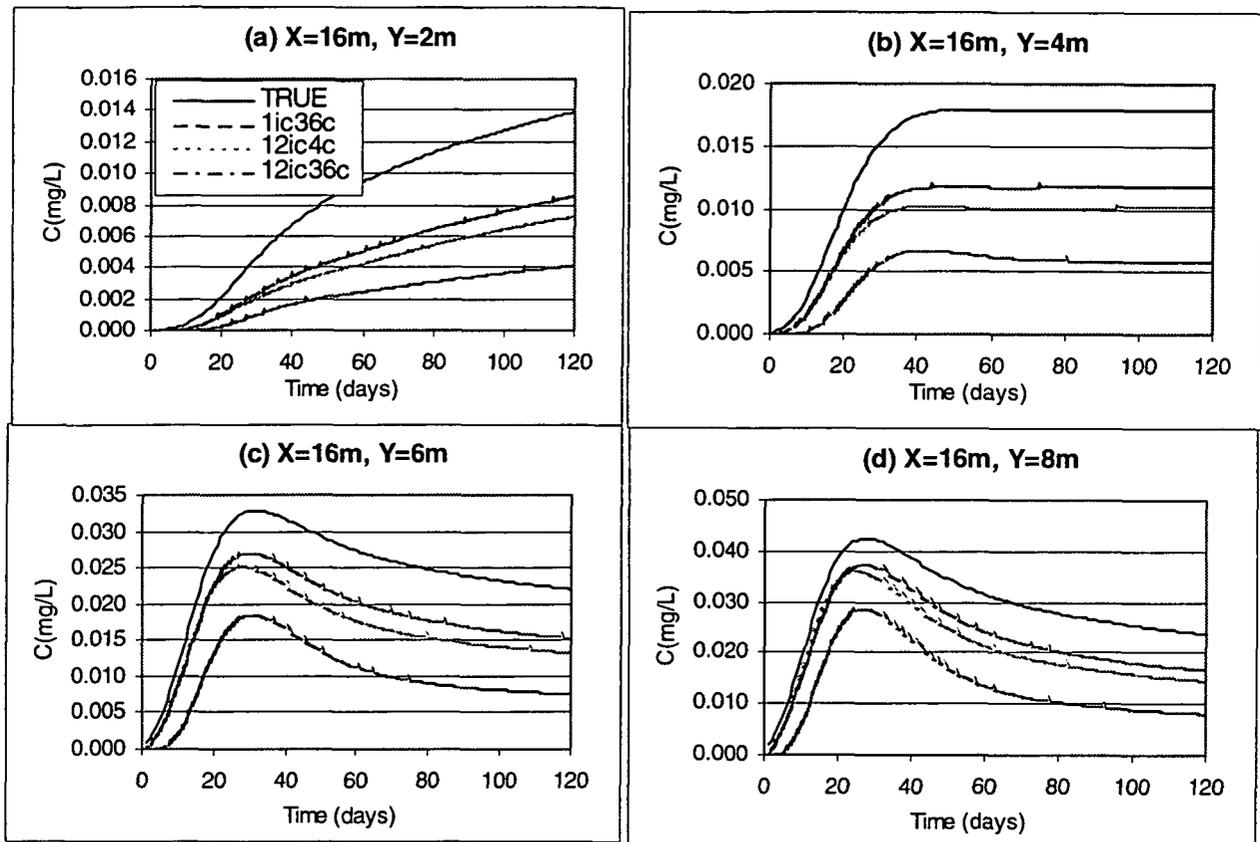


Figure 5.6 Estimation of Initial Concentrations: True and Simulated Breakthrough Curves (Four concentration measurements were made at each concentration measurement location. 1 IC 36 C implies 1 Initial Concentration and 36 (9 locations \times 4 values/each location) Concentration values, etc.)

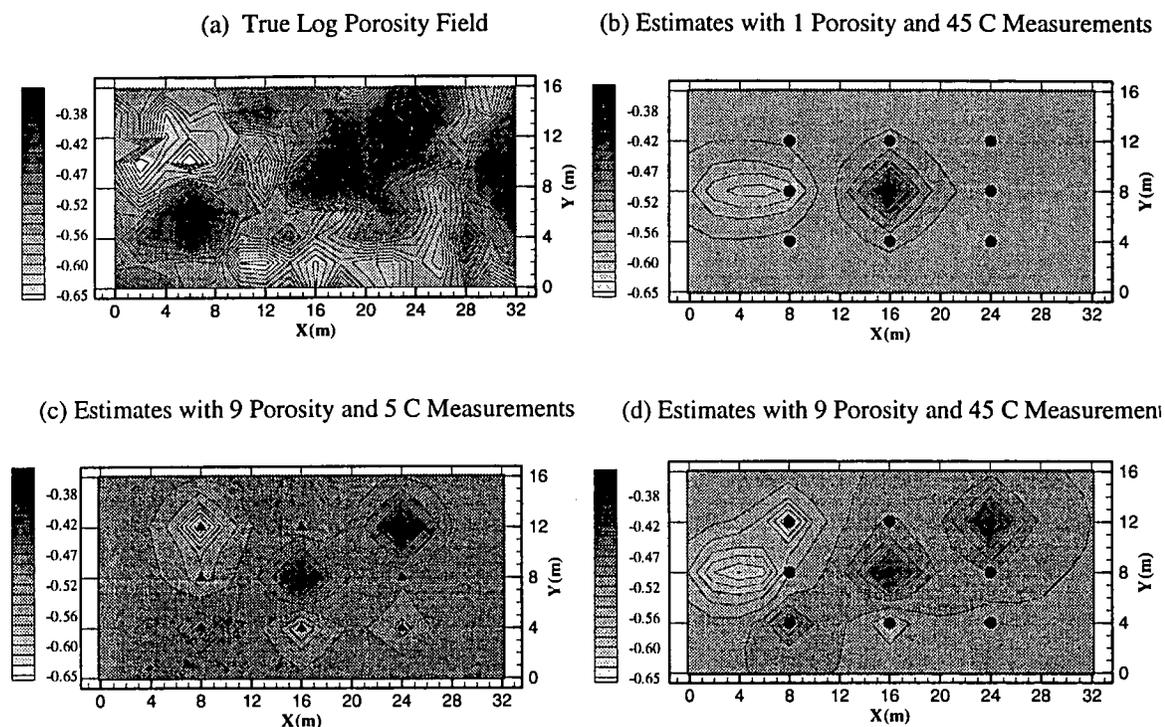
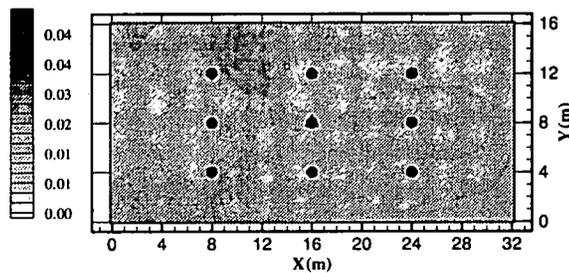
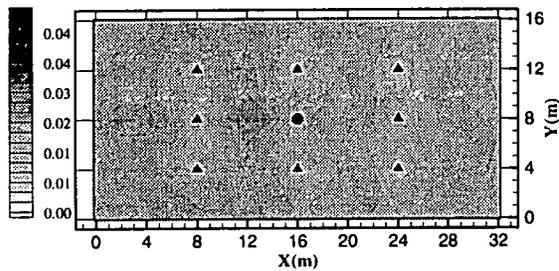


Figure 5.7 Estimation of Spatial Random Porosity with Underlying Correlation Length = 4m: True and Estimated Log Porosity Fields (Circles represent concentration measurement locations, and triangles porosity measurement locations. Five concentration measurements were made at each concentration measurement location. 1 Por 45 C implies 1 Porosity and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

(a) 1 Porosity and 45 C Measurements



(b) 9 Porosity and 5 C Measurements



(c) 9 Porosity and 45 C Measurements

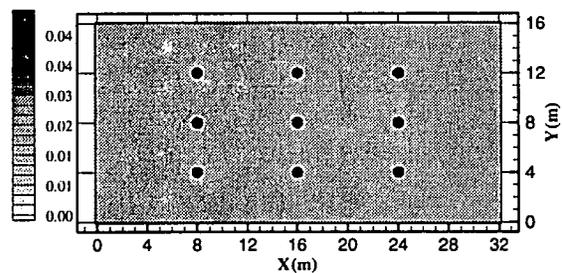


Figure 5.8 Estimation of Spatial Random Porosity with Underlying Correlation Length = 4m: Standard Deviations Associated with Estimated Log Porosity Fields (Circles represent concentration measurement locations, and triangles porosity measurement location. Five concentration measurements were made at each concentration measurement location. 1 Porosity 45 C implies 1 Porosity and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

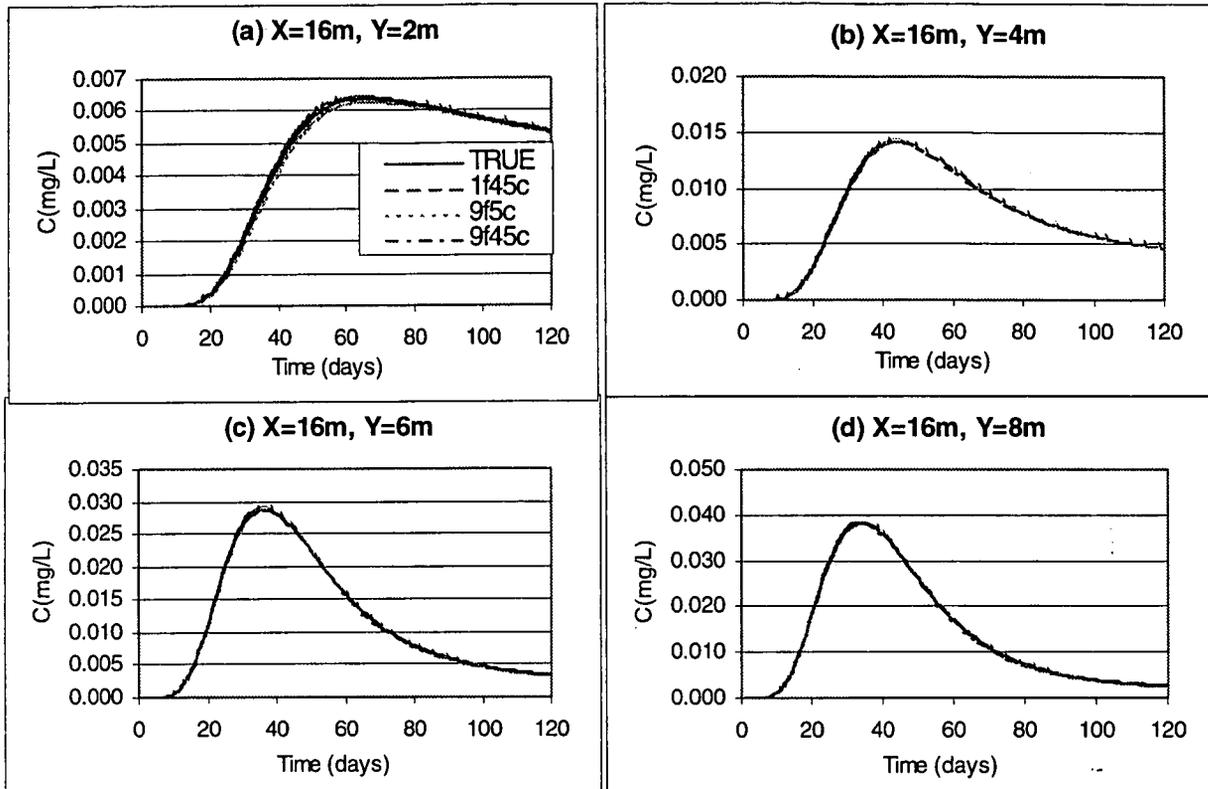


Figure 5.9 Estimation of Spatial Random Porosity with Underlying Correlation Length = 4m: True and Simulated Breakthrough Curves (Five concentration measurements were made at each concentration measurement location. 1 Por 45 C implies 1 porosity and 45 (9 locations \times 5 values/each location) concentration values, etc.).

000066

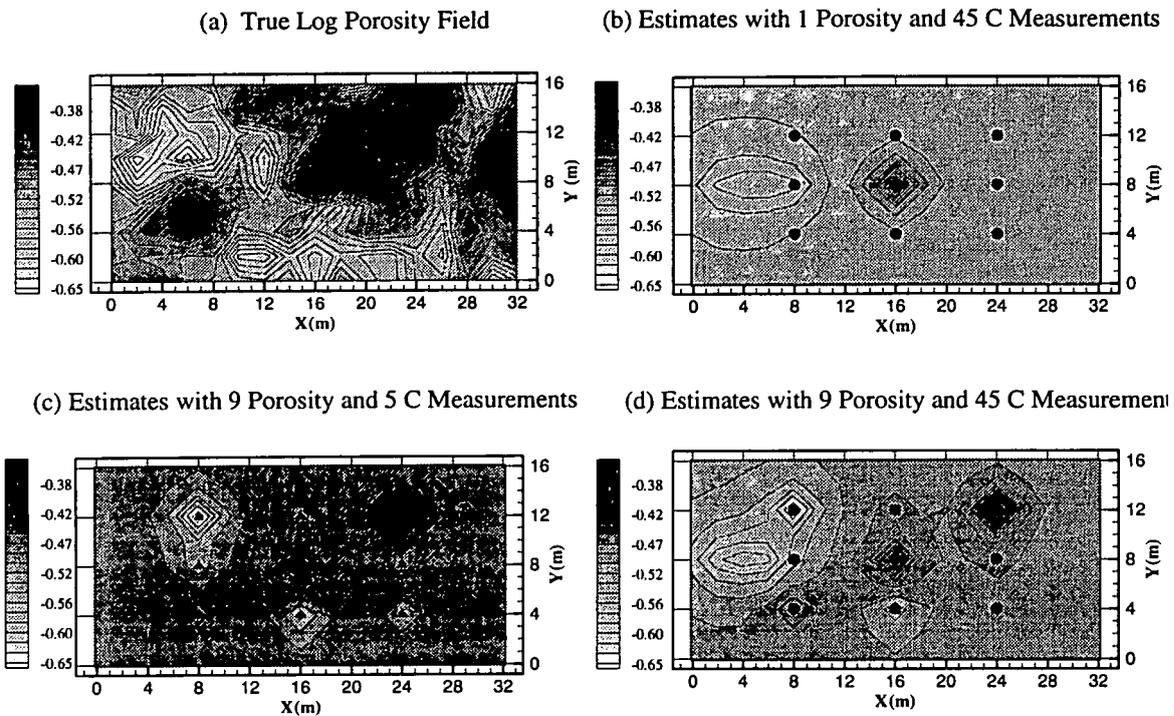
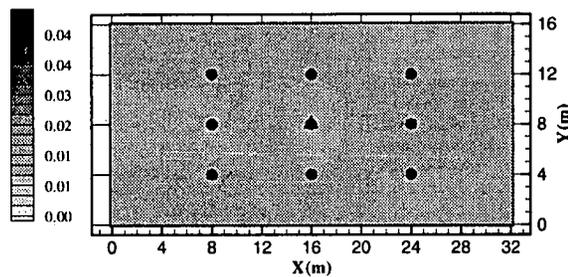
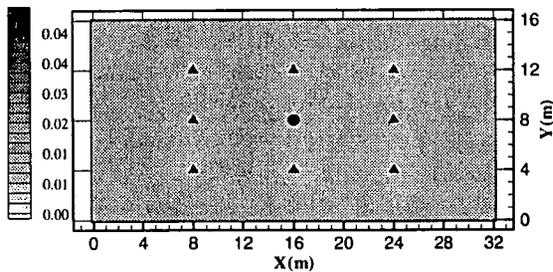


Figure 5.10 Estimation of Spatial Random Porosity with Underlying Correlation Length = 8m: True and Estimated Log Porosity Fields (Circles represent concentration measurement locations, and triangles porosity measurement locations. Five concentration measurements were made at each concentration measurement location. 1 Por 45 C implies 1 Porosity and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

(a) 1 Porosity and 45 C Measurements



(b) 9 Porosity and 5 C Measurements



(c) 9 Porosity and 45 C Measurements

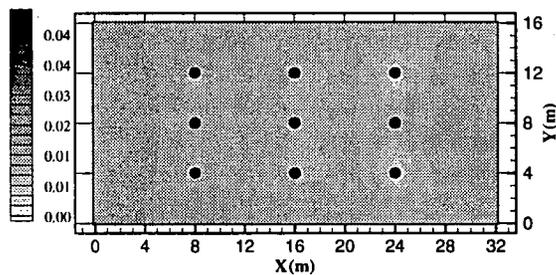


Figure 5.11 Estimation of Spatial Random Porosity with Underlying Correlation Length = 8m: Standard Deviations Associated with Estimated Log Porosity Fields (Circles represent concentration measurement locations, and triangles porosity measurement locations. Five concentration measurements were made at each concentration measurement location. 1 Porosity 45 C implies 1 Porosity and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

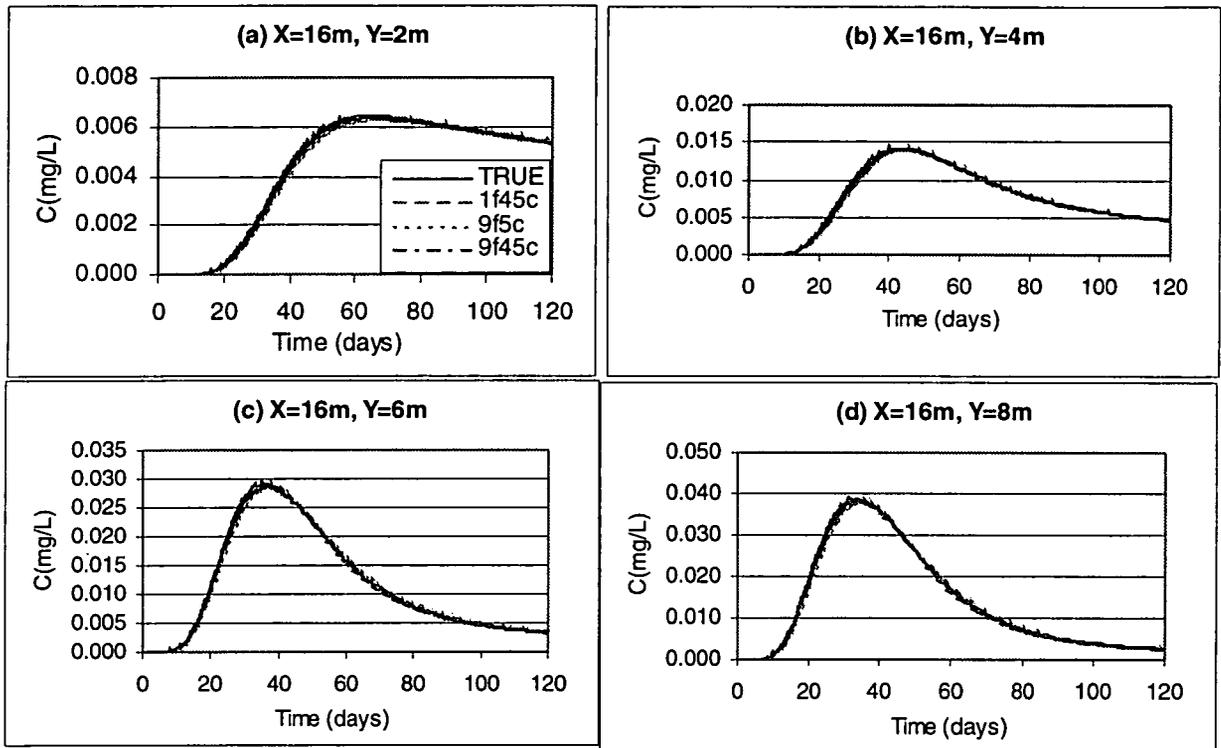


Figure 5.12 Estimation of Spatial Random Porosity with Underlying Correlation Length = 8m: True and Simulated Breakthrough Curves (Five concentration measurements were made at each concentration measurement location. 1 Por 45 C implies 1 Por and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

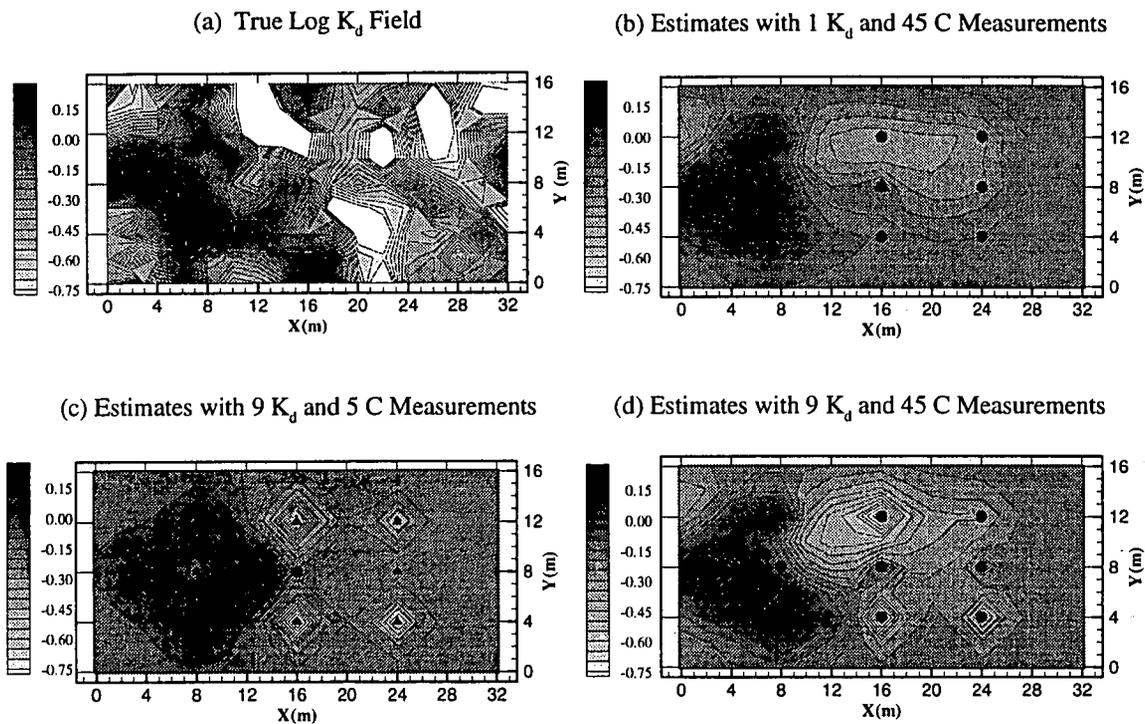
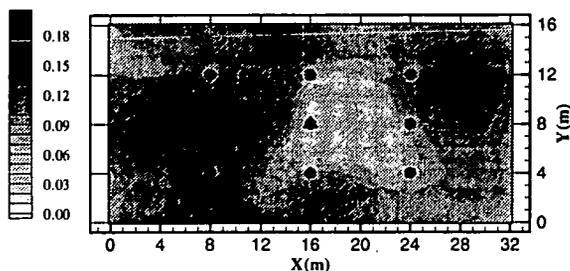


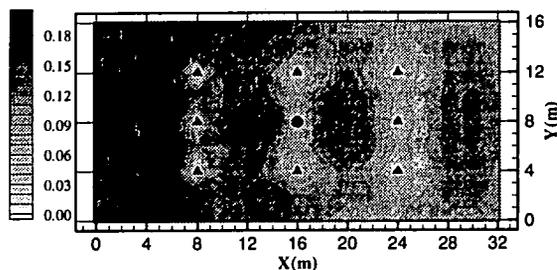
Figure 5.13 Estimation of Spatial Random K_d with Underlying Correlation Length = 4m: True and Estimated Log K_d Fields (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

000070

(a) 1 K_d and 45 C Measurements



(b) 9 K_d and 5 C Measurements



(c) 9 K_d and 45 C Measurements

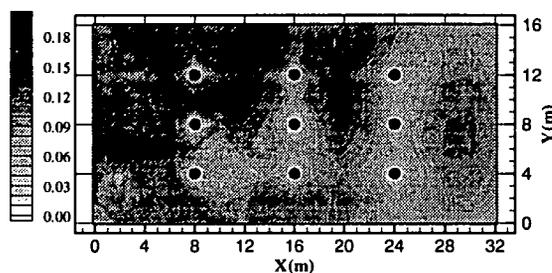


Figure 5.14 Estimation of Spatial Random K_d with Underlying Correlation Length = 4m: Standard Deviations Associated with Estimated Log K_d Fields (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.)

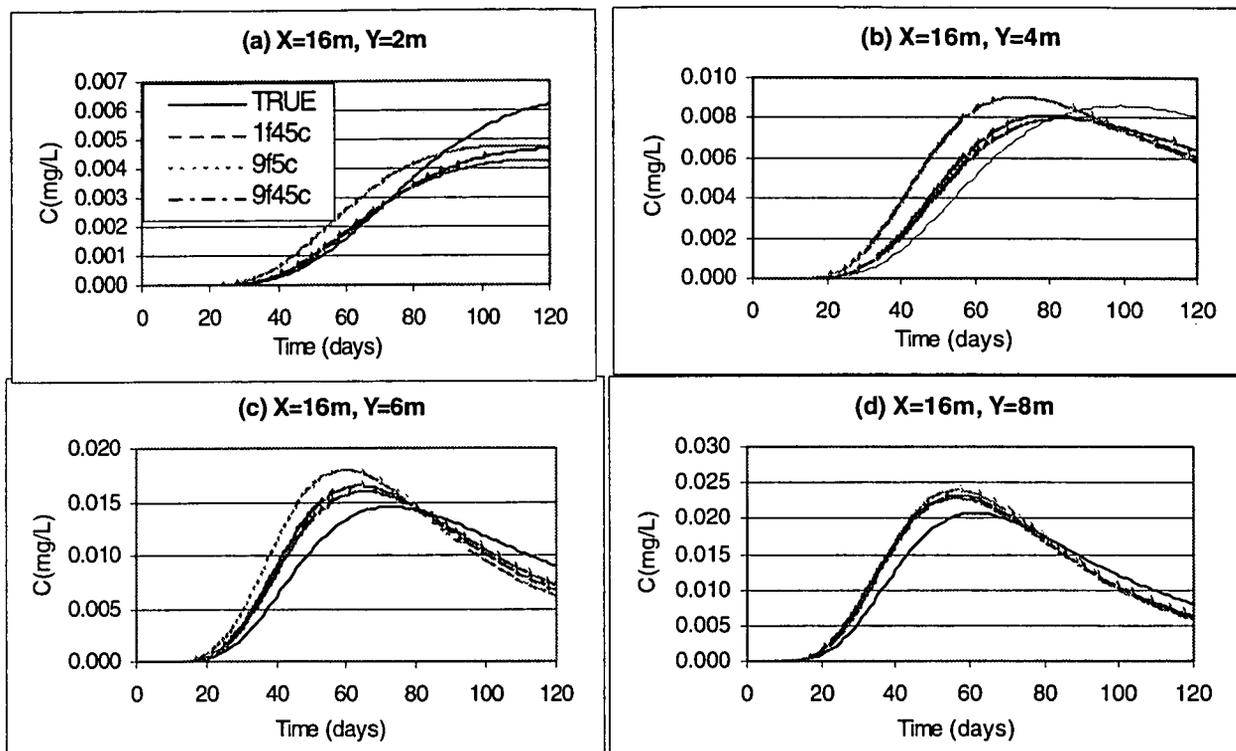


Figure 5.15 Estimation of Spatial Random K_d with Underlying Correlation Length = 4m: True and Simulated Breakthrough Curves (Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

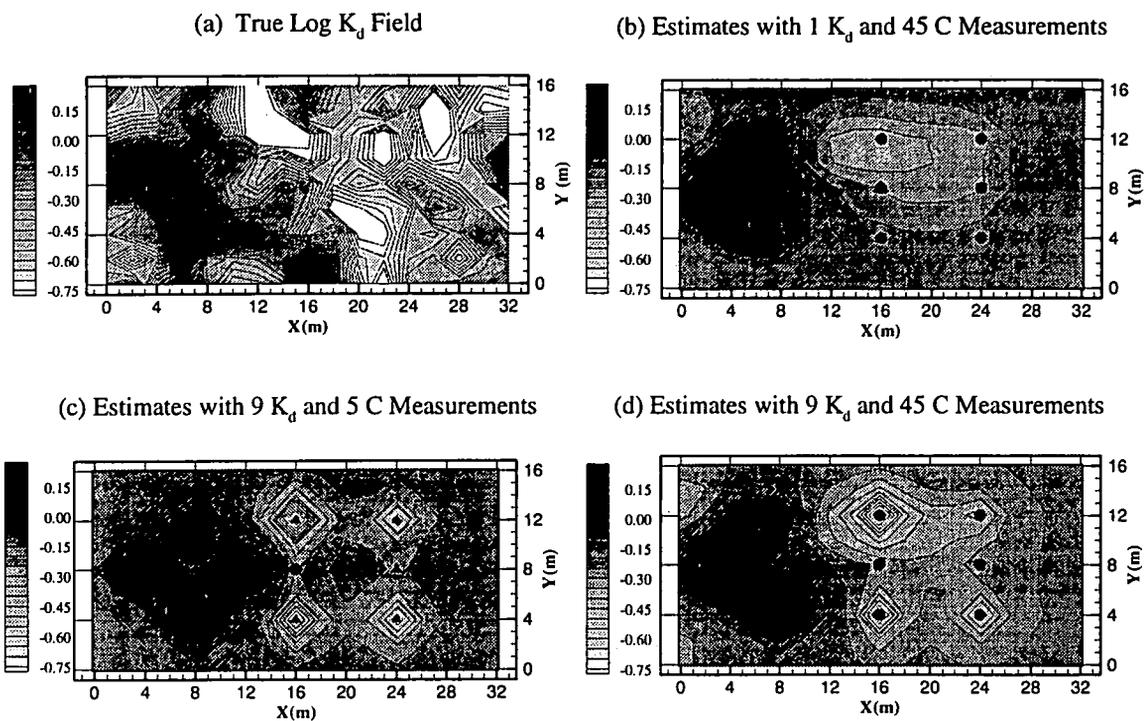


Figure 5.16 Estimation of Spatial Random K_d with Underlying Correlation Length = 8m: True and Estimated Log K_d Fields (Circles represent concentration measurement locations, and triangles K_d measurement location. Five concentration measurements were made at each concentration measurement locations. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

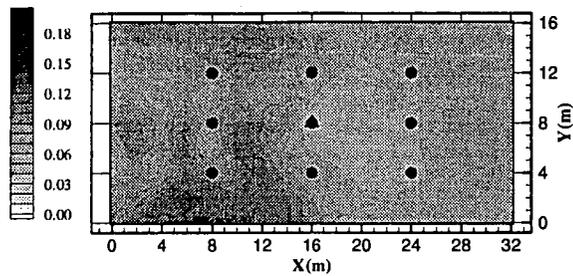
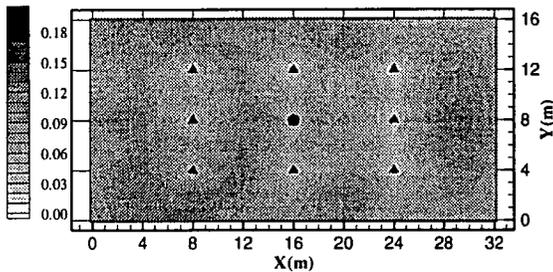
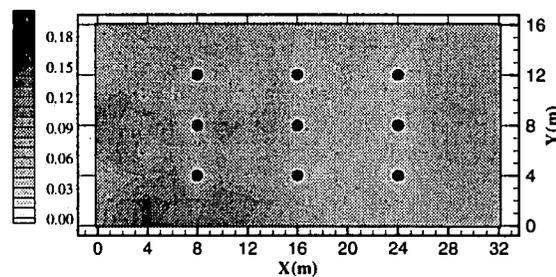
(a) 1 K_d and 45 C Measurements(b) 9 K_d and 5 C Measurements(c) 9 K_d and 45 C Measurements

Figure 5.17 Estimation of Spatial Random K_d with Underlying Correlation Length = 8m: Standard Deviations Associated with Estimated Log K_d Fields (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

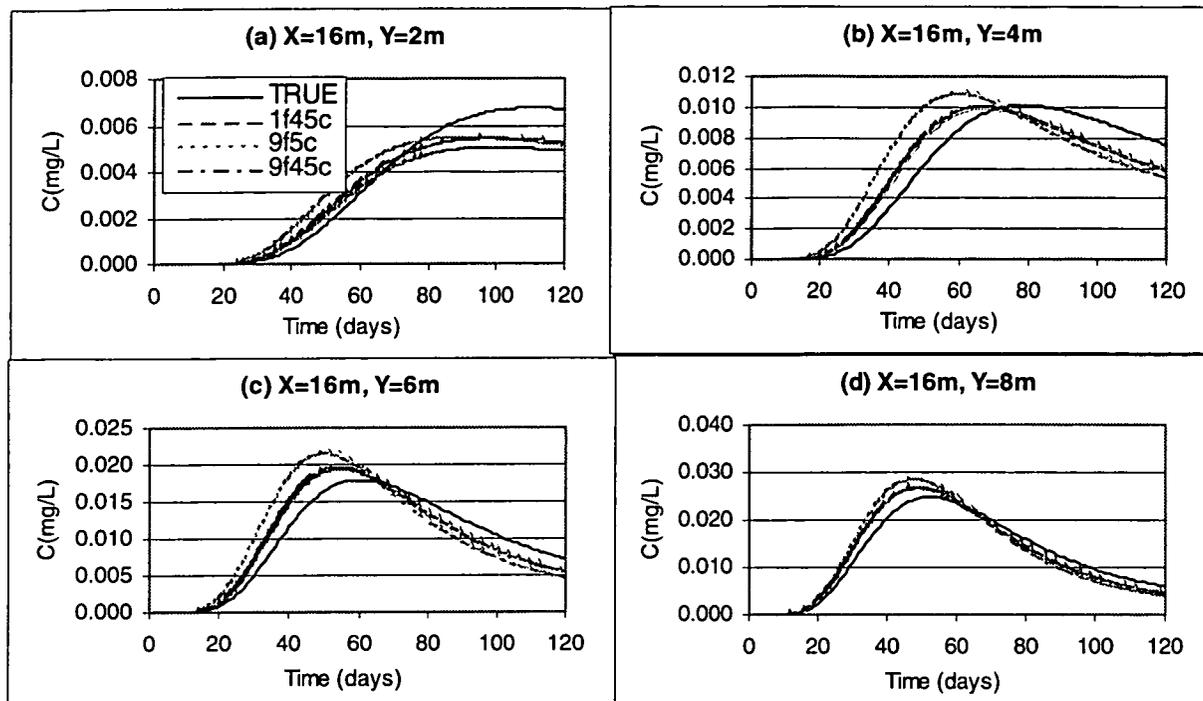


Figure 5.18 Estimation of Spatial Random K_d with Underlying Correlation Length = 8m: True and Simulated Breakthrough Curves (Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

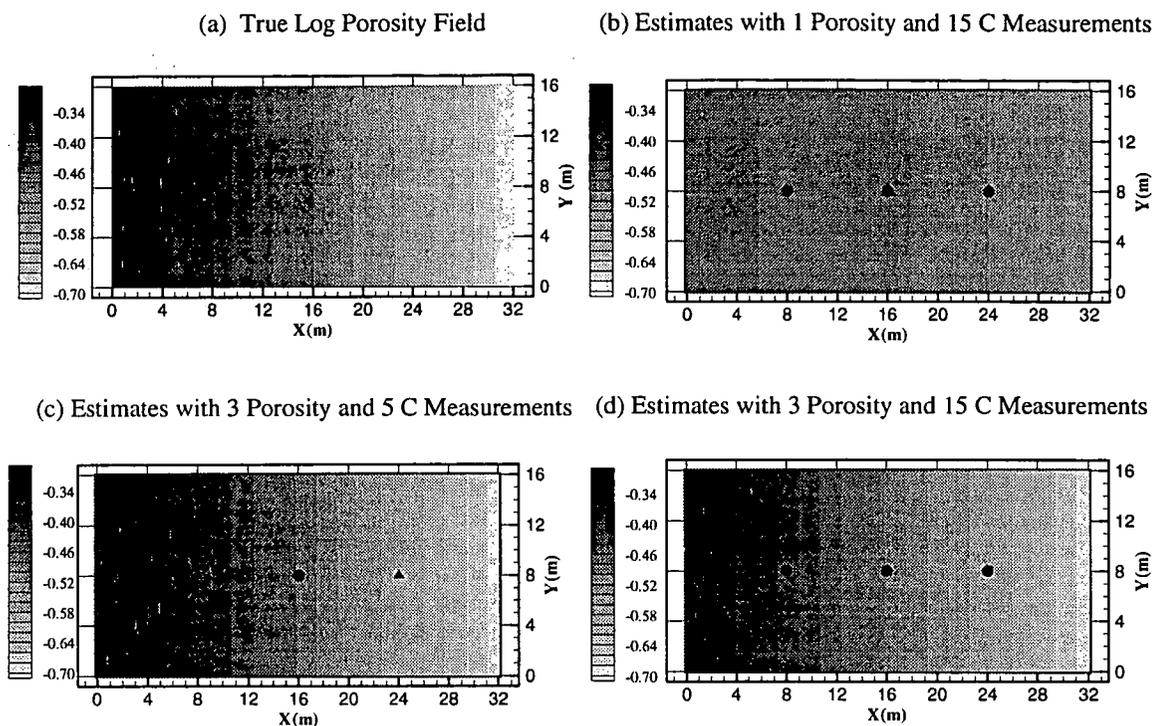


Figure 5.19 Estimation of Linear Trend in Log Porosity: True and Simulated Log Porosity (Circles represent concentration measurement locations, and triangles porosity measurement location. Five concentration measurements were made at each concentration measurement location. 1 Por 15 C implies 1 Porosity and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

000076

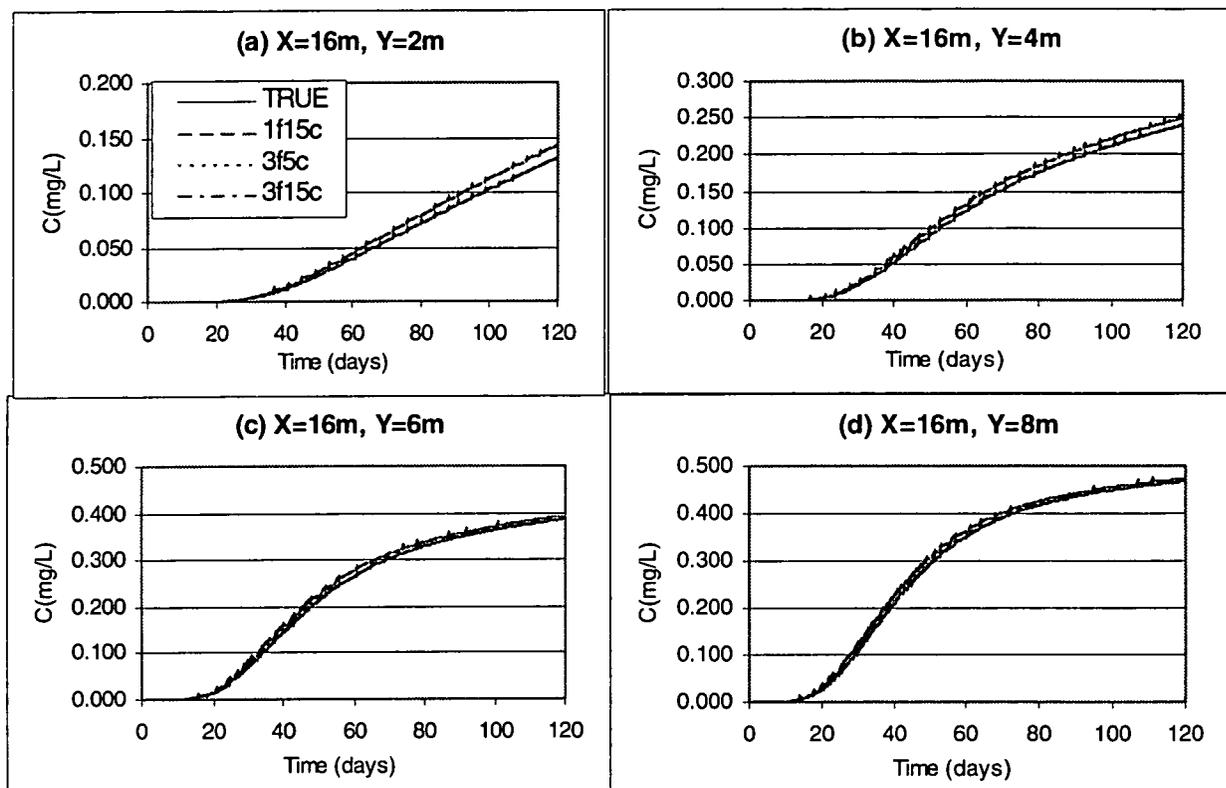


Figure 5.20 Estimation of Linear Trend in Log Porosity: True and Simulated Breakthrough Curves (Five concentration measurements were made at each concentration measurement locations. 1 Por 15 C implies 1 Porosity and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

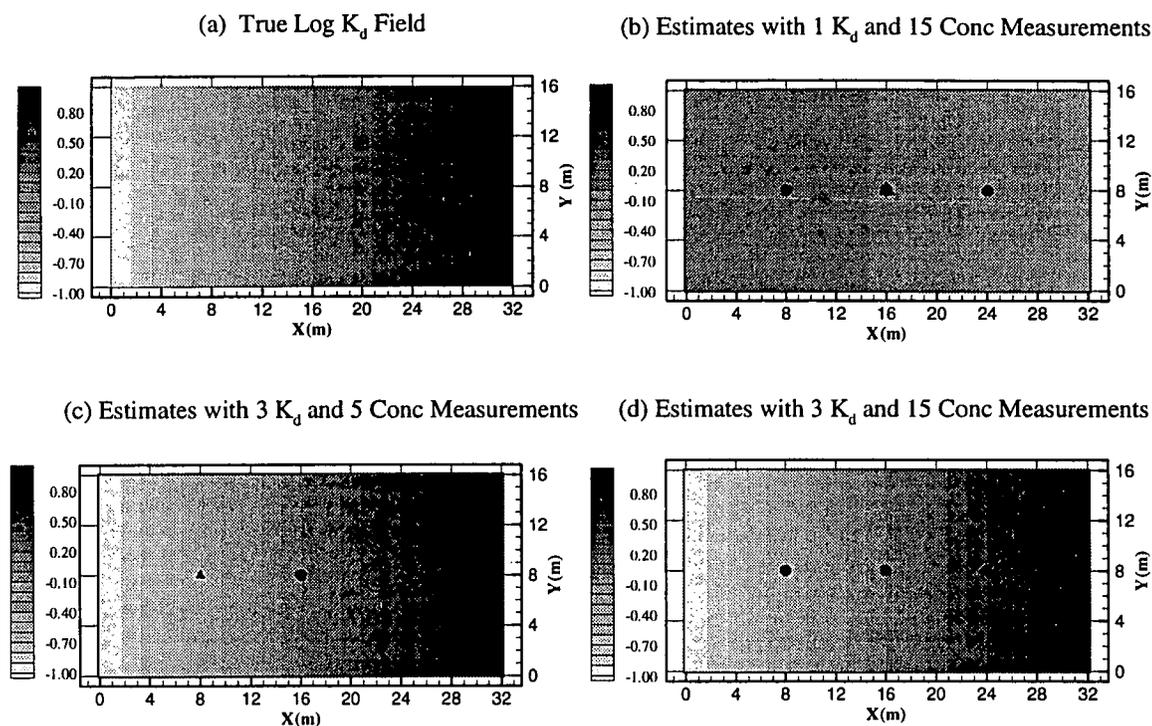


Figure 5.21 Estimation of Log K_d with a Linear Trend Based on Log Concentration Measurement Error of 0.1: True and Estimated Log K_d Fields (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 15 C implies 1 K_d and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

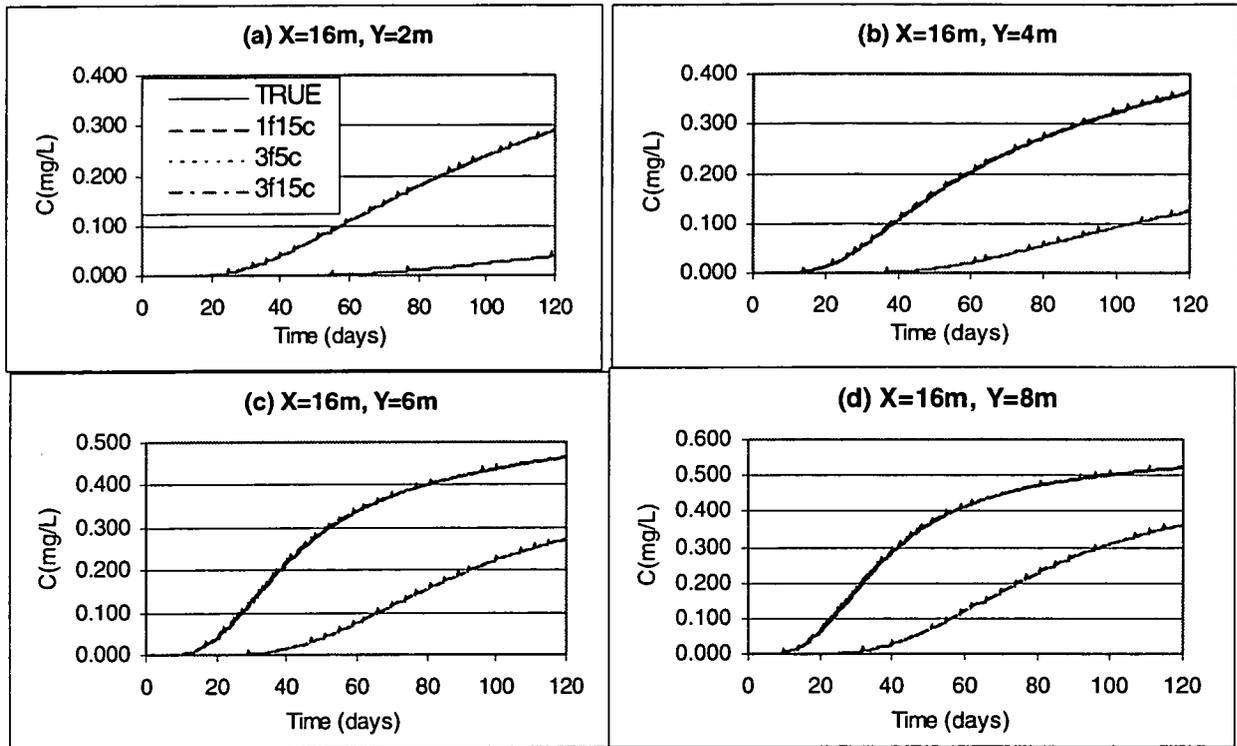


Figure 5.22 Estimation of $\text{Log } K_d$ with a Linear Trend Based on Log Concentration Measurement Error of 0.1: True and Simulated Breakthrough Curves (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. $1 K_d 15 C$ implies $1 K_d$ and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

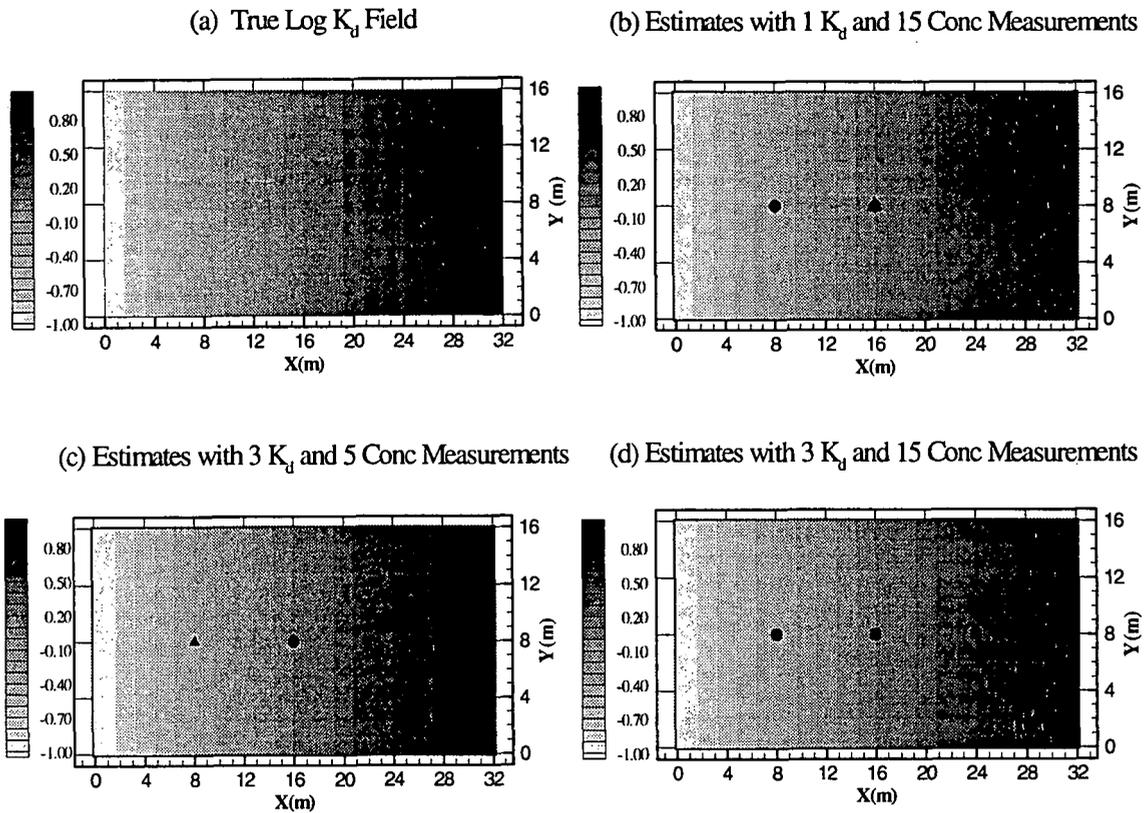


Figure 5.23 Estimation of Log K_d with a Linear Trend Based on Log Concentration Measurement Error of 0.01: True and Estimated Log K_d Fields (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 15 C implies 1 K_d and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

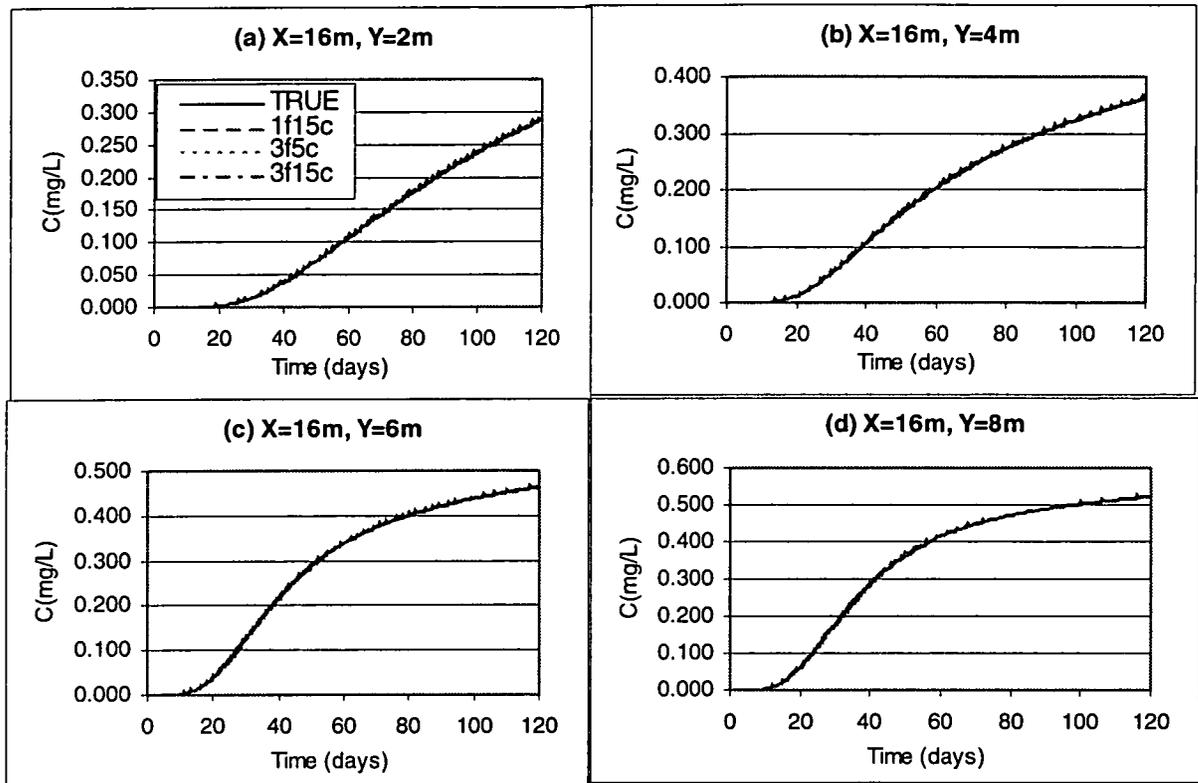


Figure 5.24 Estimation of $\text{Log } K_d$ with a Linear Trend Based on Log Concentration Measurement Error of 0.01: True and Simulated Breakthrough Curves (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 15 C implies 1 K_d and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

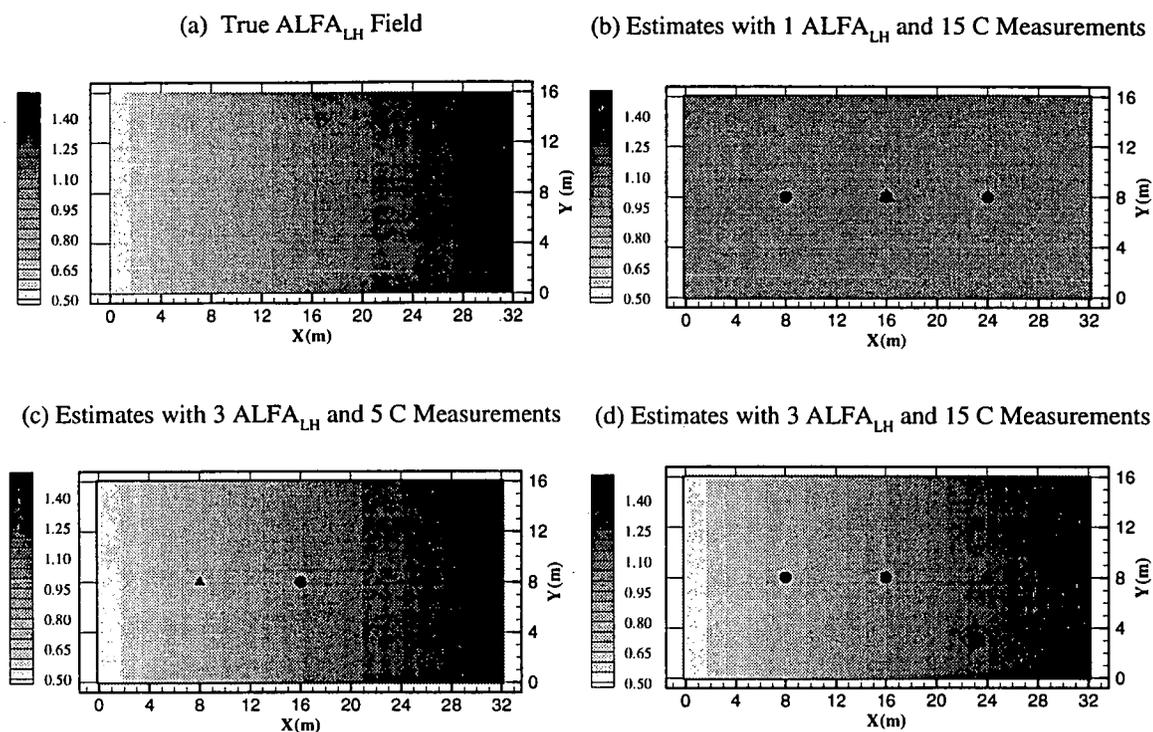


Figure 5.25 Estimation of Longitudinal Dispersivity with a Linear Trend: True and Estimated Dispersivity α_{LH} Fields (Circles represent concentration measurement locations, and triangles dispersivity measurement location. Five concentration measurements were made at each concentration measurement locations. 1 α_{LH} 15 C implies 1 dispersivity and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

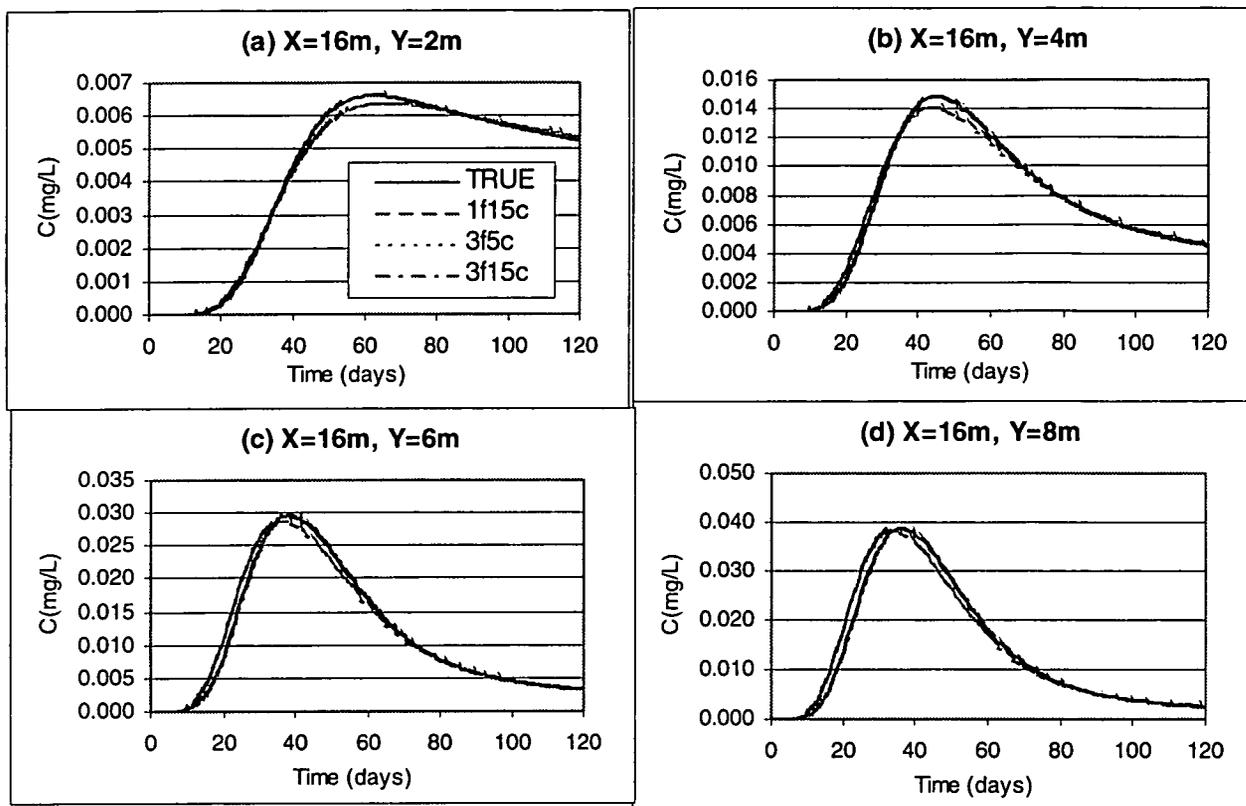


Figure 5.26 Estimation of Longitudinal Dispersivity with a Linear Trend: True and Simulated Breakthrough Curves (Circles represent concentration measurement locations, and triangles dispersivity measurement location. Five concentration measurements were made at each concentration measurement locations. $1 \alpha_{LH} 15 C$ implies 1 dispersivity and 15 (3 locations \times 5 values/each location) Concentration values, etc.).

000083

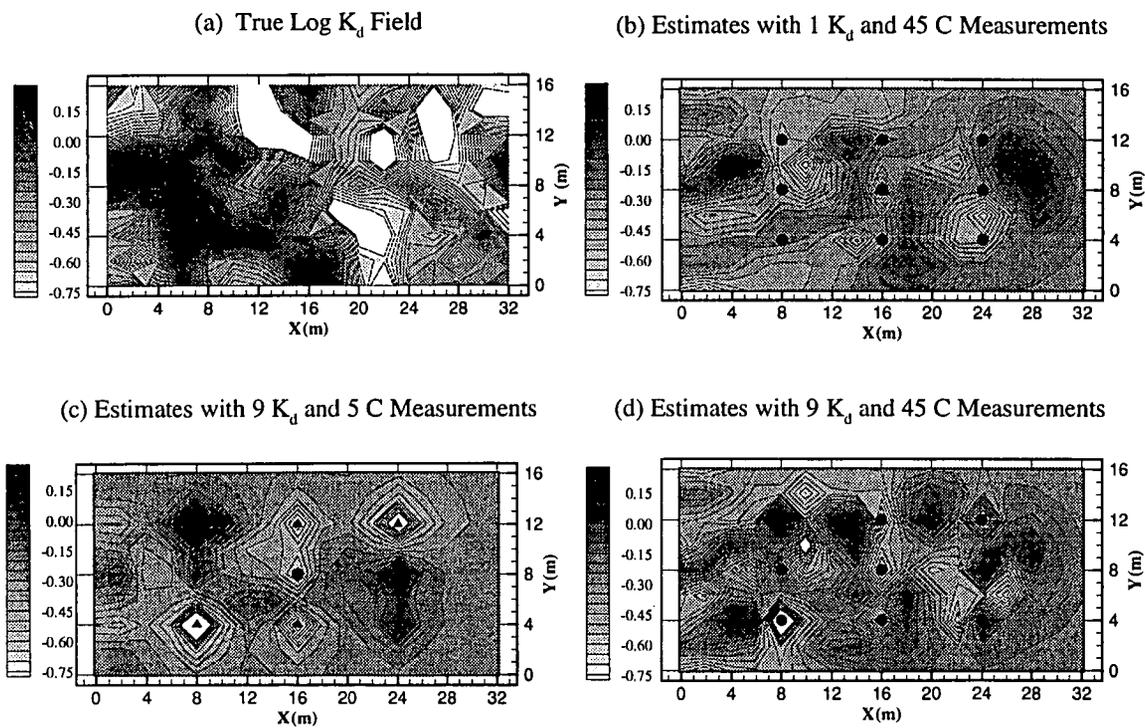
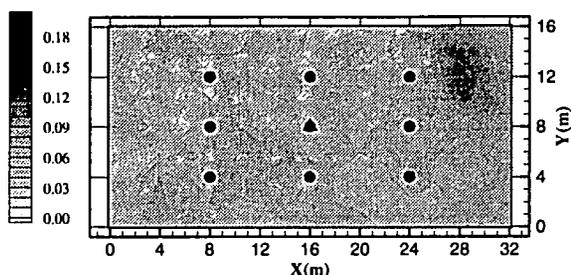
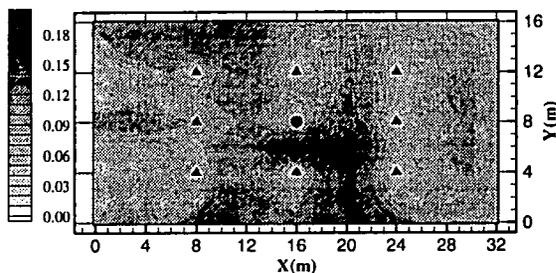


Figure 5.27 Estimation of Random Freundlich Isotherm K_d with Underlying Correlation Length = 4m: True and Estimated Log K_d Fields (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.) .

(a) 1 K_d and 45 C Measurements



(b) 9 K_d and 5 C Measurements



(c) 9 K_d and 45 C Measurements

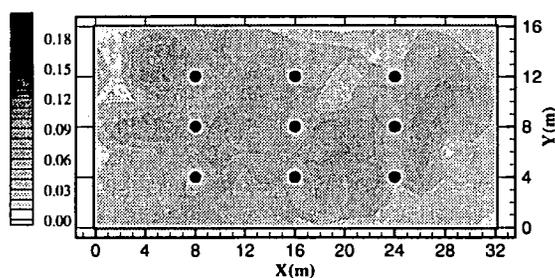


Figure 5.28 Estimation of Random Freundlich Isotherm K_d with Underlying Correlation Length=4m: Standard Deviations Associated with Log K_d Estimates (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

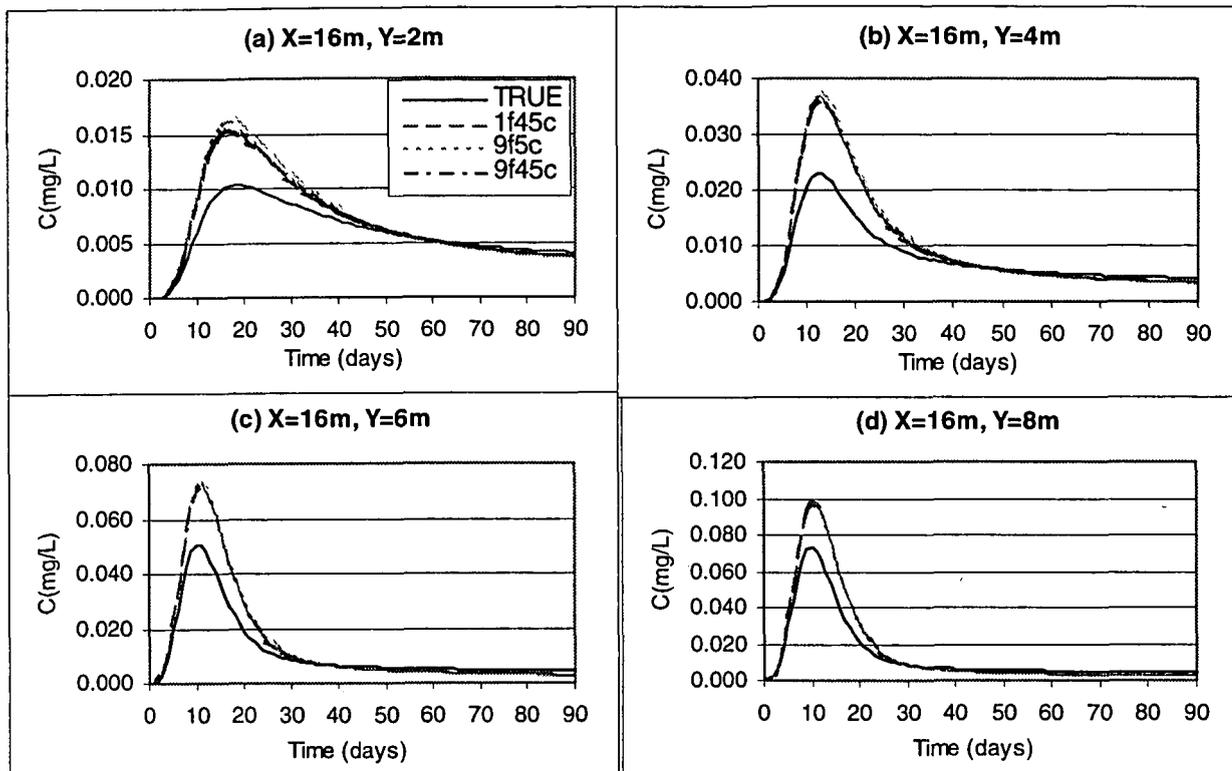
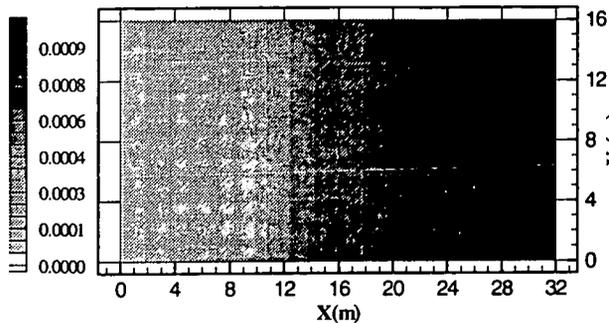


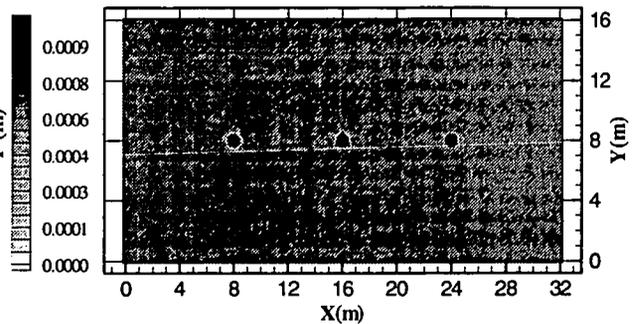
Figure 5.29 Estimation of Random Freundlich Isotherm K_d with Underlying Correlation Length=4m: True and Simulated Breakthrough Curves (Circles represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 1 K_d 45 C implies 1 K_d and 45 (9 locations \times 5 values/each location) Concentration values, etc.).

000086

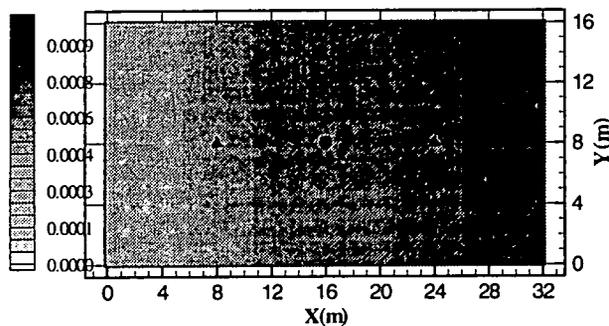
(a) True ALFA_D Field



(b) Estimates with 1 ALFA_D and 15 C Measurements



(c) Estimates with 3 ALFA_D and 5 C Measurements



(d) Estimates with 3 ALFA_D and 15 C Measurements

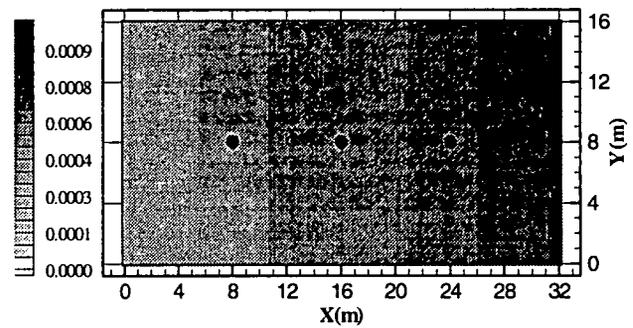


Figure 5.30 Estimation of Desorption Rate Coefficient with a Linear Trend: True and Estimated Dispersivity α_d Fields (Circles represent concentration measurement locations, and triangles dispersivity measurement location. Five concentration measurements were made at each concentration measurement locations. 1 α_d 15 C implies 1 desorption rate and 15 (3 locations \times 5 values/each location) Concentration values, etc.)

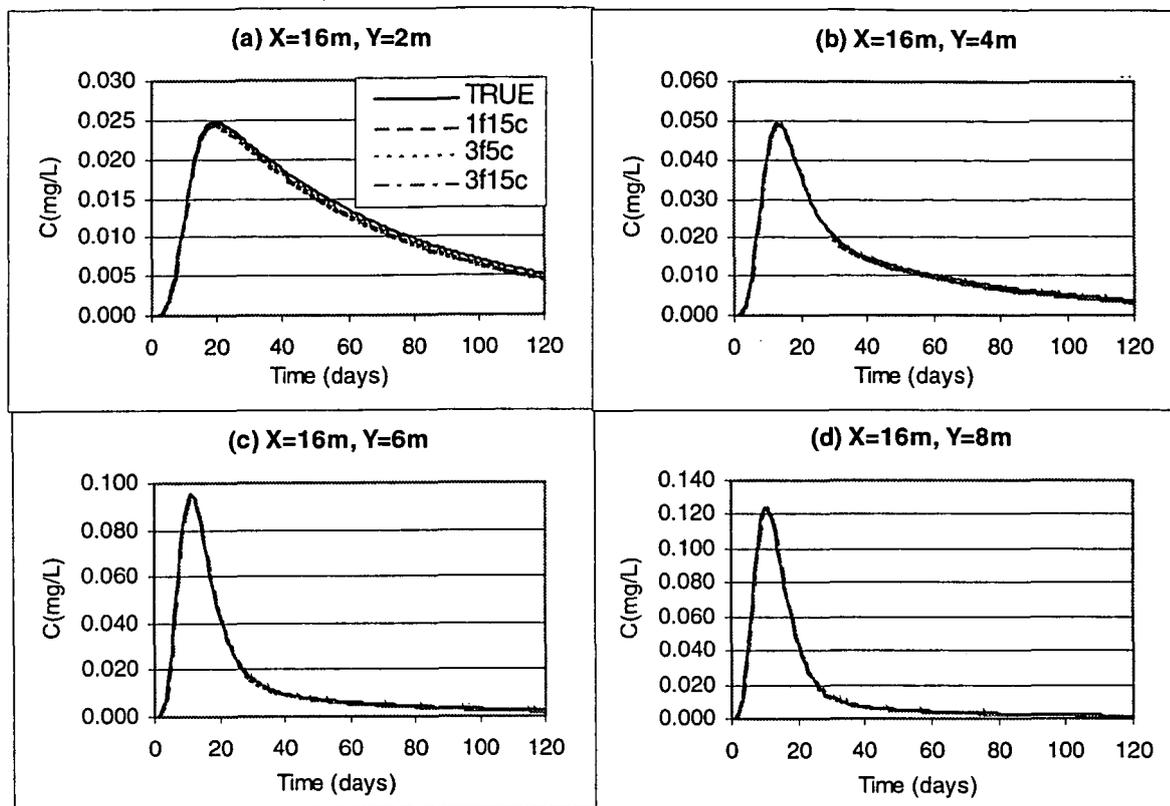
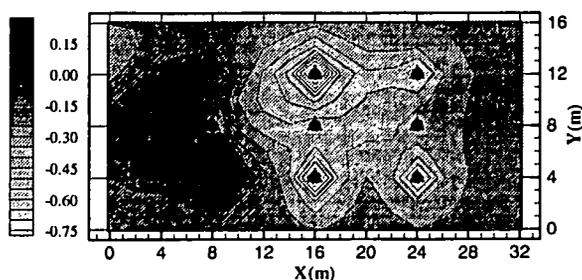
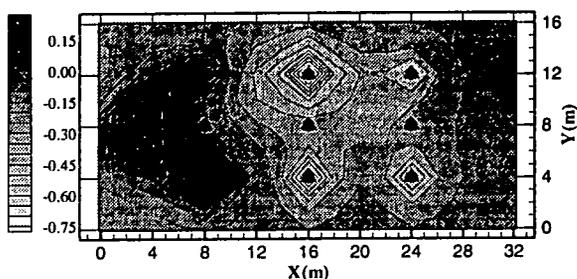


Figure 5.31 Estimation of Desorption Rate Coefficient with a Linear Trend: True and Simulated Breakthrough Curves (Circles represent concentration measurement locations, and triangles desorption rate measurement location. Five concentration measurements were made at each concentration measurement locations. $1 \alpha_d 15 C$ implies 1 desorption rate and 15 (3 locations \times 5 values/each location) Concentration values, etc.)

(a) Estimates with 9 K_d and 36 C Measurements



(b) Estimates with 9 K_d , 36 C and EKFU 9 C



(c) Estimates with 9 K_d and 45 C Measurements

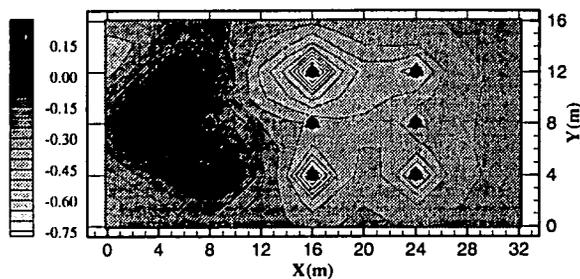
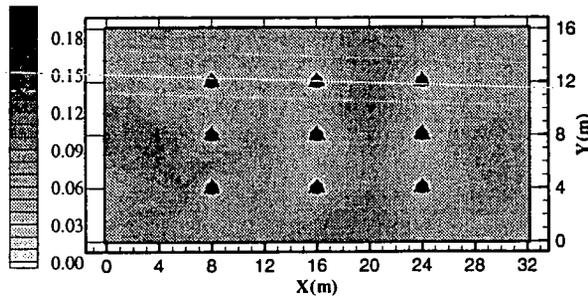


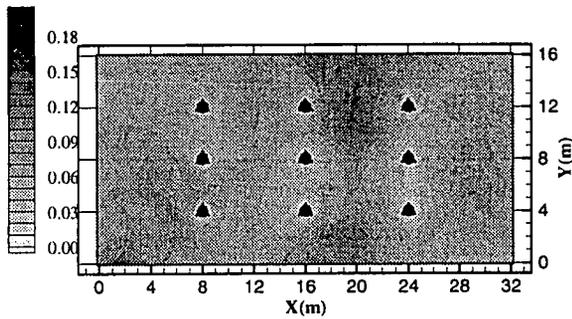
Figure 5.32 Estimation of Spatial Random K_d with Underlying Correlation Length = 8m: Estimated Log K_d Fields (Circle represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 9 K_d 45 C implies 9 K_d and 45 (9 locations X 5 values/each location) Concentration values, etc.).

000089

(a) 9 K_d and 36 C Measurements



(b) 9 K_d , 36 C and EKFU 9 C



(c) 9 K_d and 45 C Measurements

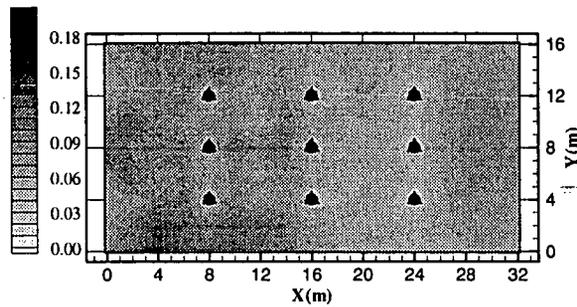


Figure 5.33 Estimation of Spatial Random K_d with Underlying Correlation Length = 8m: Standard deviations Associated with Estimated Log K_d Fields (Circle represent concentration measurement locations, and triangles K_d measurement locations. Five concentration measurements were made at each concentration measurement location. 9 K_d 45 C implies 9 K_d and 45 (9 locations X 5 values/each location) Concentration values, etc.).

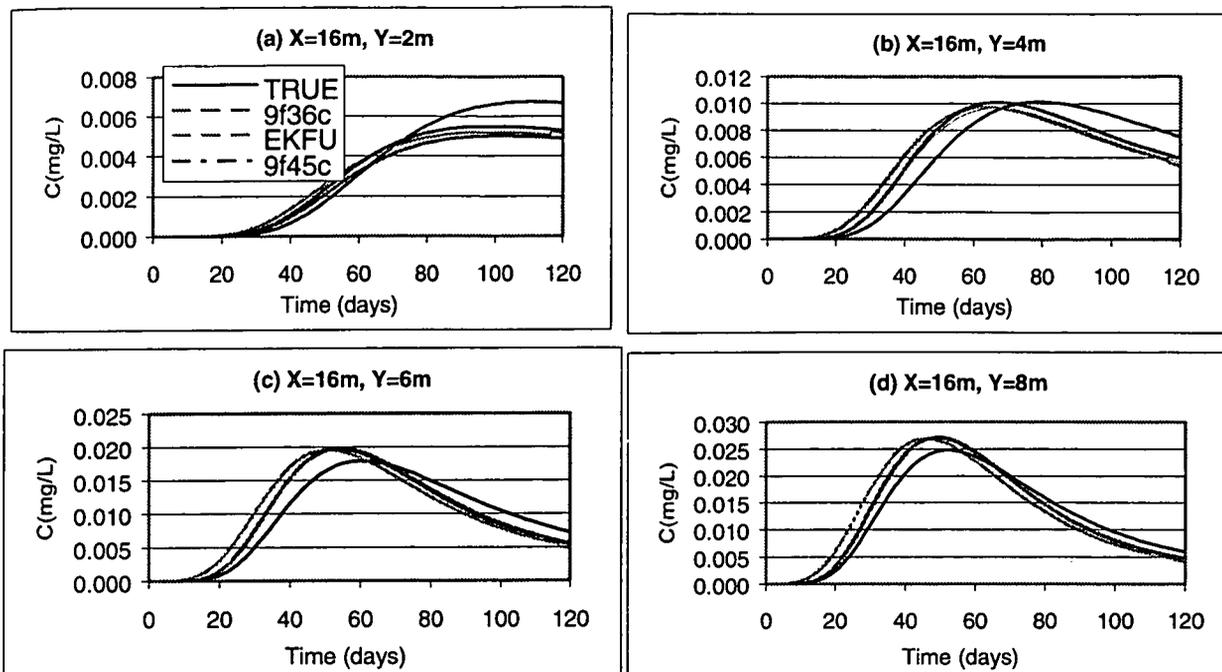


Figure 5.34 Estimation of Spatial Random K_d with Underlying Correlation Length = 4m: Simulated Breakthrough Curves (Five concentration measurements were made at each concentration measurement location. 9 K_d 45 C implies 9 K_d and 45 (9 locations X 5 values/each location) Concentration values, etc.).

6.0 SITE-SPECIFIC APPLICATION EXAMPLE: FERNALD SITE

6.1 INTRODUCTION

The DFM/VAM3DF code was applied to a local area enclosing the South Plume and the South Field areas of the Fernald Site (Figure 6.1). A local model was developed for this area to demonstrate the applicability of the DFM/VAM3DF code in identifying transport parameters by fitting total uranium concentrations from monitoring wells, extraction wells, and geoprobe data from the beginning of 1994 through 1999. Before the DFM/VAM3DF results are discussed, an overview of the data fusion modeling approach is given.

6.2 DATA FUSION MODELING METHODOLOGY

The DFM combines field data with the flow and transport model in an attempt to estimate unknown model parameters. Below are the general steps in the DFM process:

- 1) **Physical Model Description:** The flow and transport model must be constructed which represents site conditions and is consistent with field data.
- 2) **Sensitivity Analysis:** Prior to applying the DFM/VAM3DF code, sensitivity analysis should be performed to determine sensitive parameters and to identify potential parameters to be estimated.
- 3) **Statistical Model Description:** The estimated parameters, statistical grid, and prior statistics must be specified.
- 4) **DFM/VAM3DF Input:** The statistical model information and field data must be specified in the format given in Appendix A.
- 5) **Run DFM/VAM3DF**
- 6) **Error Analysis:** The spatial and temporal investigation of data fit errors (residuals) is the key to improving the DFM/VAM3DF parameter estimates. The parameter estimates must be examined to determine whether estimates are reasonable. If the data fit errors are acceptable, steps 7 and 8 may be omitted.
- 7) **Adjust Physical and Statistical Models:** The results of the error analysis may suggest improvements in the flow and transport model. Often, changes in the statistical model are necessary to improve parameter estimates.
- 8) **Go to Step 5.**

After the above DFM process is complete, Monte Carlo analysis may be performed to estimate concentration uncertainty from parameter uncertainty. The DFM/VAM3DF Monte Carlo procedure is described below.

000092

- 1) Produce multiple Monte Carlo parameter realizations, which are conditional on the physical model, prior statistics, and field data.
- 2) Run VAM3DF for each Monte Carlo realization.

For a user-defined output time, DFM/VAM3DF will output the resulting concentration variance, maximum concentration histogram, and maximum concentration cumulative distribution function.

6.3 TOTAL URANIUM CONCENTRATION DATA

Total uranium concentrations from monitoring-well, extraction-well, and geoprobe data from the beginning of 1994 through 1999 were used as calibration targets for DFM/VAM3DF. The South Plume and the South Field areas of the Fernald Site contains 68 2000-series wells, 50 3000-series wells, 6 4000-series wells, 16 extraction wells, and 34 geoprobe locations. The geoprobe data are from 10/23/96 to 5/13/99 while the monitoring well and extraction well data span the simulation period 1994-1999.

6.4 MODEL DESCRIPTION

6.4.1 Discretization

A localized model (Figure 6.1) of South Plume and the South Field areas was created by extracting information from the existing Phase I site model (HydroGeoLogic, 1998). The local model, with 38 by 74 cells, was refined to 100 feet in the x-y plane from 125 feet in the site model. In the vertical direction, the local model has 14 nodal layers, 12 nodal layers from the site model plus 2 addition nodal layers for better resolution in (Figure 6.2).

Two flow fields were constructed. The first flow field included the South Plume extraction wells pumping before July 1998 and a second flow field included the remaining extraction and reinjection wells turned on after July 1998. Total simulation time was six years from the beginning of 1994 through 1999.

6.4.2 Initial Uranium Concentrations and Boundary Conditions

Initial uranium concentrations were assumed to be an average of all 1993 sampling data at each monitoring well with water quality data in 1993. To create the three-dimensional initial concentrations for VAM3DF, the GSLIB geostatistical software (Deutsch and Journal, 1998) was utilized. The average 1993 concentrations were kriged onto a grid with 50 by 76 cells with 100-foot grid spacing and with a uniform 20-foot vertical spacing, slightly bigger than the local model grid. The kriging semivariogram model was similar to the model used to krig the 1993 uranium concentration data in DOE (1994). The kriging results were interpolated onto the local model mesh and then written in the VAM3DF input format. Figures 6.3, 6.4 and 6.5 portray the kriged average 1993 concentrations for the 2000-, 3000-, and 4000-series wells, respectively (the average concentration is posted below the well).

Constant head boundaries on the local model were interpolated from the site model head solution. Extraction wells, reinjection wells, and recharge flux were mapped from the Phase I site model to the local model. For contaminant transport, mass flux source nodes inside the local model were interpolated from the Phase I site transport model.

6.5 SENSITIVITY ANALYSIS

6.5.1 Sensitivity Description

Limited sensitivity analysis was performed to determine the model sensitivity to certain input parameters. These results were used to assist in the DFM/VAM3DF modeling.

A baseline simulation was performed using the transport parameters from the Phase I site model (HydroGeoLogic, 1998). Table 6.1 presents the baseline transport parameters. Each sensitivity simulation involved changing one input parameter and analyzing differences in overall concentration using two-dimensional plume footprints and breakthrough curves at monitoring and extraction wells. Sampling, or field information, for each extraction well and monitoring well was overlaid on simulated results to determine whether model results are consistent with field conditions. For the two-dimensional footprints, the concentrations at approximate initial conditions (1 day) and in July 1998 (1,641 days) were contoured with average sampling data posted on each plot, representing observed versus simulated conditions. Figures 6.6 and 6.7 represent the two-dimensional plume footprints at initial conditions and in July 1998 for the baseline simulation, respectively. Layer 11 was chosen for initial conditions since the highest concentrations reside in this layer. For plots containing July 1998 results, layer 10 was chosen since this layer contains more data than any other layer for this sampling date.

6.5.2 Sensitivity Results

Sensitivity simulation runs are listed in Table 6.2. Thirteen sensitivity simulations were conducted. Results of the simulation runs are presented in Figures 6.8 to 6.20 (see Table 6.2 for details). In these figures, concentration distributions at 1,641 days (7/1/98) were obtained from layer 11. The sensitivity run 1 (K_d), sensitivity run 2 (K_d without kinetics), sensitivity runs 4 (chemisorption rate), and sensitivity run 6 (Freundlich exponent) vary significantly from the baseline concentrations. In Figure 6.11, increasing the chemisorption rate resulted in lower groundwater concentration. A consistency check for the time-step size was conducted in sensitivity runs 11 and 12, by approximately doubling and halving the baseline time-step size, respectively. Concentration distributions shown in Figure 6.18 (double time-step size) are almost identical to the baseline concentration distribution in Figure 6.7, indicating a good degree of self consistency and the appropriateness of the chosen time-step size.

All breakthrough curves are shown for two extraction wells and two monitoring wells in Figures 6.21-6.24. Table 6.3 summarizes the variation from the baseline concentrations for extraction wells 31550 and 31560. The variation from the baseline is measured by summing the absolute value of the difference in concentration over all time steps. The simulated uranium concentrations are most sensitive to the Freundlich exponent, chemisorption rate, and K_d . There is minimal sensitivity to changes in the vertical grid refinement, dispersion, effective porosity, upstream

weighting, and number of time steps. When the kinetic mass transfer was not used in sensitivity run 3, variation from the baseline breakthrough curves was small. Preliminary transport simulations showed that the baseline concentrations (which uses kinetic mass transfer) can be duplicated without kinetic mass transfer if chemisorption is simulated as first-order decay.

In the breakthrough curve figures, most of the sensitivity runs result in smooth (almost linear) breakthrough curves. Since all of the sensitivity runs except sensitivity run 6 (Freundlich exponent) involve linear transport simulations, scaling the initial concentrations will shift the breakthrough curves up or down without changing their shape. This suggests that initial concentrations will have varying degree of impact on the breakthrough curves, depending on the magnitude of initial concentration errors. For example, if the initial concentration were scaled up to 300 ppb surrounding extraction well 31560, then the baseline simulation would very closely mimic the field data (ignoring the short transient response at 10/28/95).

6.6 STATISTICAL MODEL

The sensitivity results suggest that the Freundlich exponent, chemisorption rate, K_d , and initial concentrations have a significant impact on the simulation of uranium migration. The small-scale testing in Section 5.0 shows that direct measurements have a significant influence on the accuracy of the DFM/VAM3DF estimates. Section 6.4.2 presented direct measurements of the initial total uranium concentrations (1993 average concentrations) while there are no direct measurements of the Freundlich exponent and chemisorption rate. The initial concentrations were estimated by kriging the 1993 average monitoring well concentrations in Section 6.4.2, but the post-1993 geoprobe data may contain concentration information not captured by the 1993 monitoring wells. This suggests that DFM/VAM3DF should be used to enhance the estimation of the 1993 initial concentrations. There are limited site data for K_d . DFM/VAM3DF was used to estimate K_d on an MRF. There is little information on the Freundlich exponent and chemisorption rate. Sensitivity analysis within the DFM process may be used to study the Freundlich exponent and chemisorption rate.

The geostatistical model for the three-dimensional spatial variation used one hydrostratigraphic unit so that the entire model has the same statistical model. The assigned MRF standard deviation and correlation lengths are consistent with spatial variability of the 1993 initial concentrations in Section 6.4.2. The MRF standard deviation is $0.3 \log_{10}(\text{nano lb/ft}^3)$ and the correlation lengths are $(x, y, z) = (1000 \text{ ft}, 1000 \text{ ft}, 40 \text{ ft})$. The parameter grid has 10 non-uniform nodal layers corresponding to 10 layers (out of 14 layers) of the local model VAM3DF grid. Figure 6.25 shows a slice of the VAM3DF grid and labels the 10 nodal layers of the parameter grid. The parameter grid has the same (x, y) mesh as the VAM3DF numerical grid in Section 6.4.1. The parameter grid has 28,120 ($38 \times 74 \times 10$) nodes.

6.7 FIELD MEASUREMENTS

The direct measurements of the initial total uranium concentrations (1993 average concentrations) are from Section 6.4.2, Figures 6.3, 6.4, and 6.5. Table 6.4 lists the number of direct measurements by well series.

In order to limit runtime, a subset of the monitoring wells was used for concentration measurements (historical concentration data). The selected monitoring wells were sampled in 1993 and sampled several times from 1994 to 1999. The 2000- and 3000- series monitoring wells that were used for concentration measurements are identified in Figures 6.26 and 6.27, respectively. All extraction wells inside the local model were used for concentration measurements. All geoprobe data inside the local model were used as concentration measurements. The extraction wells and geoprobe data that were used for concentration measurements are identified in Figures 6.28 and 6.29, respectively. Table 6.5 lists the wells used for concentration measurements, and the wells sampled between 1994 and 1999 but not used as concentration measurements. The percentages of 2000-series wells, 3000-series wells, and 4000-series wells used for concentration measurements were 57%, 61%, and 33%, respectively.

A subset of the monitoring well data and extraction well data was used as concentration measurements. To further reduce computational burden, each monitoring well and extraction well breakthrough curve was approximated by two control points. The two control points that approximate the trend of the breakthrough curve are used as concentration measurements. Temporal fluctuations in the observed breakthrough curves were considered to be due to small-scale localized heterogeneity and short-term fluctuations in climatologic and hydrologic conditions. Only the salient characteristics of the migration of uranium were captured by the model. Table 6.4 lists the number of concentration measurements by monitoring wells, extraction wells and geoprobes.

The prior error standard deviations for the measurements represent the uncertainties or errors in the measurements. Since these uncertainties are unknown, initial standard deviations are assigned and DFM/VAM3DF is executed. The standard deviations are tuned by performing sensitivities on the DFM/VAM3DF data fit error. This requires adjusting the prior error standard deviations and running DFM/VAM3DF multiple times. For this investigation, the prior error standard deviations were not tuned. The prior error standard deviations for all 109 direct measurements of the initial concentrations were assigned an identical value of $0.05 \log_{10}(\text{nano lb/ft}^3)$. The prior error standard deviations for all 479 concentration measurements were assigned a value of $0.07 \log_{10}(\text{nano lb/ft}^3)$.

6.8 DFM/VAM3DF APPLICATION

6.8.1 Baseline Simulations

The baseline simulations were performed using the transport parameters from the Phase I site model. For the DFM/VAM3DF application, the transport simulation time was 13.5 years from the beginning of 1994 through June 2006. Three steady-state flow fields were used with pumping from Table 5-1 of the Baseline Remedial Strategy Report (DOE, 1997). The initial uranium concentrations were estimated by kriging the 1993 average monitoring well concentrations in Section 6.4.2. The uranium source terms were acquired from Fluor Fernald personnel. Figures 6.30-6.32 show the baseline uranium concentrations in parts per billion for layer 11 at 1/1/98, 1/1/02, and 5/1/06, respectively. Layer 11 is in between the 2000-series well and 3000-series well horizons. Figure 6.32 shows the simulated 5/1/06 uranium concentrations are slightly above 20 ppb in layer 11. Figures 6.33-6.35 show the baseline uranium concentrations in parts per billion

000096

for layer 13 at 1/1/98, 1/1/02, and 5/1/06, respectively. Layer 13 corresponds to the 2000-series well horizon. Figure 6.35 shows the simulated 5/1/06 uranium concentrations are slightly above 50 ppb in layer 13.

Two sensitivity simulations were performed based on the sensitivity results in Section 6.5.2. The chemisorption rate was decreased from 10^{-4} (1/day) to 10^{-6} (1/day) and the Freundlich exponent was decreased from 1 to 0.5. Figure 6.36 shows the maximum aqueous uranium concentrations for the baseline and the two sensitivity runs. The baseline, chemisorption sensitivity run, and Freundlich exponent sensitivity run, maximum concentrations at 7/1/06 are: 82 ppb, 128 ppb, and 26 ppb, respectively. Notice that the maximum concentration curves decrease rapidly when the extraction/reinjection system changes at 7/1/98 and 1/1/04. Figure 6.37 shows the uranium mass in the groundwater, uranium mass absorbed, and uranium mass chemisorbed for the baseline simulation. Figures 6.38 and 6.39 show the mass summary for the chemisorption sensitivity run and Freundlich exponent sensitivity run, respectively. Notice that decreasing the chemisorption rate increased the mass absorbed and decreased the mass chemisorbed. The Freundlich exponent sensitivity run resulted in less mass absorbed and less mass chemisorbed. The concentration measurements in Section 6.7 were compared with the baseline simulations. Table 6.6 shows the residual summary for the baseline, chemisorption sensitivity run, and Freundlich exponent sensitivity run. The residual is defined as measured concentration - simulated concentration. The chemisorption sensitivity run did not affect the residuals while the Freundlich exponent sensitivity run resulted in larger residuals.

6.8.2 Estimation of the 1993 Initial Uranium Concentrations

Two DFM/VAM3DF simulations estimating $\log_{10}(\text{IC})$ were performed. Each simulation used an MRF on the parameter grid with a constant polynomial trend. The prior trend was $-4.0 \log_{10}$ (nano lb/ft³) with a prior error standard deviation of $0.01 \log_{10}$ (nano lb/ft³). The MRF standard deviation was $0.3 \log_{10}$ (nano lb/ft³) and correlation lengths were from Section 6.6. The simulations used direct measurements and concentration measurements from Section 6.7. The first simulation used the baseline parameters while the second simulation used a chemisorption rate of 10^{-6} (1/day). Preliminary simulations showed that DFM/VAM3DF Gauss-Newton would converge in two iterations and the second iteration gave slightly improved estimates compared to the first iteration estimates. In order to reduce runtimes, only one Gauss-Newton step was used for the DFM/VAM3DF simulations.

The results of IC estimation run 1 for parameter grid layer 5, which corresponds to the 3000-series well horizon, are shown in Figure 6.40. Notice that the estimates are similar to the kriged monitoring well data in Figure 6.4. Figure 6.41 shows the IC estimation run 1 results for parameter grid layer 8, which corresponds to the 2000-series well horizon. Again, the estimates are similar to the kriged monitoring well data in Figure 6.3. IC estimation run 2 gave very similar estimates. Table 6.7 shows the range of the estimates and Table 6.8 shows the maximum concentration and mass summary at 7/1/06. Table 6.9 shows a residual summary. A comparison with the baseline simulation residuals in Table 6.6 shows that the DFM/VAM3DF concentration measurement residuals are larger. The negative average residual suggests that the DFM/VAM3DF concentrations are too high. The results may be improved by iterating steps 6) through 8) of the DFM process of Section 6.2.

6.8.3 Estimation of the Distribution Coefficient

Three DFM/VAM3DF simulations estimating $\log_{10}(K_d)$ were performed. Each simulation used an MRF on the parameter grid with a simple linear trend ($C_0 + C_1x + C_2y + C_3z$). For all three simulations, the prior constant trend was $-1.5 \log_{10}(\text{lb}/\text{ft}^3)$ with a prior error standard deviation of $1.0 \log_{10}(\text{lb}/\text{ft}^3)$ and the remaining coefficients had a prior estimate of $0.0 \log_{10}(\text{lb}/\text{ft}^3)$ and a prior error standard deviation of $0.05 \log_{10}(\text{lb}/\text{ft}^3)$. The MRF standard deviation was $0.05 \log_{10}(\text{lb}/\text{ft}^3)$ and correlation lengths were from Section 6.6. The simulations used concentration measurements only. To reduce computational burden, the VAM3DF transport model was used without kinetic mass transfer and chemisorption was simulated as first-order decay (as discussed in Section 6.5.2). The first simulation used the initial uranium concentrations from Section 6.4.2. The second simulation used the initial concentrations from the IC estimation run 1 and the third simulation used the initial concentrations from the IC estimation run 2 and a chemisorption rate of 10^{-6} (1/day). Preliminary simulations showed that DFM/VAM3DF Gauss-Newton would converge in two iterations and the second iteration gave slightly improved estimates compared to the first iteration estimates. In order to reduce run times, only one Gauss-Newton step was used for the DFM/VAM3DF simulations.

The results of K_d estimation run 1 for parameter grid layer 5 (which corresponds to the 3000-series well horizon) are shown in Figure 6.42. Figure 6.43 shows the K_d estimation run 1 results for parameter grid layer 8, which corresponds to the 2000-series well horizon. The results of K_d estimation run 2 for parameter grid layer 5 are shown in Figure 6.44. Figure 6.45 shows the K_d estimation run 2 results for parameter grid layer 8. Notice that K_d estimation run 2 gave higher K_d s than K_d estimation run 1. Both estimates have large K_d s south of Willey Rd. K_d estimation run 3 gave very similar estimates in Figures 6.46 and 6.47. Table 6.7 shows the range of the K_d estimates and Table 6.8 shows the maximum concentration and mass summary at 7/1/06. The lower chemisorption rate in K_d estimation run 3 gave a higher maximum concentration. Table 6.10 shows a residual summary. A comparison with the baseline simulation residuals in Table 6.6 shows that the DFM/VAM3DF concentration measurement residuals are smaller. Table 6.11 shows the estimated trend coefficients and Table 6.12 shows the SOS summary. Notice that K_d estimation run 1 has the lowest SOS. Figures in Appendix B compare the DFM/VAM3DF results with the monitoring-well, extraction-well, and geoprobe concentrations. The results may be improved by iterating steps 6) through 8) of the DFM process in Section 6.2.

6.8.4 Uranium Migration Response to K_d Estimates

In this section, the change in uranium concentrations due to the estimated K_d fields will be investigated. Figures 6.48-6.50 show the K_d estimation run 1 uranium concentrations in parts per billion for layer 11 at 1/1/98, 1/1/02, and 5/1/06, respectively. Layer 11 is located between the 2000-series well and 3000-series well horizons. Figure 6.50 shows the simulated 5/1/06 uranium concentrations are above 20 ppb in layer 11. These concentrations are higher than the baseline concentrations in Figures 6.30-6.32, since the estimated K_d field has areas of high K_d . Figures 6.51-6.53 show the K_d estimation run 1 uranium concentrations in parts per billion for layer 13 at 1/1/98, 1/1/02, and 5/1/06, respectively. Layer 13 corresponds to the 2000-series well horizon. Figure 6.53 shows the simulated 5/1/06 uranium concentrations are slightly above 140 ppb in layer 13. Again, the concentrations are higher than the baseline concentrations in Figures 6.33-6.35.

The K_d estimation run 2 uranium concentrations are shown in Figure 6.54-6.59. The concentrations are higher than the K_d estimation run 1 uranium concentrations since the K_d s are higher and the initial concentrations are higher. K_d estimation run 3 gave very similar uranium concentrations as K_d estimation run 2.

6.8.5 Monte Carlo Analysis

Monte Carlo analysis was performed to transfer parameter uncertainty to concentration uncertainty. This analysis does not account for flow field uncertainties and uranium source uncertainties. The DFM/VAM3DF Monte Carlo procedure is given in Section 6.2. For K_d estimation run 1, 200 realizations of the K_d field were simulated. The realizations are conditional on the local model, prior statistics, and concentration measurements. Figure 6.60 shows the Monte Carlo results for K_d estimation run 1. The figure includes a histogram of $\log_{10}(K_d)$ at monitoring well 2015, a histogram of maximum concentration at 7/1/06, and the maximum concentration cumulative distribution function. The CDF suggests that the median maximum concentration is 150 ppb and that there is a 100% chance that the maximum concentration at 7/1/06 is less than 175 ppb.

For K_d estimation run 2, 200 realizations of the K_d field were simulated and 200 realizations of the initial concentrations were simulated from IC estimation run 1. The realizations are conditional on the local model, prior statistics, and concentration measurements. Figure 6.61 shows the Monte Carlo results for K_d estimation run 2. The figure includes a histogram of $\log_{10}(IC)$ at monitoring well 2015, a histogram of $\log_{10}(K_d)$ at monitoring well 2015, a histogram of maximum concentration at 7/1/06, and the maximum concentration cumulative distribution function. The CDF suggests that the median maximum concentration is 160 ppb and there is a 100% chance that the maximum concentration at 7/1/06 is less than 292 ppb. The introduction of the initial concentration uncertainties has increased the range of maximum concentrations.

For K_d estimation run 3, 200 realizations of the K_d field were simulated and 200 realizations of the initial concentrations were simulated from IC estimation run 2. The realizations are conditional on the local model, prior statistics, and concentration measurements. Figure 6.62 shows the Monte Carlo results for K_d estimation run 3. The figure includes a histogram of $\log_{10}(IC)$ at monitoring well 2015, a histogram of $\log_{10}(K_d)$ at monitoring well 2015, a histogram of maximum concentration at 7/1/06, and the maximum concentration cumulative distribution function. Since the random numbers generator used the same seed as in the K_d estimation run 2 Monte Carlo analysis, the histograms and CDF have the same shape. The CDF suggests that the median maximum concentration is 261 ppb, and that there is a 100% chance that the maximum concentration at 7/1/06 is less than 483 ppb. Decreasing the chemisorption rate resulted in higher maximum concentrations compared to K_d estimation run 2.

The concentration variances at 7/1/06 for K_d estimation run 1 are shown in Figure 6.63 and 6.64 for local model layers 11 and 13, respectively. These variances were derived from simulated concentrations using 200 realizations of the K_d field. Notice that the variances are larger in layer 13. Comparing with simulated concentrations in figures 6.50 and 6.53, shows that the largest variances occur near the highest uranium concentrations. The concentration variances at 7/1/06 for K_d estimation run 2 are shown in Figures 6.65 and 6.66 for local model layers 11 and 13,

respectively. These variances were derived from simulated concentrations using 200 realizations of the initial concentrations and K_d field. The variances are larger than K_d estimation run 1 case since initial concentration uncertainty was included. A comparison with simulated concentrations in Figures 6.56 dn 6.59 shows that the largest variances occur near the highest uranium concentration. Figure 6.63 shows low concentration variances near in the reije3ction wells along Willey Rd.

6.8.6 Computational Requirements

The DFM/VAM3DF runtime is dominated by the concentration measurement sensitivity matrix computation and the least squares system solution. All DFM/VAM3DF simulations used the LSQR solver or the representer method. Table 6.13 shows the CPU usage for each simulation. The IC estimation run 2 and K_d estimation run 3 simulations have shorter runtimes since a 300 MHz Pentium II was used.

6.8.7 DFM/VAM3DF Application Summary

The DFM/VAM3DF software was tested using a site-specific test to: demonstrate that the software is functional, show that it is efficient for application to the FEMP, and establish performance benchmarks. This test was based on the VAM3DF/GMA model developed in Phase I. The DFM/VAM3DF software was used to estimate the 1993 average uranium concentration and the spatially varying distribution coefficient in the South Plume and the South Field areas. DFM/VAM3DF used the baseline flow and transport simulation from Section 6.8.1 to fit selected 1994-1999 monitoring-well data, extraction-well data, and geoprobe data.

The first K_d estimation provides the best match to the concentration measurements. Future effort should include completing steps 6) through 8) of the DFM process of Section 6.2, which are designed to improve the DFM/VAM3DF estimates. Monte Carlo analysis was performed to transfer parameter uncertainty to concentration uncertainty. The DFM/VAM3DF application demonstrates that the software is applicable to the FEMP.

**Table 6.1
Parameter Values**

Soil-solute transport parameter	Value in model
Longitudinal Dispersivity	100 (ft)
Transverse Dispersivity	10 (ft.)
Apparent Molecular Diffusion Coefficient	0 (ft.)
Exponent of Freundlich Sorption Isotherm	1 (unitless)
Effective Porosity	0.2-0.3 (unitless)
Bulk Density	124.0-116.0 (lb/ft ³)
Exponent m for Molecular Diffusion	0 (unitless)
Vertical Transverse Dispersivity	1 (ft.)
Vertical Longitudinal Dispersivity	0 (ft.)
Decay Coefficient	0 (1/day)
Distribution Coefficient	0.0285 (ft ³ /lb)
Desorption Rate Coefficient	0.01 (1/day)
Chemisorption Rate Coefficient	0.0001 (1/day)
Precipitation Rate Coefficient	0 (1/day)
Zeroeth Order Dissolution Rate Constant	0 (nano lb/ft ³ /day)
Solubility Limit	0 (nano lb/ft ³)

Table 6.2
Sensitivity Simulations

Sensitivity	Parameter modified in Sensitivity	Perturbed value	Value in phase I site model	Figures
1	K_d	0.285	0.0285	Figure 6.8
2	K_d (No kinetics)	0.285 (No kinetics)	0.0285, Kinetics parameters	Figure 6.9
3	K_d (No kinetics)	0.0285 (No kinetics)	0.0285, Kinetics parameters	Figure 6.10
4	Chemisorption Rate Coefficient	0.001	0.0001	Figure 6.11
5	Chemisorption Rate Coefficient	0.00001	0.0001	Figure 6.12
6	Freundlich Exponent (n)	0.5	1	Figure 6.13
7	Desorption Rate Coefficient	0.1	0.01	Figure 6.14
8	Longitudinal/Transverse Dispersivities	10/1	100/10	Figure 6.15
9	Effective Porosity	0.15	0.25 to 0.3	Figure 6.16
10	Upstream Weighting Factors in X, Y, Z directions	0.5	1	Figure 6.17
11	Number of Time Steps	17	32	Figure 6.18
12	Number of Time Steps	65	32	Figure 6.19
13	Vertical Refinement to surfaces in model	16	12	Figure 6.20

000102

Table 6.3
Summary of Sensitivity Run Results

Sensitivity Run Description	Well ID/Concentration Variation from Baseline ppb			
	Extraction 31550	Rank*	Extraction 31560	Rank*
Sensitivity 1 - Kd=0.285	2.95E+02	8	1.05E+02	8
Sensitivity 2 - No Kinetics, Kd=0.285	4.33E+02	10	1.72E+02	9
Sensitivity 3 - No Kinetics, Kd=0.0285	1.10E+02	7	5.26E+01	7
Sensitivity 4 - Chemisorption Rate Coefficient = 0.001	4.41E+02	9	2.41E+02	10
Sensitivity 5 - Chemisorption Rate Coefficient = 0.00001	7.78E+01	6	4.25E+01	6
Sensitivity 6 - Freundlich Exponent (n) = 0.5	4.87E+02	11	6.55E+02	11
Sensitivity 7 - Desorption Rate Coefficient = 0.1	2.45E+01	4	9.54E+00	4
Sensitivity 8 - Longitudinal/Transverse Dispersivities to 10/1	4.23E+01	5	1.60E+01	5
Sensitivity 9 - Effective Porosity = 0.15	1.44E+01	3	6.11E+00	3
Sensitivity 10 - Upstream Weighting Factor(s) = 0.5	1.03E+01	2	3.27E+00	1
Sensitivity 11 - Number of Time Steps = 17	Not ranked, performed for self-consistency check			
Sensitivity 12 - Number of Time Steps = 65	Not ranked, performed for self-consistency check			
Sensitivity 13 - Refine Vertical Layers to 16 layers	8.98E+00	1	3.67E+00	2

Note: *Rank = 11, most sensitive; rank = 1, least sensitive.

Table 6.4
Number of Measurements

Number of Direct Measurements of IC			Number Concentration Measurements				
2000 wells	3000 wells	4000 wells	2000 wells	3000 wells	4000 wells	Extraction wells	Geoprobes
57	46	6	78	60	4	32	305

Table 6.5

Wells used for Concentration Measurements				Wells not used for Concentration Measurements		
2000 Wells	3000 Wells	4000 Wells	Extraction Wells	2000 Wells	3000 Wells	4000 Wells
2002	3014	4125	3924	2006	3016	4014
2014	3015	4398	3925	2007	3017	4015
2015	3045	Total = 2	3926	2008	3062	4016
2017	3046		3927	2016	3065	4920
2045	3049		31550	2048	32305	Total = 4
2046	3068		31560	2065	32307	
2047	3069		31561	21065	3391	
2049	3070		31562	21192	3624	
2070	3093		31563	21193	3689	
2093	3095		31564	2126	3910	
2095	3106		31565	22299	3911	
21033	3125		31567	22300	3912	
2106	3128		32276	22301	3916	
21063	3385		31566	22302	3917	
2125	3387		32308	2391	3918	
2128	3390		32309	2394	3921	
2166	3396		Total = 16	2395	3922	
2385	3397			2399	3923	
2386	3398			2400	3928	
2387	3402			2401	Total = 19	
2390	3550			2546		
2396	3551			2549		
2397	3552			2553		
2398	3636			2624		
2402	3880			2943		
2434	3881			2944		
2544	3897			2945		
2545	3898			2954		
2550	3899			2955		
2551	3900			Total = 29		
2552	Total = 30					
2625						
2636						
2880						
2881						
2897						
2898						
2899						
2900						
Total = 39						

Table 6.6
Baseline Simulations Residual Summary

Simulation	Concentration Measurement Residuals										Total	
	2000-series wells		3000-series wells		4000-series wells		Extraction wells		Geoprobos			
	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS
Baseline	0.05	0.40	-0.28	0.57	-0.18	0.27	0.38	0.64	-0.03	0.72	-2.3 x 10 ⁻²	0.65
Baseline with $\alpha_c = 10^{-6}$	0.005	0.39	-0.32	0.60	-0.23	0.30	0.32	0.60	-0.09	0.72	-7.7 x 10 ⁻²	0.65
Baseline with n = 0.5	0.13	0.69	-0.48	0.86	-0.6	0.80	0.37	0.87	-0.13	0.90	-0.1	0.86

Table 6.7
Range of Values for Estimated Parameters

Simulation Name	Minimum Value	Maximum Value
IC Estimation run 1	2.7 x 10 ⁻⁶ ppb	3196 ppb
IC Estimation run 2	2.7 x 10 ⁻⁶ ppb	3196 ppb
K _d Estimation run 1	0.013 ft ³ /1b	0.56 ft ³ /1b
K _d Estimation run 2	0.009 ft ³ /1b	0.97 ft ³ /1b
K _d Estimation run 3	0.011 ft ³ /1b	1.33 ft ³ /1b

Table 6.8
Maximum Concentration and Uranium Mass Summary at 7/1/06

Simulation Name	Maximum Concentration (ppb)	Mass in Groundwater (pounds)	Mass absorbed to soil (pounds)	Mass chemisorbed (pounds)	Total Mass (pounds)
IC Estimation run 1	104	171	2893	2307	5371
IC Estimation run 2	163	255	4356	28	4639
K _d Estimation run 1	148	189	16025	9637	25851
K _d Estimation run 2	155	183	10908	6940	18031
K _d Estimation run 3	259	270	18670	94	19034

Table 6.9
IC Estimation Residual Summary

IC Estimation	Direct measurement residuals						Total		Concentration Measurement Residuals										Total	
	2000-series wells		3000-series wells		4000-series wells				2000-series wells		3000-series wells		4000-series wells		Extraction wells		Geoprobos			
	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS
run 1	0.05	0.19	0.04	0.14	0.11	0.14	0.05	0.17	-0.03	0.43	-0.26	0.62	-0.42	0.60	0.20	0.54	-0.26	0.80	-0.19	0.71
run 2	0.05	0.19	0.04	0.14	0.11	0.14	0.05	0.17	-0.07	0.44	-0.30	0.64	-0.46	0.63	0.15	0.52	-0.32	0.81	-0.25	0.73

Table 6.10
K_d Estimation Residual Summary

K _d Estimation	Concentration Measurement Residuals										Total	
	2000-series wells		3000-series wells		4000-series wells		Extraction wells		Geoprobe			
	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS	Avg.	RMS
run 1	0.03	0.34	-0.18	0.48	-0.07	0.27	0.41	0.64	0.06	0.67	0.05	0.60
run 2	-0.006	0.33	-0.14	0.46	-0.38	0.56	0.18	0.50	-0.12	0.69	-0.09	0.61
run 3	-0.05	0.33	-0.17	0.46	-0.42	0.59	0.13	0.49	-0.16	0.70	-0.13	0.61

Table 6.11
Estimated Trend Coefficients

Simulation Name	Constant Coefficient	X Coefficient	Y Coefficient	Z Coefficient
IC Estimation run 1	-3.8			
IC Estimation run 2	-3.8			
K _d Estimation run 1	-0.92	-0.14	-0.16	0.06
K _d Estimation run 2	-0.98	0.33	-0.32	0.03
K _d Estimation run 3	-0.90	0.38	-0.33	0.01

Table 6.12
Sum-of-Squares Summary

Simulation Name	Spatial Variability SOS	Trend Polynomial SOS	Direct Measurement SOS	Conc. Measurement SOS	Total SOS
IC Estimation run 1	13880.9	450.0	1222.8	49552.7	65106.4
IC Estimation run 2	13881.0	450.0	1222.8	51477.5	67031.2
K _d Estimation run 1	1704.4	20.1		35022.5	36747.0
K _d Estimation run 2	3660.5	85.6		35776.9	39523.0
K _d Estimation run 3	4064.8	103.5		36495.1	40663.4

Table 6.13
CPU Usage

Simulation Name	Pentium II Clock Speed (MHz)	Number of Gauss- Newton Iterations	CPU Usage for each Jacobian Hours:Minutes	CPU Usage for each LSQR/RM Solve Hours:Minutes	Total CPU Usage Hours:Minutes
IC Estimation run 1	300	1	38:10	1:35 (LSQR)	40:28
IC Estimation run 2	400	1	25:09	:56 (LSQR)	26:34
K _d Estimation run 1	200	1	42:33	4:28 (RM)	47:16
K _d Estimation run 2	300	1	27:45	3:13 (RM)	31:10
K _d Estimation run 3	400	1	18:15	1:59 (RM)	20:22

Site Model vs. Local Model Area

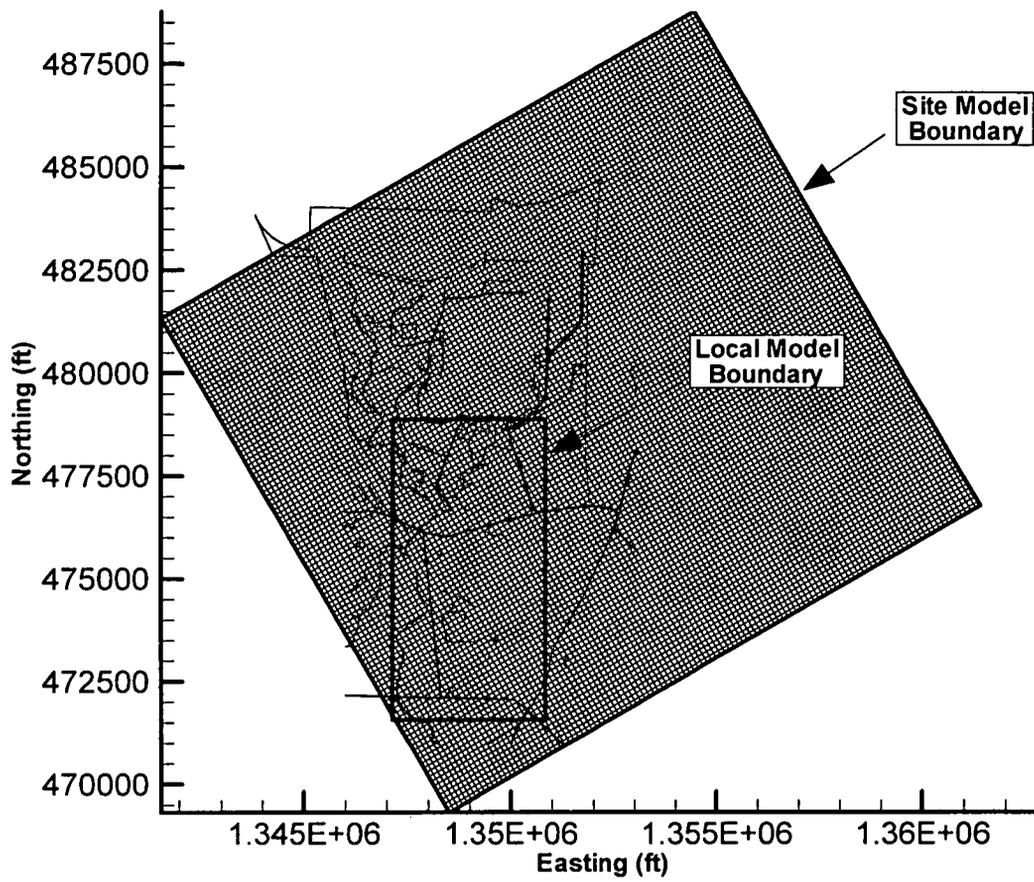


Figure 6.1 Regional and Local Model Domains.

Vertical Cross-section of Site Model Grid Showing Additional Nodal Layers in Local Model

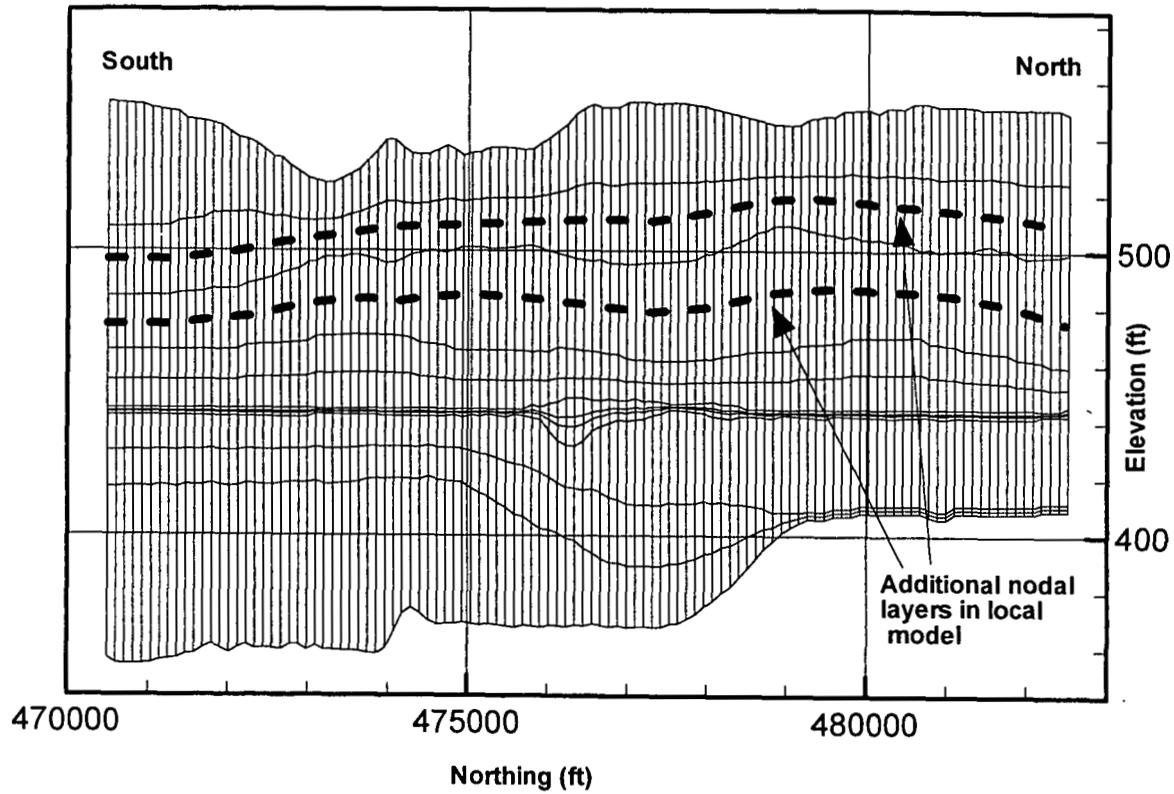


Figure 6.2 Vertical Cross-section of Site Model Grid Showing Additional Nodal Layers in Local Model

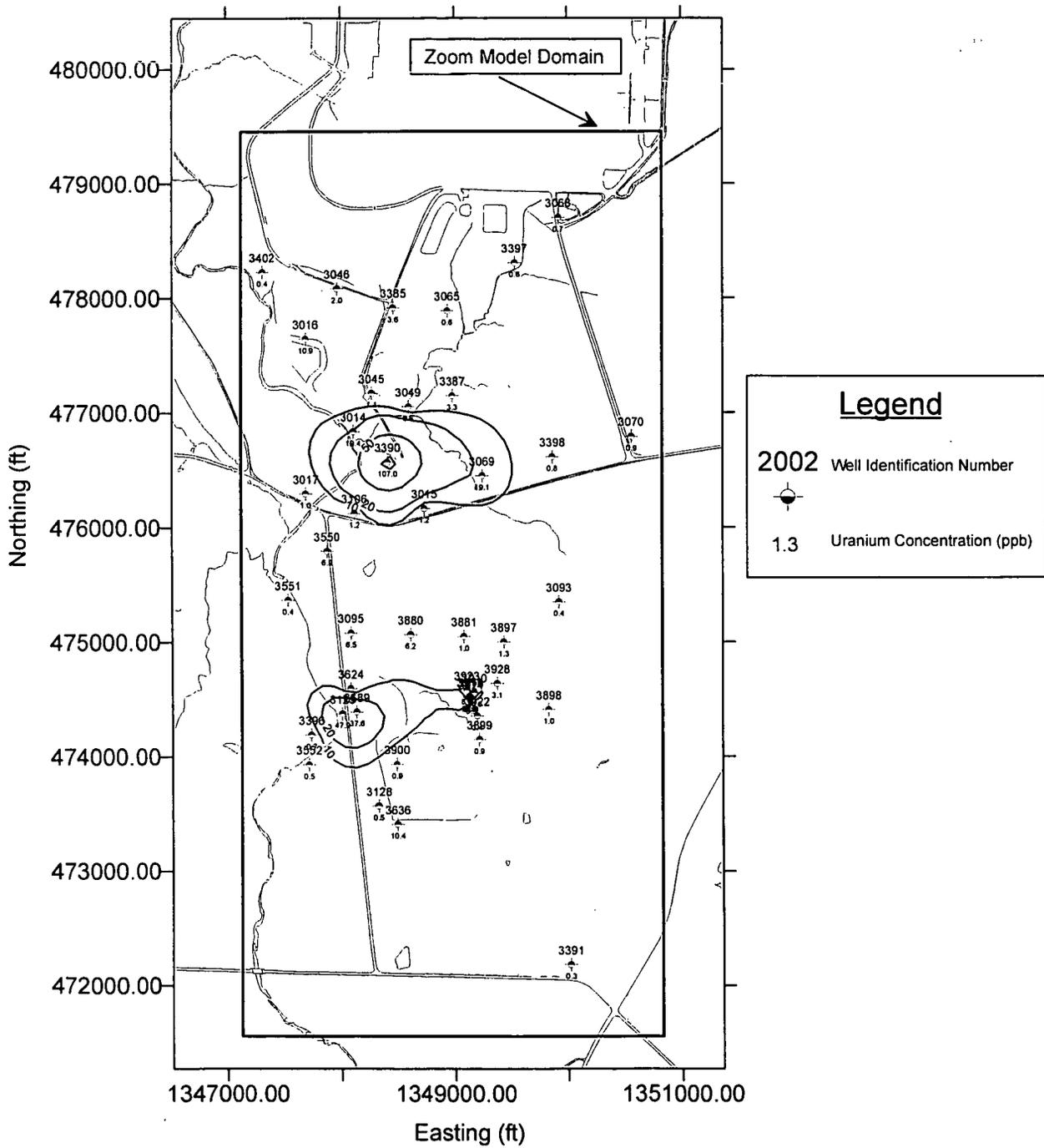


Figure 6.4 Average Uranium Concentrations for 3000 wells in 1993 Used for Initial Conditions.

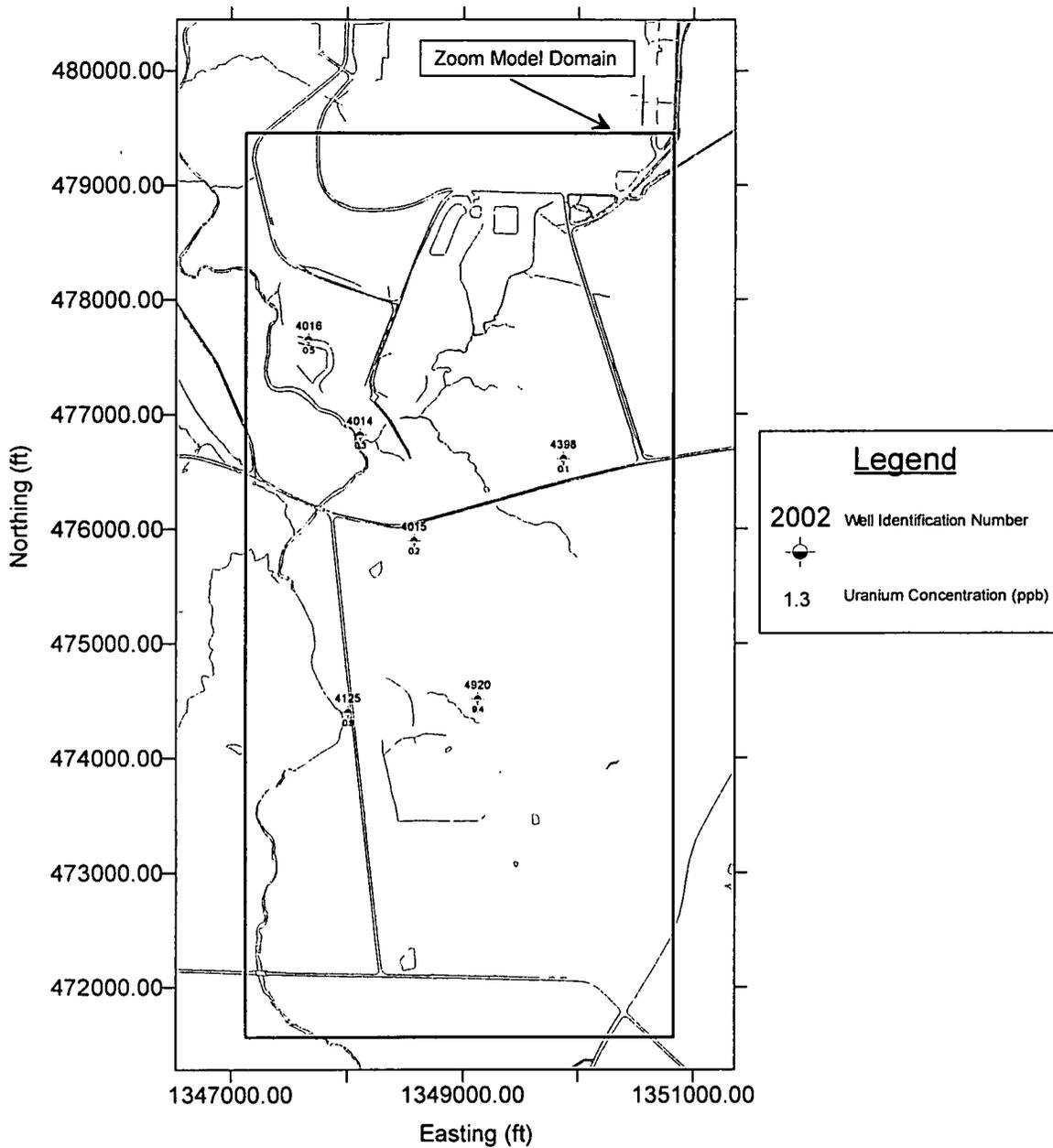


Figure 6.5 Average Uranium Concentrations for 4000 wells in 1993 Used for Initial Conditions.

000113

Observed vs. Simulated Concentration Baseline

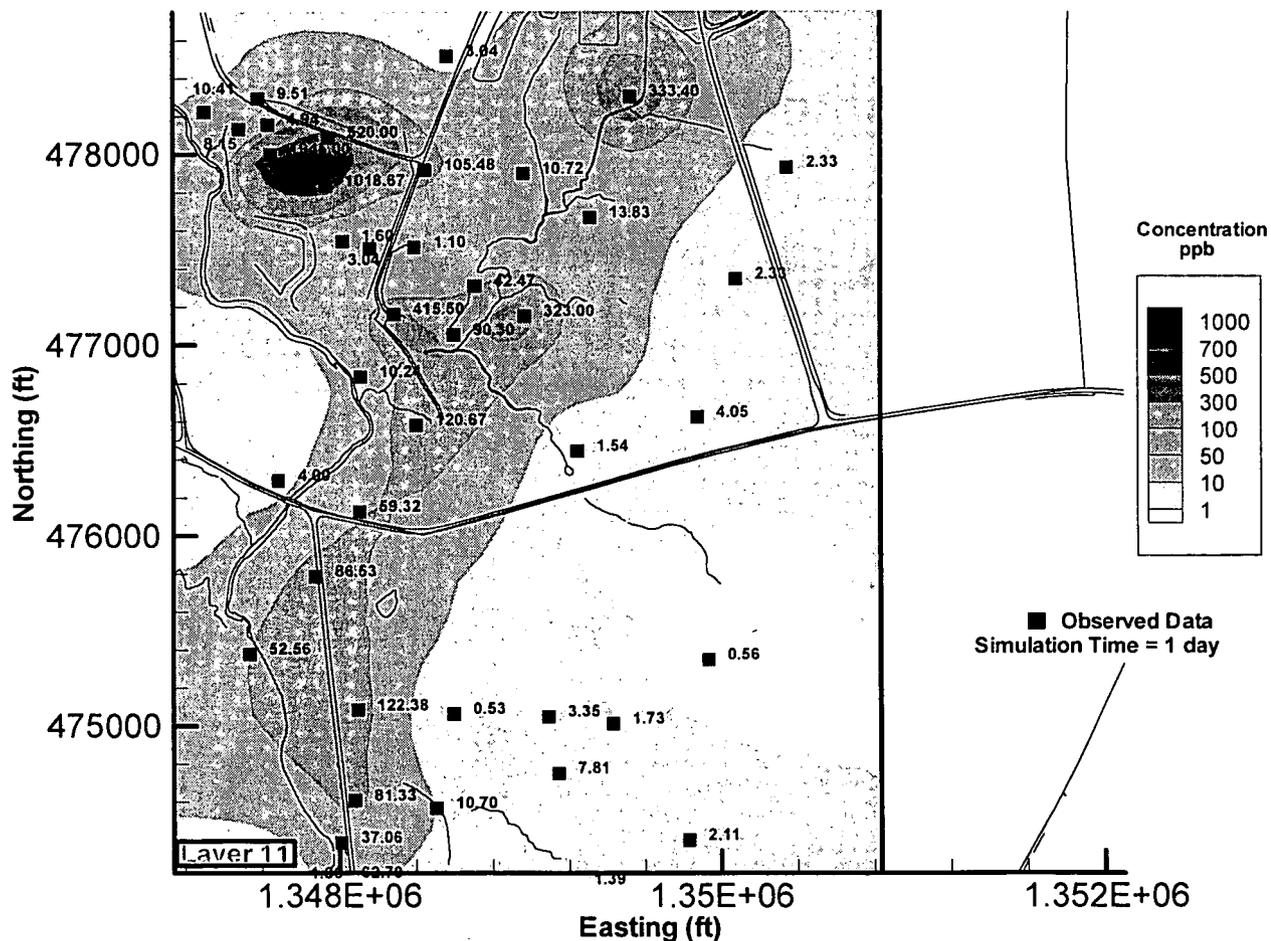


Figure 6.6 Baseline Conditions for Time = 1 day.

000114

Observed vs. Simulated Concentration Baseline

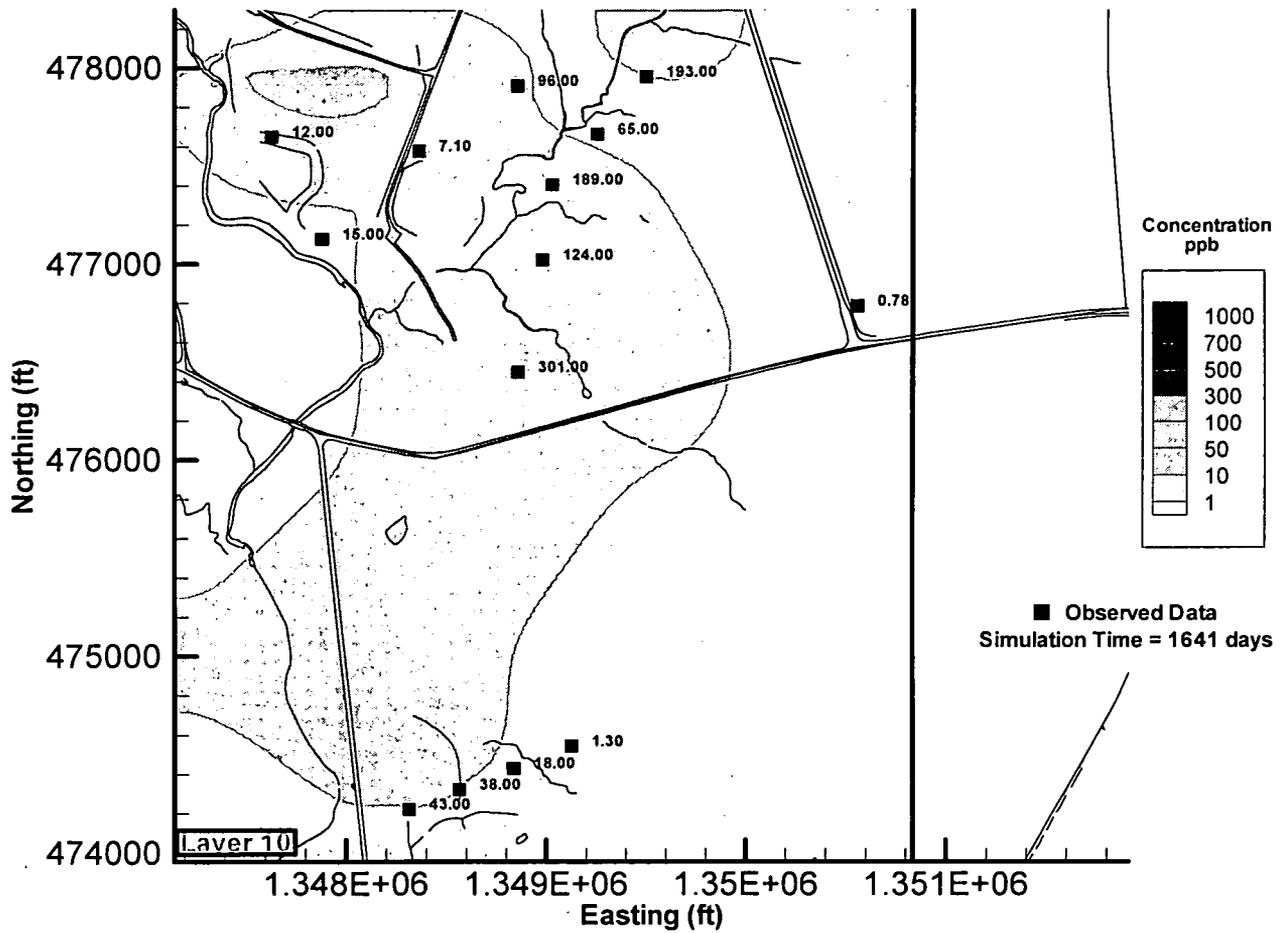


Figure 6.7 Baseline Conditions for Time = 1641 days.

Observed vs. Simulated Concentration Sensitivity 1 -- Kd=0.29

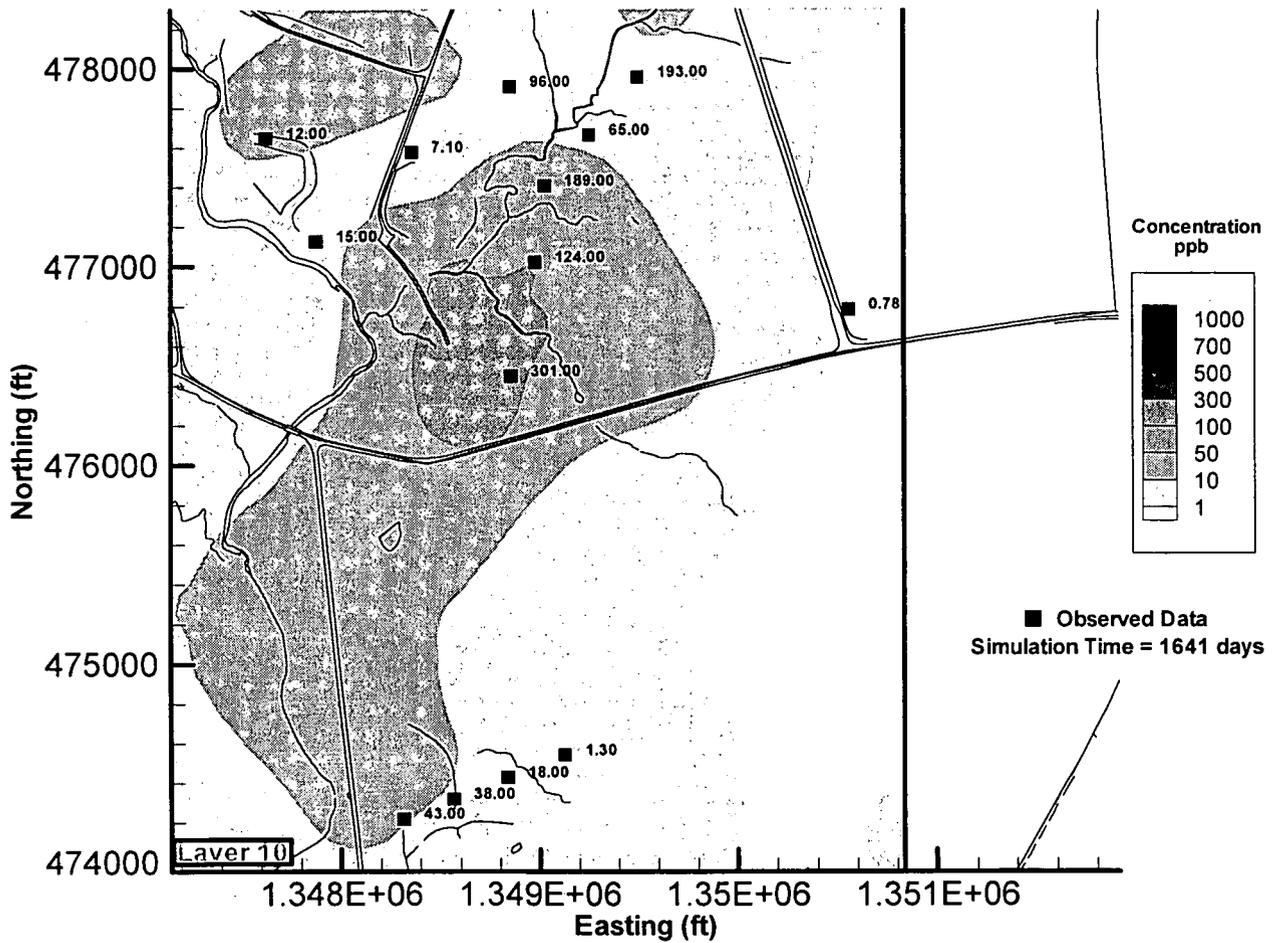


Figure 6.8 Sensitivity Simulation 1 at Time = 1641 days.

**Observed vs. Simulated Concentration
Sensitivity 2 -- Kd=0.29 with no Kinetics**

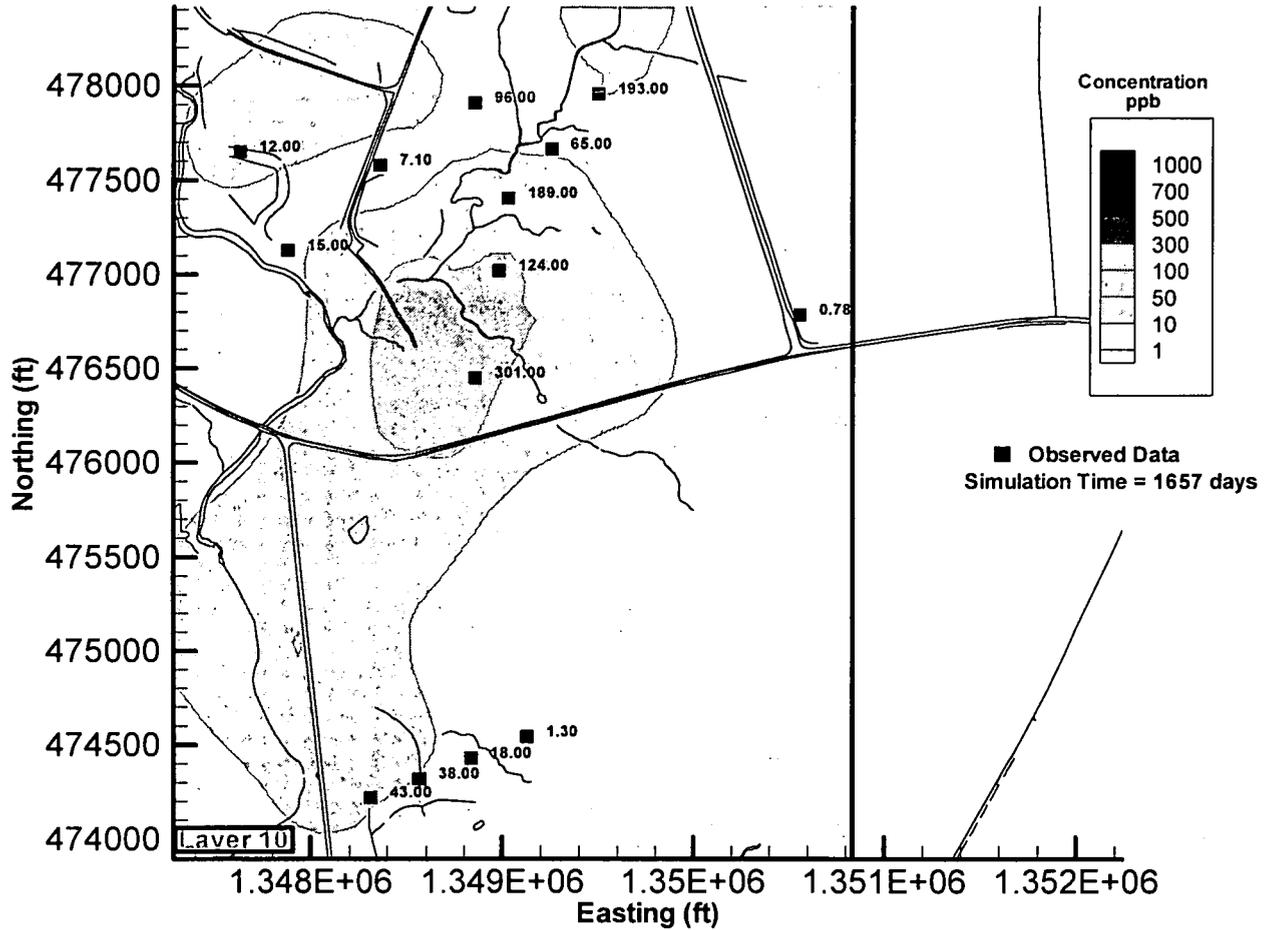


Figure 6.9 Sensitivity Simulation 2 at Time = 1641 days.

**Observed vs. Simulated Concentration
Sensitivity 3 -- Kd=0.029 with no Kinetics**

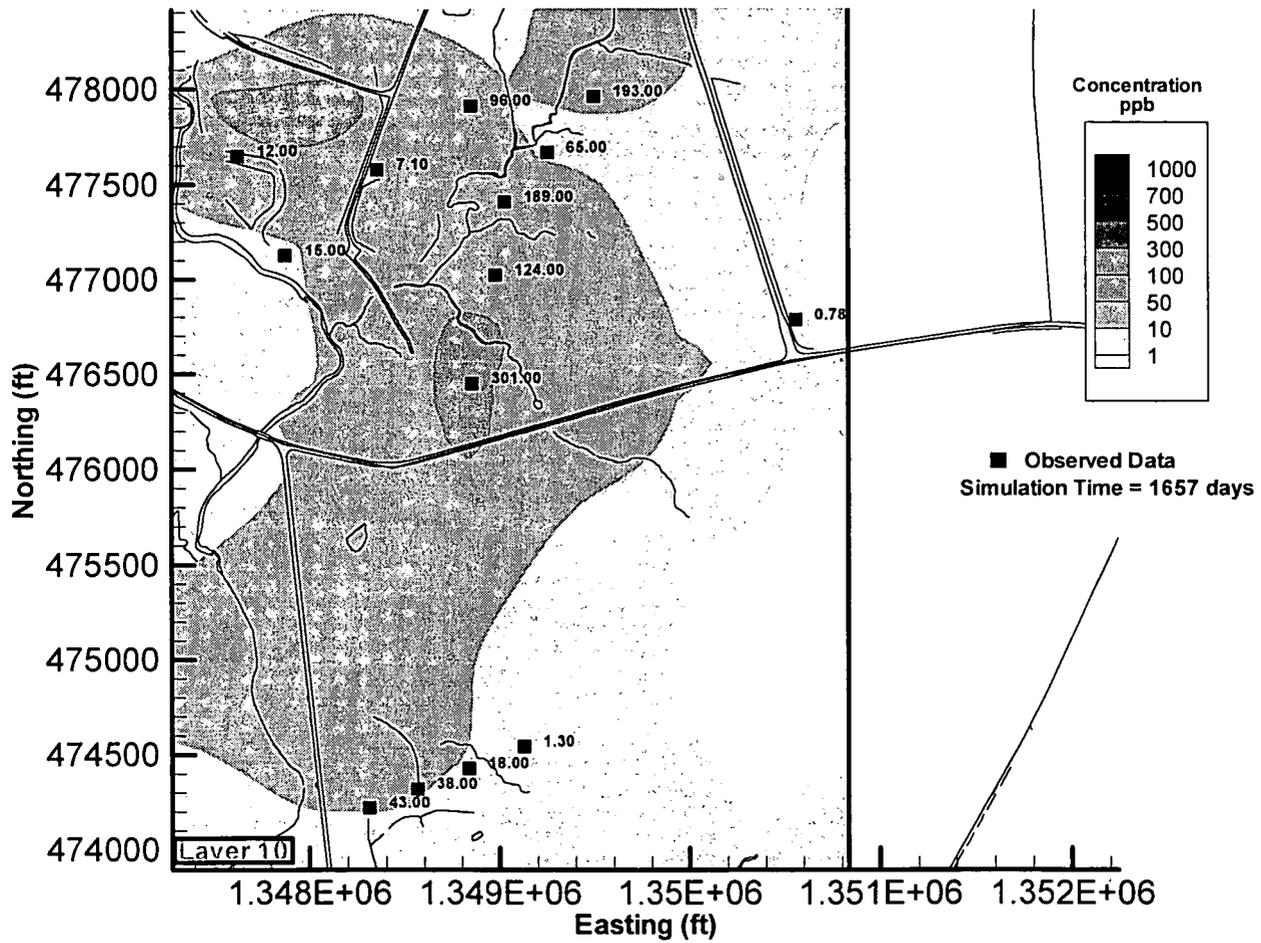


Figure 6.10 Sensitivity Simulation 3 at Time = 1641 days.

**Observed vs. Simulated Concentration
Sensitivity 4 -- Chemisorption Rate Coefficient = 0.001**

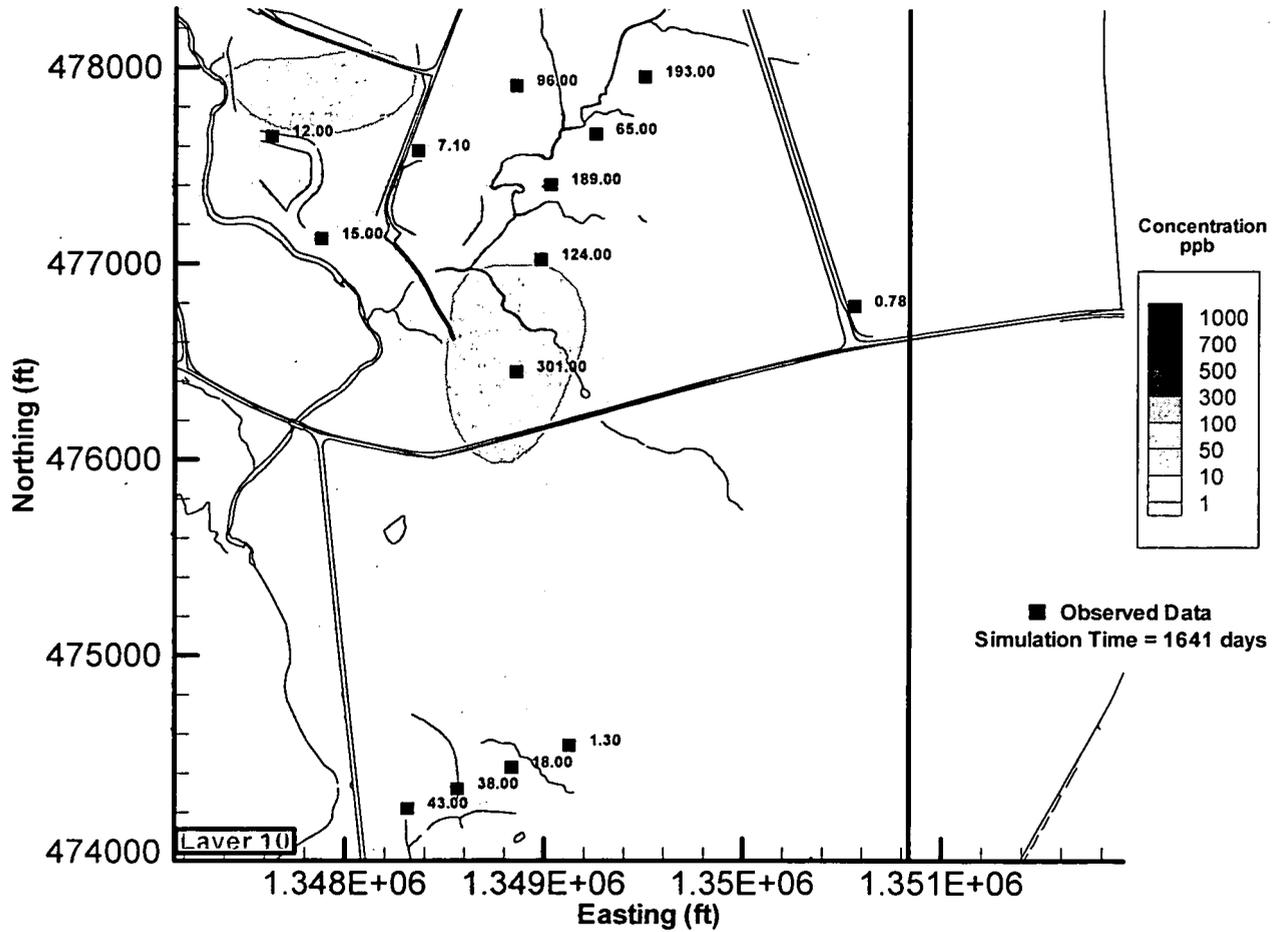


Figure 6.11 Sensitivity Simulation 4 at Time = 1641 days.

Observed vs. Simulated Concentration Sensitivity 5 -- Chemisorption Rate Coefficient = 0.00001

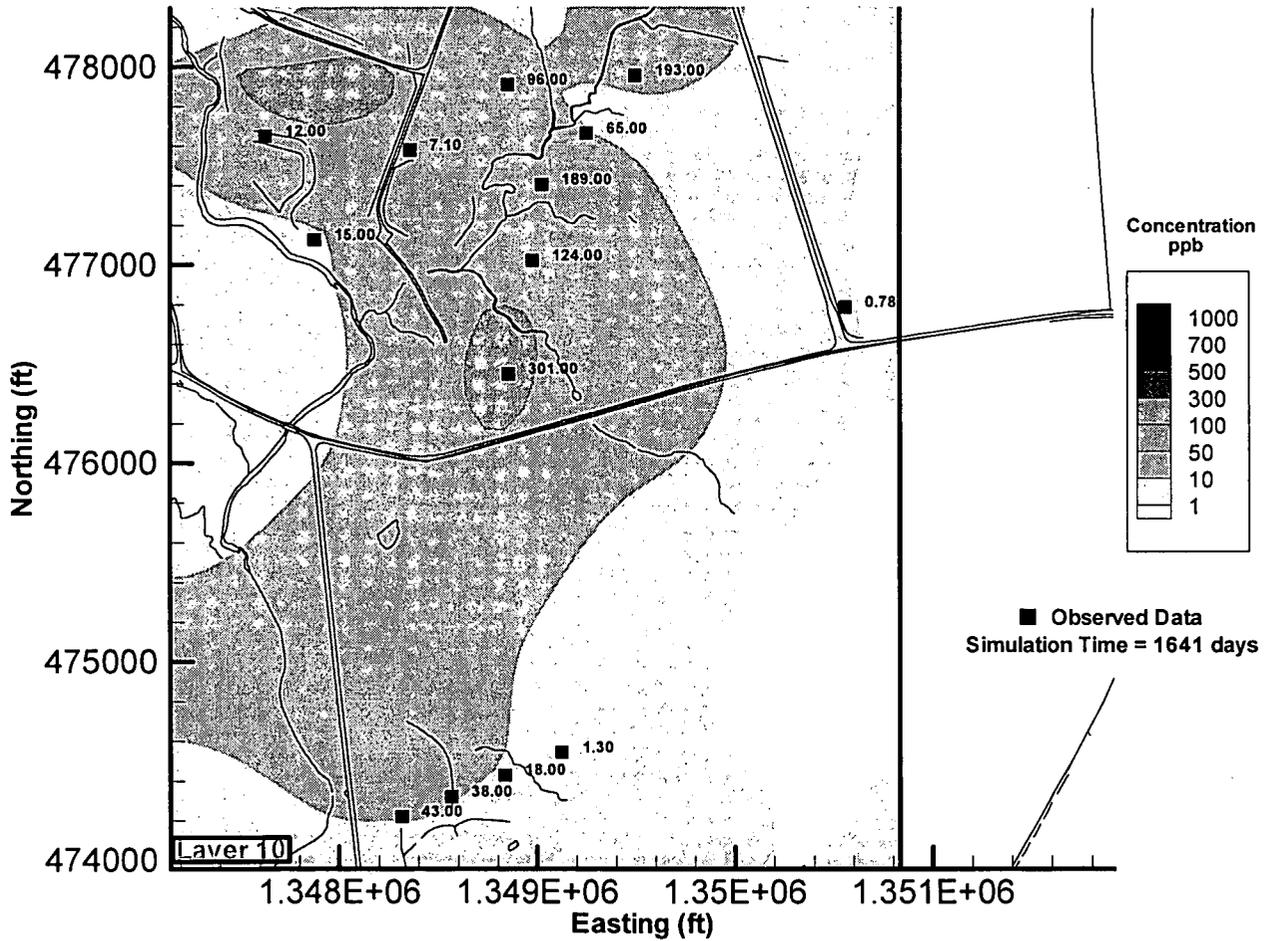


Figure 6.12 Sensitivity Simulation 5 at Time = 1641 days.

000120

**Observed vs. Simulated Concentration
Sensitivity 6 -- Freundlich Exponent=0.5**

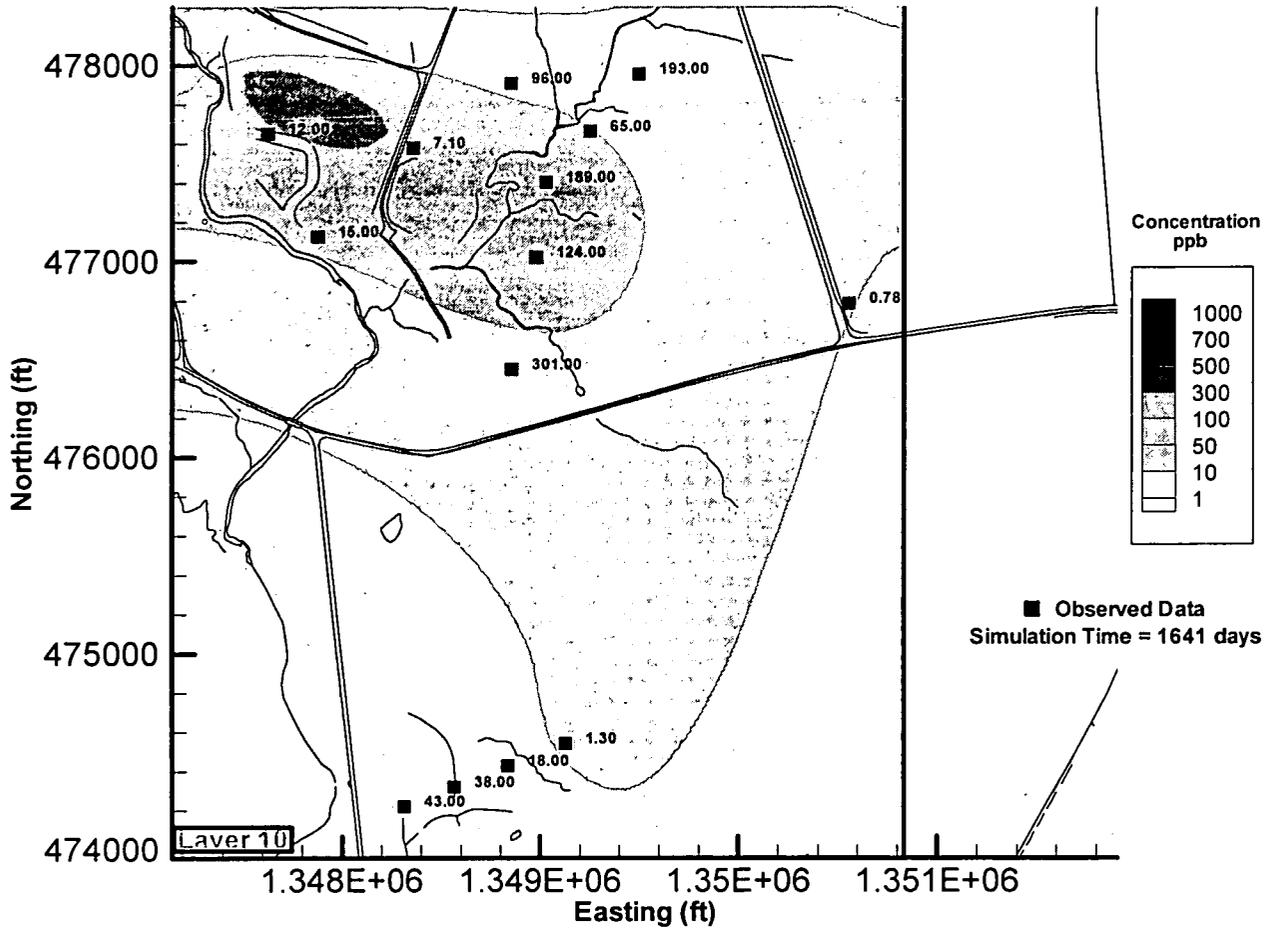


Figure 6.13 Sensitivity Simulation 6 at Time = 1641 days.

Observed vs. Simulated Concentration Sensitivity 7 -- Desorption Rate Coefficient = 0.1

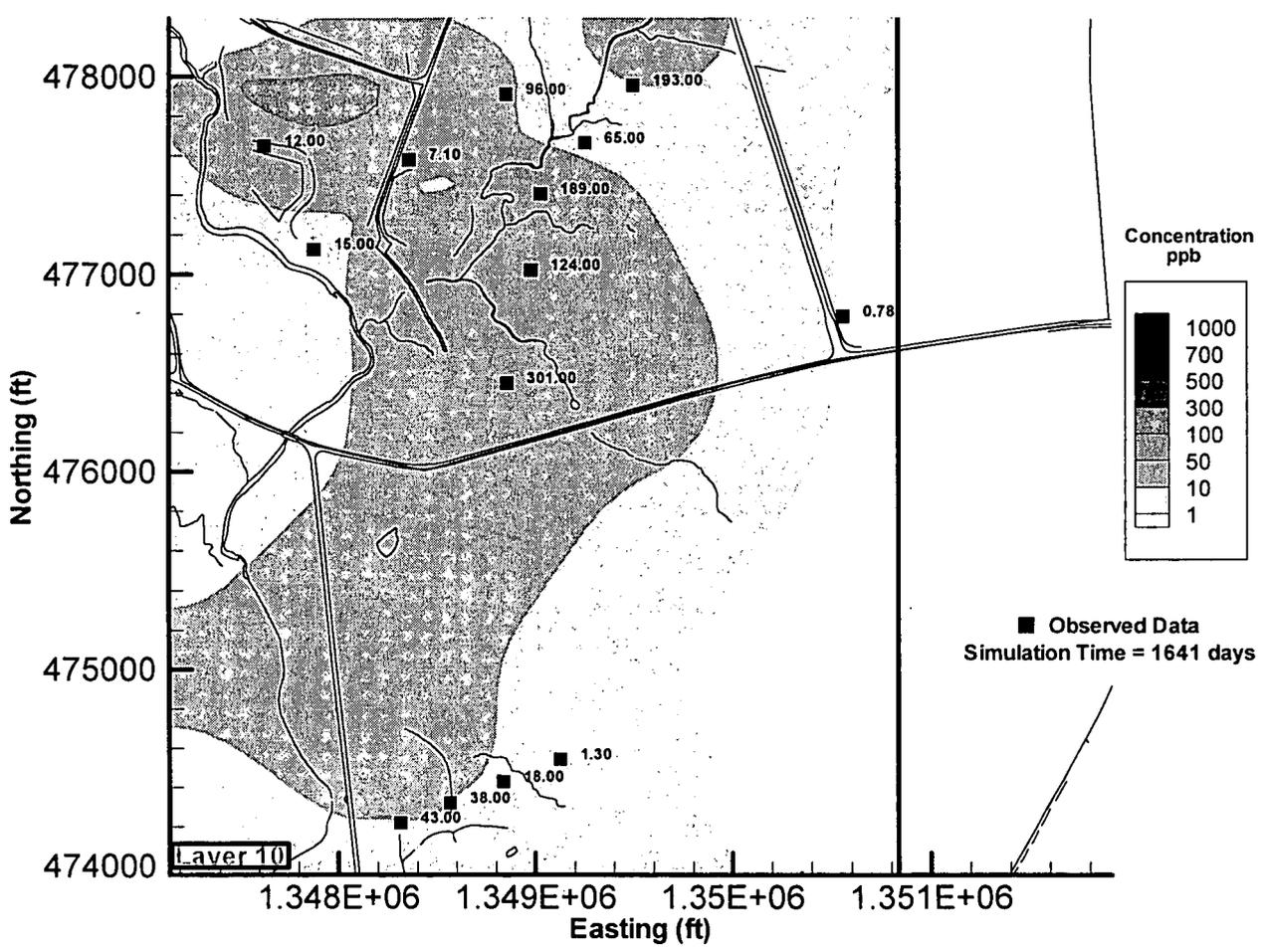


Figure 6.14 Sensitivity Simulation 7 at Time = 1641 days.

**Observed vs. Simulated Concentration
Sensitivity 8 -- Longitudinal/Transverse Dispersion = 10/1**

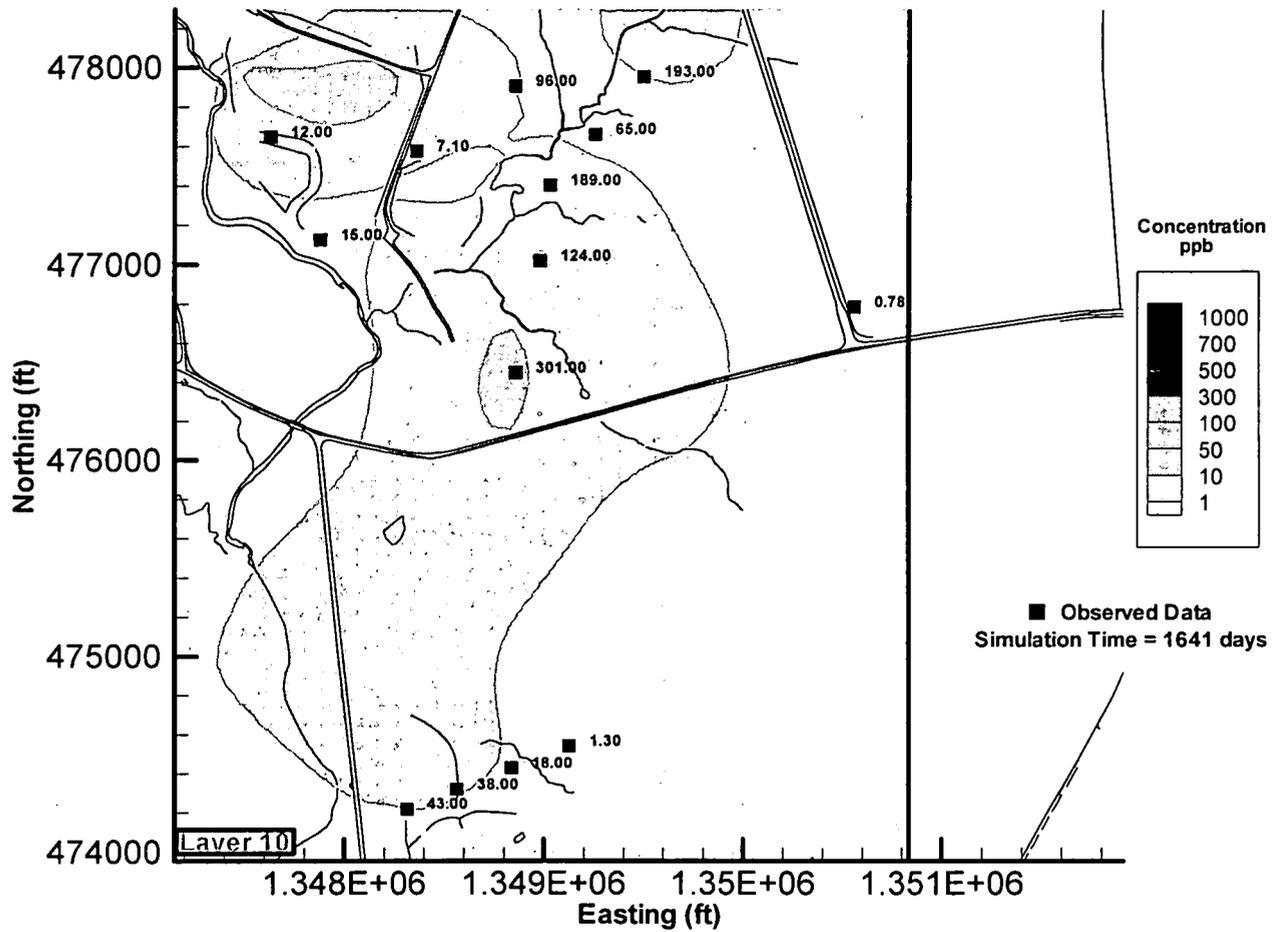


Figure 6.15 Sensitivity Simulation 8 at Time = 1641 days.

000123

**Observed vs. Simulated Concentration
Sensitivity 9 -- Effective Porosity = 0.15**

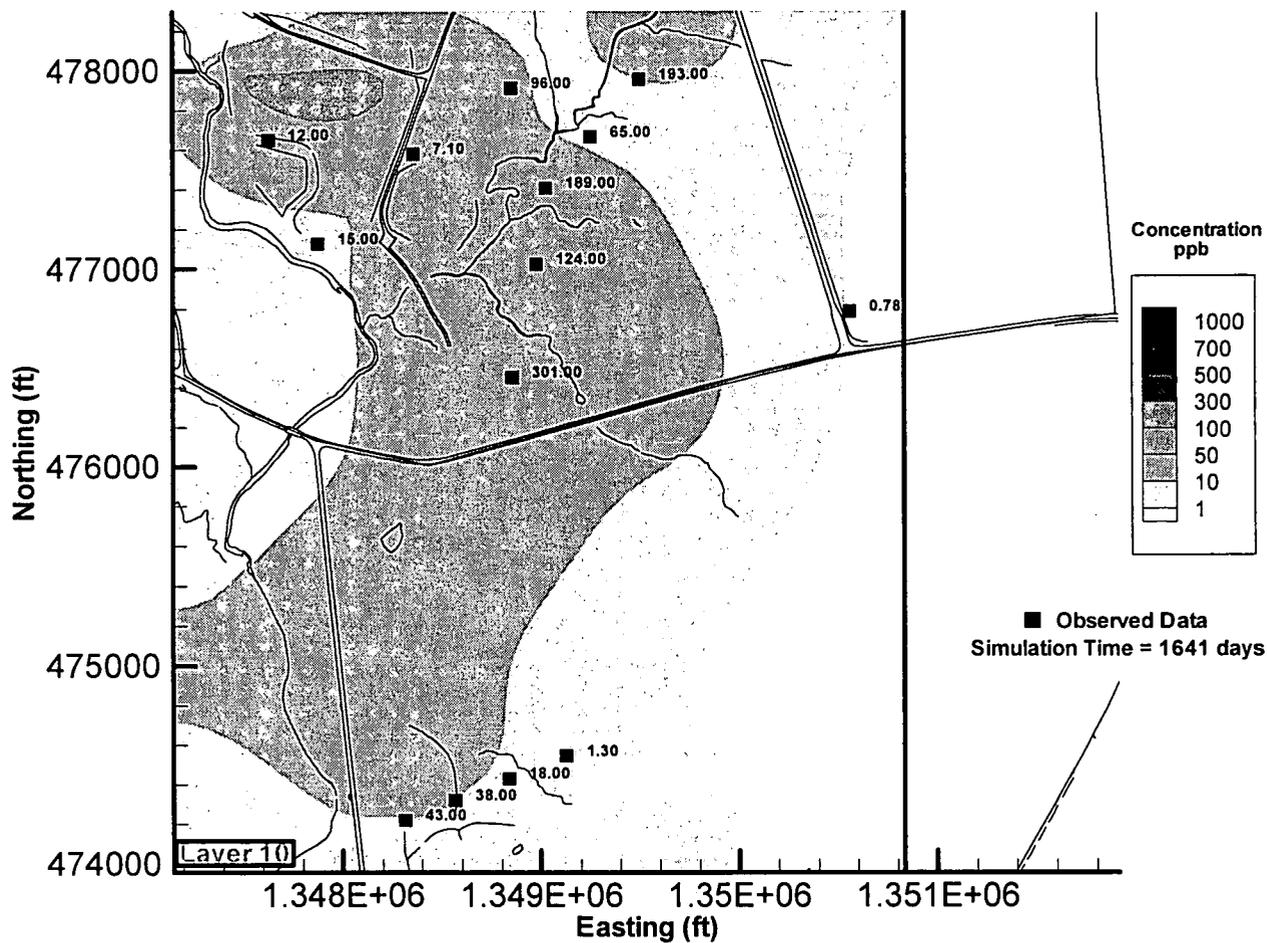


Figure 6.16 Sensitivity Simulation 9 at Time = 1641 days.

**Observed vs. Simulated Concentration
Sensitivity 10 -- Upstream Weighting = 0.5**

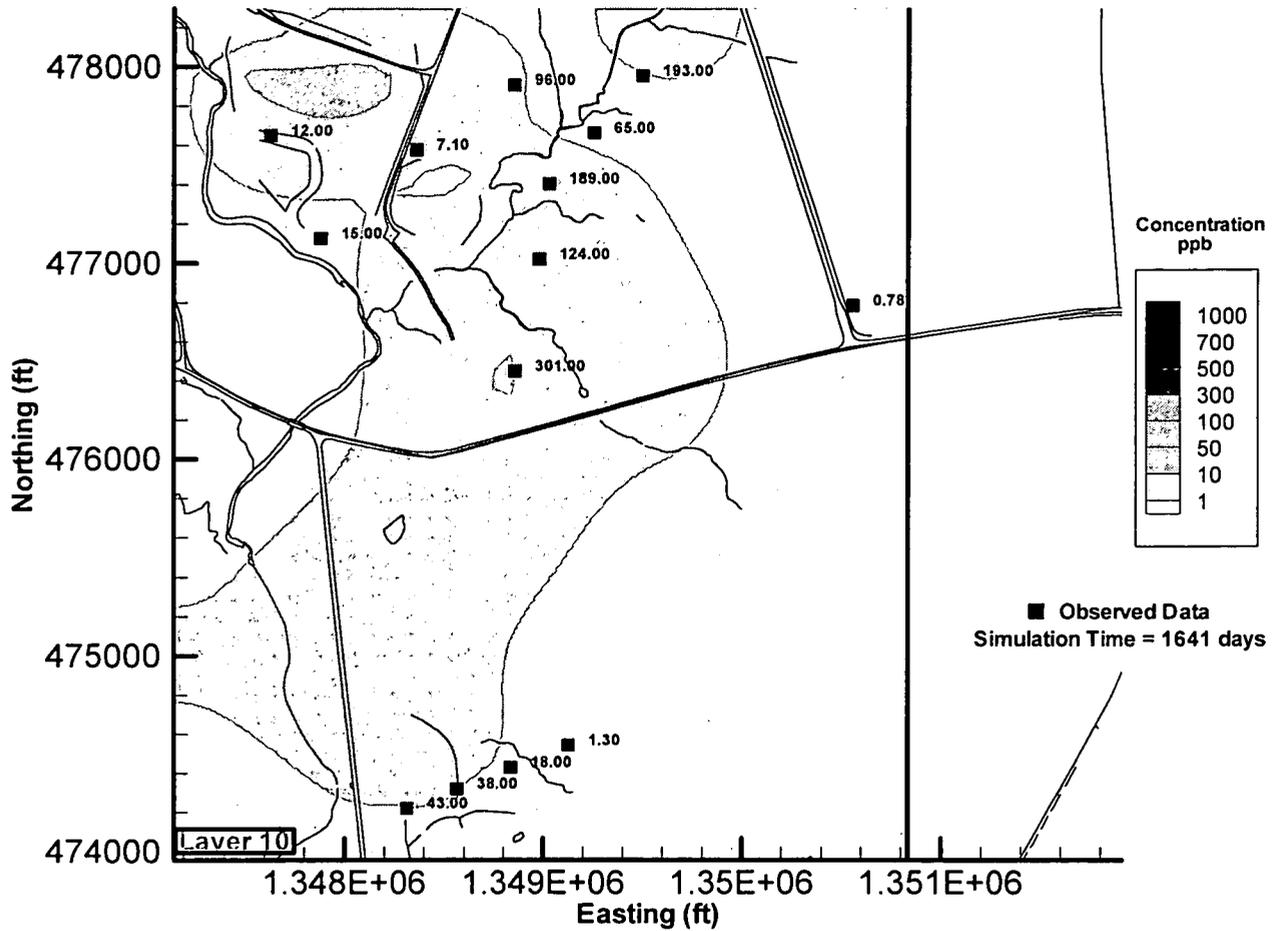


Figure 6.17 Sensitivity Simulation 10 at Time = 1641 days.

Observed vs. Simulated Concentration Sensitivity 11 -- Number of Time Steps = 17

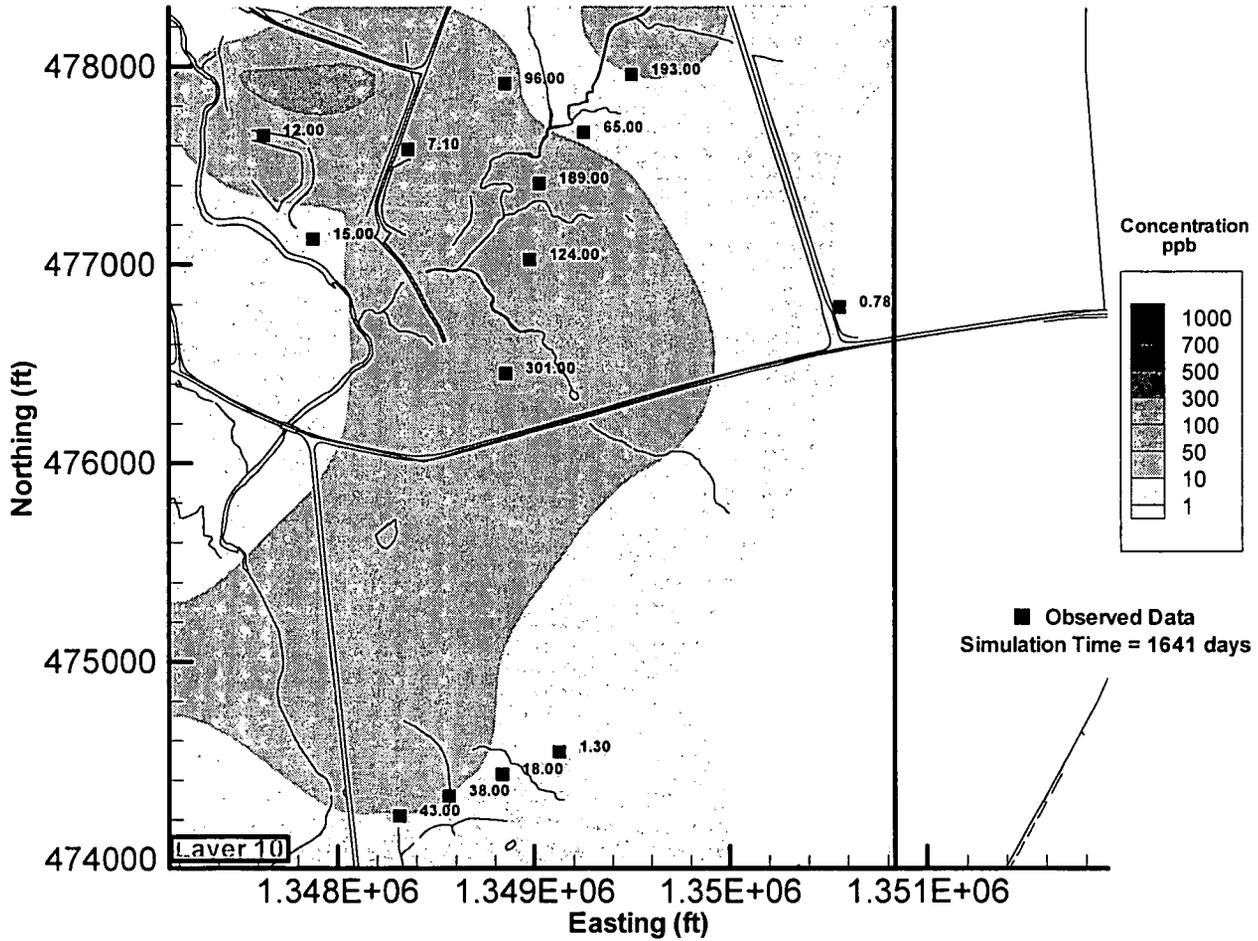


Figure 6.18 Sensitivity Simulation 11 at Time = 1641 days.

**Observed vs. Simulated Concentration
Sensitivity 12 -- Number of Time Steps = 65**

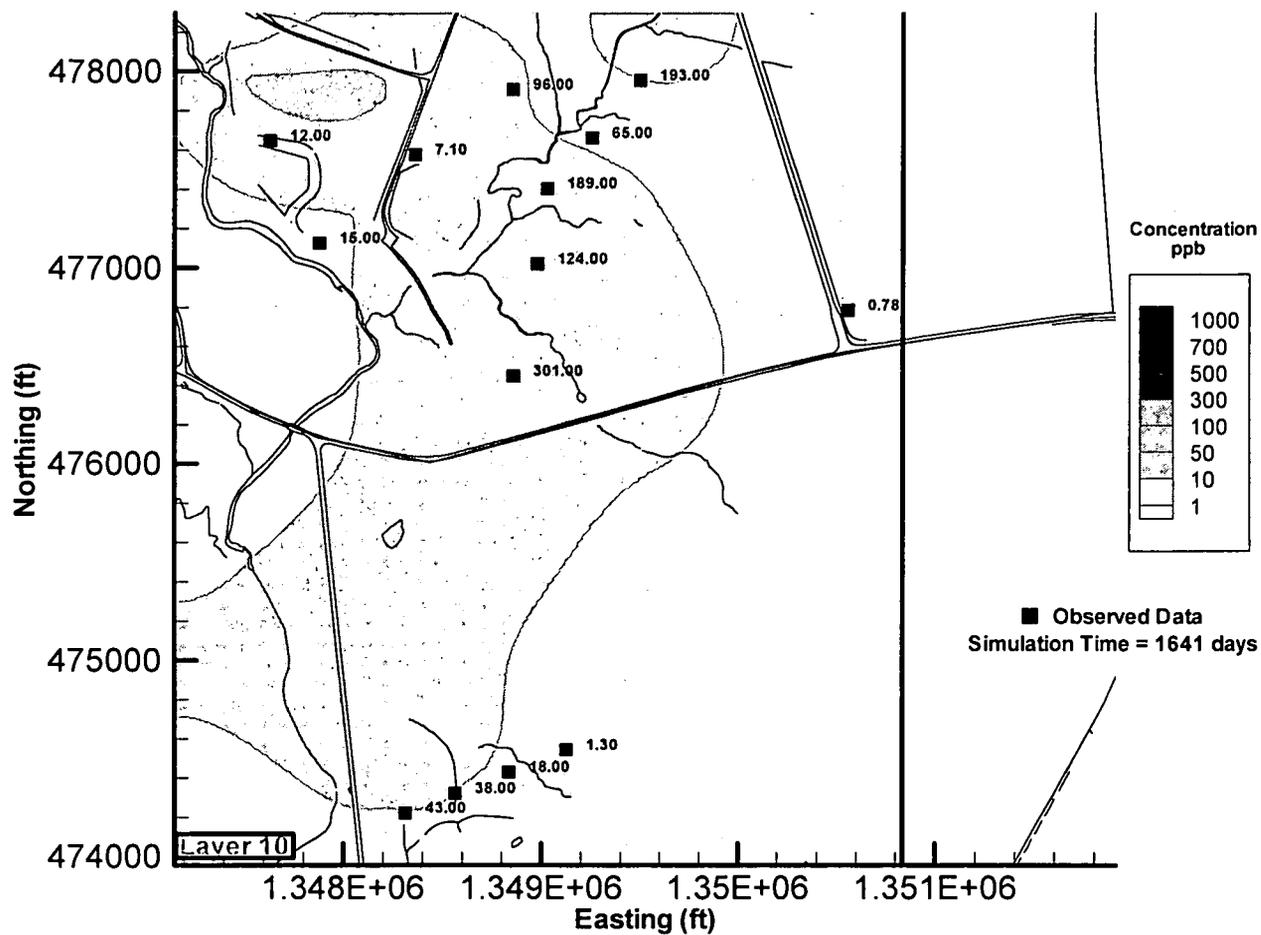


Figure 6.19 Sensitivity Simulation 12 at Time = 1641 days.

Observed vs. Simulated Concentration Vertical Refinement to 16 layers

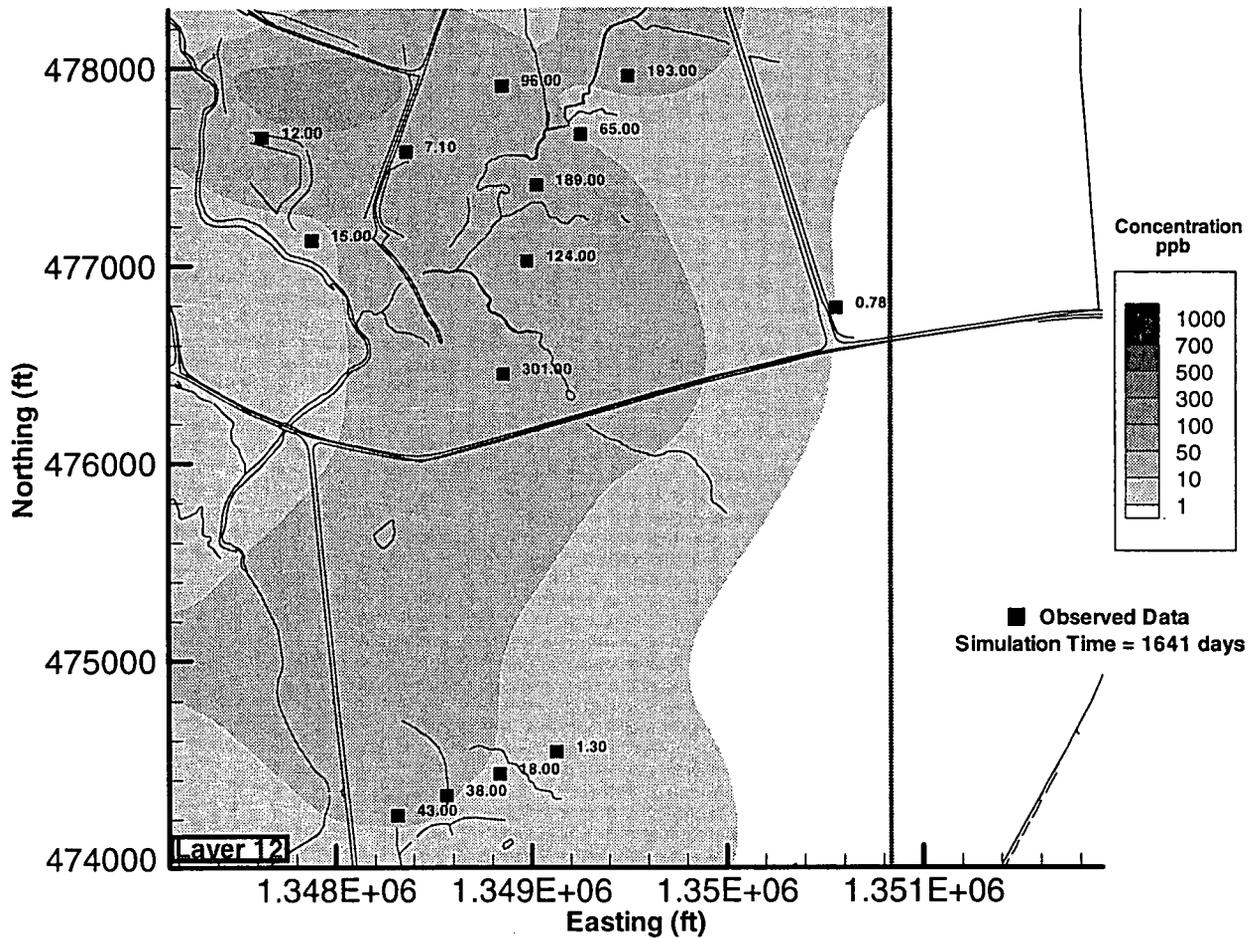


Figure 6.20 Sensitivity Simulation 13 (changing Vertical Discretization) at Time = 1641 days.

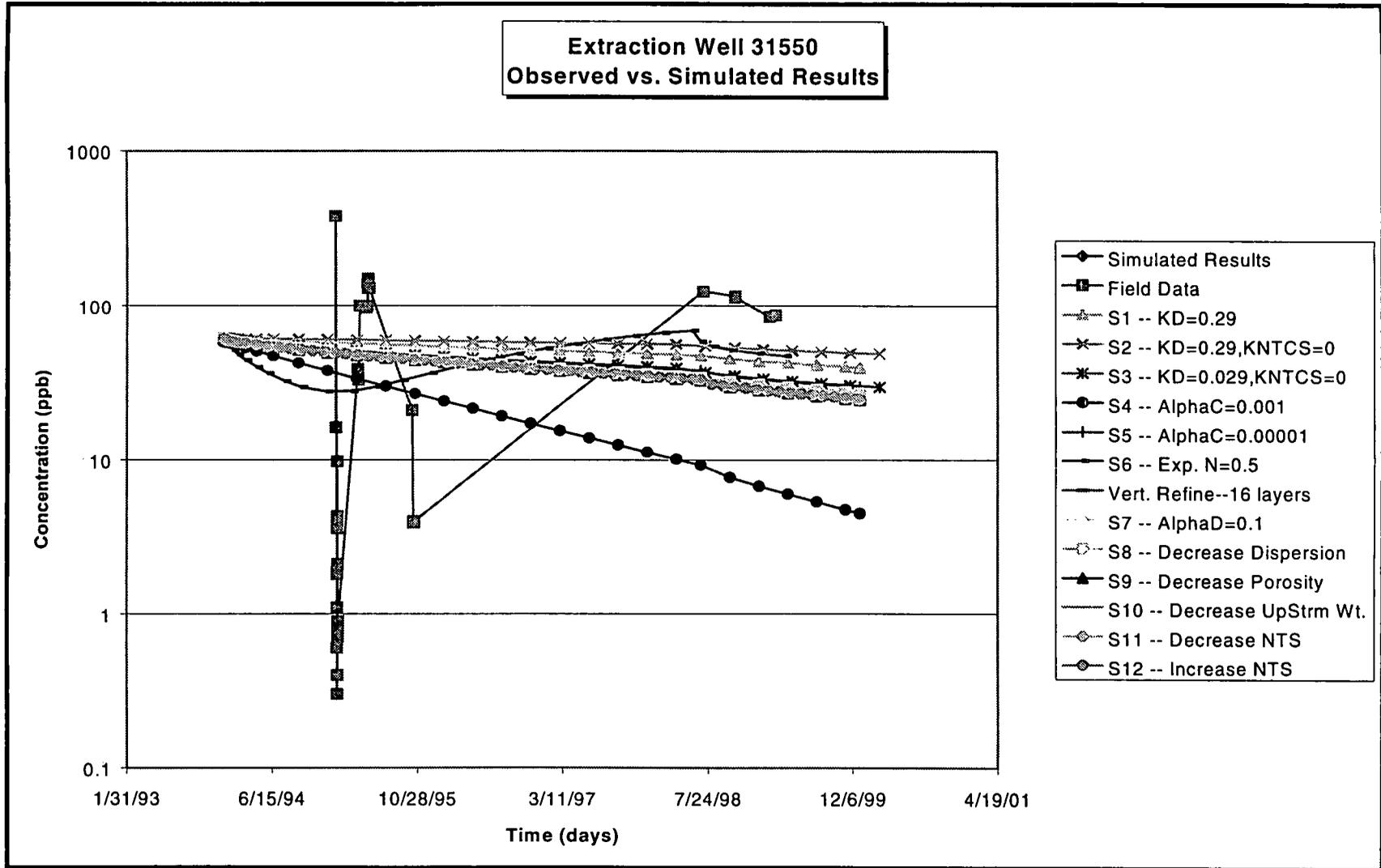


Figure 6.21 Breakthrough Curves at Extraction Well 31550

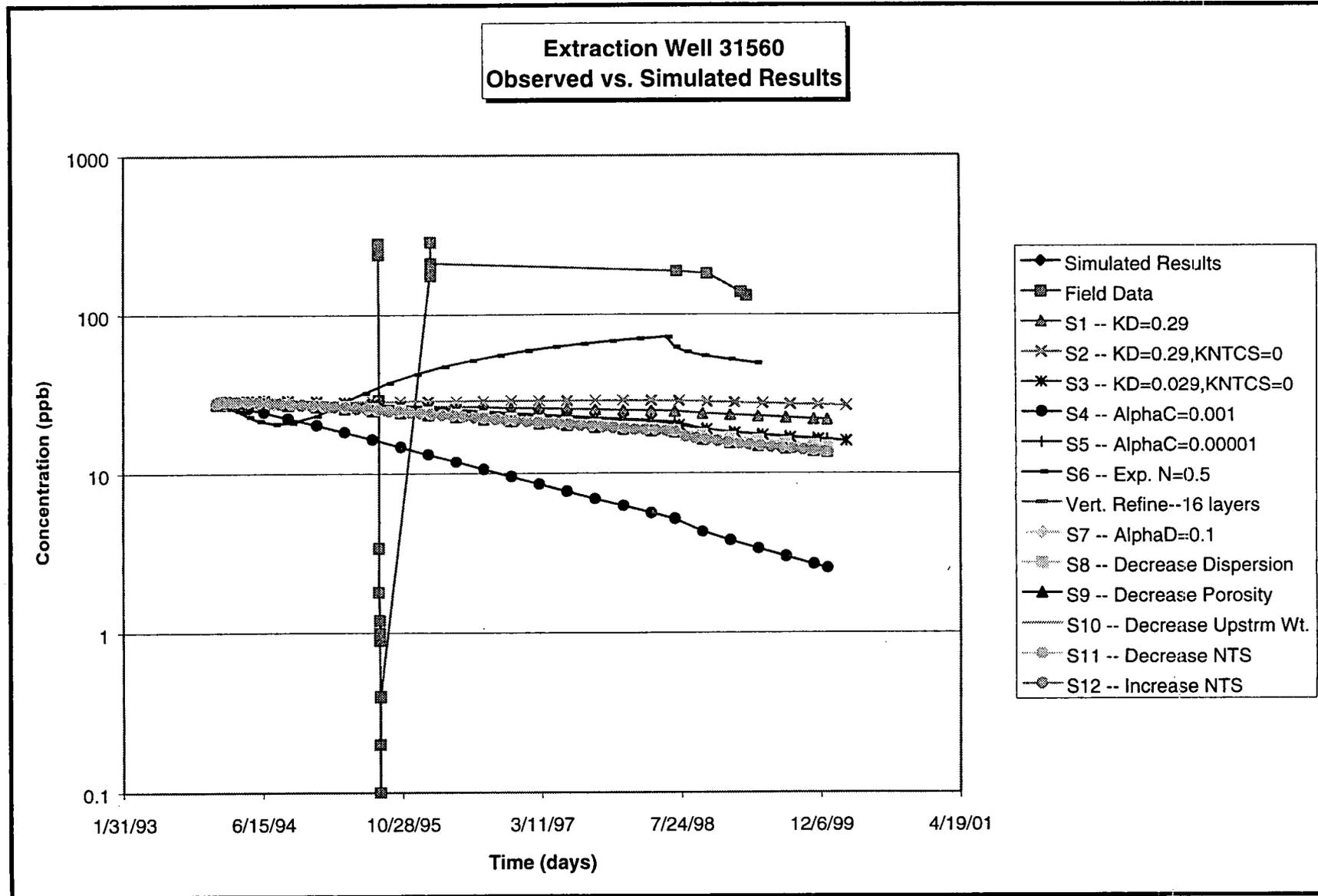


Figure 6.22 Breakthrough Curves at Extraction Well 31560.

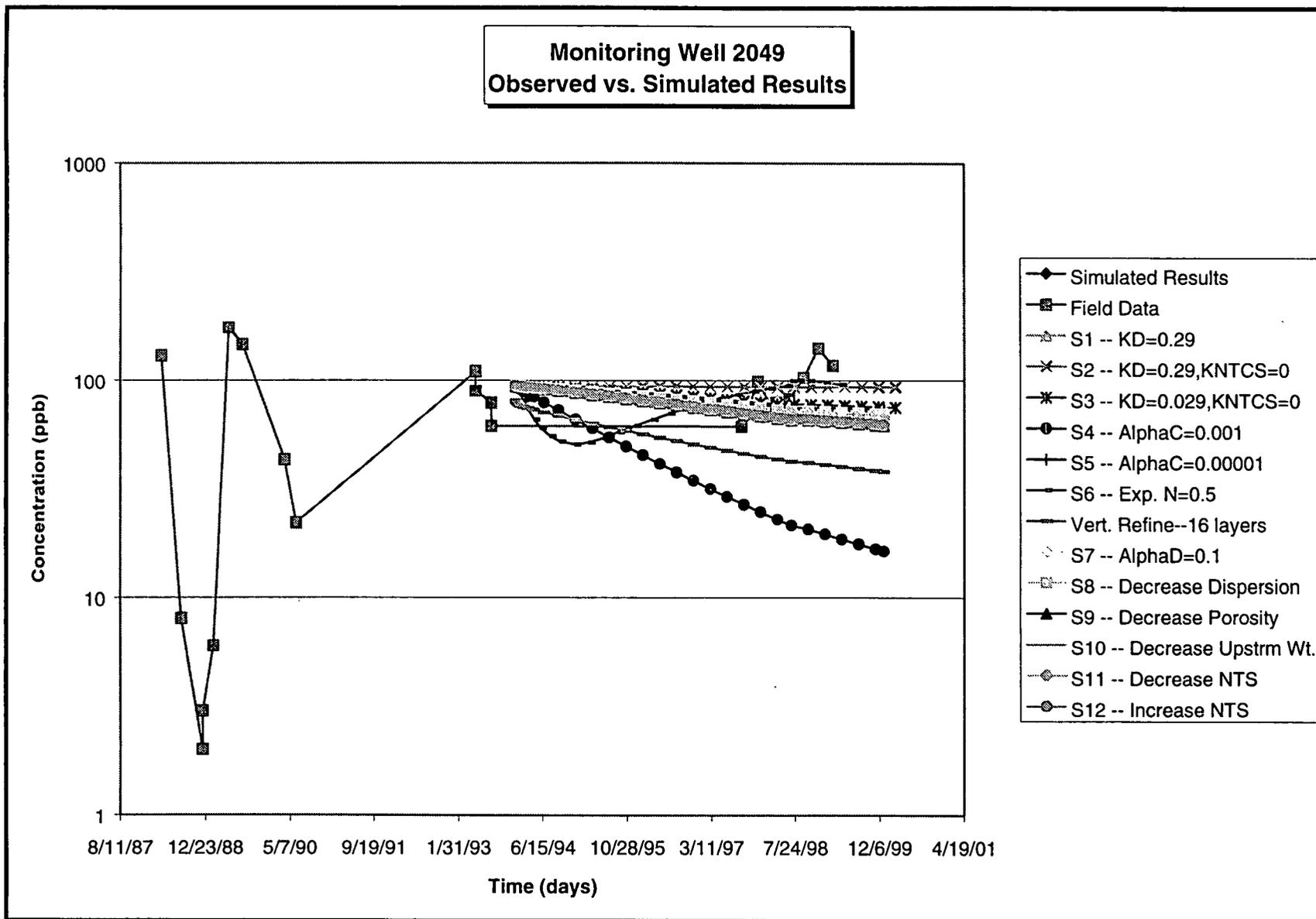


Figure 6.23 Breakthrough Curves at Monitoring Well 2049.

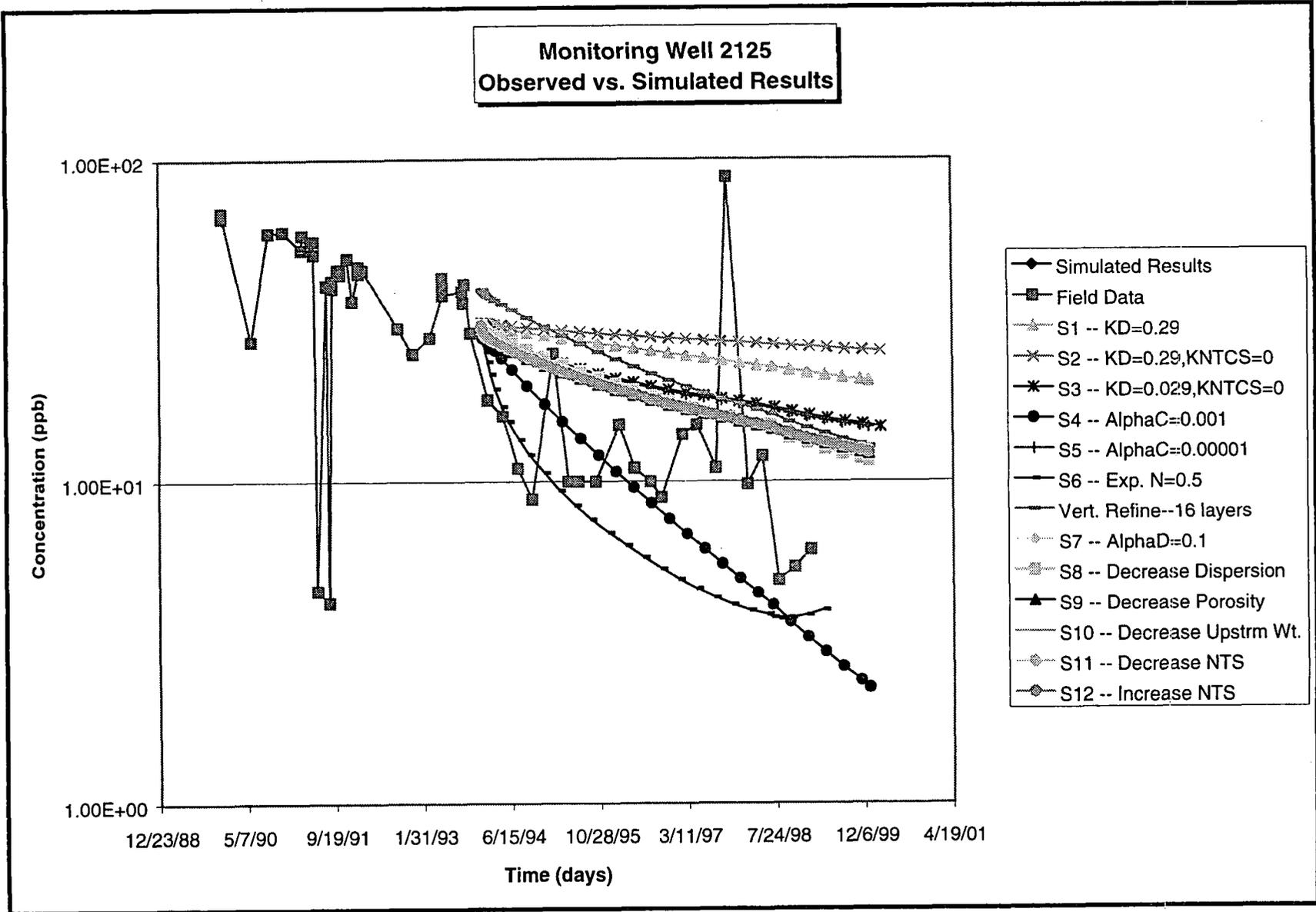


Figure 6.24 Breakthrough Curves at Monitoring Well 2125.

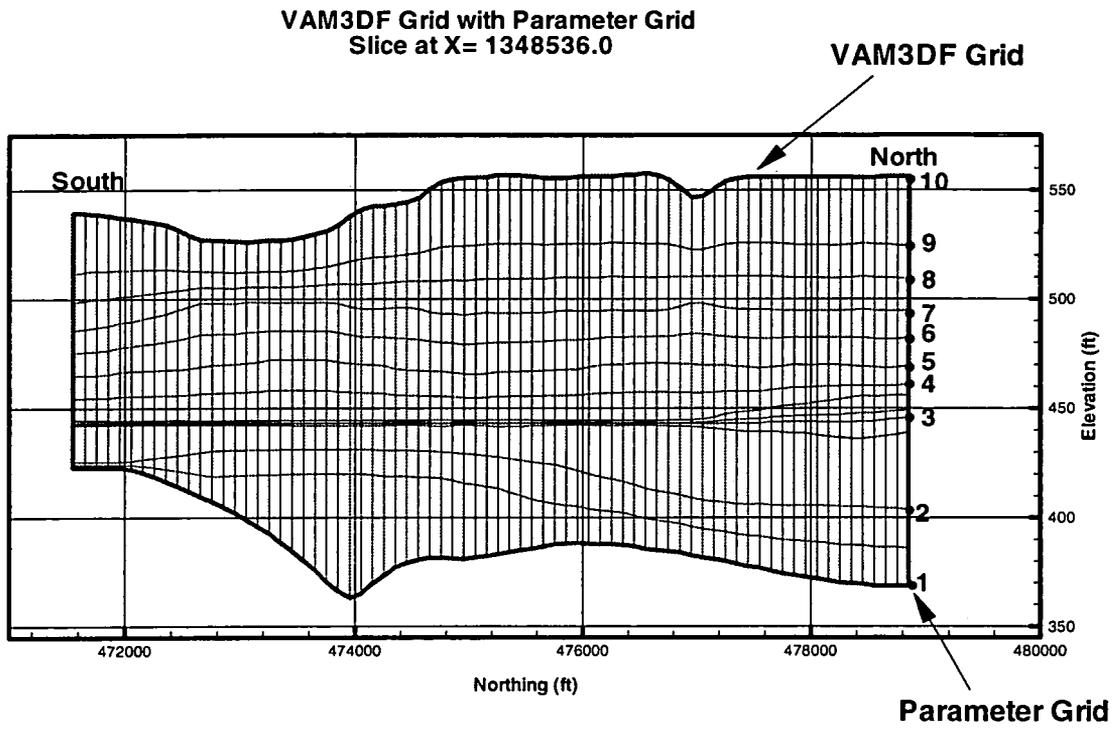


Figure 6.25 Slice of the VAM3DF Parameter Grid.

000133

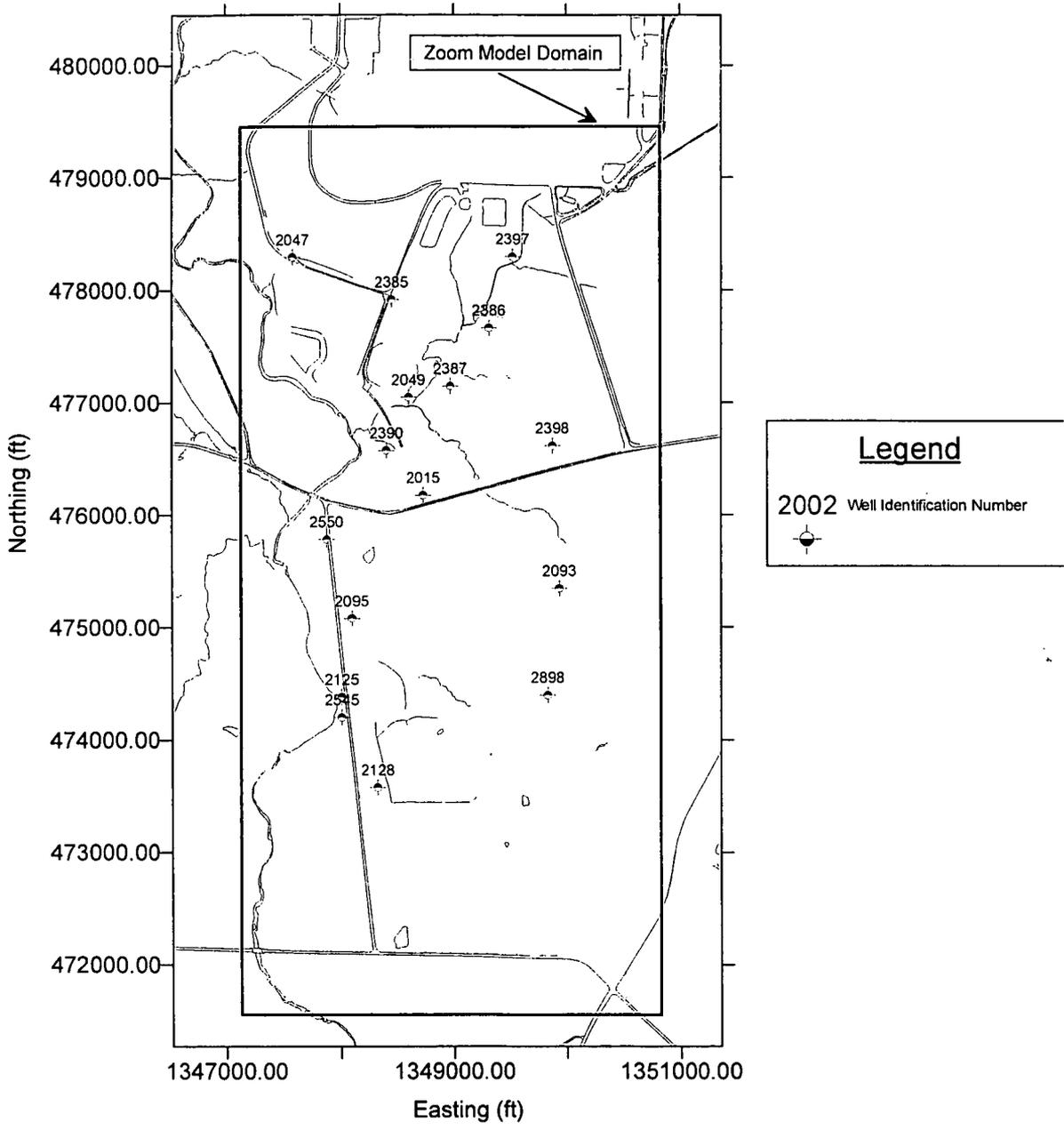


Figure 6.26 2000 Series Monitoring Wells.

000134

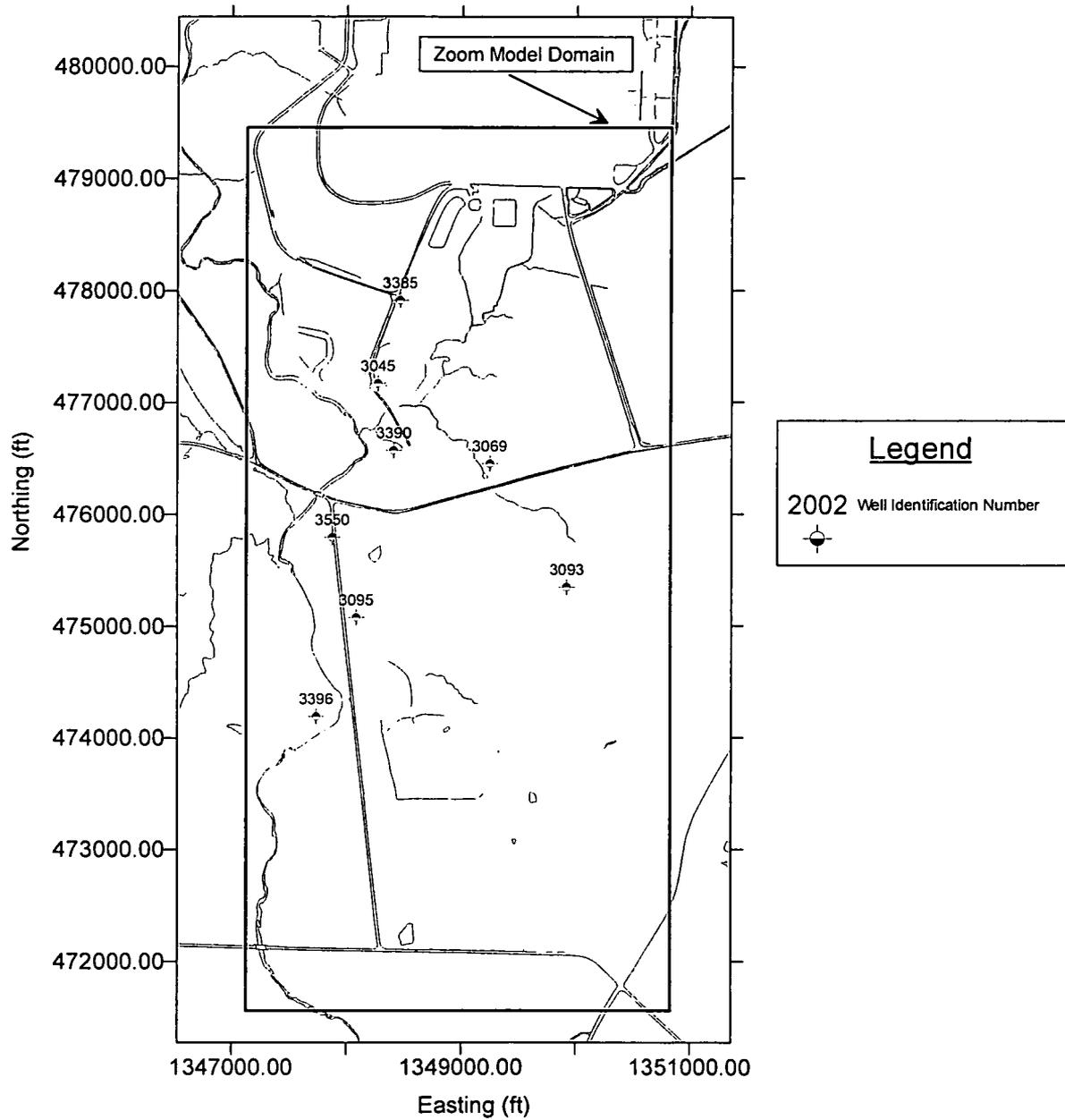


Figure 6.27 3000 Series Monitoring Wells.

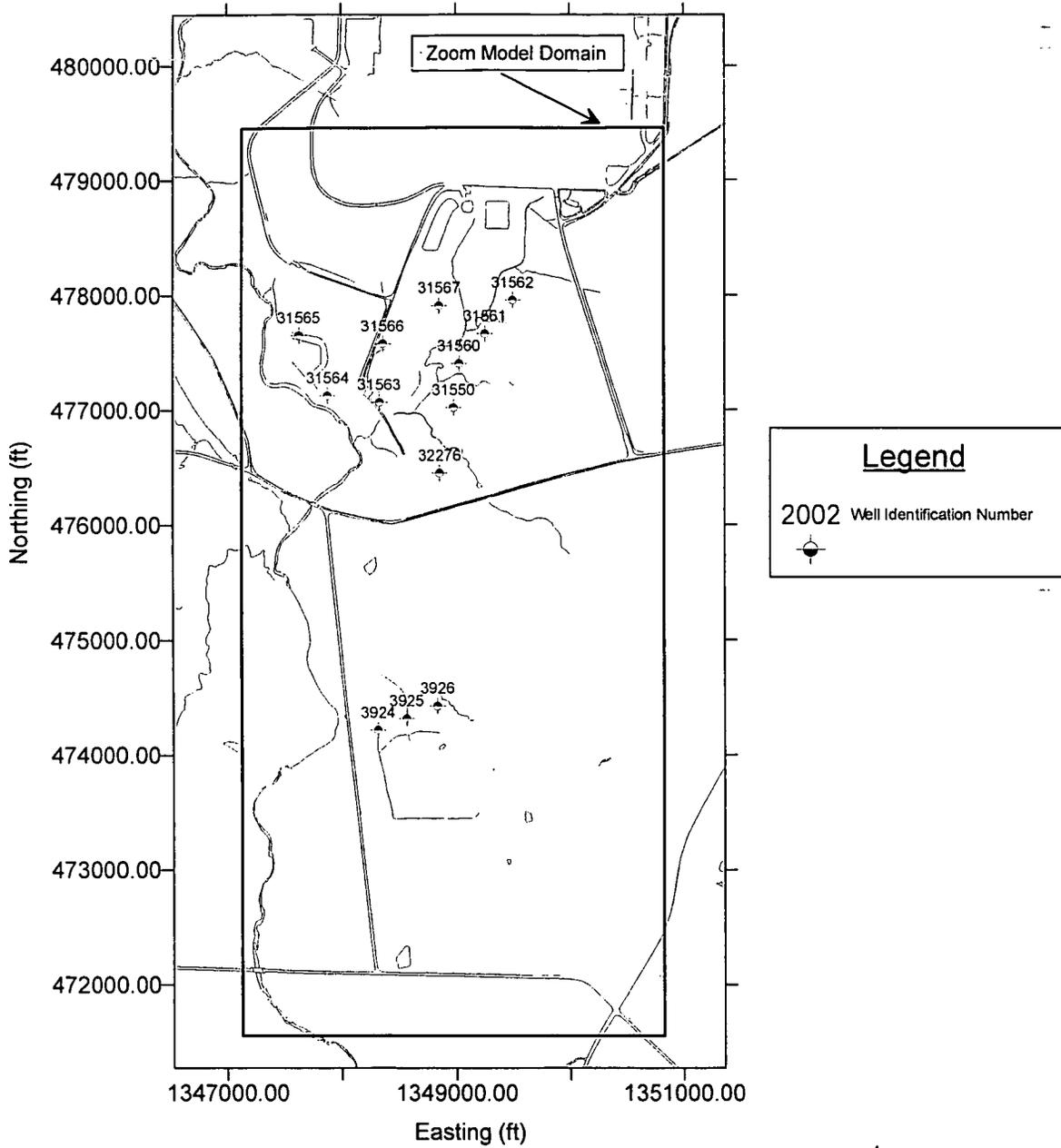


Figure 6.28 Extraction Wells.

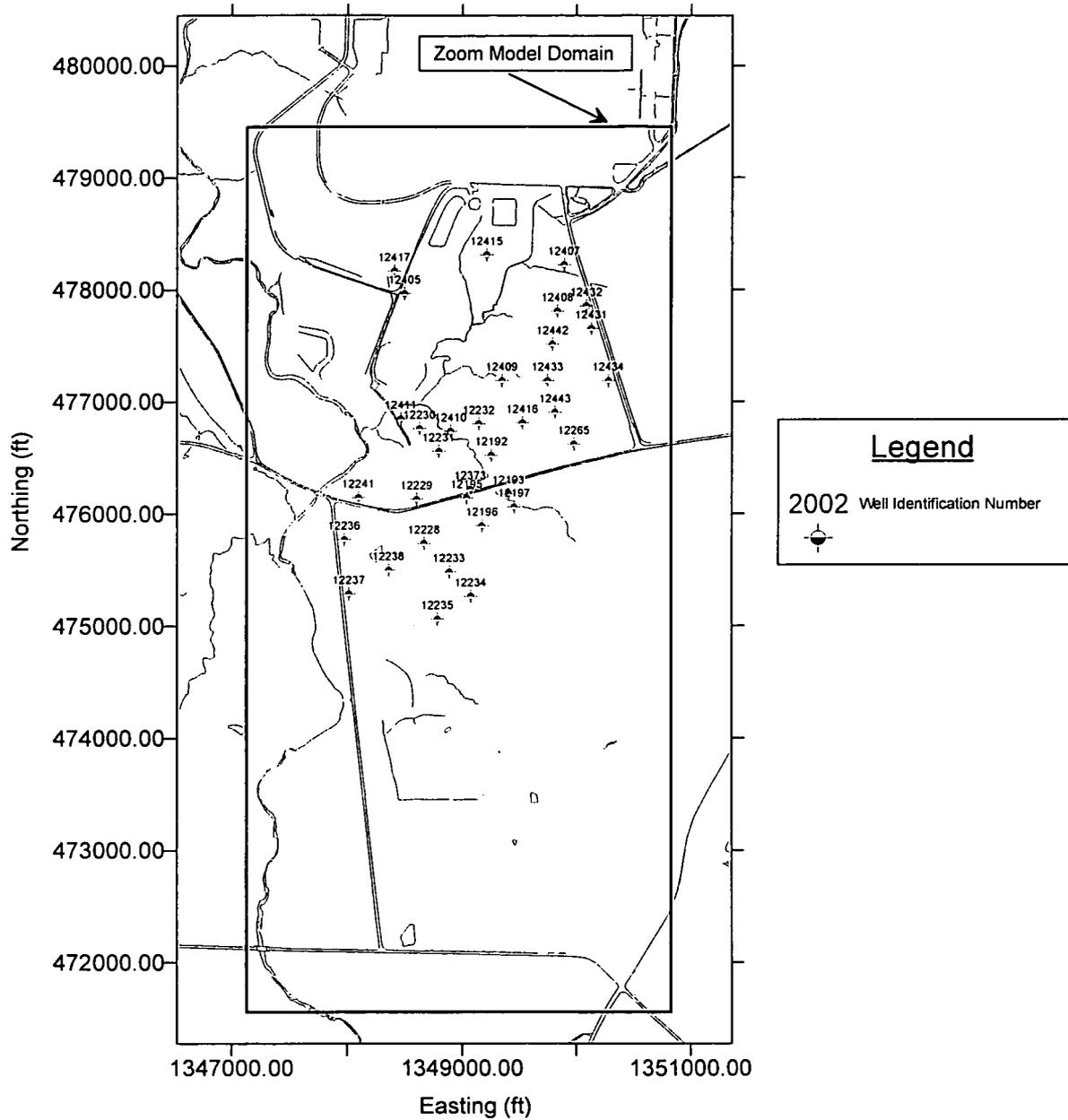


Figure 6.29 Geoprobe Locations.

Baseline Concentrations for Layer 11 at 1/1/98

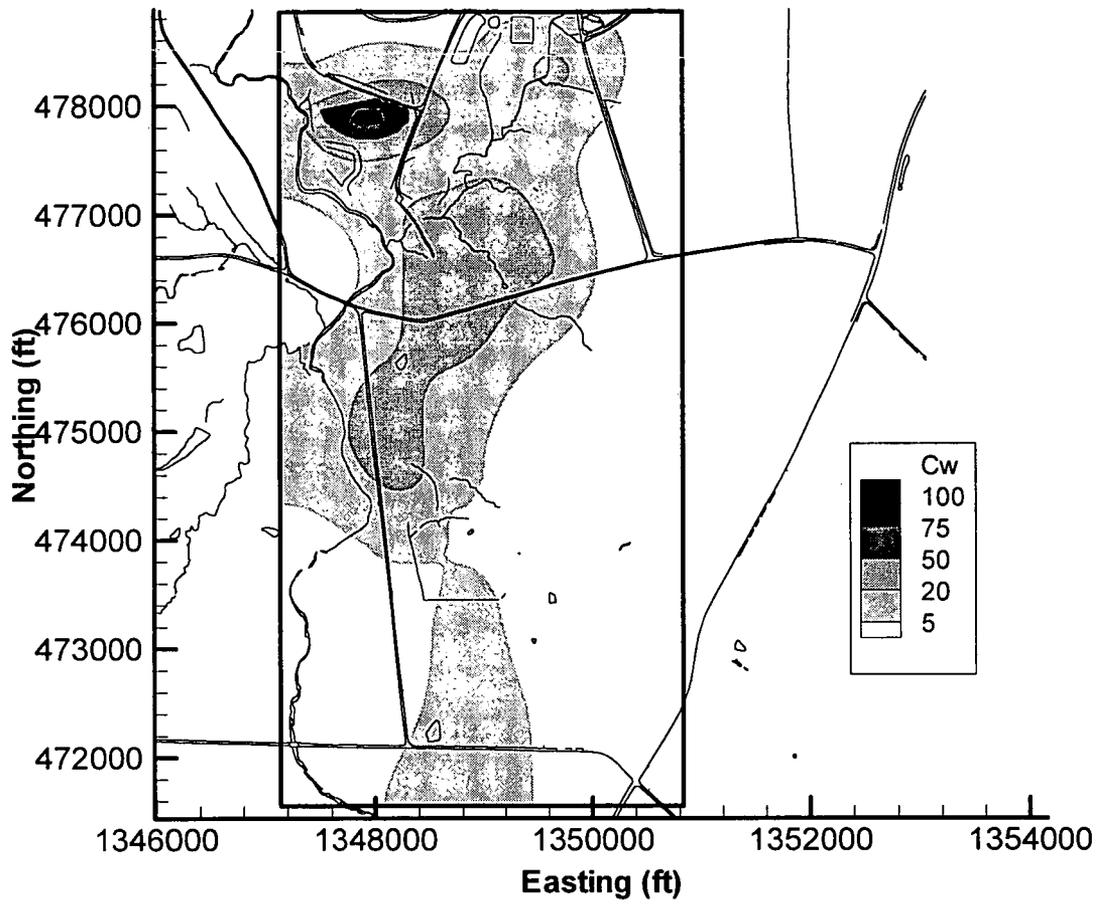


Figure 6.30 Baseline Concentrations for Layer 11 at 1/1/98.

Baseline Concentrations for Layer 11 at 1/1/02

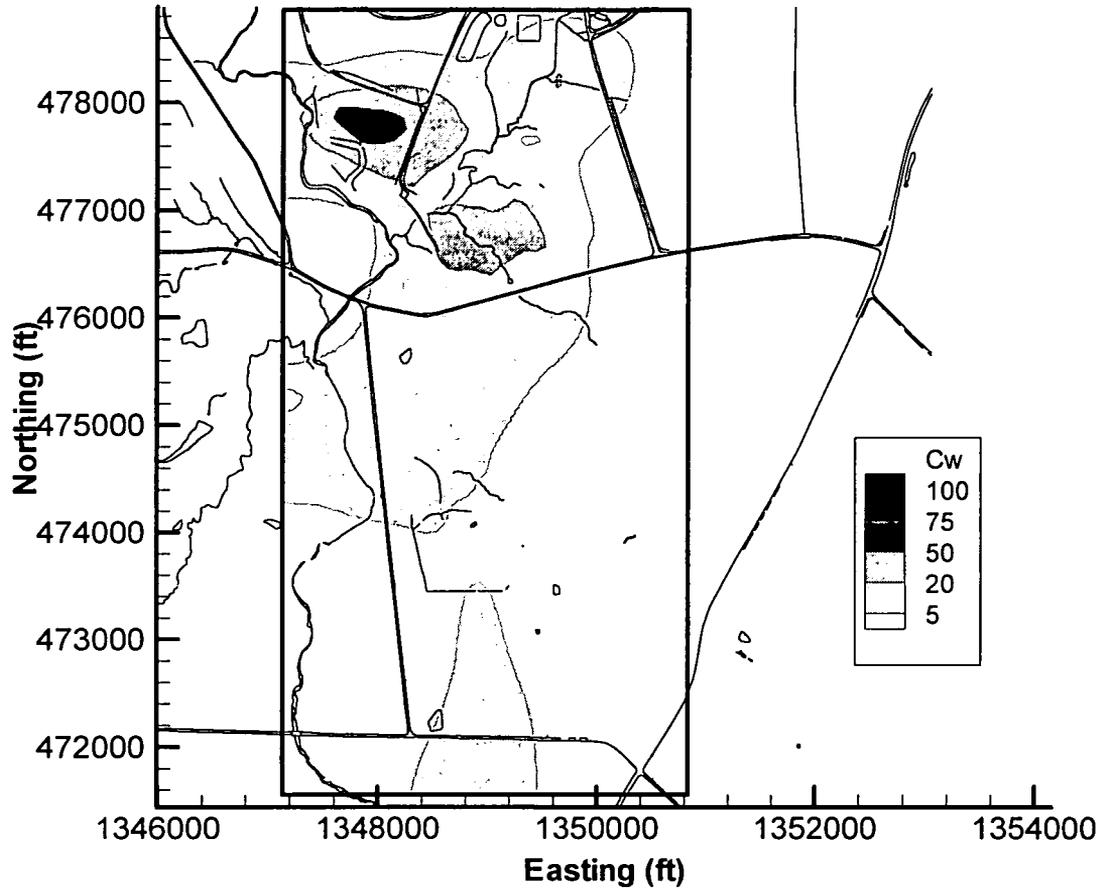


Figure 6.31 Baseline Concentrations for Layer 11 at 1/1/02.

000139

Baseline Concentrations for Layer 11 at 5/1/06

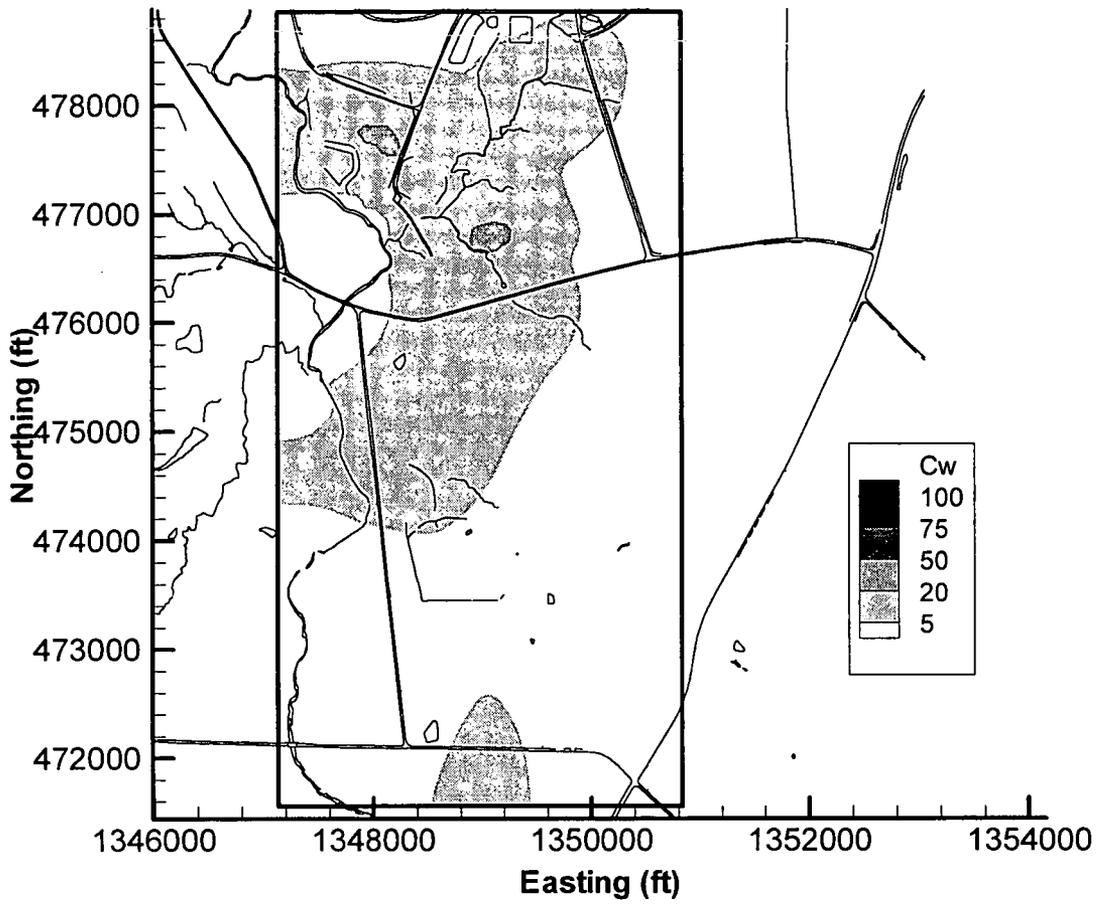


Figure 6.32 Baseline Concentrations for Layer 11 at 5/1/06.

000140

Baseline Concentrations for Layer 13 at 1/1/98

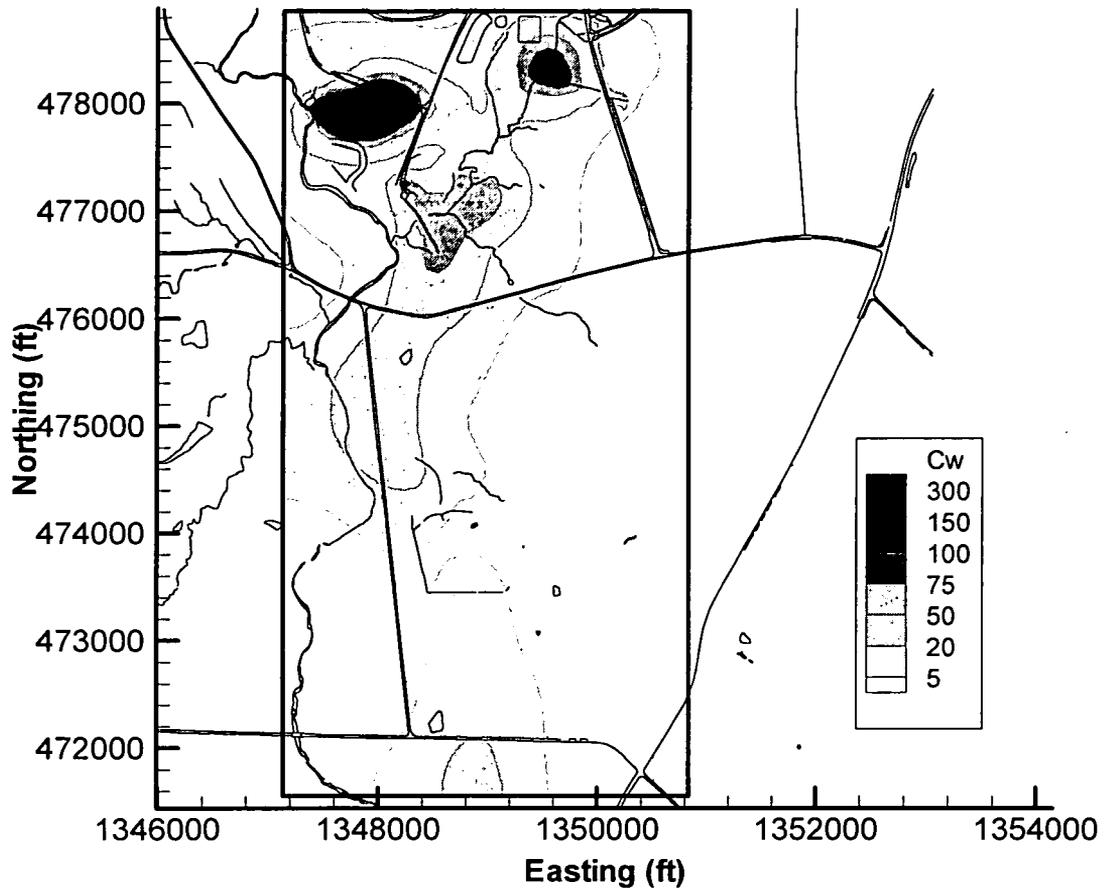


Figure 6.33 Baseline Concentrations for Layer 13 at 1/1/98.

Baseline Concentrations for Layer 13 at 1/1/02

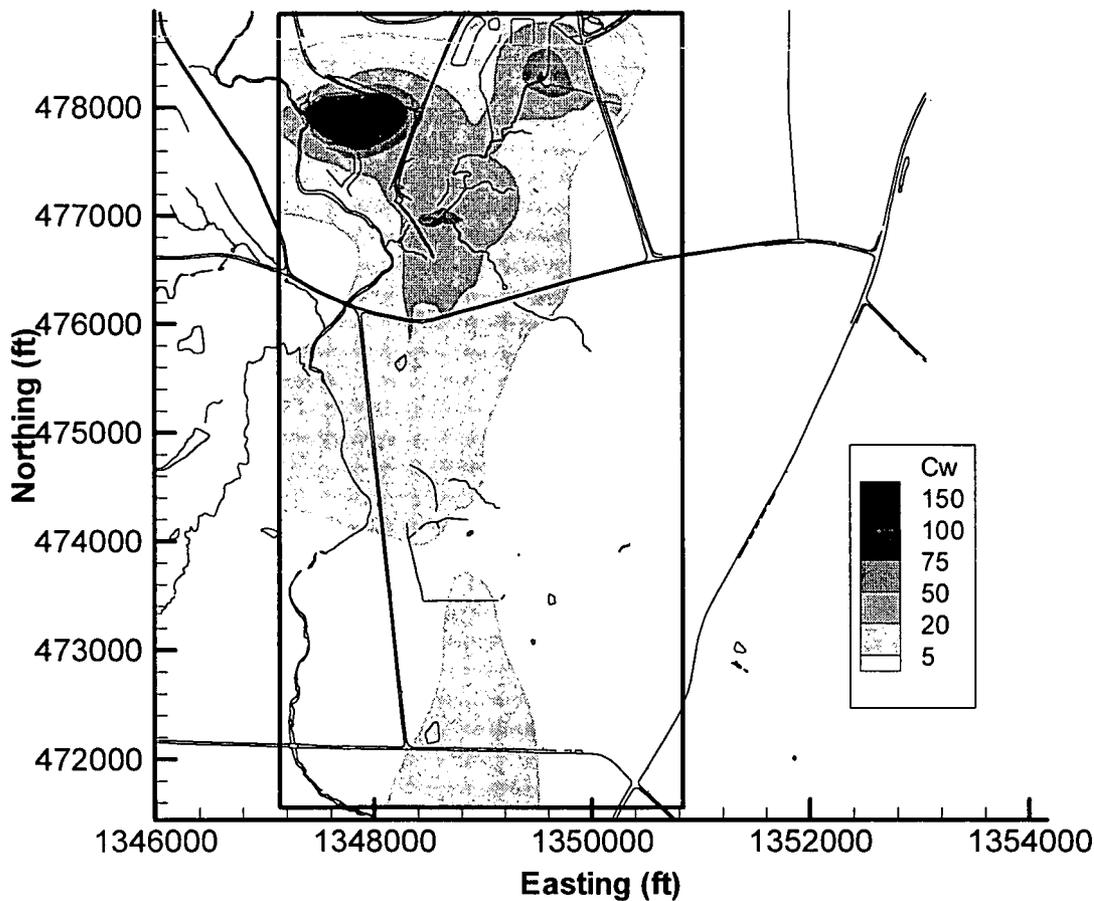


Figure 6.34 Baseline Concentrations for Layer 13 at 1/1/02.

Baseline Concentrations for Layer 13 at 5/1/06

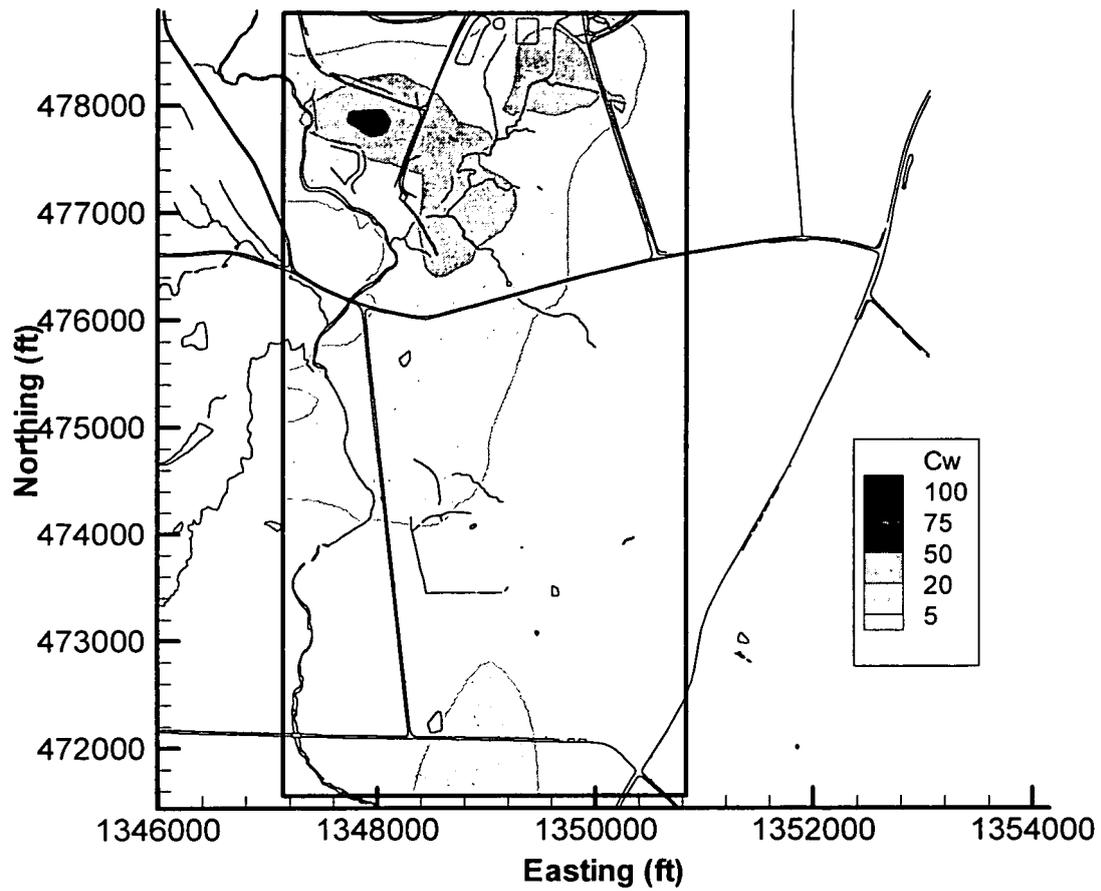


Figure 6.35 Baseline Concentrations for Layer 13 at 5/1/06.

Maximum Concentration for Baseline Simulations

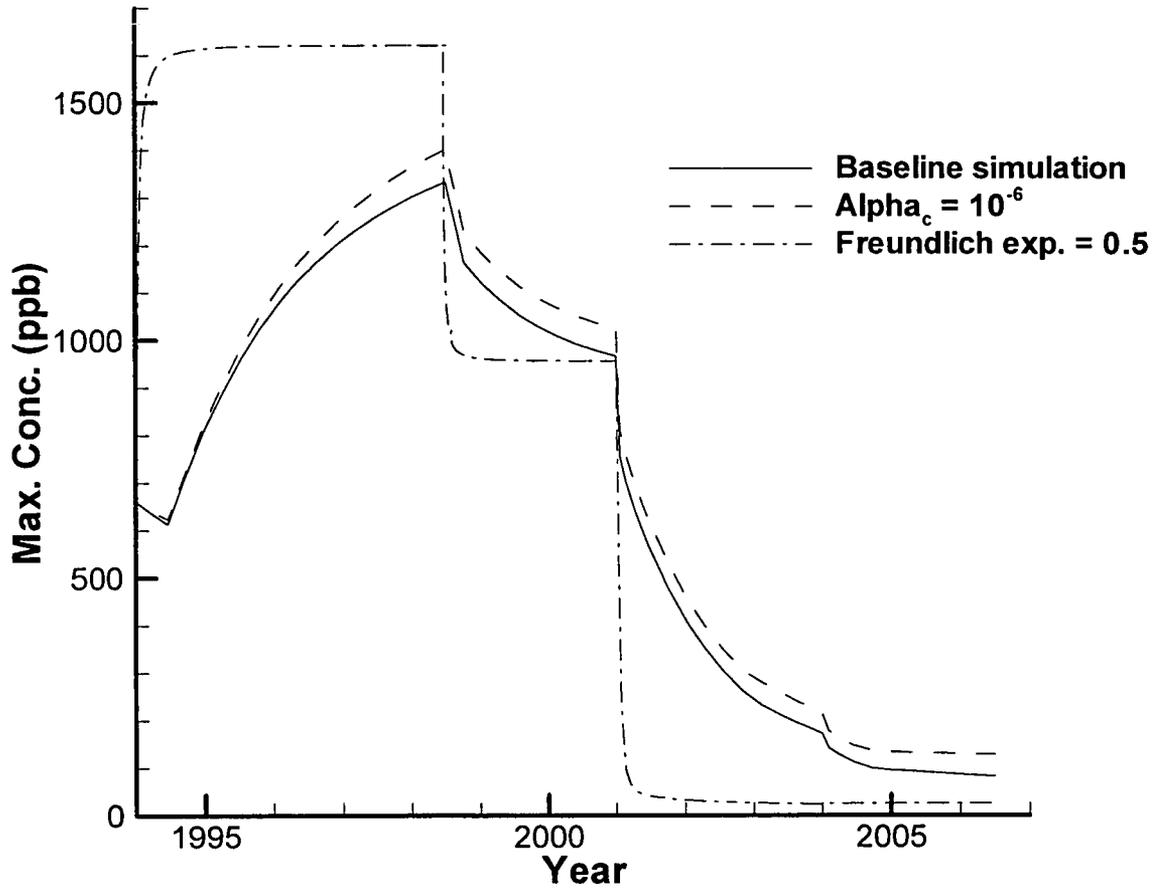


Figure 6.36 Temporal Variation of Maximum Concentrations for Baseline Simulations.

Baseline Transport Simulation Mass Summary

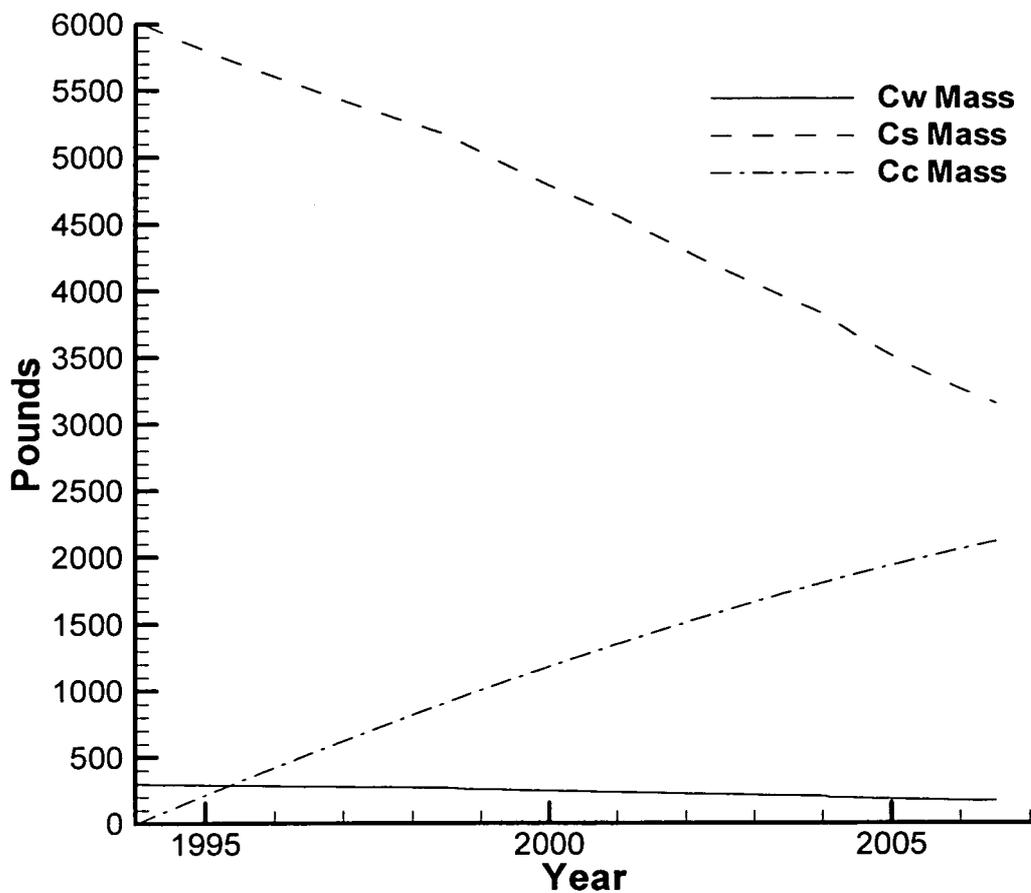


Figure 6.37 Baseline Transport Simulation Uranium Mass Summary. Cw = Aqueous Mass, Cs = Mass Sorbed in solid Phase, and Cc = Chemisorbed Mass.

Baseline Simulation Mass Summary with $\alpha_c = 10^{-6}$

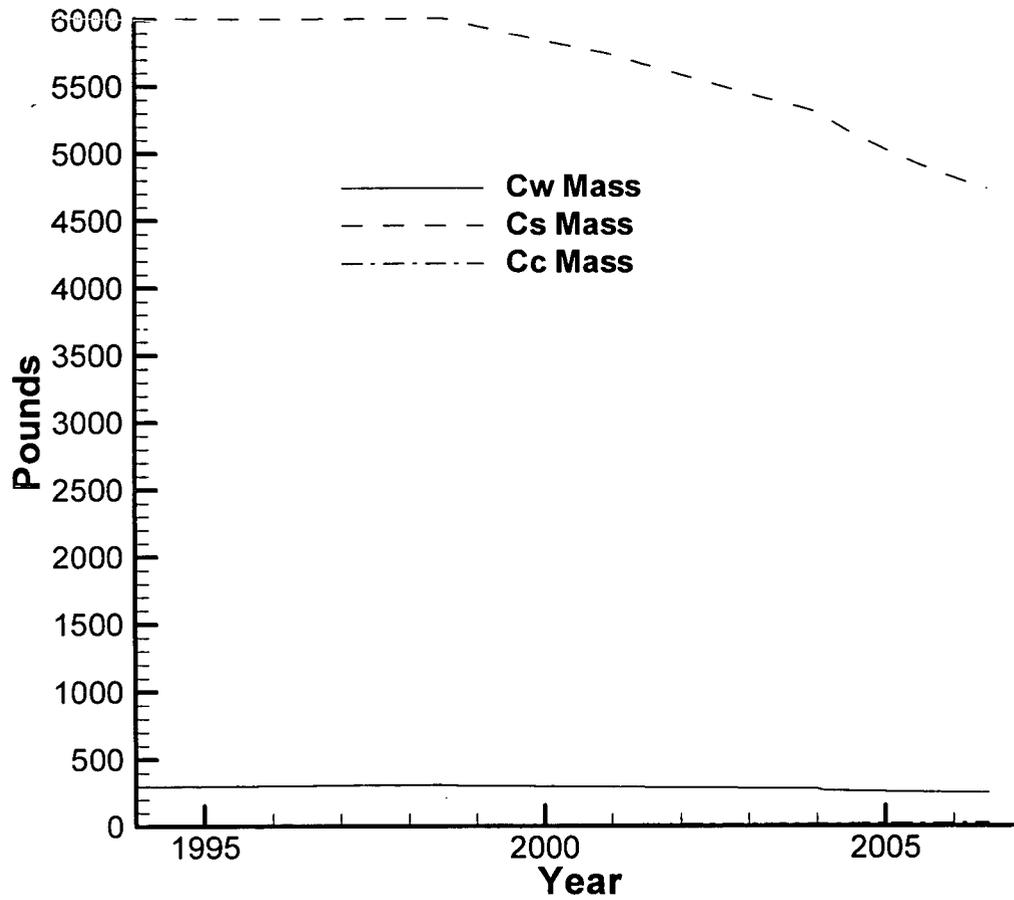


Figure 6.38 Baseline Simulation Uranium Mass Summary with $\alpha_c = 10^{-6}$. (See Figure 6.37 caption for nomenclature).

Baseline Simulation Mass Summary with Freundlich exp. = 0.5

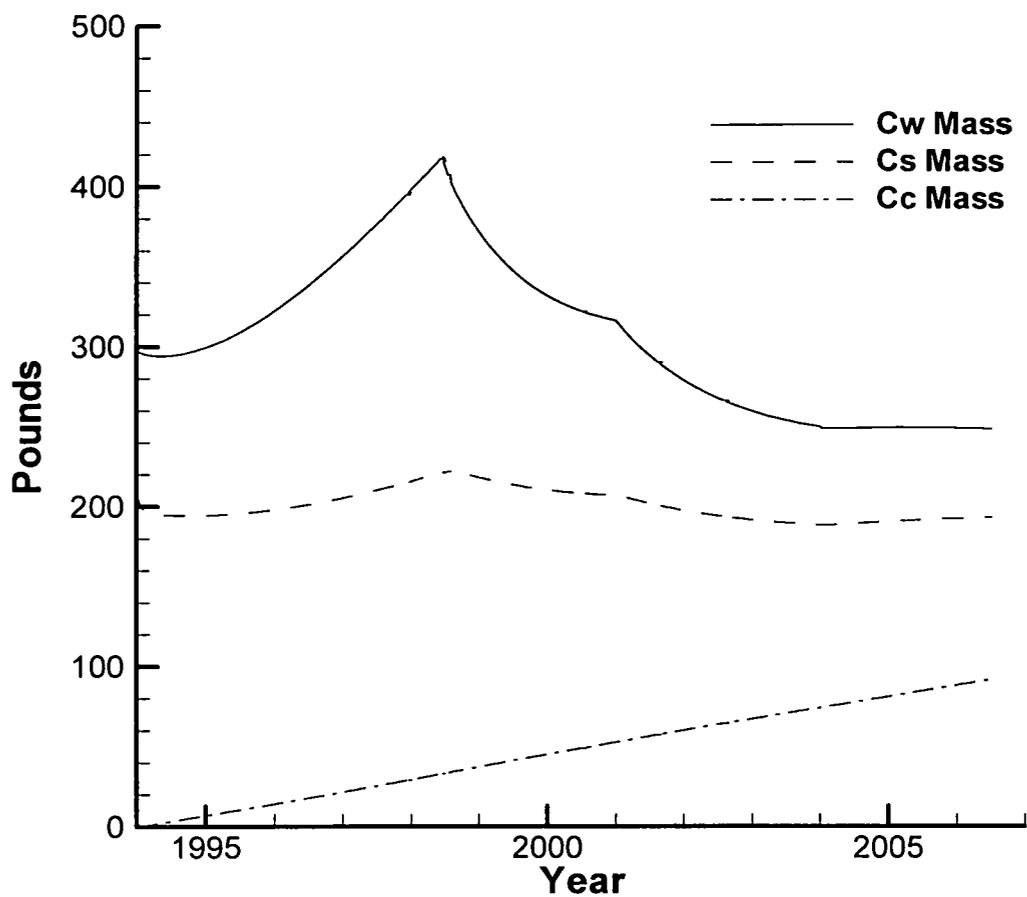


Figure 6.39 Baseline Simulation Uranium Mass Summary with Freundlich exp. = 0.5. (See Figure 6.37 caption for nomenclature).

DFM/VAM3DF Estimates of Initial Conditions -- Model Layer 5

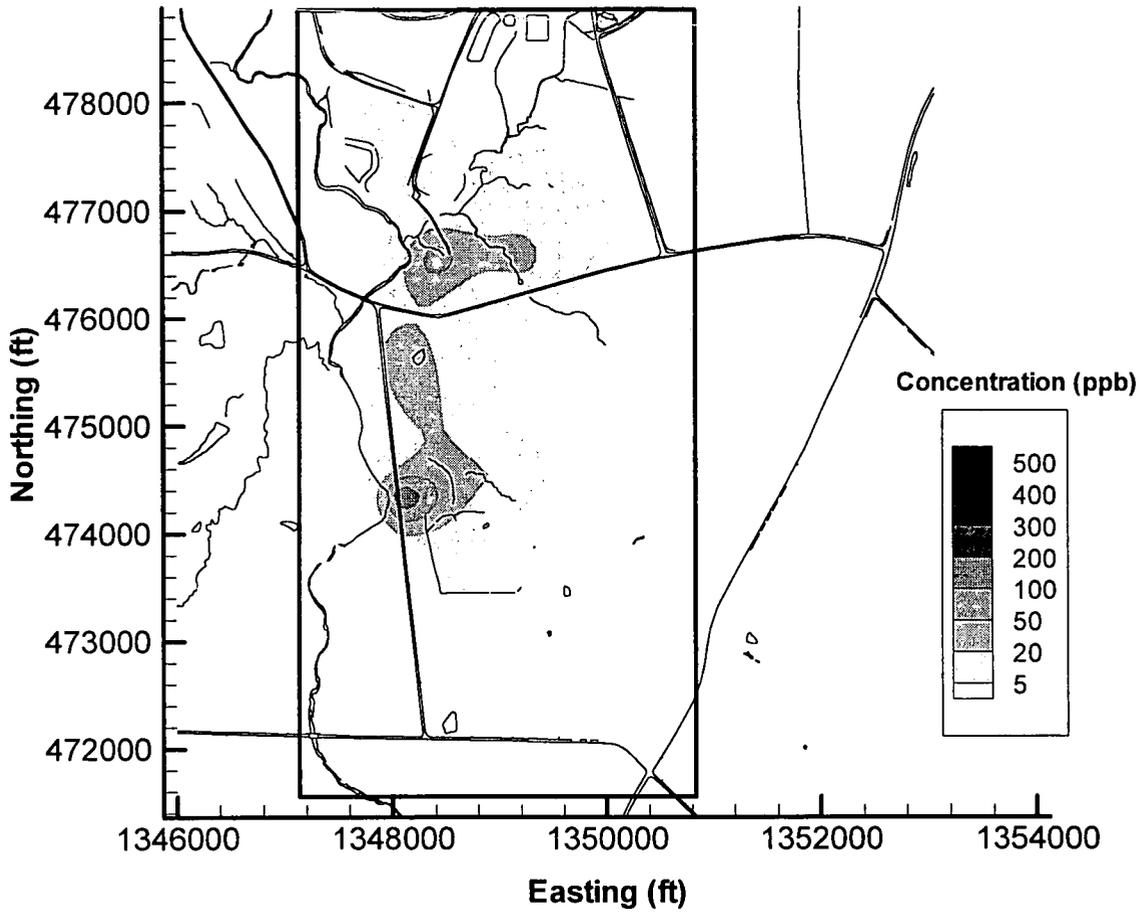


Figure 6.40 IC Estimation Run 1 -- Model Layer 5.

DFM/VAM3DF Estimates of Initial Conditions -- Model Layer 8

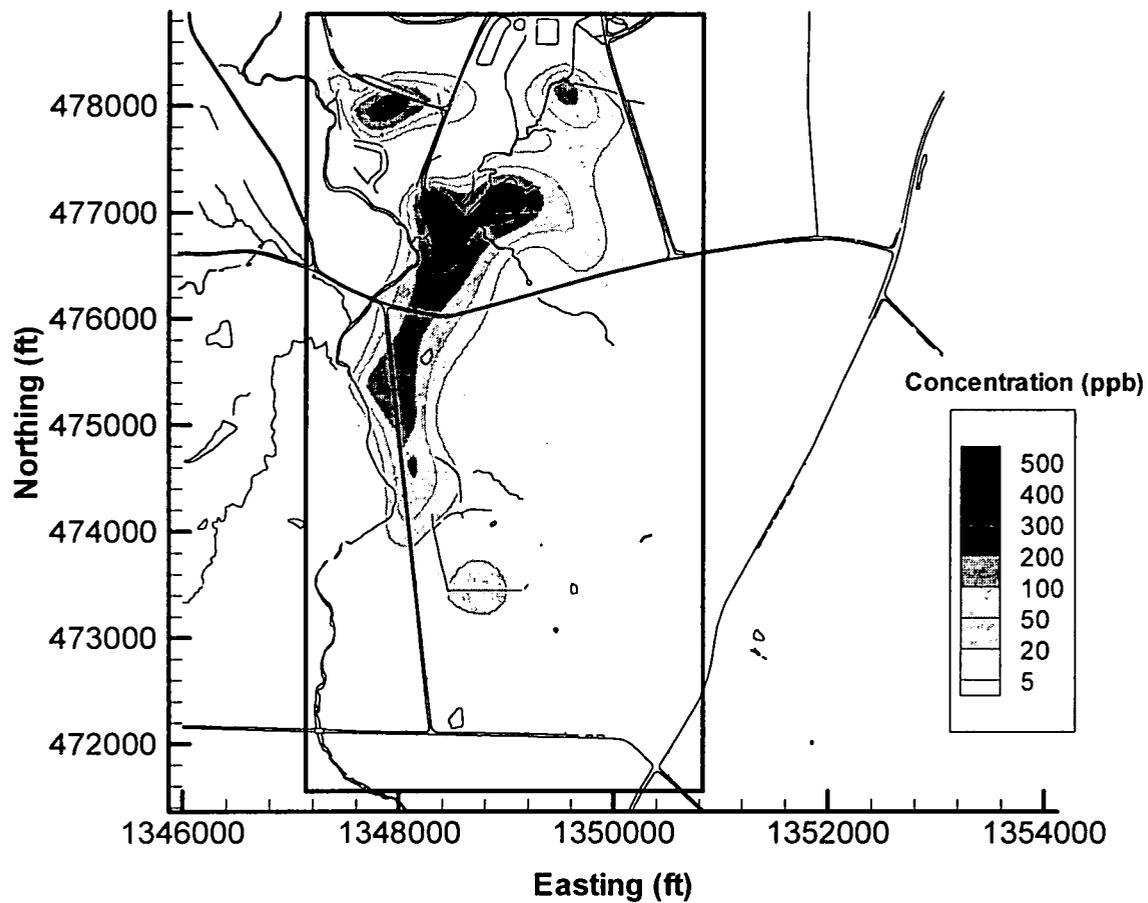


Figure 6.41 IC Estimation Run 1 -- Model Layer 8.

DFM/VAM3DF Estimates of K_d -- Model Layer 5

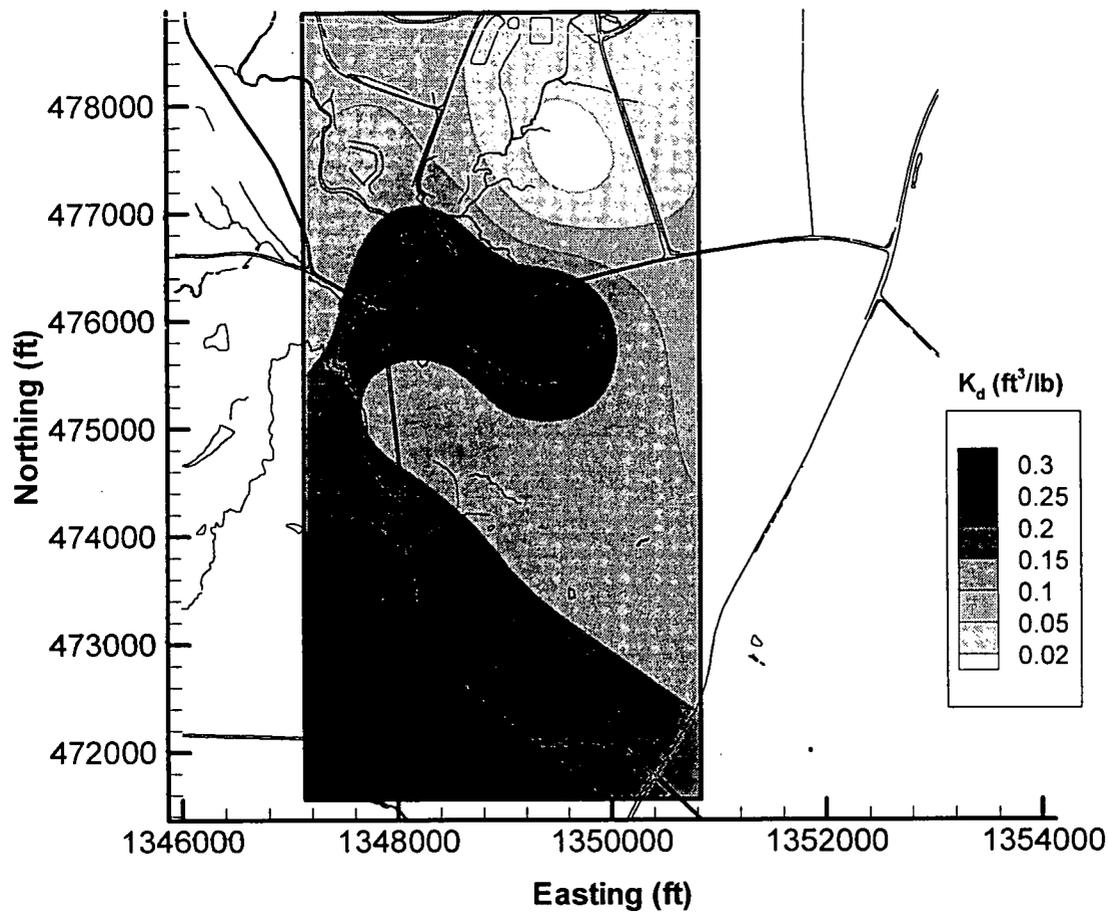


Figure 6.42 K_d Estimation Run 1 -- Model Layer 5.

000150

DFM/VAM3DF Estimates of K_d -- Model Layer 8

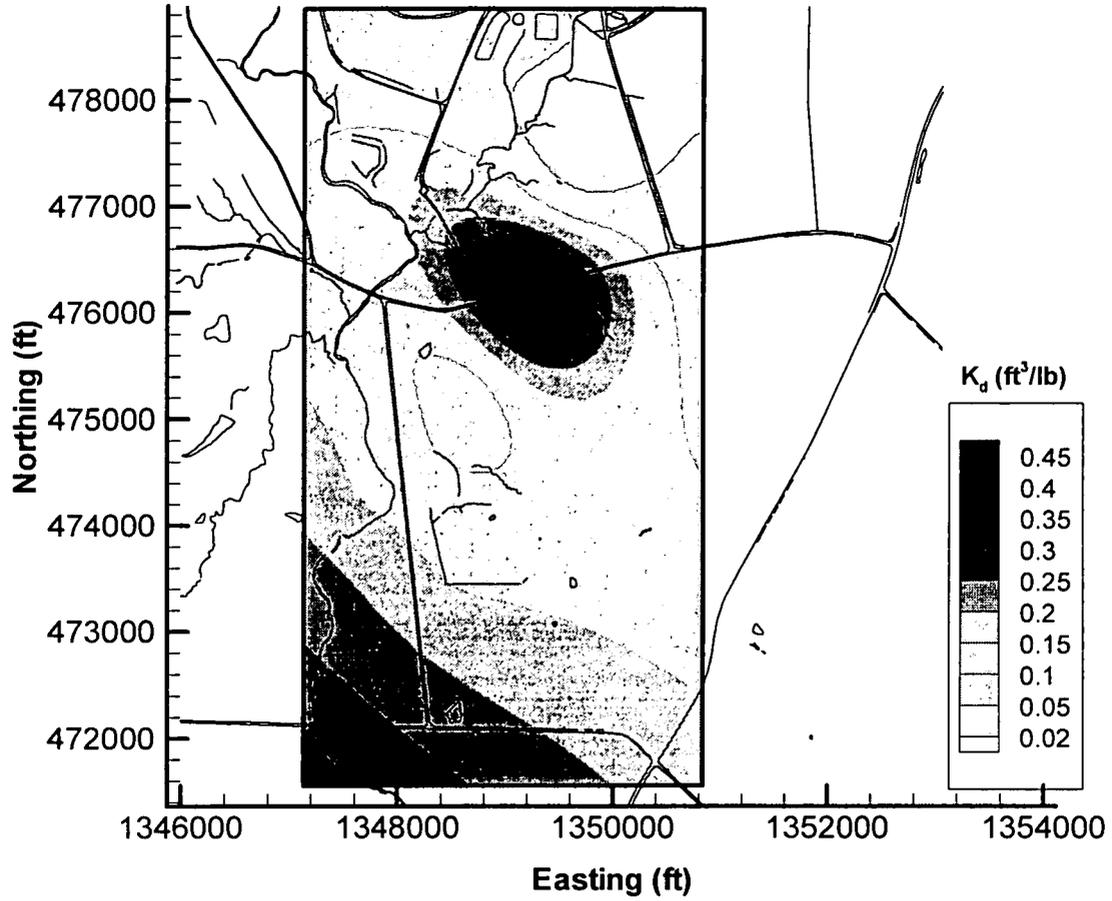


Figure 6.43 K_d Estimation Run 1 -- Model Layer 8.

000151

DFM/VAM3DF Estimates of K_d -- Model Layer 5

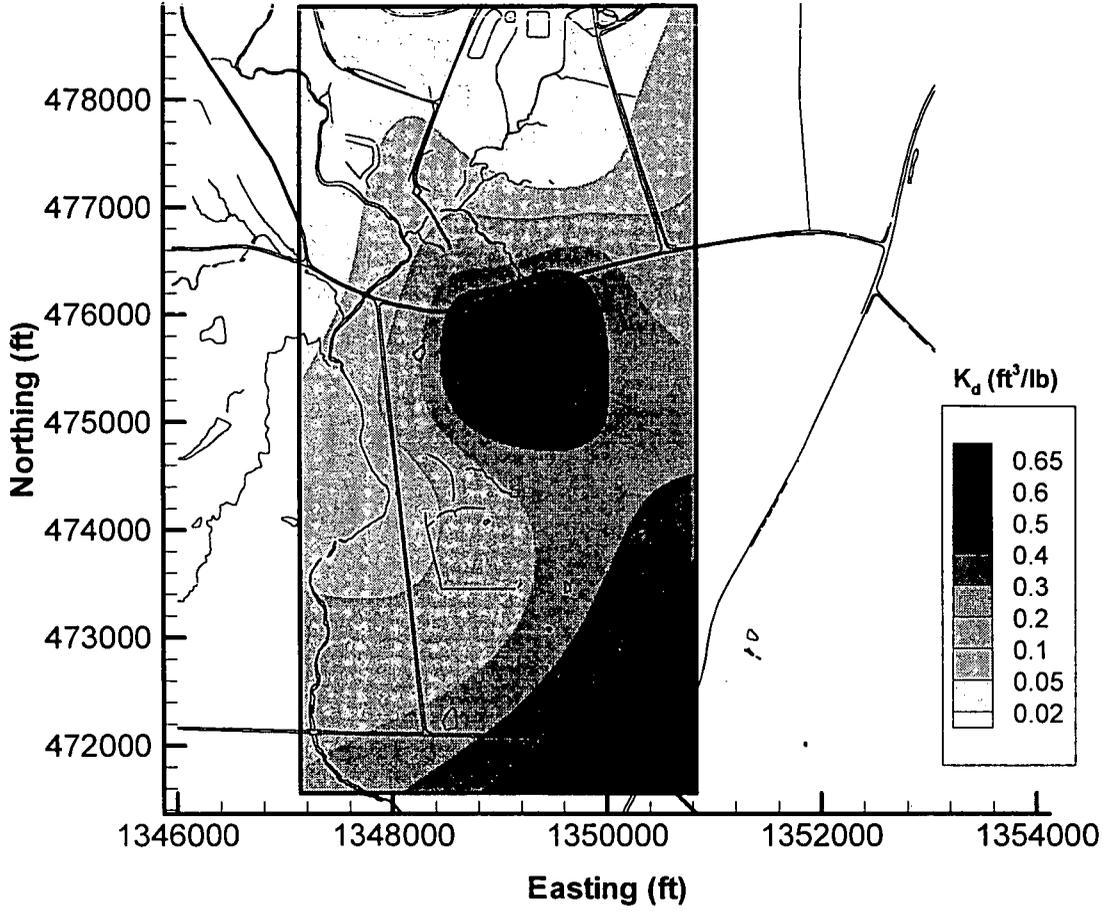


Figure 6.44 K_d Estimation Run 2 -- Model Layer 5.

000152

DFM/VAM3DF Estimates of K_d -- Model Layer 8

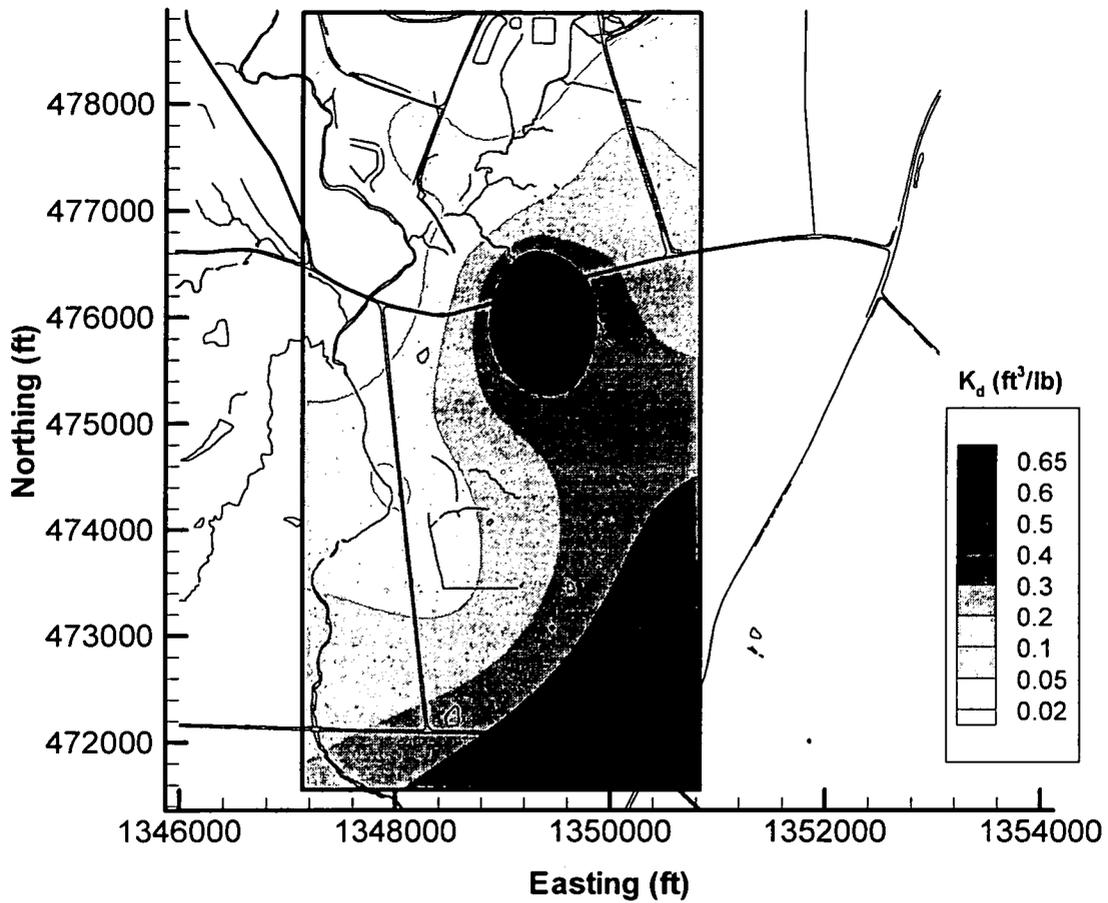


Figure 6.45 K_d Estimation Run 2 -- Model Layer 8.

000153

DFM/VAM3DF Estimates of K_d -- Model Layer 5

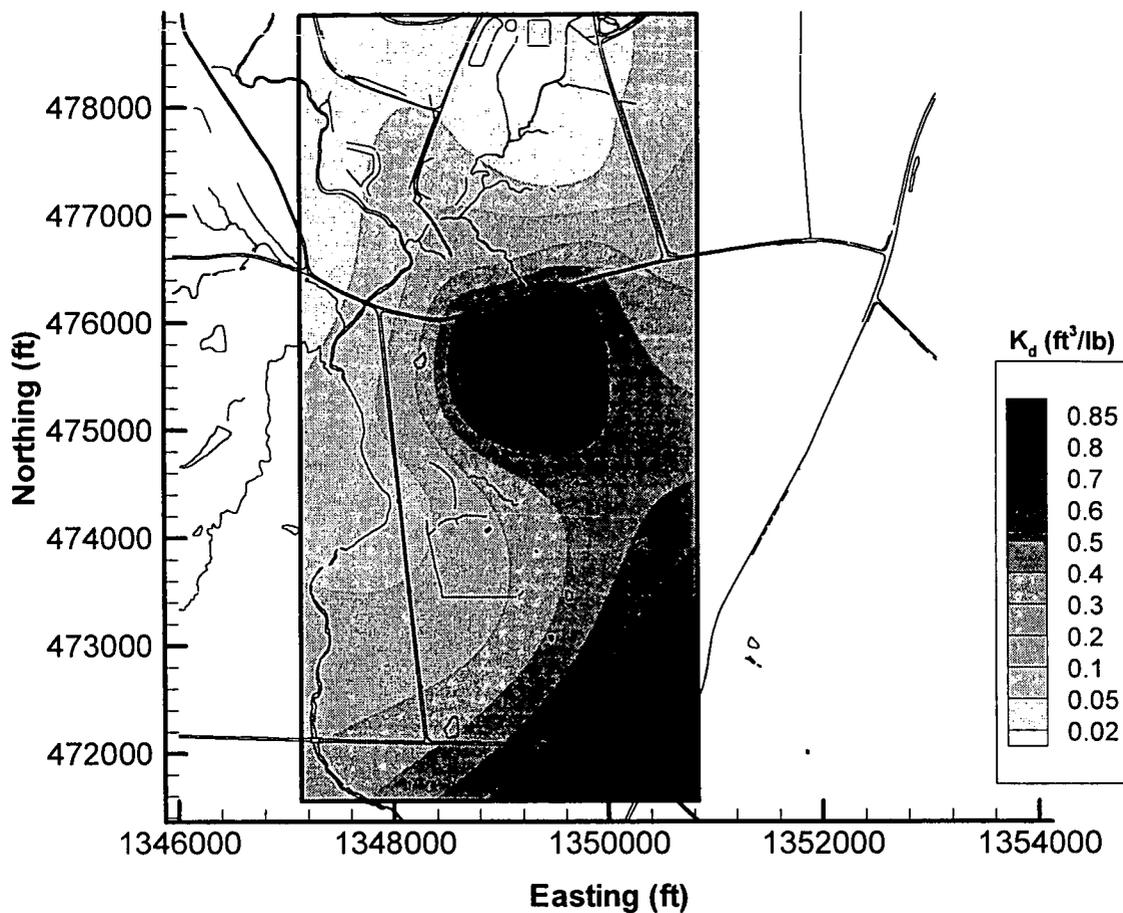


Figure 6.46 K_d Estimation Run 3 -- Model Layer 5.

000154

DFM/VAM3DF Estimates of K_d -- Model Layer 8

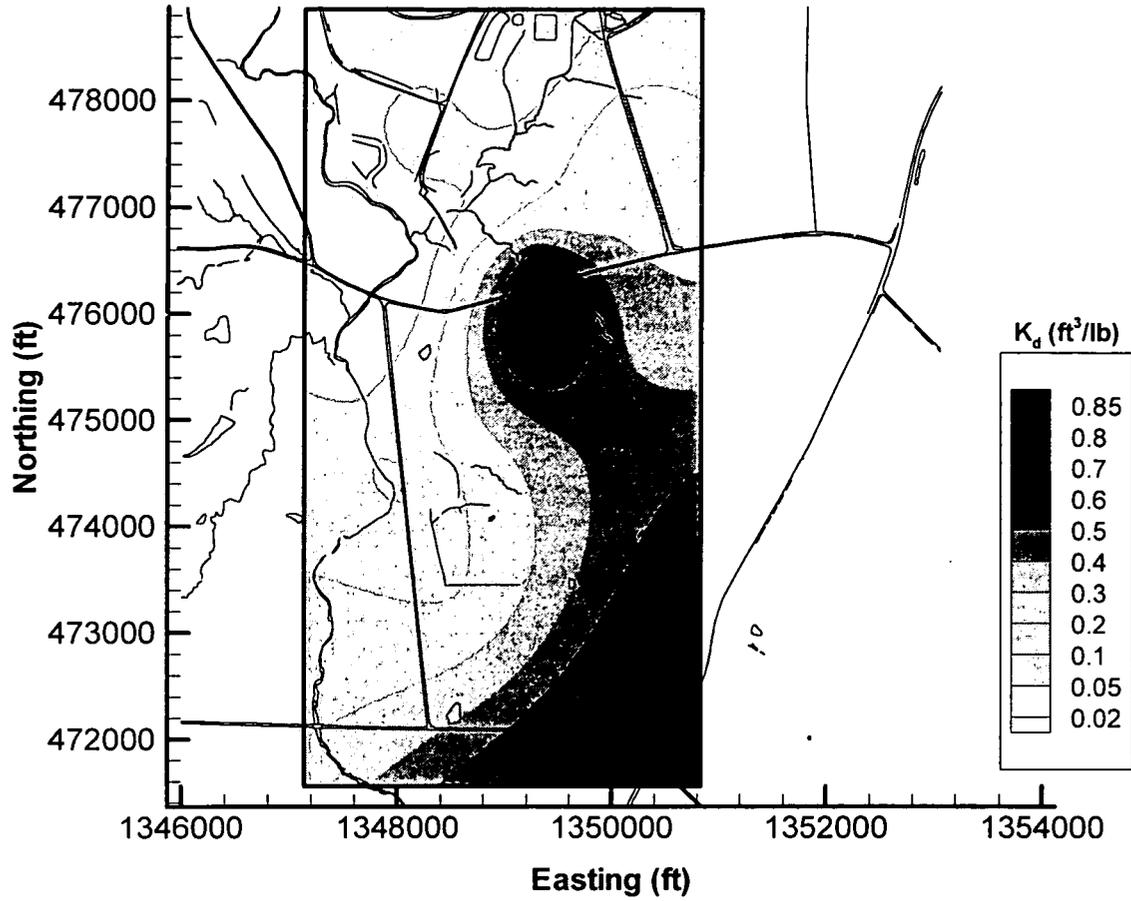


Figure 6.47 K_d Estimation Run 3 -- Model Layer 8.

K_d Estimation #1 Concentrations for Layer 11 at 1/1/98

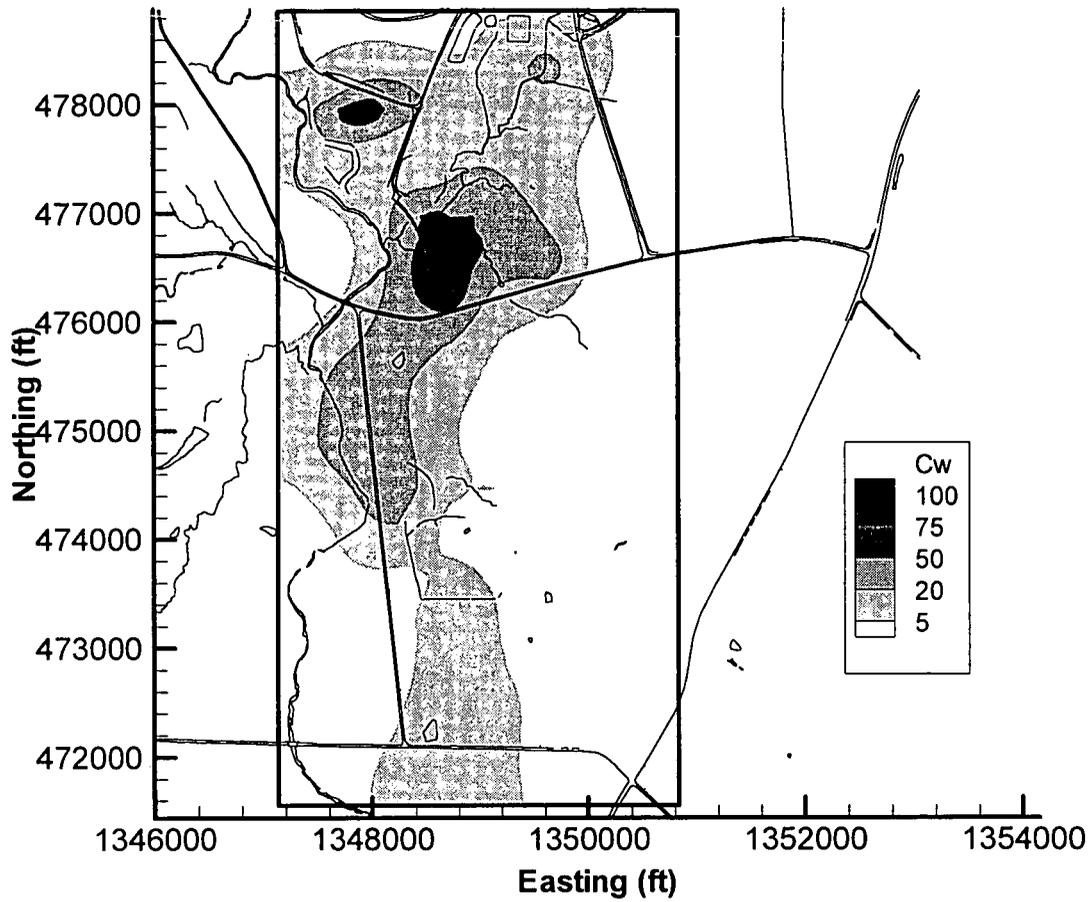


Figure 6.48 K_d Estimation Run 1 Concentrations for Layer 11 at 1/1/98.

000156

K_d Estimation #1 Concentrations for Layer 11 at 1/1/02

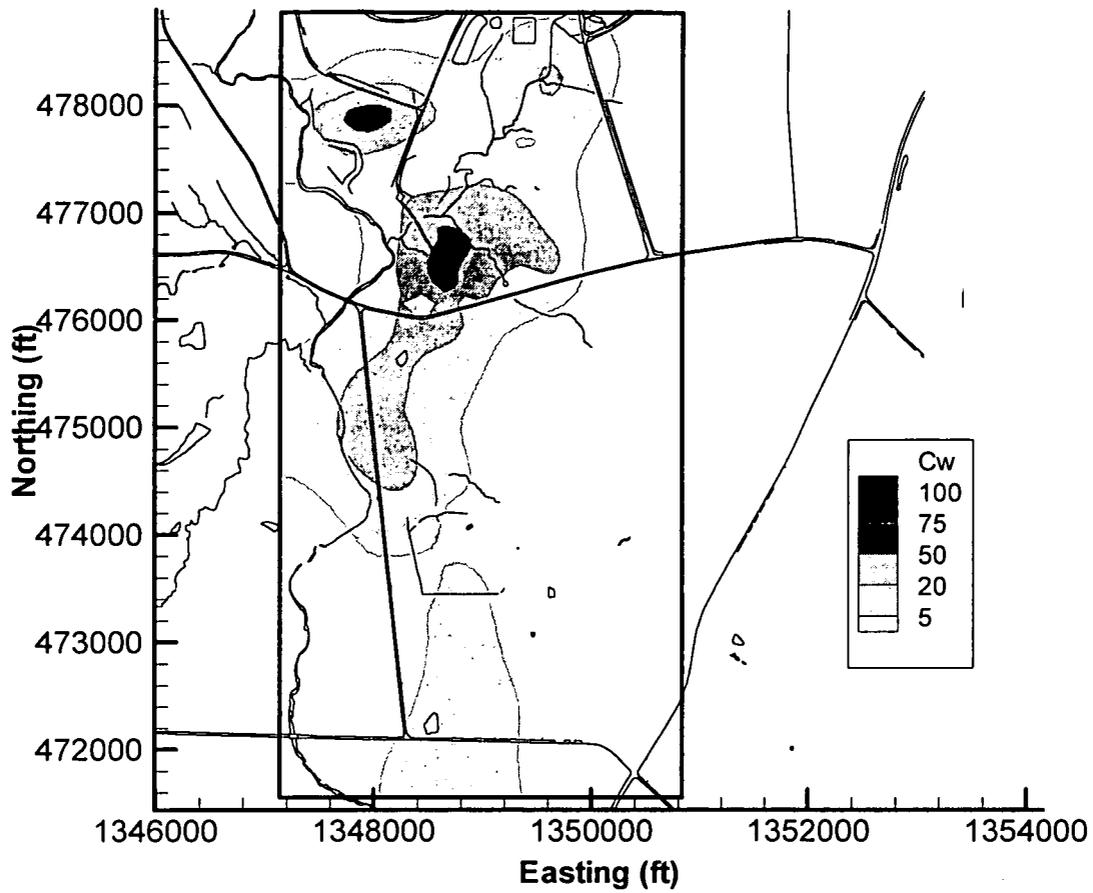


Figure 6.49 K_d Estimation Run 1 Concentrations for Layer 11 at 1/1/02.

000157

K_d Estimation #1 Concentrations for Layer 11 at 5/1/06

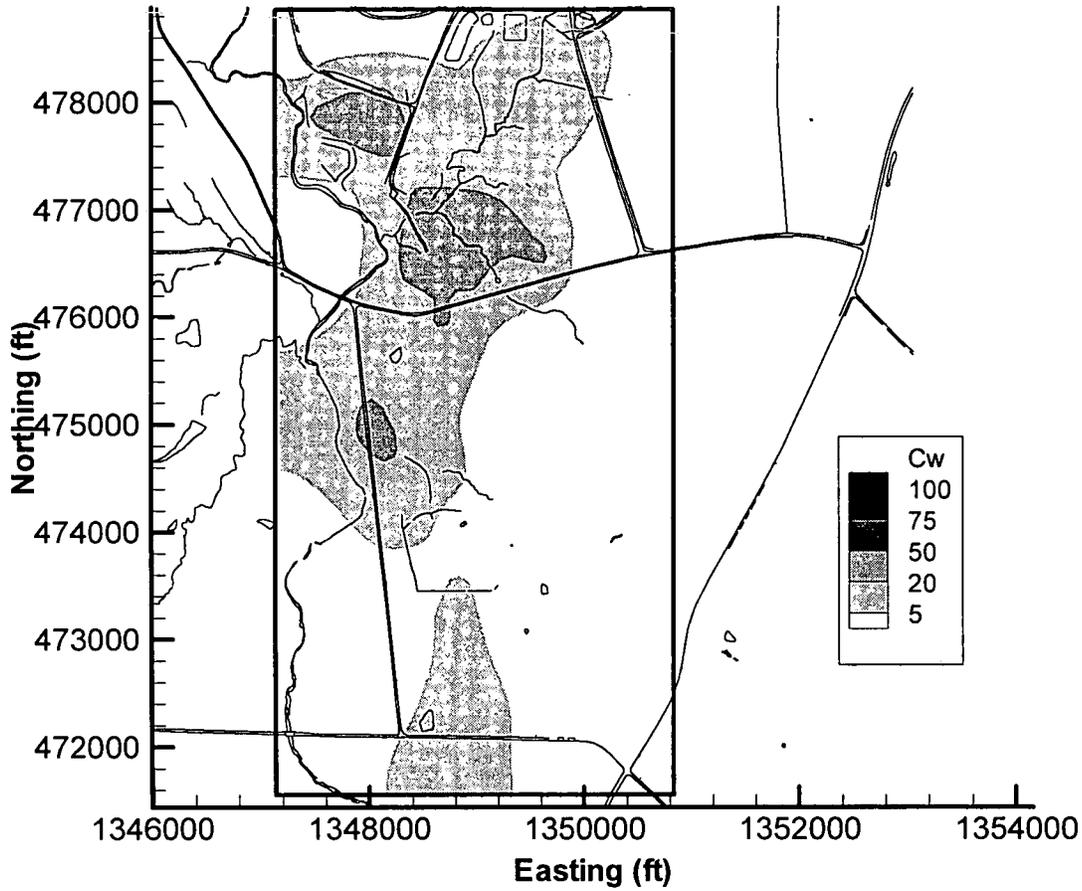


Figure 6.50 K_d Estimation Run 1 Concentrations for Layer 11 at 5/1/06.

000158

K_d Estimation #1 Concentrations for Layer 13 at 1/1/98

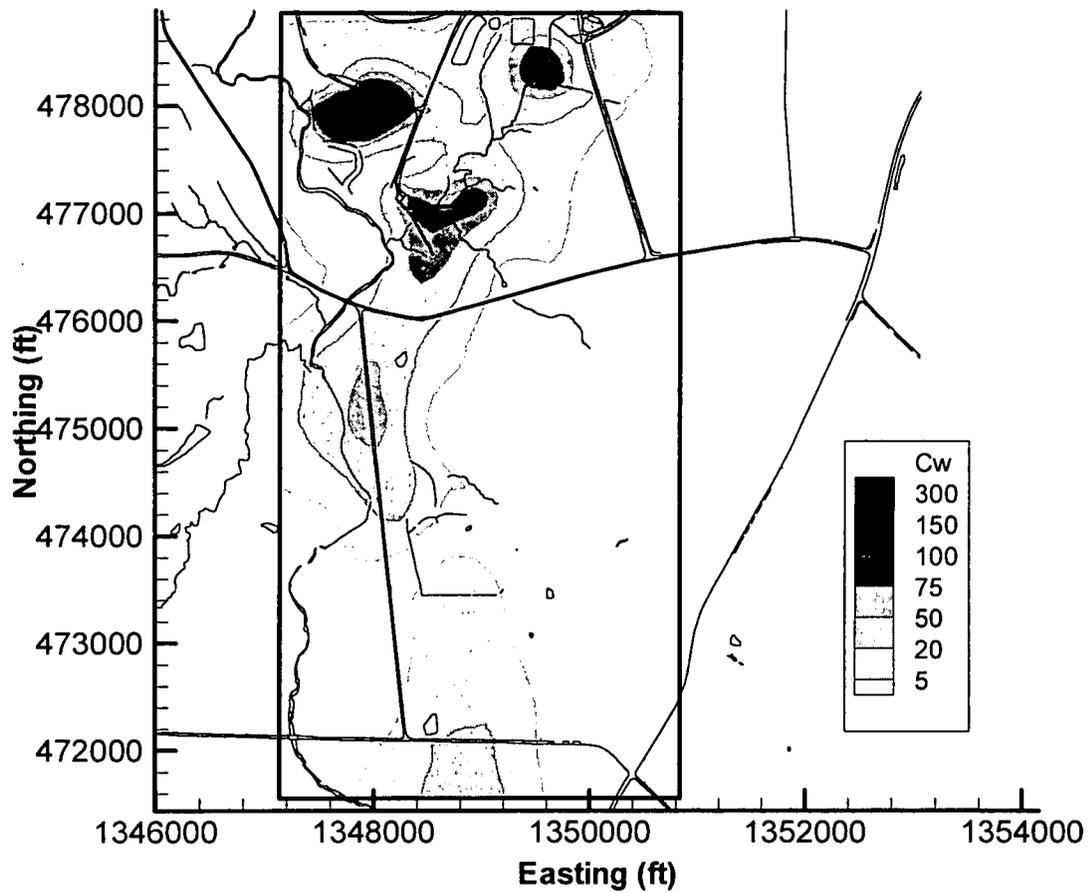


Figure 6.51 K_d Estimation Run 1 Concentrations for Layer 13 at 1/1/98.

K_d Estimation #1 Concentrations for Layer 13 at 1/1/02

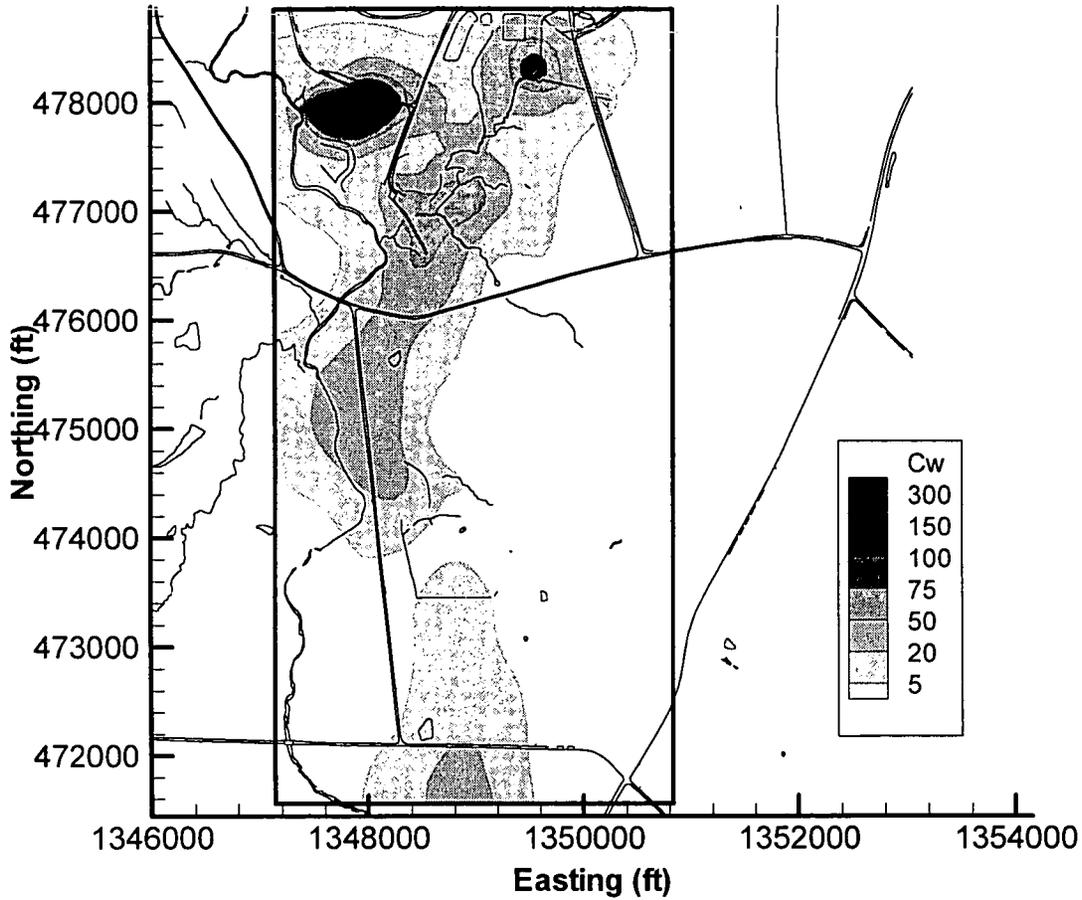


Figure 6.52 K_d Estimation Run 1 Concentrations for Layer 13 at 1/1/02.

000160

K_d Estimation #1 Concentrations for Layer 13 at 5/1/06

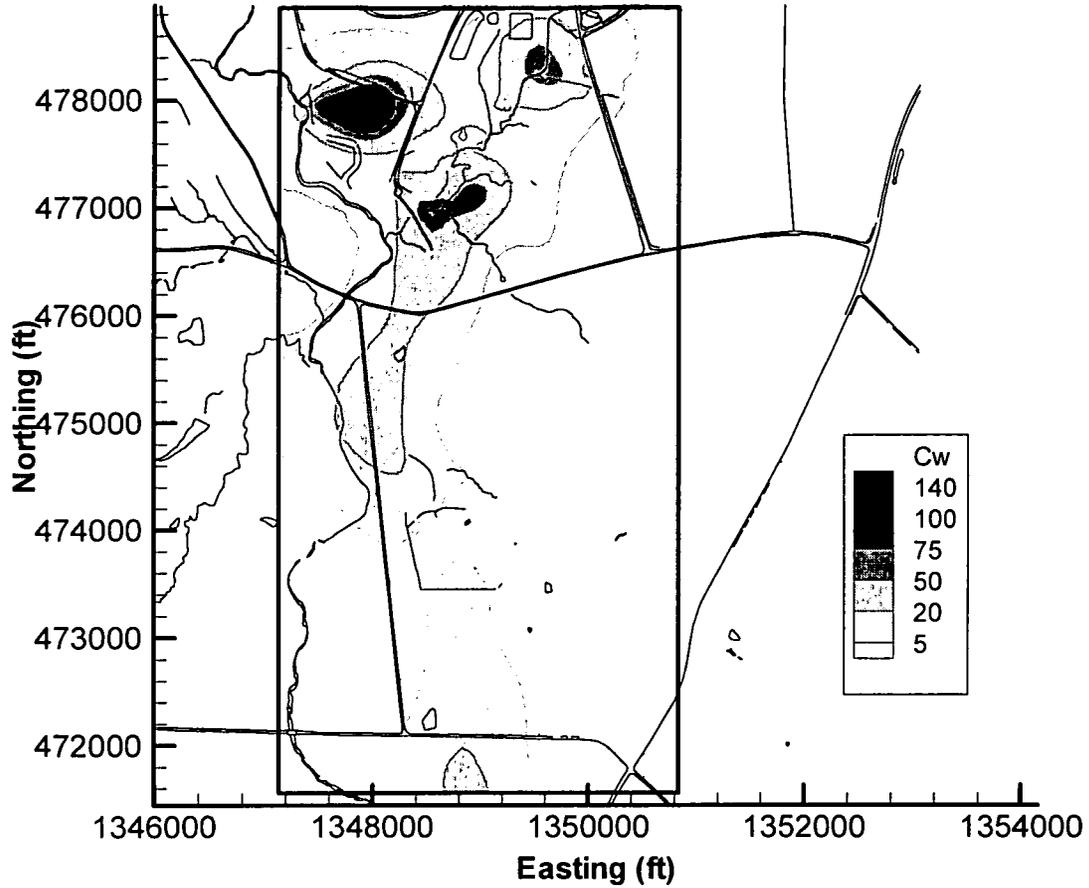


Figure 6.53 K_d Estimation Run 1 Concentrations for Layer 13 at 5/1/06.

000161

K_d Estimation #2 Concentrations for Layer 11 at 1/1/98

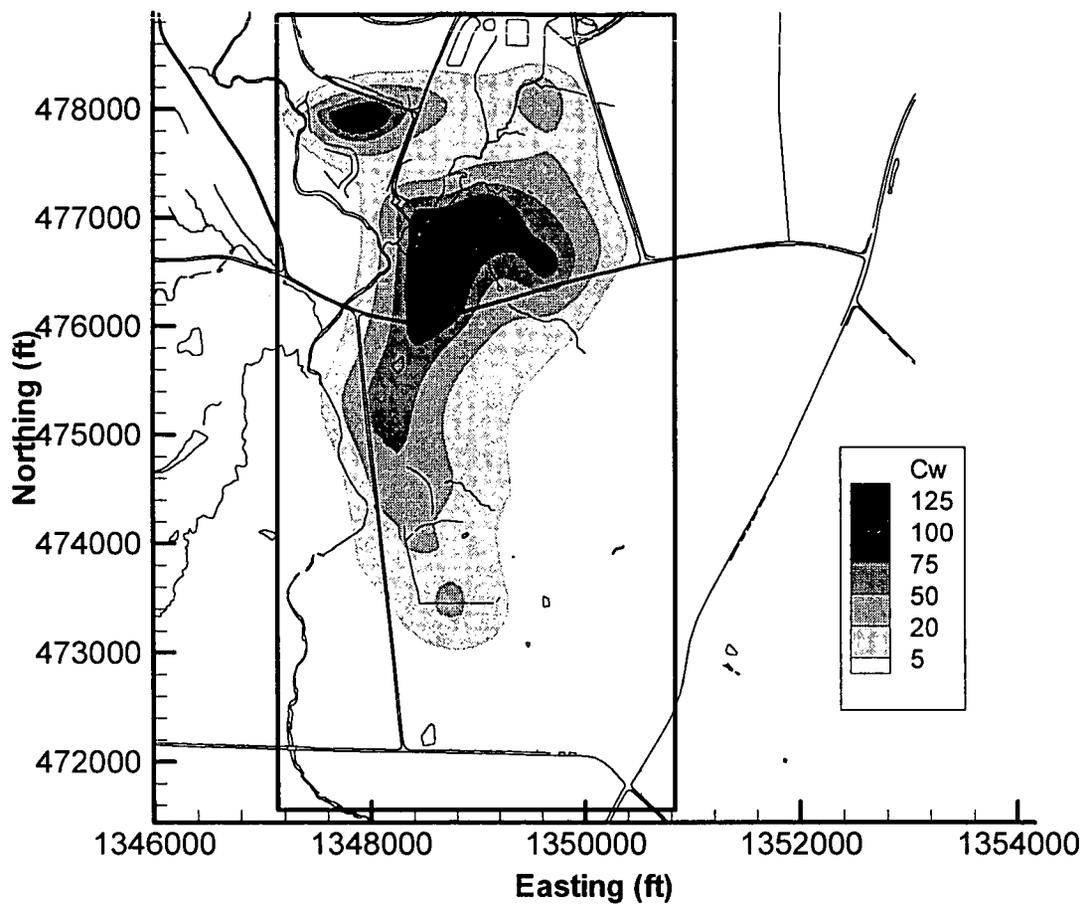


Figure 6.54 K_d Estimation Run 2 Concentrations for Layer 11 at 1/1/98.

K_d Estimation #2 Concentrations for Layer 11 at 1/1/02

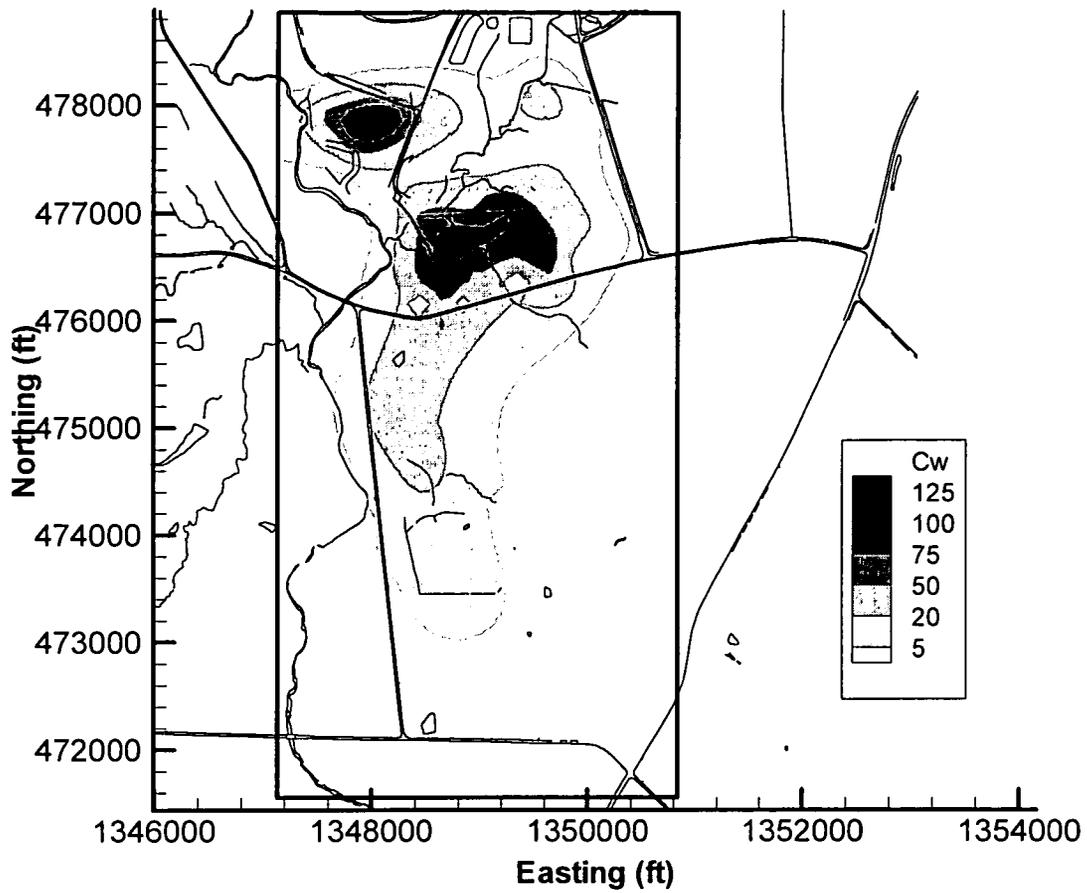


Figure 6.55 K_d Estimation Run 2 Concentrations for Layer 11 at 1/1/02.

000163

K_d Estimation #2 Concentrations for Layer 11 at 5/1/06

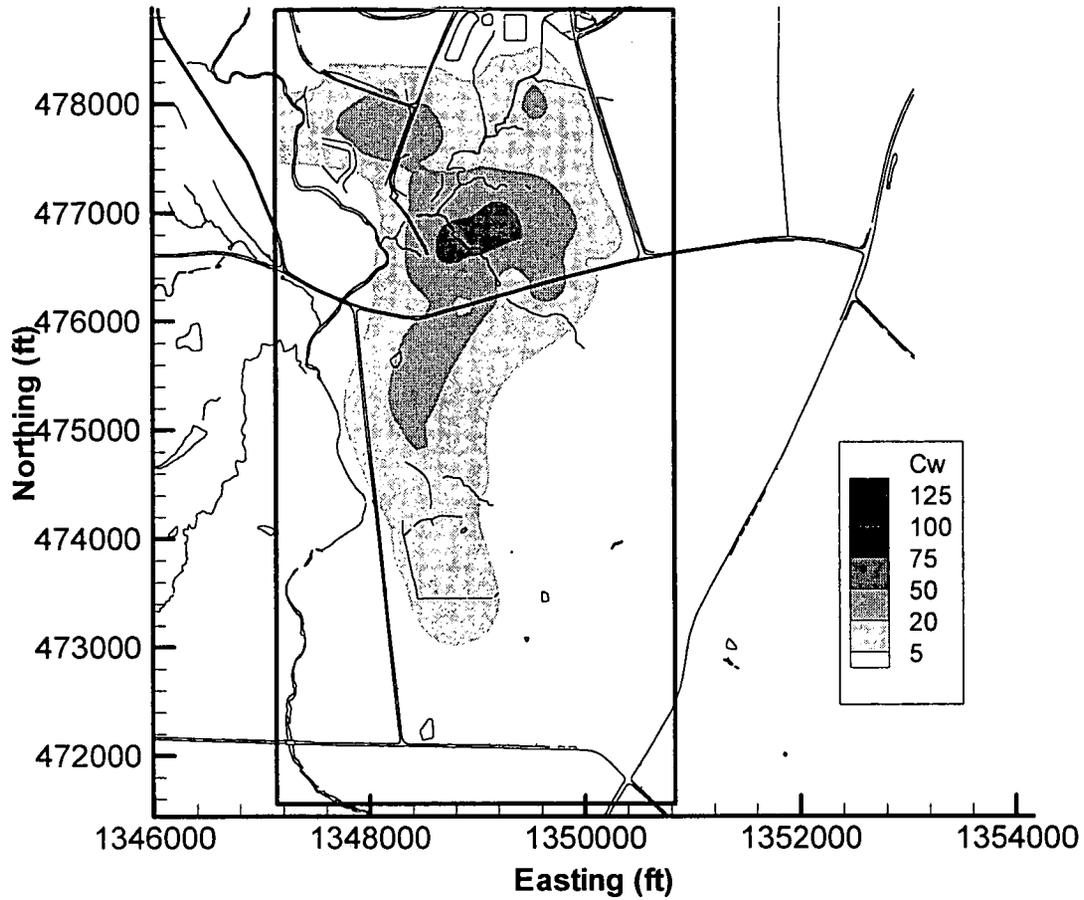


Figure 6.56 K_d Estimation Run 2 Concentrations for Layer 11 at 5/1/06.

K_d Estimation #2 Concentrations for Layer 13 at 1/1/98

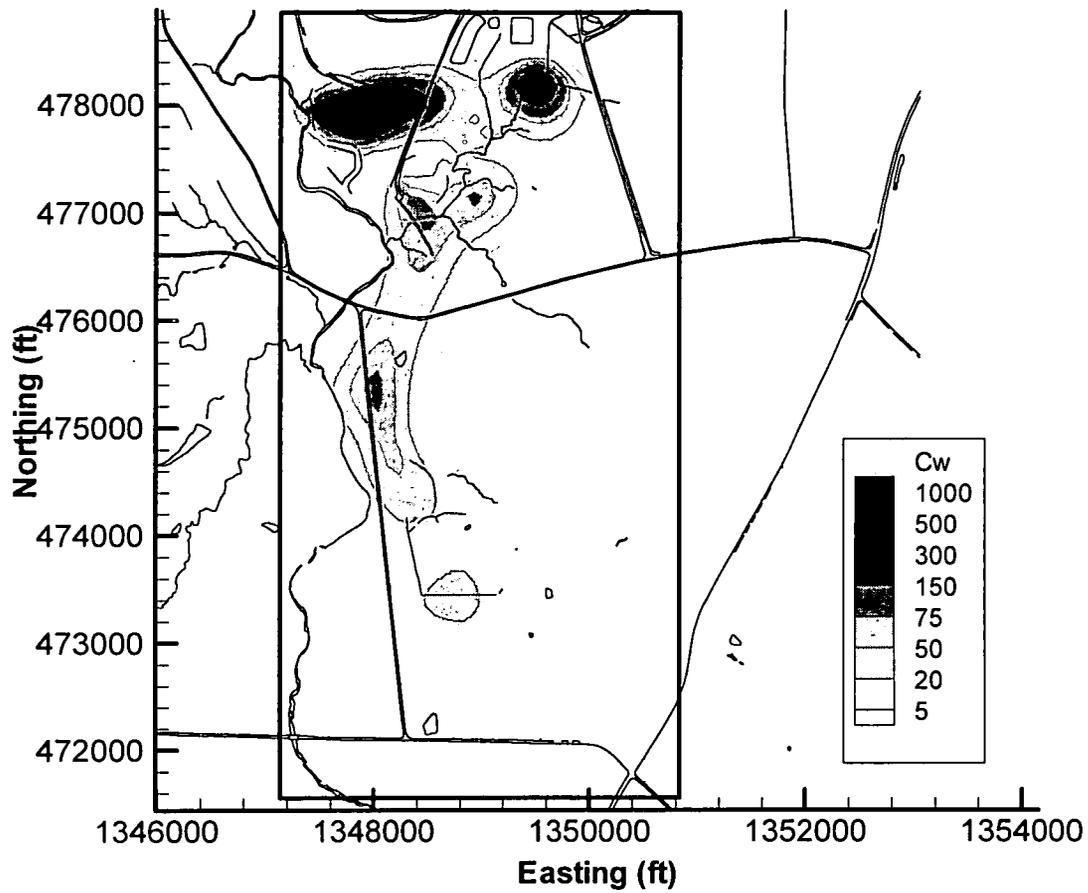


Figure 6.57 K_d Estimation Run 2 Concentrations for Layer 13 at 1/1/98.

K_d Estimation #2 Concentrations for Layer 13 at 1/1/02

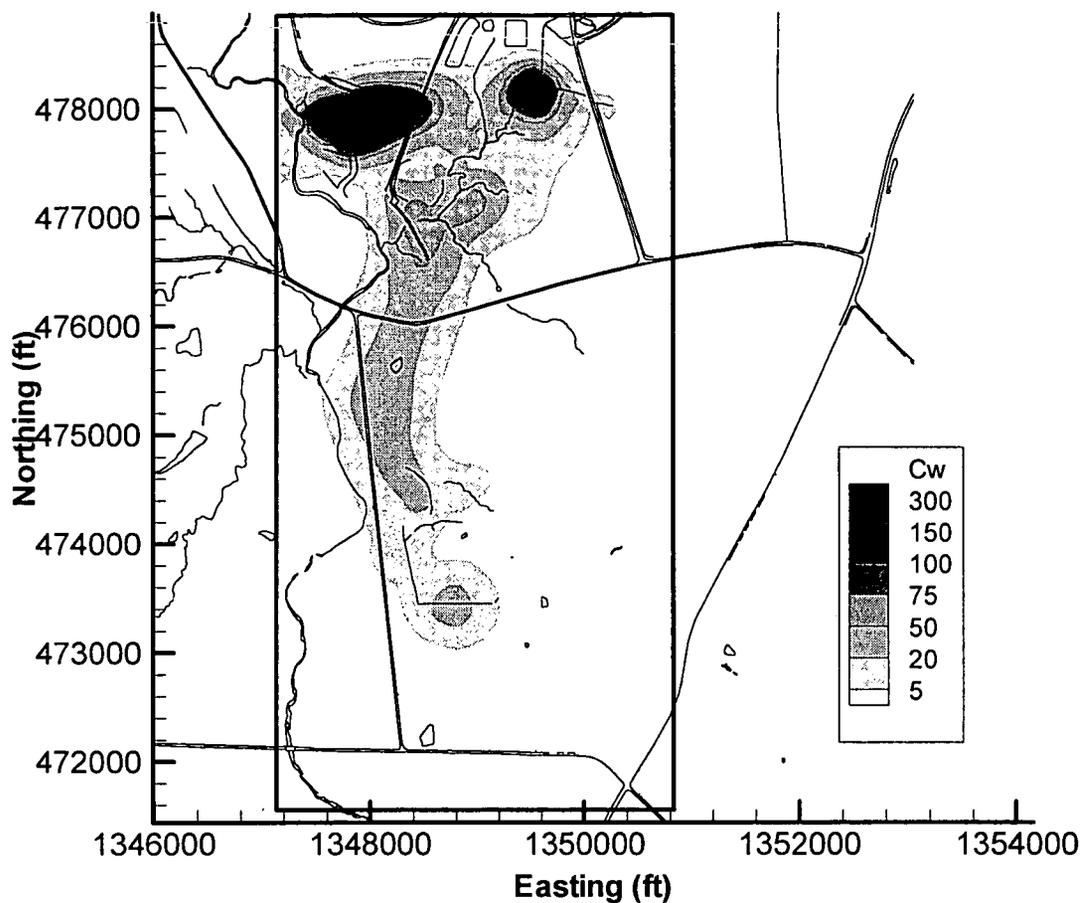


Figure 6.58 K_d Estimation Run 2 Concentrations for Layer 13 at 1/1/02.

K_d Estimation #2 Concentrations for Layer 13 at 5/1/06

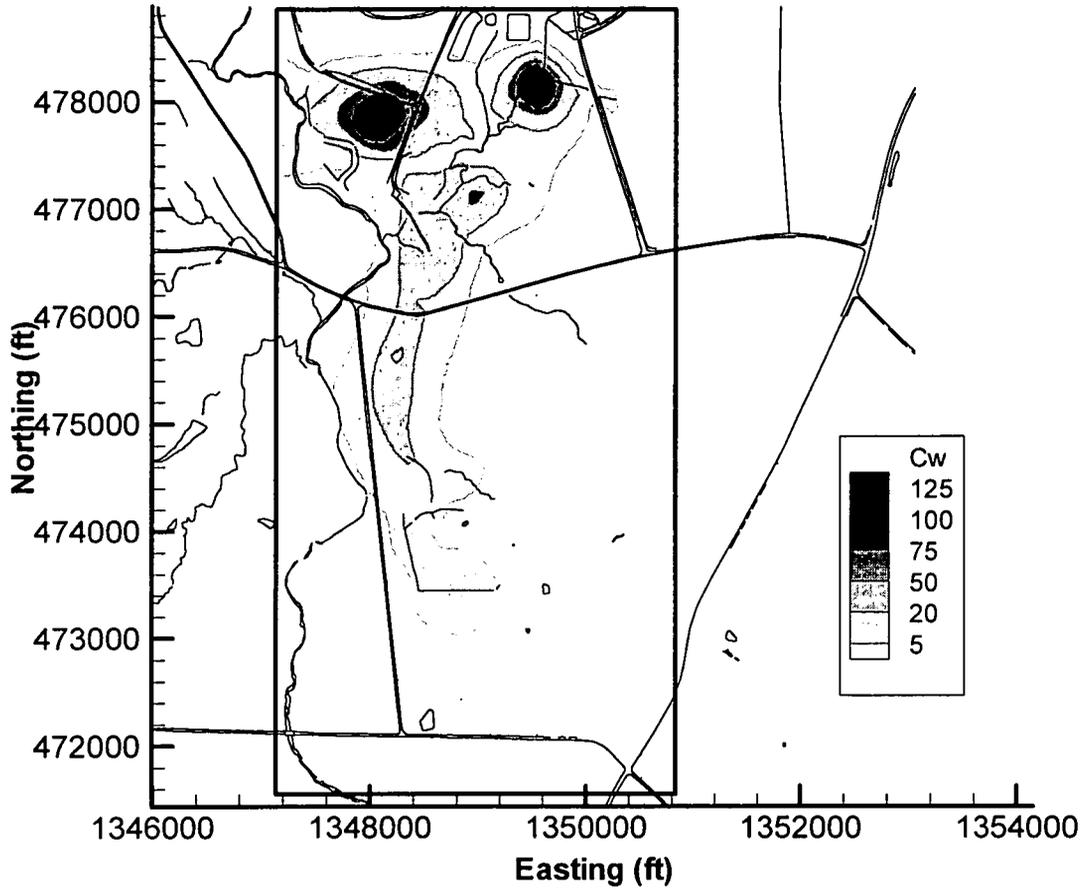


Figure 6.59 K_d Estimation Run 2 Concentrations for Layer 13 at 5/1/06.

000167

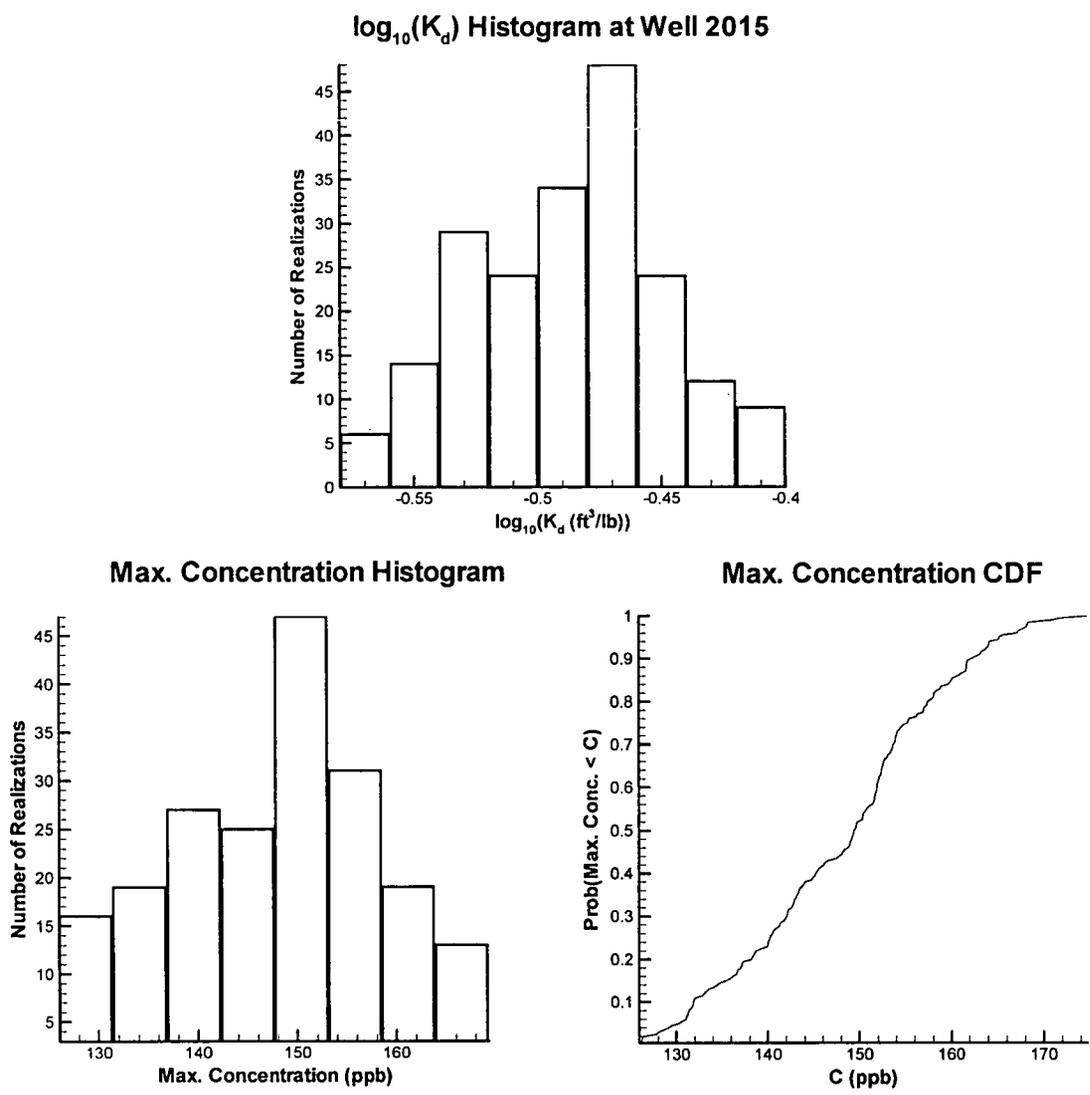


Figure 6.60 Monte Carlo Results for K_d Estimation Run 1 (200 Realization).

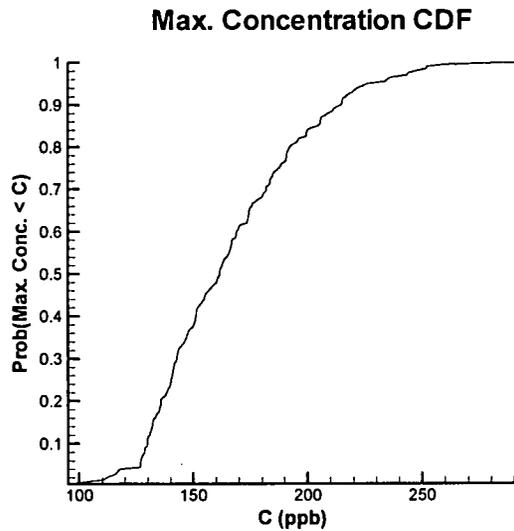
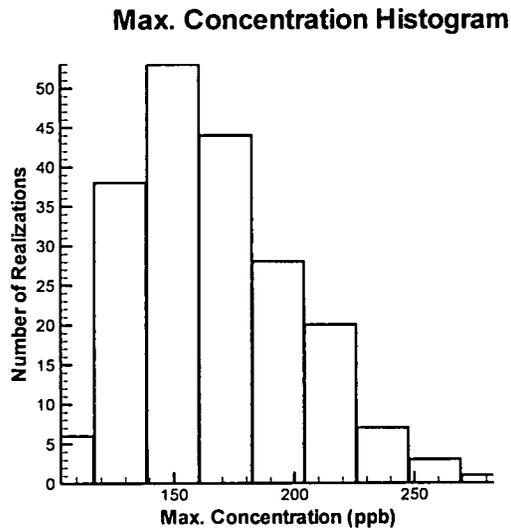
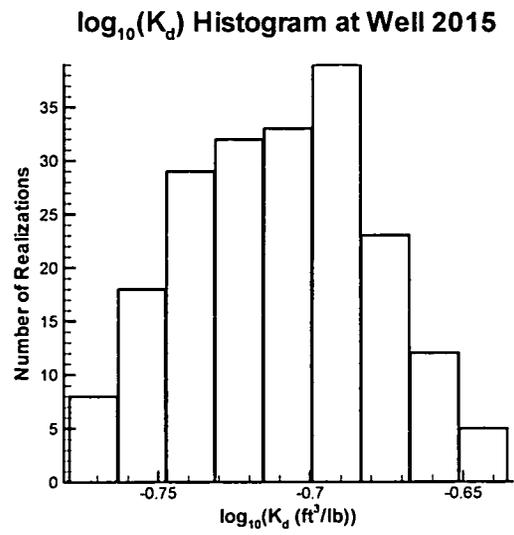
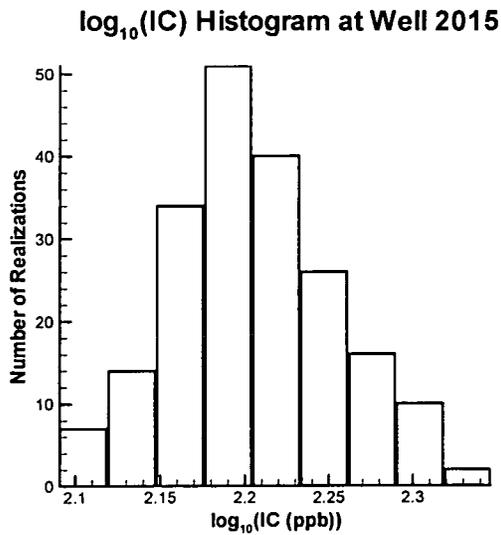


Figure 6.61 Monte Carlo Results for K_d Estimation Run 2 (200 Realizations).

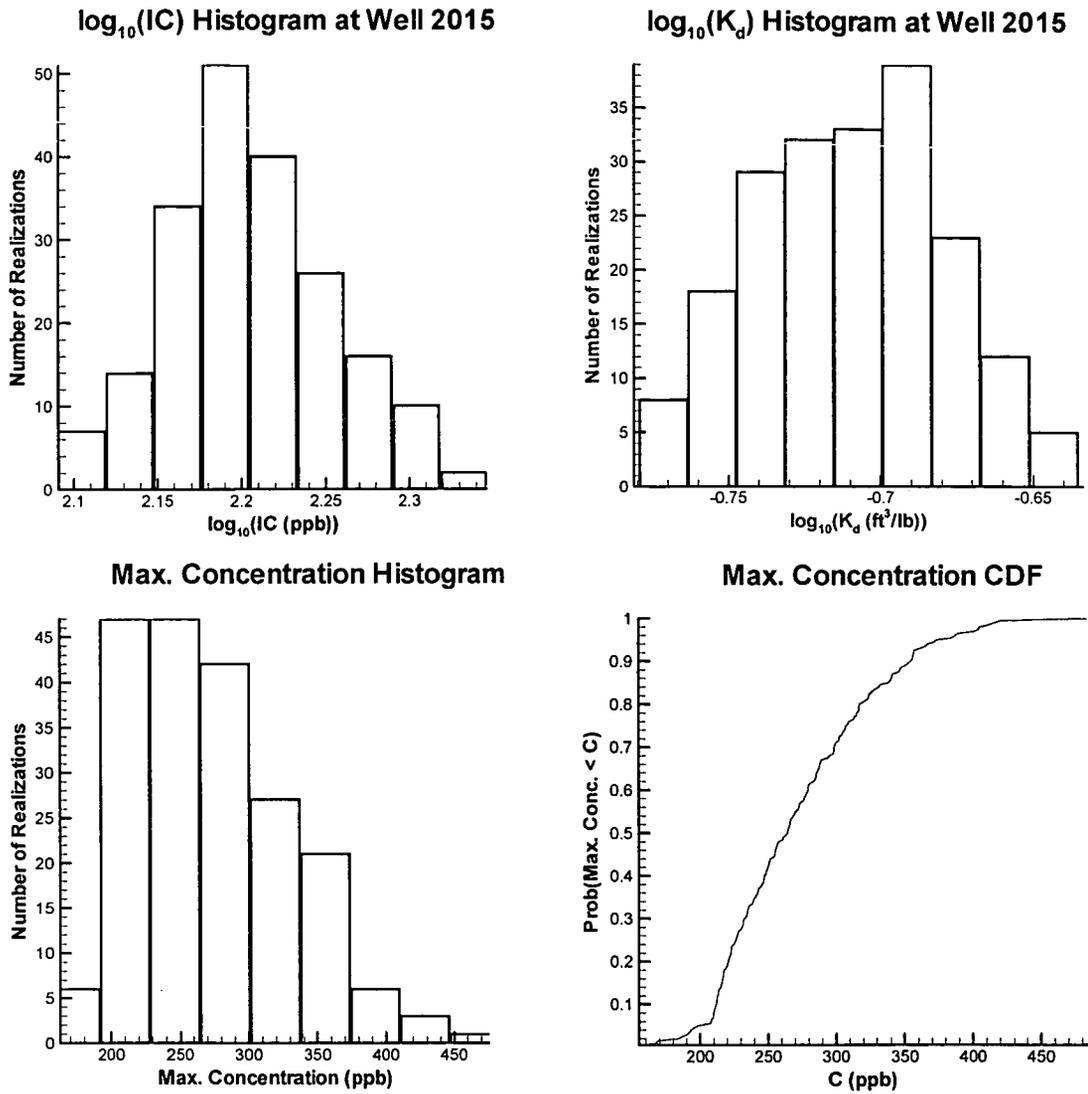


Figure 6.62 Monte Carlo Results for K_d Estimation Run 3(200 Realizations).

Variance of Concentration at 7/1/06 -- Model Layer 11

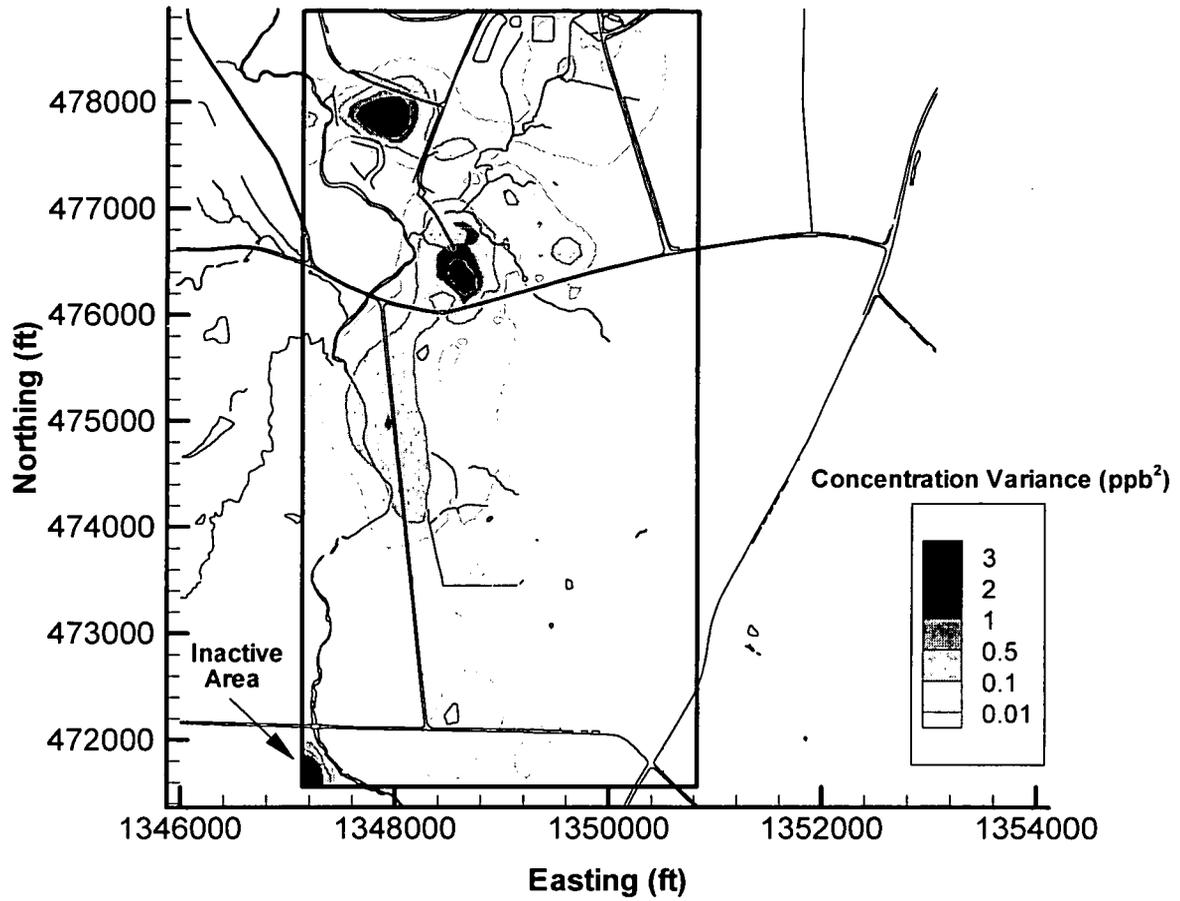


Figure 6.63 K_d Estimation Run 1, Variance of Concentration at 7/1/06 -- Model Layer 11.

000171

Variance of Concentration at 7/1/06 -- Model Layer 13

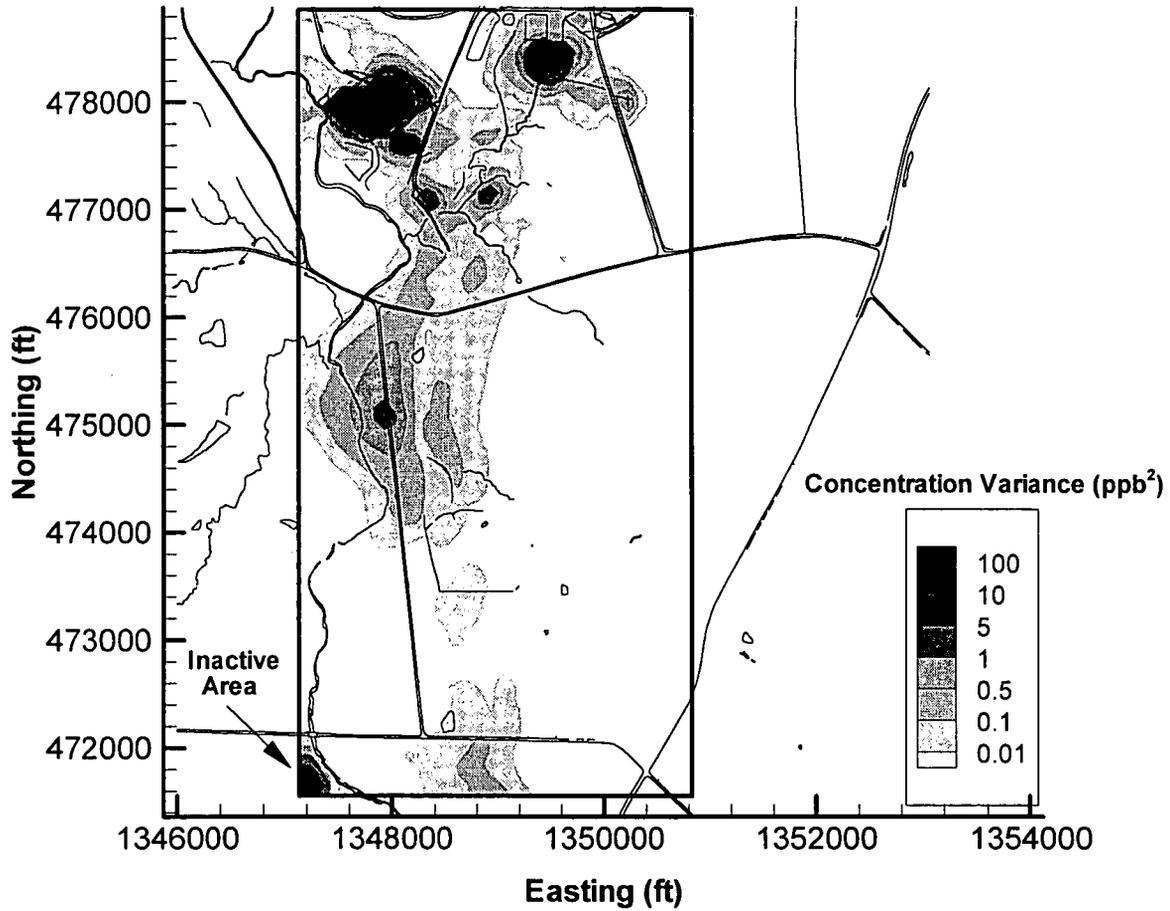


Figure 6.64 K_d Estimation Run 1, Variance of Concentration at 7/1/06 -- Model Layer 13.

Variance of Concentration at 7/1/06 -- Model Layer 11

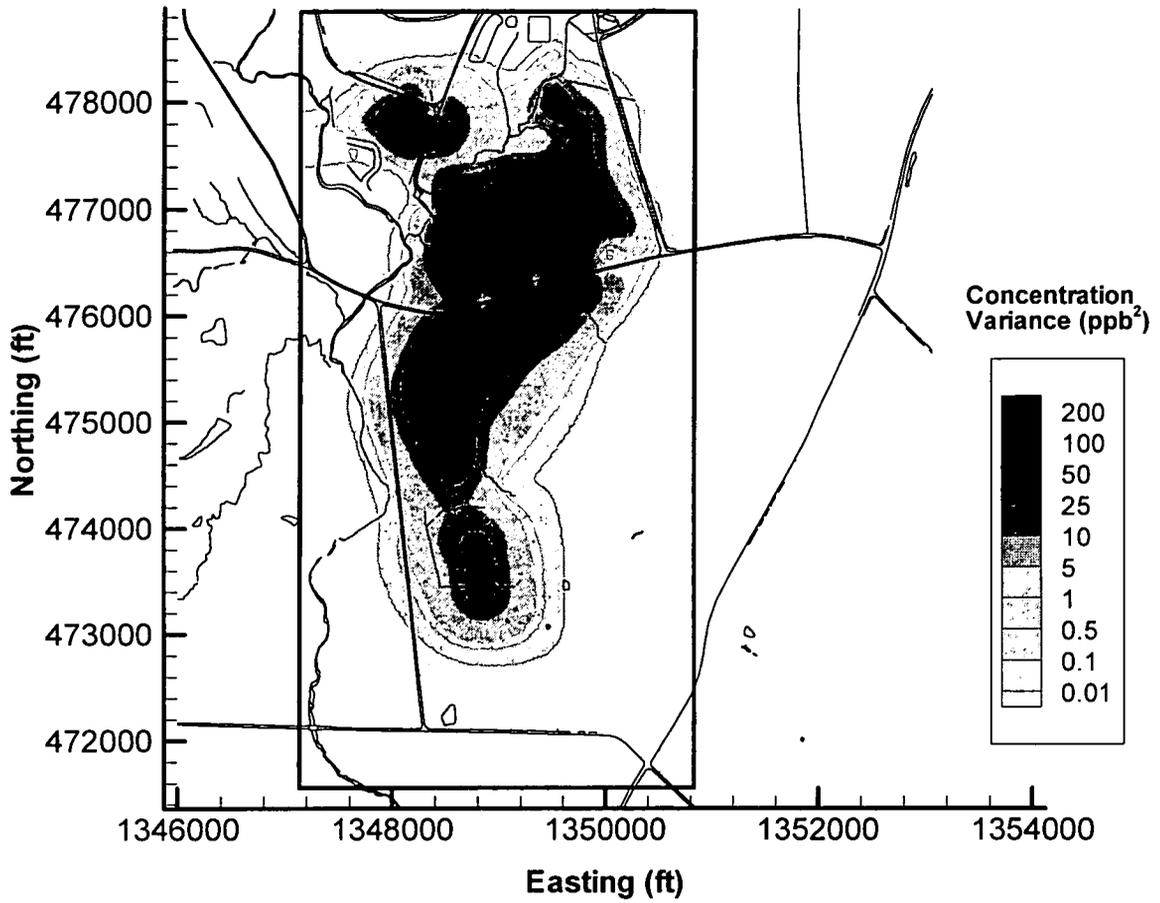


Figure 6.65 K_d Estimation Run 2, Variance of Concentration at 7/1/06 -- Model Layer 11.

Variance of Concentration at 7/1/06 -- Model Layer 13

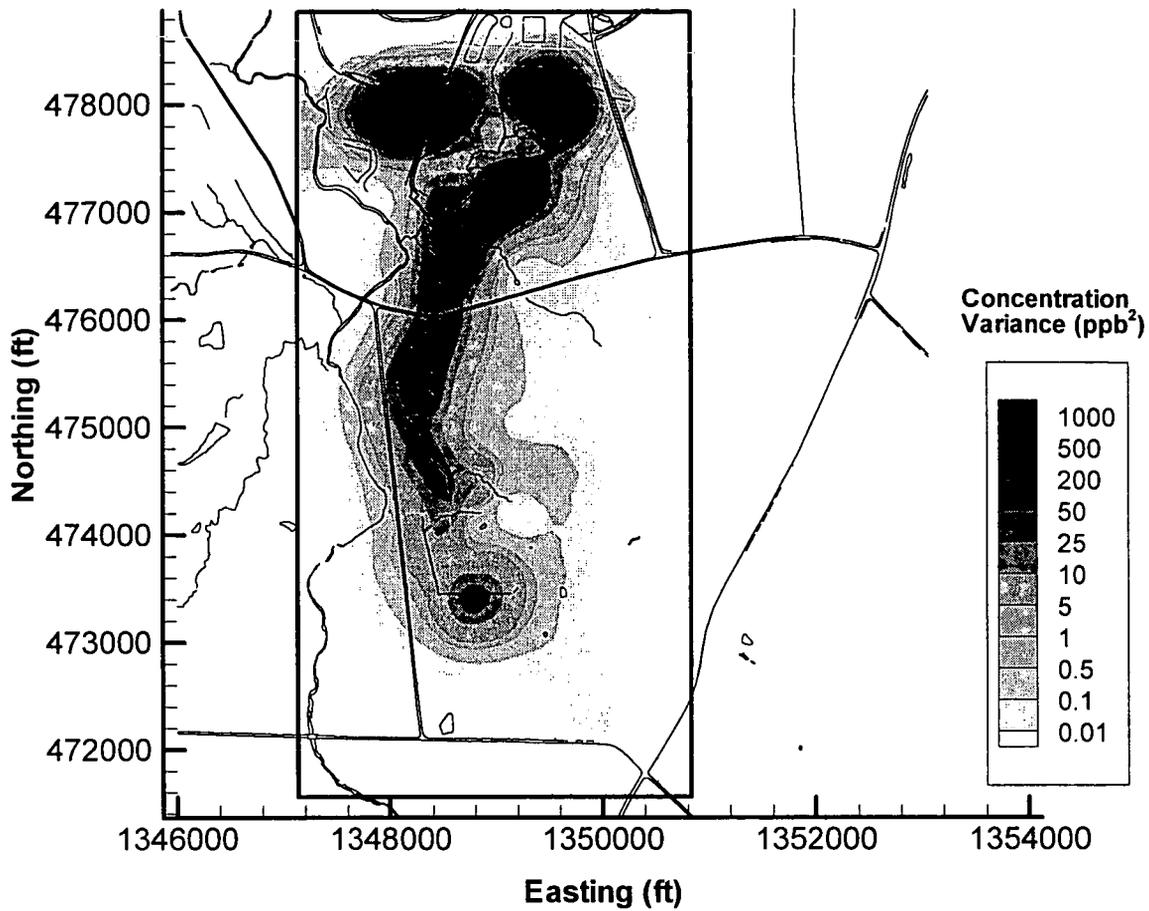


Figure 6.66 K_d Estimation Run 2, Variance of Concentration at 7/1/06 -- Model Layer 11.

000174

7.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

7.1 SUMMARY AND CONCLUSIONS

This document presents the software capabilities, input/output specification and test results for the DFM/VAM3DF software. The software provides the basis for model predictions and quantified uncertainties that will be used in the tools developed in Phase III to optimize monitoring and remediation.

The DFM/VAM3DF software combines information from solute concentration time histories, the VAM3DF transport model, geostatistical variability, and prior knowledge of parameter values/distributions. The software computes parameter estimates and model predictions and quantifies uncertainty in terms of estimate and prediction error covariances.

Testing and performance benchmarking has been performed to demonstrate the software functionality and to show that it can be efficiently applied to future groundwater remediation effort as part of the FEMP.

The testing was divided into two phases: small-scale testing; site-specific testing. Small-scale tests were first performed to verify DFM/VAM3DF and demonstrate the software functionality. The small-scale tests demonstrated that the DFM/VAM3DF software can provide estimates for the following parameters:

- Kinetic mass transfer parameters;
- Effective porosity;
- Initial uranium solute concentration; and
- Dispersivities.

A site-specific test of the DFM/VAM3DF software was conducted to demonstrate the software applicability to the FEMP, and to establish performance benchmarks. This test was based on the VAM3DF/GMA model developed in Phase I. The DFM/VAM3DF software was used to estimate 1993 average uranium concentration and the spatially varying distribution coefficient in the South Plume and the South Field areas. DFM/VAM3DF used the baseline flow and transport simulation from Section 6.3.1 to fit selected 1994-1999 monitoring well data, extraction well data, and geoprobe data. Estimating $\log_{10}(K_d)$ provides the best match to 479 concentration measurements. The site-specific parameters identification work was carried out for demonstration and benchmarking purposes only. The results reported herein should be considered preliminary.

7.2 RECOMMENDATIONS

Based on the results of the DFM/VAM3DF tests, the following future enhancements are recommended:

- Additional investigation is needed to improve the model. Specifically, the statistical parameters must be adjusted and the areas with large residuals should be investigated.

000175

- All significant stresses and climatological periods should be simulated to further improve the model. The DFM/VAM3DF software has the capability to use a sequence of steady state flow models.
- Laboratory and, if possible, field investigations should be conducted to determine kinetic mass transfer parameters, especially the chemisorption rate coefficient, for which there are no direct measurements available.

8.0 REFERENCES

- Bennett, A.F., Inverse Methods in Physical Oceanography, Cambridge University Press Cambridge, 1992.
- Bertsekas, D.P., Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York, 1982.
- Bierman, G.J., Factorization Methods for Discrete Sequential Estimation, Academic Press, New York, 1977.
- Carrera, J., A. Medina, C. Axness, and T. Zimmerman, "Formulations and Computational Issues of the Inversion of Random Fields," from G. Dagan and S.P. Neuman, Subsurface Flow and Transport: A Stochastic Approach, International Hydrology Series, Cambridge University Press, Cambridge, 1977.
- Dempster, A.P., N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Ser. B.* 39 (1), 1-22, 1977.
- Deutsch, C.V., and A.G. Journel, GSLIB Geostatistical Software Library and User's Guide, Oxford University Press, New York, 1992.
- Duncan, D.B. and S.D. Horn, "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis," *Journal of the American Statistical Association*, 67 (340), 815-821, 1972.
- Gill, P.E., W. Murray and M.H. Wright, Practical Optimization, Academic Press, New York, 1981.
- HydroGeoLogic, Inc., 1998, Development and Verification of VAM3DF, A Numerical Flow and Transport Modeling Code, Herndon, VA.
- Jazwinski, A.H., Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.
- Kitanidis, P.K., "Comments on "A Reassessment of the Groundwater Inverse Problem" by D. McLaughlin and L.R. Townley," *Water Resources Research*, Vol. 33, No.9, pp. 2199-2202, September 1997.
- Lawson, C.L. and R.J. Hanson, Solving Least Squares Problems, Prentice Hall, Englewood Cliffs, NJ, 1974.
- Levy, L.J. and D.W. Porter, "Large-Scale System Performance Prediction with Confidence from Limited Field Testing Using Parameter Identification," *the Johns Hopkins APL Technical Digest*, 13 (2), 300-308, 1992.

000177

- McLaughlin, D. and L.R. Townley, "A Reassessment of the Groundwater Inverse Problem," *Water Resources Research*, Vol.32, No. 5, pp.1131-1161, May 1996.
- Paige, C.C. and M.A. Saunders, *LSQR. An algorithm for sparse linear equations and sparse least squares*, *ACM Trans. Math. Software*, 8 (1982), pp. 43-71.
- Porter, D. W., "Multi-Object Tracking Via Recursive Generalized Likelihood Approach," *Proceedings 18th IEEE Conference on Decision and Control*, December 1979.
- Porter, D.W., B.P. Gibbs, W.F. Jones, P.S. Huyakorn, L.L. Hamm, and G.P. Flach, "Data Fusion Modeling for Groundwater Systems," *Journal of Contaminant Hydrology*, Vol. 42, pp 303-335, March 2000.
- Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1992.
- Rao, C. R., *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.
- Shumway, R.H., *Applied Statistical Time Series Analysis*, Prentice Hall, Englewood Cliffs, 1998.
- Snodgrass, M.F. and P.K. Kitanidis, "A Geostatistical Approach to Contaminant Source Identification," *Water Resources Research*, Vol.33, No. 4, pp. 537-546, April 1997.
- Sun, N.Z., and W.W.G. Yeh, "A Stochastic Inverse Solution for Transient Groundwater Flow: Parameter Identification and Reliability Analysis," *Water Resources Research*, 28 (12), 3269-3280, 1992.
- U.S. Department of Energy, 1997, Baseline remedial strategy report: Remedial design for aquifer restoration (Task 1), Fernald Environmental Management Project, DOE, Fernald Area Office, Cincinnati, OH.
- U.S. Department of Energy, 1994, SWIFT Great Miami Aquifer model - summary of improvements, report, Fernald Environmental Management Project, DOE, Fernald Area Office, Cincinnati, OH.
- Vandergraft, J. S., "Efficient Optimization Methods for Maximum Likelihood Parameter Estimation," *Proceedings of 24th Conference on Decision and Control*, Ft. Lauderdale, 1985.
- Whittle, P., "On Stationary Processes in the Plane," *Biometrika*, Vol. 41, 434-449, 1954.
- Yucel, Z.T. and R.H. Shumway, "A Spectral Approach to Estimation and Smoothing of Continuous Spatial Processes," *Stochastic Hydrology and Hydraulics*, Vol.10, pp. 107-126, 1996.

APPENDIX A
EXAMPLES OF INPUT FILES

APPENDIX A

All keywords are upper case and left justified.

Remaining input is free format unless specified otherwise.

All variables starting with 'i', 'j', 'k', 'l', 'm' or 'n' integers and all remaining variables are real numbers.

Input Format: Comments (not part of input):

XGRID #Keyword

nxg #number of X subdivisions groups

nx(i) delx(i) #number of X subdivisions in ith group, width of each X subdivision
#the above line is specified for I=1,nxg

YGRID #Keyword

nyg #number of Y subdivisions groups

ny(i) dely(i) #number of Y subdivisions in ith group, width of each Y subdivision
#the above line is specified for I=1,nyg

ZGRID #Keyword (optional)

nzg #number of Z subdivisions groups

nz(i) delz(i) #number of Z subdivisions in ith group, width of each Z subdivision
#the above line is specified for I=1,nzg

ZGRID3D #Keyword (use only if ZGRID is not used)

nnz #number of nodal planes in 3D mesh
(((zgrid(i,j,k),i=1,nnx),j=1,nnny),k=1,nnnz)
#where nnx is the number of x-direction nodes and nny is the
#number of y-direction nodes

SITE2MODEL #Keyword Transformation from VAM3DF grid to parameter grid

xoff yoff zoff angle scale

NRFNTRD #Keyword

nrf ntrd #number of parameters estimated by a random field with a
#polynomial trend and number of parameters estimated by a
#polynomial trend

PARAMS #Keyword list of parameters to be estimated

paramlabel(i) #format('(a8)'), parameter label for I=1,(nrf+ntrd)
#with nrf parameters first
#(POROSITY,KD,IC,ALPHA_D,ALPHA_LH, ALPHA_TH, ALPHA_LV,
ALPHA_TV)

SCALECONC#Keyword

scalec #scale FIELDDATA and CONCDATA conc. data to match
#VAM3DF units (i.e. ug/L to nanolb/ft^3)

FIELDDATA #Keyword direct measurements

nfdg #number of field data subdivisions groups

nfd(1) nfd(2) .. nfd(nfdg)#number of field data in each of the nfdg subdivisions

nparam x y z fd sigma #nparam is the parameter number in the set {1,2,...,(nrf+ntrd)}
#x,y,z are the VAM3DF grid coordinates of the measurement
#fd is the direct measurement of parameter number nparam
#sigma is the prior standard deviation of the error in the
#measurement
#the above input line is specified for each subdivision of each group

NSCREENPTS #Keyword (optional but before CONCDATA)

000180

nwspts #number of well screen integration points

CONCDATA #Keyword (optional) Concentration measurements

ncdg #number of conc. data subdivisions groups

ncd(1) ncd(2) .. ncd(ncdg)#number of conc. data in each of the ncdg subdivisions

x,y,ztop,zbot,t,cd,sigma #x,y are the VAM3DF grid coordinates of the measurement
 #ztop and zbot are the top and bottom of the screen in
 #VAM3DF grid coordinates
 #t is the time of the measurement
 #cd is the log10 of the conc. measurement
 #sigma is the prior standard deviation of the error in the
 #measurement
 #the above input line is specified for each subdivision of each group

HYDROUNITSi #Keyword for i=1, (nrf+ntrd)

nunits(i) # number of vertically isolated units for parameter #i

lay1(i,j) lay2(i,j) #lower layer number and upper layer number for the jth unit
 #for parameter #i
 #the above input line is specified for j=1, nunits(i)
 #(put HYDROUNITS with largest number of units first)

VARIOGRAMi #Keyword for i=1, (nrf+ntrd) corresponding with HYDROUNITS

s(i,j) tx(i,j) ty(i,j) tz(i,j)
 #s is the spatial variability standard deviation for the jth
 #unit for parameter #i
 #tx is the x-direction correlation distance for the jth
 #unit for parameter #i
 #ty is the x-direction correlation distance for the jth
 #unit for parameter #i
 #tz is the x-direction correlation distance for the jth
 #unit for parameter #i

TRENDi #Keyword for i=1, (nrf+ntrd) corresponding with HYDROUNITS

nt(i) #number of trend polynomial coefficients for parameter #i

pv(i,j,k) ps(i,j,k) #prior est. and prior error standard deviation for the kth
 #coefficient for the jth unit for parameter #i
 #the above input line is specified for k=1, nt(i)

EDITTHRES #Keyword

thres #data editing threshold, large for no editing of data

SOLVER #Keyword

stext #equal to 'DIRECT' for direct method
 #equal to 'LSQR' for lsqr solver
 #equal to 'REPM' for representer method

NITER #Keyword(optional, if NITER is not given, then niter=8)

niter #max. Gauss-Newton iterations allowed

CONVTEST #Keyword(optional, if CONVTEST is not given, then conv=0.2)

conv #Gauss-Newton iteration convergence criterion

COVARIANCE #Keyword (optional) for covariance output

nct #number of covariance terms

ic(i) jc(i) #(I,J) for ist covariance term
 #the above input line is specified for i=1, nct

MONTECARLO #Keyword (optional) for Monte Carlo output (must use REPM)
 nr iseed #number of Monte Carlo realizations, negative integer seed
HISTOGRAM#Keyword (optional) for Monte Carlo histogram plot (must use MONTECARLO)
 nhist nbar #number of histograms, number of bars in each histogram
 title(i) #format(a80), title on histogram plot
 nparam(i) x(i) y(i) z(i)#nparam is the parameter number in the set {1,2,...,(nrf+ntrd)}
 #x,y,z are the VAM3DF grid coordinates of the histogram location
 #the above two input lines are specified for each histogram
MONTECRUN #Keyword (optional) run vam3df for each realizations
 timeout #output statistics at given at time timeout
 filename(i) #format(a80), binary file name containing realizations
 #the above input line is specified for parameter in
 #the set {1,2,...,(nrf)}
RESTART #Keyword (optional) for restarting with previous results
 rstname #format('a30') file name of binary restart file
 #containing current states and J matrix
EKFUPDATE #Keyword (optional, can not use DIRECT)Extended Kalman Filter
 #update. Use RESTART, if not CONCDATA terms of LHS will
 #be updated before the 1st iteration (first iteration like batch
 #run and subsequent iterations are EKF updates).
CONCDATAE #Keyword (optional) Concentration measurements for EKF update
 ncdg #number of conc. data subdivisions groups
 ncd(1) ncd(2) .. ncd(ncdg)#number of conc. data in each of the ncdg subdivisions
 x,y,ztop,zbot,t,cd,sigma #x,y are the VAM3DF grid coordinates of the measurement
 #ztop and zbot are the top and bottom of the screen in
 #VAM3DF grid coordinates
 #t is the time of the measurement
 #cd is the log10 of the conc. measurement
 #sigma is the prior standard deviation of the error in the
 #measurement
 #the above input line is specified for each subdivision of each group
 #Note: Add new field data to FIELDDATA
DEBUG #Keyword (optional) when estimating IC, this option will
 #write the tecplot file debug.tec containing the VAM3DF
 #initial conditions.
CHEM4LINRKD #Keyword (optional) chemsorption for linear Kd,
 Alpha_c #1st order decay becomes chemsorption by setting
 #1st order decay = alpha_c/(1.0 + (poro*Sw/r_bKd)).
 #This option applies when estimation Kd.

SAMPLE INPUT

```

XGRID          #Keyword
1              #number of X subdivisions groups
26 100.0      #number of X subdivisions, width of each X subdivision
YGRID          #Keyword
1              #number of Y subdivisions groups
16 100.0      #number of Y subdivisions, width of each Y subdivision
ZGRID          #Keyword or ZGRID3D
1              #number of Z subdivisions groups
10 10.0       #number of Z subdivisions, width of each Z subdivision
ZGRID3D       #Keyword or ZGRID
2              #number of nodal planes (nnz) in 3D mesh
131.2 133.5   #free format read(((zgrid(I,j,k),I=1,nnx),j=1,nnz),k=1,nnz)
SITE2MODEL    #Keyword Transformation from transport grid to statistical grid
573750.0 836850.0 280.0 0.0 1.0 #xoff,yoff,zoff,angle,scale
NRFNTRD       #Keyword number of parameters with random field and trend
2 2           #and number of parameters with polynomial trend
PARAMS        #Keyword
KD             #Parameter to be estimated(POROSITY,KD,IC,ALPHA_D,ALPHA_LH)
POROSITY      #number of parameters listed = NRF + NTRD with
ALPHA_D       #NRF parameters first
ALPHA_LH
SCALECONC     #Keyword scale FIELDDATA and CONCDATA conc. data to match
62.43         #VAM3DF units (i.e. ug/L to nanolb/ft^3)
FIELDDATA    #Keyword (optional)
2             #number of field data subdivisions groups
1 1           #number of field data in each subdivisions
1 574675.10 837828.60 310.0 1.1139 0.01 #para#,x,y,z,fd,sigma (vam coord.)
4 575083.00 837941.70 311.1 -1.7447 0.01 #para#,x,y,z,fd,sigma
NSCREENPTS   #Keyword (optional but before CONCDATA)
3            #number of well screen integration points
CONCDATA     #Keyword (optional)
2            #number of conc. field data subdivisions groups
1 1         #number of conc. field data in each subdivisions
572675.10 837728.60 290.0 280.0 3.3 20.1 0.01
#x,y,ztop,zbot,t,log10(conc),sigma (vam coord.)
571083.00 836041.70 301.1 291.1 23.2 1.7 0.01
#x,y,ztop,zbot,t,log10(conc),sigma
HYDROUNITS1  #Keyword (put HYDROUNITS with largest number of units first)
2            #number of vertically isolated units for parameter #1
1 6         #1st unit layers 1-6
7 11       #2nd unit layers 7-11
HYDROUNITS2  #Keyword
1            #number of vertically isolated units for parameter #2
1 11       #1st unit layers 1-11

```

HYDROUNITS3 #Keyword
 1 #number of vertically isolated units for parameter #3
 1 11 #1st unit layers 1-11
 HYDROUNITS4 #Keyword
 1 #number of vertically isolated units for parameter #4
 1 11 #1st unit layers 1-11
 VARIOGRAM1 #Keyword for parameter #1
 1.0 300.0 200.0 20.0 #sigma, x-correlation, y-correlation ,z-coor for 1st unit
 1.0 300.0 200.0 20.0 #sigma, x-correlation, y-correlation ,z-coor for 2nd unit
 VARIOGRAM2 #Keyword for parameter #2
 1.0 300.0 200.0 20.0 #sigma, x-correlation, y-correlation ,z-coor for 1st unit
 VARIOGRAM3 #Keyword for parameter #3
 1.0 300.0 200.0 20.0 #sigma, x-correlation, y-correlation ,z-coor for 1st unit
 VARIOGRAM4 #Keyword for parameter #4
 1.0 300.0 200.0 20.0 #sigma, x-correlation, y-correlation ,z-coor for 1st unit
 TREND1 #Keyword for parameter #1
 4 #number of trend coefficients for 1st unit
 1.0 1.0 #constant term value and sigma
 0.0 0.05 #x term value and sigma
 0.0 0.05 #y term value and sigma
 0.0 0.05 #z term value and sigma
 4 #number of trend coefficients for 2nd unit
 1.0 1.0 #constant term value and sigma
 0.0 0.05 #x term value and sigma
 0.0 0.05 #y term value and sigma
 0.0 0.05 #z term value and sigma
 TREND2 #Keyword for parameter #2
 1 #number of trend coefficients for 1st unit
 1.0 1.0 #constant term value and sigma
 TREND3 #Keyword for parameter #3
 1 #number of trend coefficients for 1st unit
 1.0 1.0 #constant term value and sigma
 TREND4 #Keyword for parameter #4
 1 #number of trend coefficients for 1st unit
 1.0 1.0 #constant term value and sigma
 EDITTHRES #Keyword
 10.0 #data editing threshold, large for no editing of data
 SOLVER #Keyword put DIRECT on next line for direct solve
 LSQR #Keyword put LSQR for lsqr solver, for representer method put REPM
 NITER #Keyword(optional, if NITER is not given, then NITER=8)
 3 #max. nonlinear iterations allowed
 CONVTEST #Keyword(optional, if CONVTEST is not given, then CONVTEST=0.2)
 0.1 #nonlinear iteration convergence criterion
 COVARIANCE #Keyword (optional) for covariance output
 3 #number of covariance terms
 1 1 #(I,J) for 1st covariance term
 2 2 #(I,J) for 2nd covariance term

000184

3 3 # (I,J) for 3rd covariance term
 MONTECARLO # Keyword (optional) for Monte Carlo output (must use REPM)
 3 -133 # number of Monte Carlo realizations, negative integer seed
 HISTOGRAM # Keyword (optional) for Monte Carlo histogram plot (must use MONTECARLO)
 2 9 # number of histograms, number of bars in each histogram
 title #1 # format(a80), title on histogram plot
 1 572675.10 837728.60 290.0 # para#,x,y,z for location of histogram
 title #2 # format(a80), title on histogram plot
 1 572675.10 837728.60 260.0 # para#,x,y,z for location of histogram
 MONTECRUN # Keyword (optional) run vam3df for realizations
 4545.0 # output statistics at given time
 run1.bin # format(a80), binary file name containing realizations
 RESTART # Keyword (optional) for restarting with previous results
 Restart.bin # file name of binary restart file containing cstat and M_c
 EKFUPDATE # Keyword (optional, can not use DIRECT) Extended Kalman Filter update.
 # use RESTART, if not CONCDATA terms of LHS will be updated before
 # the 1st iteration (first iteration like batch run and subsequent
 # iterations are EKF updates).
 CONCDATAE # Keyword (optional) Note: Add new field data to FIELDDATA
 2 # number of new conc. field data subdivisions groups
 1 1 # number of new conc. field data in each subdivisions
 572675.10 837728.60 290.0 280.0 3.3 20.1 0.01
 # x,y,z,t,log10(conc),sigma (vam coor.)
 571083.00 836041.70 301.1 291.1 23.2 1.7 0.01
 # x,y,ztop,zbot,t,log10(conc),sigma
 DEBUG # Keyword (optional) when estimating IC, this option will
 # write the tecplot file debug.tec containing the VAM3DF
 # initial conditions.
 CHEM4LINRKD # Keyword (optional) chemsorption for linear Kd,
 0.01 # alpha_c, 1st order decay becomes chemsorption by setting
 # 1st order decay = alpha_c/(1.0 + (poro*Sw/r_bKd)).

APPENDIX B
BREAKTHROUGH CURVES AND CONCENTRATION PROFILES
FOR K_d ESTIMATION RUNS

000187

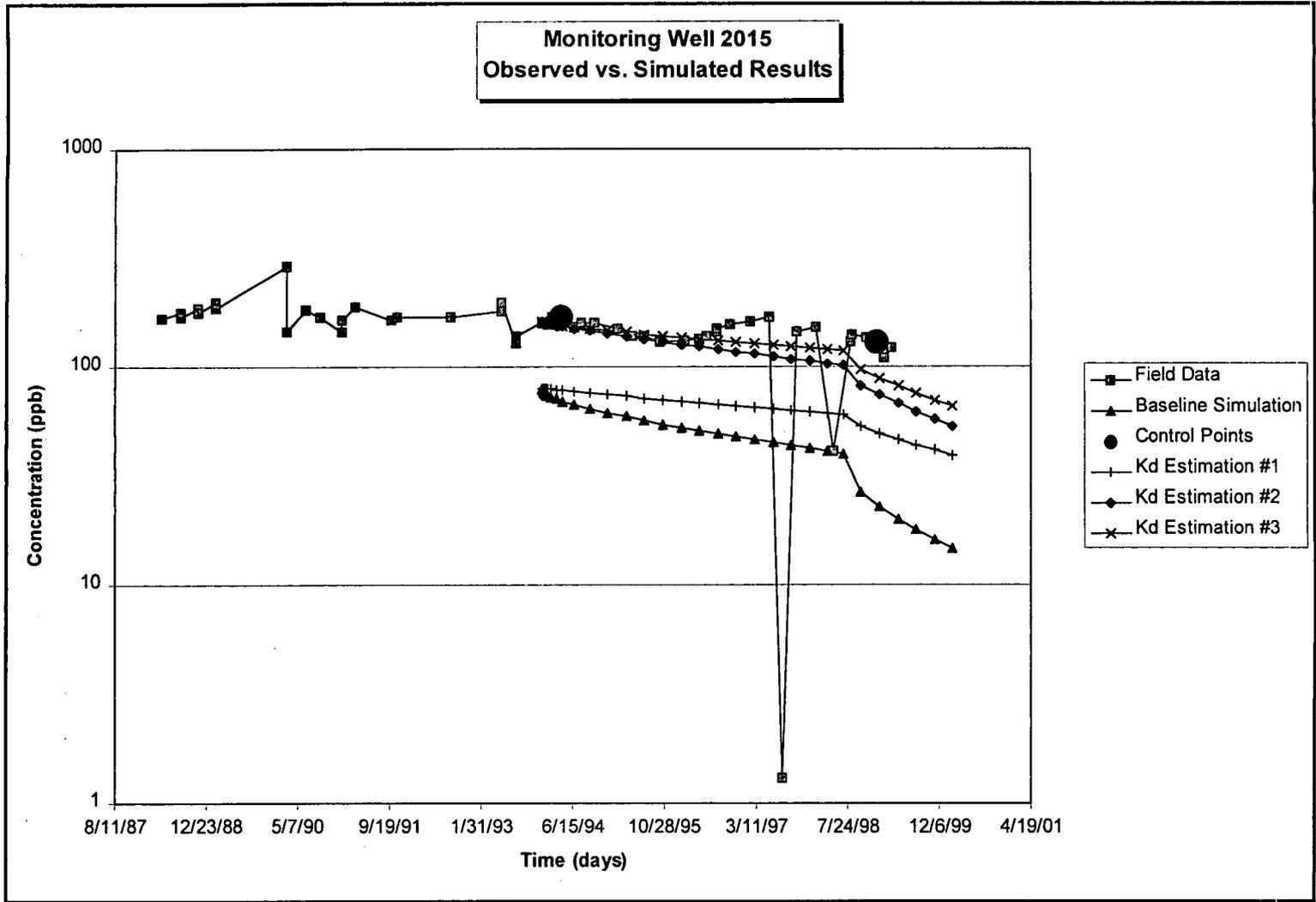


Figure B-1 Breakthrough Summary for Monitoring Well 2015.

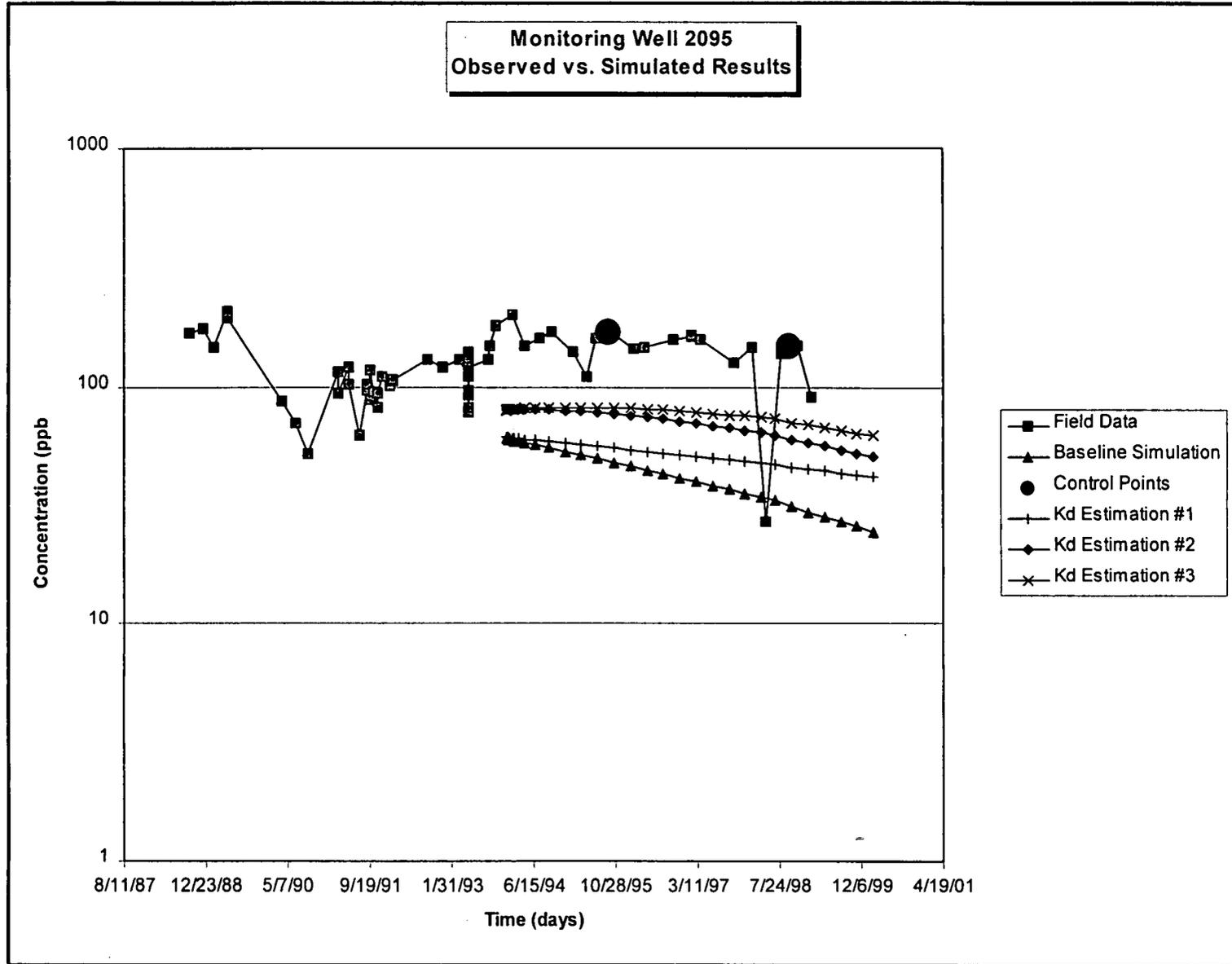


Figure B-2 Breakthrough Summary for Monitoring Well 2095.

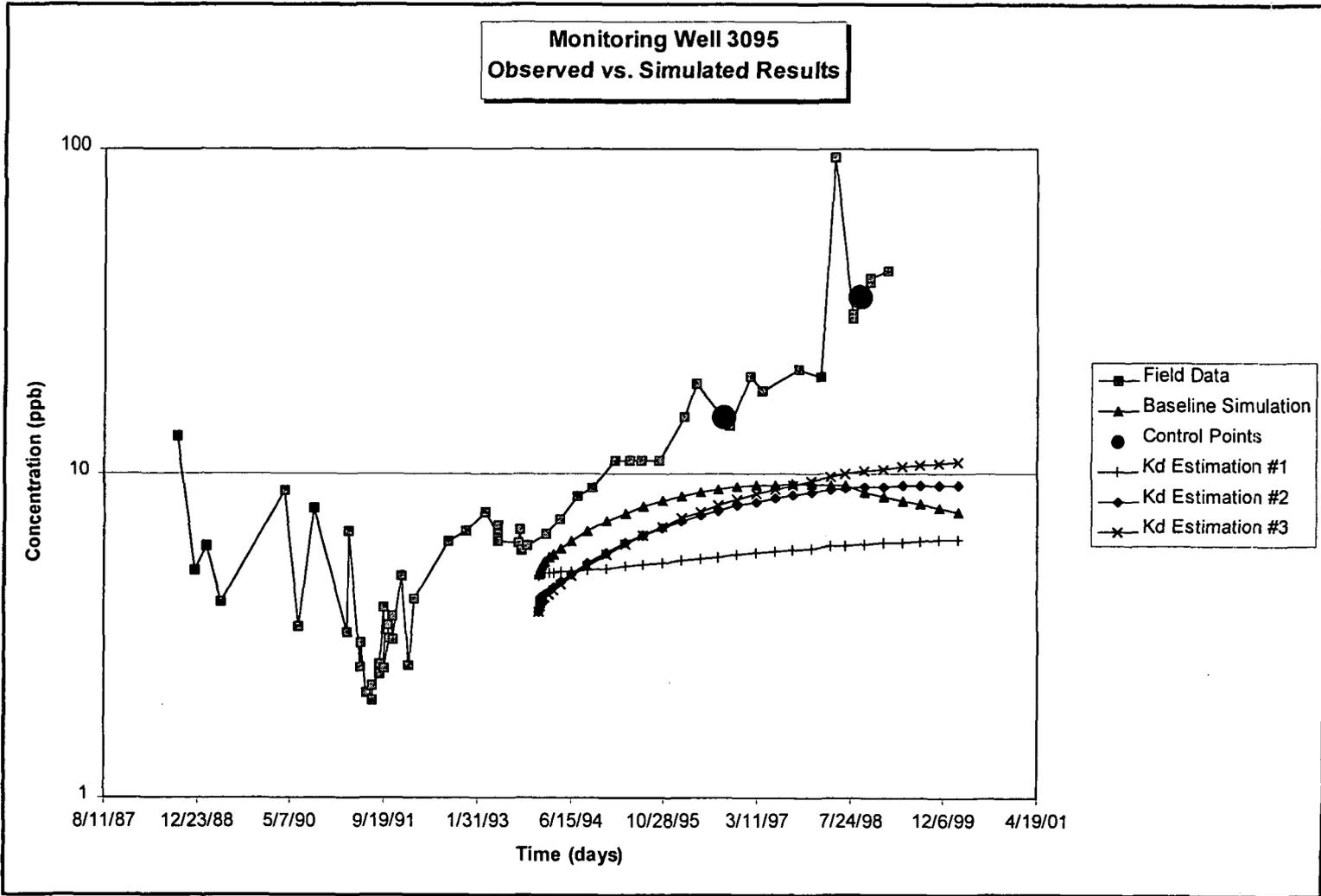


Figure B-3 Breakthrough Summary for Monitoring Well 3095.

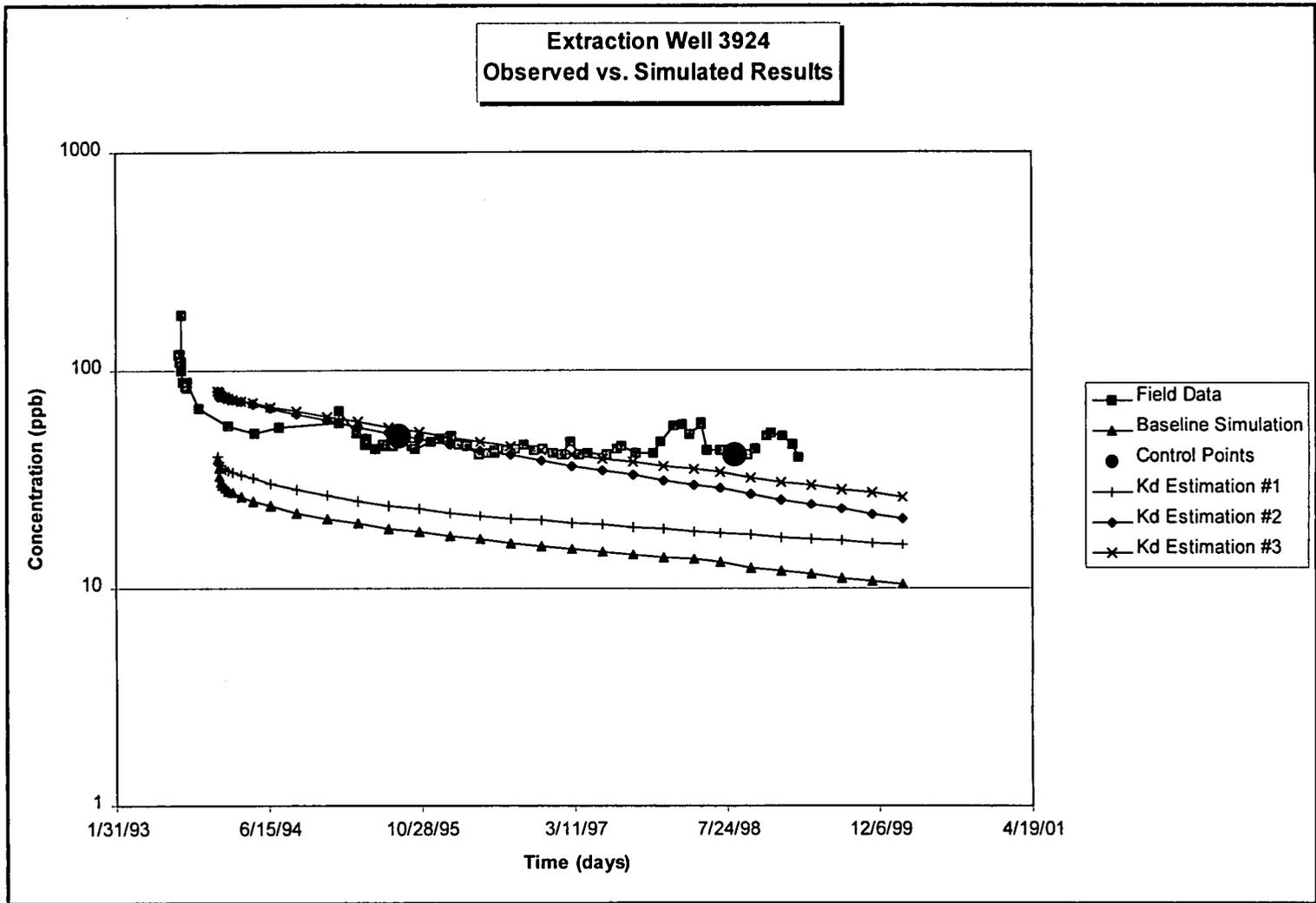


Figure B-4 Breakthrough Summary for Extraction Well 3924.

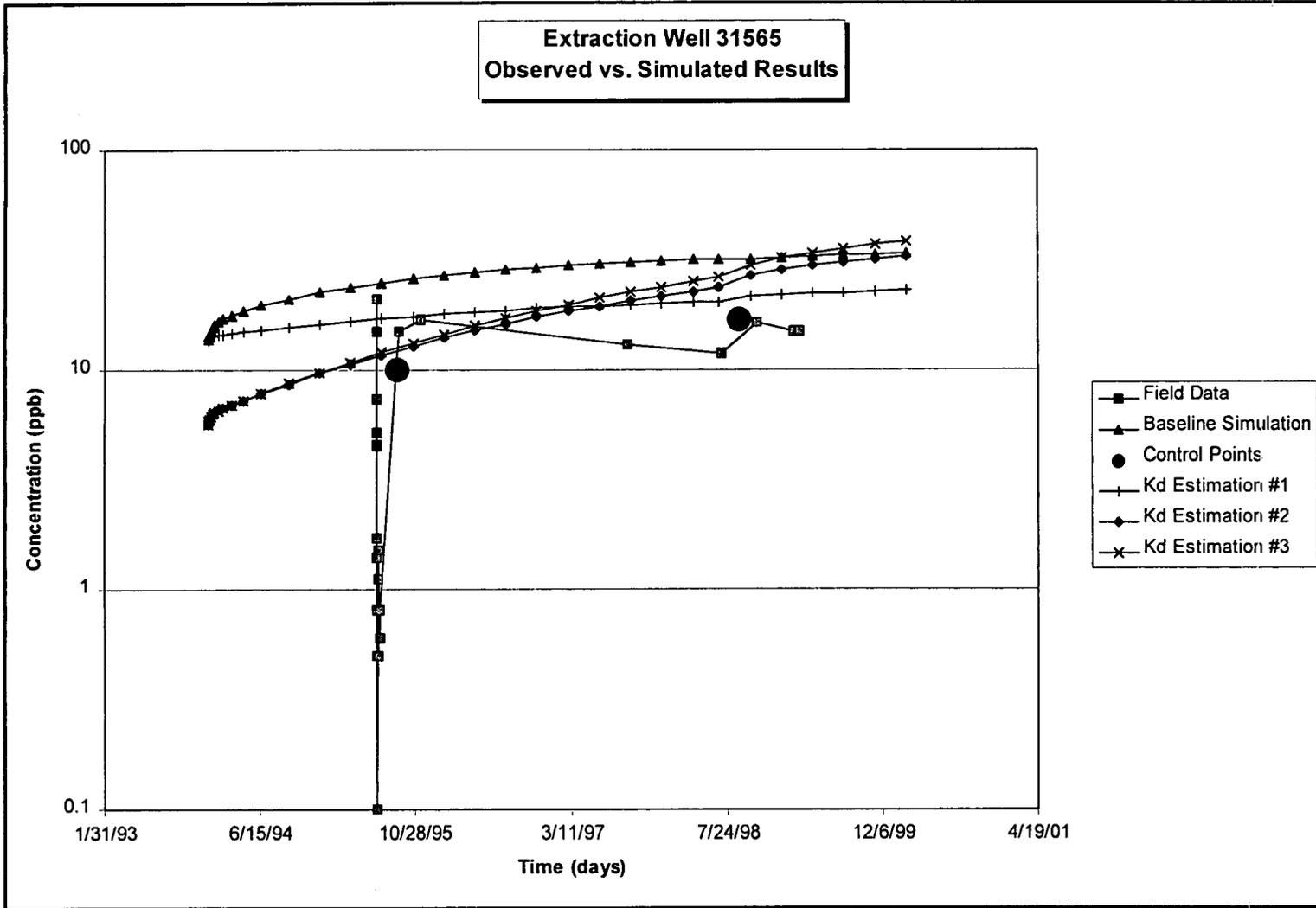


Figure B-5 Breakthrough Summary for Extraction Well 31565.

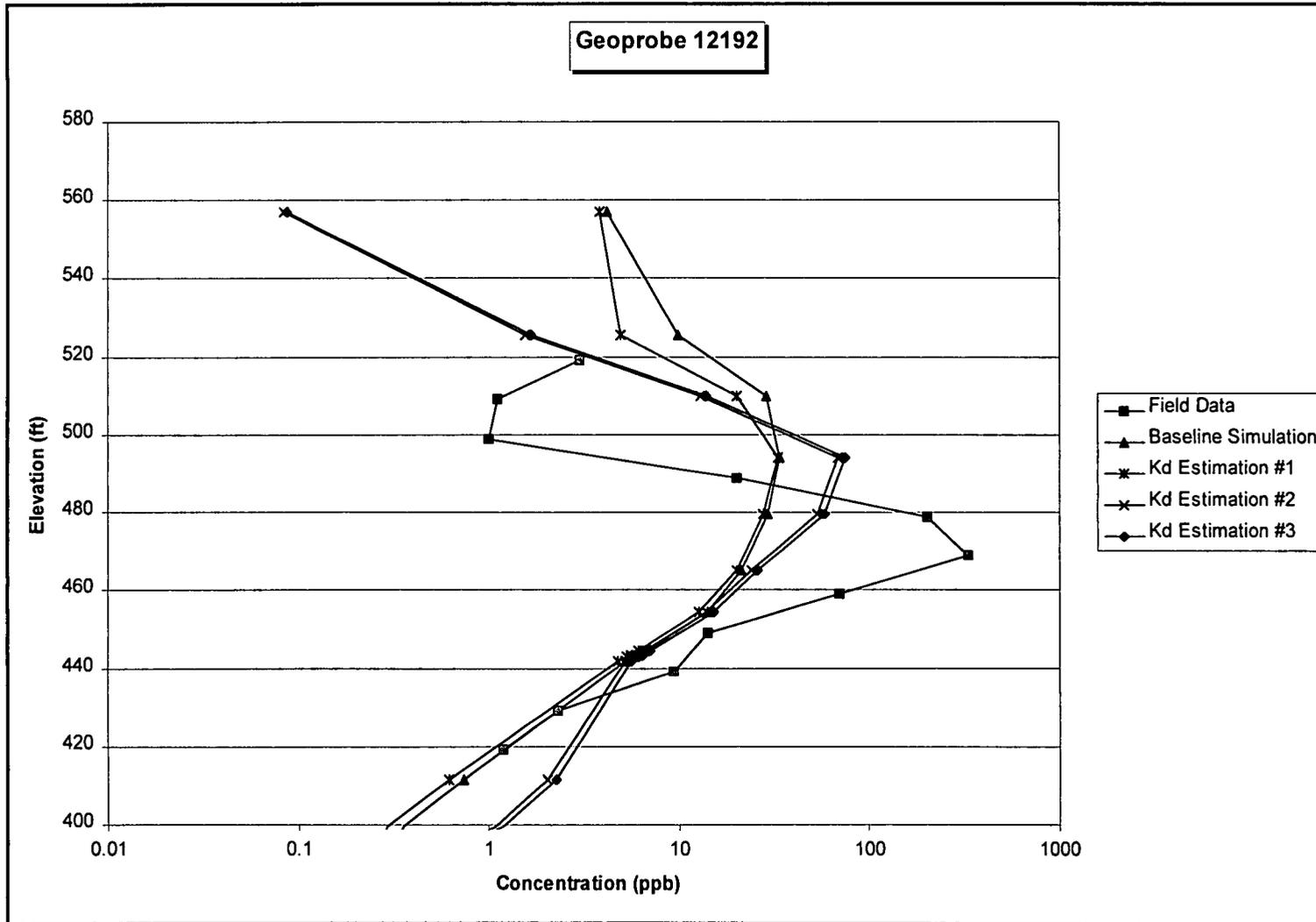


Figure B-6 Concentration Profile Summary for Geoprobe 2015.

000193

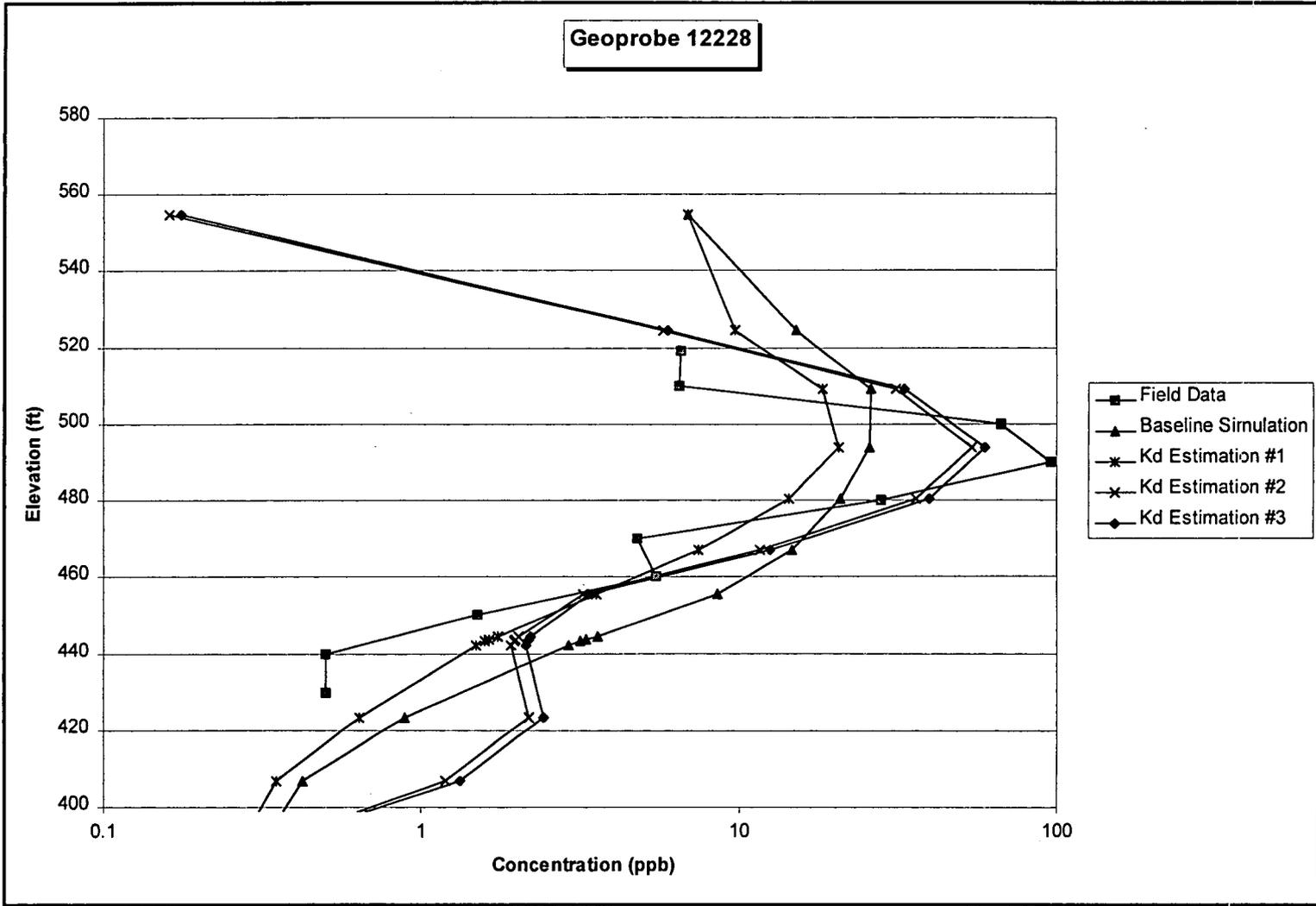


Figure B-7 Concentration Profile Summary for Geoprobe 2015.